

NORTHWESTERN UNIVERSITY

Structure and Assembly of Macromolecular Compartments

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Materials Science and Engineering

By

Curt Waltmann

EVANSTON, ILLINOIS

September 2023

© Copyright by Curt Waltmann 2023

All Rights Reserved

ABSTRACT

Structure and Assembly of Macromolecular Compartments

Curt Waltmann

The relationship between the structure and function of proteins is a fundamental problem in biology with implications for the future of biotechnology and global health. For example, changes to the structure of a coronavirus spike protein led to a global pandemic where our best defenses were vaccines that could only be designed with an intricate knowledge of the virus's structure. Proteins, such as those that comprise a virus's shell, are often thought of as completely distinct from synthetic polymers, due to their rigid, folded structure. However, unfolded, intrinsically disordered proteins and proteins with unfolded domains known as intrinsically disordered regions show semi-flexible polymer behavior, challenging this classification. Thus, it is necessary to understand the function of amino acid sequences that lack structure, such as the intrinsically disordered proteins found in membraneless organelles, which are involved in spatiotemporal control of cell metabolism. This dissertation will investigate how the properties of macromolecular assemblies are determined by the structural properties of their components both biological and synthetic. It includes collaborations with experimentalists, which show how these properties impact

material function in applications where species are compartmentalized including water filtration, plastic upcycling, gene therapy, and bio-nanoreactors.

Acknowledgements

There are so many people to acknowledge for their many different types of contributions to this thesis. First and foremost I would like to acknowledge my advisor, Monica Olvera de la Cruz. She gave me a lot of freedom to try different projects and techniques within the wide range of topics, but was still there to reign me back in when I inevitably went off the rails a few times. This includes connecting me to and allowing me to connect with many different experimental collaborators both within the group and Northwestern, also at other universities and even countries. She also allowed me to become a trainee for two different training programs, the Northwestern Biotechnology Training Program and NSF Center for Molecularly Optimized Networks. I would also like to thank the entire committee, Prof. Danielle Tullman-Ercek, Prof. Todd Gingrich, Prof. Nathan Gianneschi, and Prof. Peter Voorhees for taking the time to review this thesis. I believe all of these experiences with external labs, departments, and universities allowed me to grow a lot as a scientist over the past 5 years. I believe I was well-prepared to operate with this amount of freedom thanks to my undergraduate advisor at Iowa State, Prof. Alex Travesset. He was the number one reason I did a PhD at all and prepared me very well to do so.

As mentioned I had many experimental collaborators who I would like to specifically acknowledge here. I will start with Dr. Jeremy Wang whose work is featured throughout part 1 of this thesis. We worked together throughout both of our PhDs and I think we

both enjoyed the experience. I would also like to thank Prof. Carolyn Mills and Dr. Nolan Kennedy of the Tullman-Ercek lab. They were both super helpful for me working in the bio-engineering field which I was not very familiar with starting out and were also more than willing to do experiments I recommended even though they did not always work. Some of these experiments are shown in Part 2. The Nelson lab at the University of Washington also featured many collaborators who I would like to acknowledge even though the work did not fit in this document. Finally, I would like to acknowledge Prof. Uri Raviv and Dr. Roi Asor at the Hebrew University of Jerusalem who worked with me on what would become Chapter 6 of this thesis despite the fact that I was very early on in my PhD and was not familiar with a lot of best practices for writing scientific manuscripts.

There are also many people I would like to acknowledge within the Olvera group. The entire group is incredibly supportive, knowledgeable, and friendly. I would like to especially thank all of the officemates I had throughout the years, Dr. Yaohua Li, Dr. Wei Li, Dingwen Qian, Dr. Debadutta Prusty, Dr. Jianshe Xia, and Ali Ehlen who sat at the desk across from mine for the entirety of the 5 years I spent doing my PhD. Also our honorary officemate Hector Lopez, who despite not having a desk, was always in the office spreading his many thought-provoking opinions. I would also like to thank the "lunch crew" for excellent discussions including Dr. Debarshee Bagchi, Dr. Wei Li, Dr. Hang Yuan, Dr. Felipe Jimenez-Angeles, Dr. Leticia Lopez-Flores, Dr. Chase Brisbois, Dr. Jeremy Wang, and Han Umana-Kossio among others who dropped in from time to time. From a scientific standpoint, I must acknowledge the help of Prof. Baofu Qiao who taught me a lot about Martini and all-atom simulations, as well as Dr. Trung Nyguen and Dr. Felipe Jimenez-Angeles who were always willing to answer any questions I had. I

would also like to acknowledge Ali Ehlen and Nick Pogharian for reading and editing this thesis.

Outside of lab there are many people at Northwestern who made my time here so enjoyable and who I would like to acknowledge. Everyone who has been a part of the different organizations I was involved in at Northwestern (both official and unofficial) including coding club, MSSA executive board, BTP committees, intramurals, ect. A special shoutout to everyone who would play basketball on Saturday mornings and Karl Olson for being my co-president of Coding Club.

A final thanks to my friends and family whose contributions to my life extend incredibly far beyond my PhD and the contents of this document.

Table of Contents

ABSTRACT	3
Acknowledgements	5
Table of Contents	8
List of Tables	11
List of Figures	12
Chapter 1. Introduction	44
1.1. Structure of Individual Macromolecules	44
1.2. Assemblies I: Disordered Polyelectrolyte Complexes	46
1.3. Assemblies II: Crystalline Protein Shells	51
Part 1. Disordered Polyelectrolyte Complexes	56
Chapter 2. Heterogeneous Charged Complexes of Random Copolymers for the Segregation of Organic Molecules	57
2.1. Abstract	57
2.2. Introduction	58
2.3. Results and Discussion	63
2.4. Future Outlook	80

	9
2.5. Methods	82
Chapter 3. Tailoring Interactions of Random Copolymer Polyelectrolyte Complexes to Remove Nanoplastic Contaminants from Water	87
3.1. Abstract	87
3.2. Introduction	88
3.3. Results and Discussion	93
3.4. Conclusions	98
3.5. Methods	99
Chapter 4. Functional Enzyme-polymer Complexes	104
4.1. Abstract	104
4.2. Introduction	105
4.3. Results and Discussion	109
4.4. Conclusions and Outlook	121
4.5. Methods	122
4.6. Appendix	128
Part 2. Crystalline Protein Shells	134
Chapter 5. Vertex Protein PduN Tunes Encapsulated Pathway Performance by Dictating Bacterial Metabolosome Morphology	135
5.1. Abstract	135
5.2. Introduction	136
5.3. Results	139
5.4. Conclusions	153

	10
5.5. Methods	155
Chapter 6. Assembly and Stability of Simian Virus 40 Polymorphs	162
6.1. Abstract	162
6.2. Introduction	163
6.3. Results and Discussion	166
6.4. Conclusions	183
6.5. Methods	185
6.6. Appendix	196
Chapter 7. Kinetic Growth of Multi-Component Microcompartment Shells	202
7.1. Abstract	202
7.2. Introduction	203
7.3. Coarse-Grained Simulation Model	208
7.4. Results and Discussion	210
7.5. Conclusions and Outlook	233
7.6. Methods	235
7.7. Appendix	242
Chapter 8. Conclusions	246
References	250

List of Tables

2.1	Copolymer Characterization: Mole Fractions, Apparent Average Molecular Weights, Apparent Dispersity, and Apparent Average Degrees of Polymerization	63
2.2	Non-bonded parameters for the Coarse-Grained Forcefield	84
2.3	Bond parameters for the Martini Forcefield	86
2.4	Angle parameters for the Martini Forcefield	86
3.1	Anionic Copolymer Compositions	93

List of Figures

2.1 (a) Chemical structure, Martini parameterization, coarse grain model, illustration, and simulation snapshot of the anionic random copolymer. (b) Chemical structure, illustration, and coarse grain model of the cationic random copolymer. (c) Illustration and simulation snapshot of complexation between the anionic random copolymer and the cationic random copolymer. (d) Chemical structure of organic molecules used in experiments. (e) Coarse grain model of organic molecules used in simulations. Red and green beads correspond to positive and negative charges respectively, while tan beads are hydrophobic, and blue beads are hydrophilic.

62

2.2 (a) Filtration results. For each of the three dyes, the results are averages from three separate filtration samples. Crystal violet and phenolphthalein are quantitatively removed. For perfluorooctanoic acid, results for one and three filtrations on a sample of perfluorooctanoic acid are shown. Error bars are standard deviations from three runs of a single sample. (b) Images of 2 $\mu\text{g}/\text{mL}$ aqueous solutions of crystal violet before (left vial) and after (right vial) addition of 100 μL of anionic copolymer solution (66 mg/mL). (c) Images of 2 $\mu\text{g}/\text{mL}$ aqueous solutions of

crystal violet before (left vial) and after (right vial) encapsulation in a polyelectrolyte complex of anionic and cationic copolymer. (d) Visible absorbance spectra of a $2 \mu\text{g}/\text{mL}$ solution of crystal violet in water as a function of added anionic copolymer solution ($66 \text{ mg}/\text{mL}$). A significant solvatochromic shift is observed upon addition of trace levels of copolymer solution.

68

2.3 Distributions of different bead types in the Martini (right) and Coarse Grain (left) versions of the micelles. Both models show that the anionic polymers can form spherical micelles with dimensions of 4-5 nm. Distributions of beads which are hydrophobic are colored gold in both models and compose the hydrophobic cores of the micelles. Hydrophilic beads in both models are colored blue and form the corona of the micelles. The negatively charged beads in both models are colored teal and sit at the edge of the hydrophobic core about 2 nm from the center of mass. In the Coarse Grained model, more stretched conformations are observed when interacting with the cationic copolymers, similar to the Martini model.

70

2.4 Probability of a chain being in a certain sized cluster, where the cluster size is measured as a fraction of the finite sized simulation, as a function of polymer charge ratio, which is defined as the ratio of the number of positive charges to the number of negative charges in polymer clusters (the system is electrically neutral due to counterions). At low charge ratio, free micelles, small clusters, medium-sized clusters, and large

clusters comprised of nearly every chain are observed (left). As the polymer charge ratio is increased above 1, there is an electrostatic driving force for the free micelles to enter the dense phase; the free micelles should be incorporated in the dense phase even if it is not connected through hydrophobic interactions (center). As the charge ratio is further increased to 2 and above, the medium-sized clusters effectively disappear, leaving the polymers in one dense phase (red test tube). This agrees well with experiments where a polymer charge ratio above 3 is necessary to drive all the polymers into a macroscopic phase. 74

2.5 (a) Calculations of charge heterogeneity for charges on polymers. The simulation box is split into smaller cells of different lengths, L , and then the effective charge from the polymers in these boxes is calculated according to eq 2.2. Box sizes of 2.5 and 5 nm show two peaks where the effective charge is ± 1 . This shows how the hydrophobic energy of the polymers leads to local charge segregation in these polyelectrolyte complexes. (b) Calculations of charge heterogeneity for charges on polymers and counterions. The same calculation is performed as described above. Free counterions help to negate some of this charge segregation, but at small length scales, the two peaks are still observed. At larger length scales, with a Gaussian distribution centered at 0, effective charge is observed. (c) Electrostatic driving force for dye segregation. The electrostatic energy of different dyes when they are

free in solution and segregated in the complexes as well as the energy difference between the states.

2.6 (a) Percentage segregation for the seven simulated dyes. The ratio of dyes to polymer monomers is 1:130, and the volume fraction of polymers is $\approx 10\%$. Segregation is defined as any hydrophobic dye bead being within a cutoff distance from a hydrophobic polymer bead. Almost all condensation occurs on the anionic polymers, which form stretched micelle-like structures with hydrophobic cores. The more hydrophobic a dye is, as measured by the number of hydrophobic beads, the higher the percentage segregated. Adding a charged bead to a hydrophobic dye always decreases the percentage segregated. (b) Circular variance as a measurement of dye location within hydrophobic domains. The circular variance is used to measure the degree of hydrophobic burial of a dye. It is calculated by taking the length of the vector sum of all the unit vectors from the dye to hydrophobic beads that are within the cutoff distance. This is then divided by the number of vectors and subtracted from 1. Thus, 1 is the maximum burial, and 0 is the minimum. (c) The position of dyes within hydrophobic domains as measured by circular variance. Left and center show specific comparisons for dyes with and without charges. The uncharged dyes are always much more buried. The right shows how the degree of burial continues to increase as the number of hydrophobic beads in the dye is increased.

- 2.7 A full description of the bonded parameters for the random anionic copolymers. Angle names (see Table 2.4) are italicized and bond names (see Table 2.3) are underlined. Bead names control which the Lennard-Jones parameters all come from the martini forcefield. 85
- 3.1 Schematic of polymer complexes and polymer adsorption onto PET. (a) Diagram of polyelectrolyte complex of random copolymers with adsorption of PET nanoparticles. (b) Chemical composition of the anionic and cationic copolymers. (c) Coarse-grained molecular dynamics simulation of anionic copolymer adsorbing onto PET. (d) Picture of nanoplastic spiked solution (left) and the polyelectrolyte complex formed after addition of random copolymer solutions, which has accumulated on a stir bar. (right) 92
- 3.2 Image of PET nanoplastics and remediation rates from water. (a) SEM image of PET nanoplastics synthesized through a nanoprecipitation technique. (b) Remediation rates of nanoplastics for complexes that include polymer samples 1-4, which have a similar charge fraction. F_H denotes the fraction of hydrophobic monomer of the polymer sample. (c) Remediation rates of nanoplastics for complexes that include polymer samples 1, 5, and 6, which have roughly 50% more hydrophobic monomer than hydrophilic monomer. F_C denotes the fraction of charged monomer of the polymer sample 94

- 3.3 Polymer adsorption onto PET vs hydrophobic fraction F_H with experiments and simulation. (a) Mass of polymer adsorbed by samples 1-4 onto a film of PET as measured by QCM. The error bars are approximately the size of the markers used in the plot. (b) Simulation results of hydrophobic contacts between anionic copolymers and PET surface. 96
- 3.4 Simulation results for polyelectrolyte complexes with varying hydrophobic fraction of the anionic copolymer, $F_{H,anionic}$. (a) Scattering functions for $F_{H,anionic} = 25, 35, 45,$ and 55 all show peaks at $q = 0.56 \text{ nm}^{-1}$ and some show a smaller peak at $q = 0.78 \text{ nm}^{-1}$. These peaks correspond to the spacing between domains, which $\approx 10\text{nm}$. (b, d) Snapshots of the most and least hydrophobic complexes. (c,e) Local hydrophobic monomer density distributions show multiple peaks for the least hydrophobic complexes at length scales smaller than the interdomain spacing. 97
- 4.1 Description and models of PETase and the random copolymers. (a) GoMartini model of PETase in magenta with the active site in gray. (b) The secondary structure of PETase is shown with the same color scheme as in (a). The active site is shown using the van der Waals representation to highlight the cleft-like binding pocket for PET. (c) Surface representation of PETase. (d) Chemical and Martini description of the methacrylate-based random copolymers. Hydrophobic beads are tan, while hydrophilic beads are blue, and negatively charged beads

are cyan. F_H , F_L , and F_- refer to the percentage of the respective monomer in the random copolymers. (e) Snapshot of the Martini random copolymer model with the colors corresponding to (d). 108

4.2 Comparison of the GoMartini PETase (b) alone to atomistic simulations (a) and to the GoMartini PETase complexed with random compolymers as the temperature is increased (b). RMSD is used to measure the conformation relative to the crystal structure with higher values signifying more deformation. (a) Results for the atomistic model show a general increase in the RMSD of the protein as well as the active site as is expected based on known decreases in activity with increasing temperature. (b) The GoMartini model shows an increase in RMSD for the whole protein and the active site at 320K, but at 350K the active site RMSD decreases, unlike the entire protein backbone. The active site RMSD may be inaccurate at high temperatures, but can still be used as a baseline to measure the thermal stability of the PETase-polymer complexes. In these complexes, temperature dependence of the whole protein and active site conformation is nearly eliminated. 110

4.3 Random copolymer complexation with the PETase protein for different mean polymer compositions. F_H is the percentage of hydrophobic EHMA, F_- is the percentage of negatively charged SPMA and F_L is the percentage of hydrophilic OEGMA-9. These percentages sum to 100 and thus F_L , which is not displayed, is $100 - F_H - F_-$. (a) The number of contacts is shown as a function of composition at 298K. A maximum

is observed at very low percentages of EHMA and higher percentages of SPMA, while a local minimum is observed at $F_H=20\%$. (b) Simulation snapshot of a wrapped polymer conformation, which occurs at low values of F_H and is characterized by a high percentage of contacts between the enzyme and polymer backbone. (c) Simulation snapshot of a globular polymer conformation, which occurs at high values of F_H and is characterized by micelle-like behavior of the amphiphilic polymers. 112

4.4 Hydrophobic interactions affect PETase-random copolymer complexation. (a) The fraction of contacts that involve hydrophobic polymer beads. There are two local maxima, one at very low F_H where the polymer backbone wraps around the protein surface and one at high F_H where the EHMA increases the baseline hydrophobic fraction of the polymer. (b) The fraction of these hydrophobic contacts that occur on the hydrophobic surface of PETase is lower than the hydrophobic surface fraction of PETase. This struggle to optimize the interaction could be related to ill-defined hydrophobic domains due to partially hydrophobic amino acids in the Martini model. However, polar-polar interactions seem to be optimized and this increases as charge is added, while the opposite occurs for hydrophobic-hydrophobic interactions. 113

4.5 Including negatively charged monomers increases contact with positive surface domains. (a) and (b) show PETase with only the surface potential and with additional contacts (ACs) overlaid respectively. One can see that the yellow and cyan sites preferred when $F_- = 10\%$

are overwhelmingly on the positive part of the protein due to the addition of negative charge to the polymer. (c) and (d) show the a rotated orientation of the protein. The full data set for this figure, more method description, and higher temperature results can be found in the Appendix Figure 4.9. 115

4.6 Less perturbed active sites are stabilized by additional contacts near the active site. (a) The distribution of active site RMSDs at 298K and 320K. The less perturbed active sites are colored in blue while the more perturbed active sites are colored in red. Each point refers to a different polymer composition and the line refers to PETase alone at 298K. These are the groups being compared in (c) and (d). (b) Surface representation of PETase with no excess contacts shown for comparison. (c) Comparison at 298K shows many additional contacts (ACs) around the active sites for less perturbed compositions. These contacts stabilize the active site instead of further perturbing it at 320K (d) as well. The full data set for this figure and more method description can be found in the Appendix, Figure 4.13. 118

4.7 Activity of PETase and PETase/copolymer complexes at two different PETase:copolymer molar ratios. (a) Specific activity against small molecule substrate p-nitrophenyl acetate after one hour incubation at various temperatures. Error bars represent standard deviation over 3 replicate experiments (b) PET degradation activity over five hours at 35 °C. Error bars represent the 95% confidence interval on activity values. 119

- 4.8 Contacts at Elevated Temperature. At 320K and 350K the same trends are observed in terms of which compositions lead to the most contacts and the fraction of those contacts that are hydrophobic. In general, the total number of contacts does decrease for every composition as temperature is increased, while the fraction of those contacts that are hydrophobic increases. 128
- 4.9 Charged Polymers Affect Spatial Distribution of Contacts. In Figure 4.5 in the main text, we show that the positive section of the protein is more contacted when negative charges are present on the polymers. Here, we explain the full methodology. The first step is calculate the average number of contacts between the random copolymers and every single bead that makes up the GoMartini PETase. Then we average that over all compositions with $F_- = 0\%$ and do a separate average over all compositions with $F_- = 10\%$. Then we subtract the $F_- = 0\%$ data from the $F_- = 10\%$ data. That is what is shown at the top of (a), (b), and (c) with the active site shown in dashed gray lines. At the bottom of (a), (b), and (c) We then take that data and graph the spatial position of those bead in spherical coordinates if they meet the cut off threshold of having an absolute value equal to or greater than 1 as described in the legend at the bottom of (a). In Figure 4.5 in the main text, any residue featuring a bead that reaches a threshold value at the bottom of (a) at 298K is colored over the electrostatic surface potential of the protein. This shows that these excess contacts for $F_- = 10\%$ occur on the positive

part of the protein. In (b) and (c), we also show this spatial distribution for 320K and 350K. This shows that although the exact beads that break the threshold value change, the preference for the same region of the protein in spherical coordinates is clear.

129

4.10 The distribution of RMSD values when polymers are attached. The data for attached polymers is shown as a boxplot with individual data points corresponding to different polymer compositions. The box extends from the first quartile of the data to the third quartile meaning half of the distribution is inside the box. A line is drawn at the median. The protein by itself is shown in red as a single average point and the distribution represented by the standard deviation and thus 66% of the data is inside the red bars. These data as a function of composition can be found in the Appendix Figures 4.11 and 4.12 respectively. (a) The distribution of the active site RMSD values at 298K and 320K. 350K is not shown due to inconsistencies in the active site behavior of the atomistic and GoMartini models at 350K. The distribution is much wider when polymers attached suggesting the active site conformation is modified and not just “trapped” by the presence of the polymers. (b) The distribution of the whole protein RMSD values at 298K, 320K, and 350K. This distribution when polymers are attached is narrower than the active site RMSD in (a).

130

4.11 Active site RMSD as a function of polymer composition at 298K (a), 320K (b), and 350K (c). For each temperature the colorbar is adjusted

to be centered at the active site RMSD value of the GoMartini PETase active site with no polymers bound. This is done to provide a direct comparison between the bound states and unbound state of PETase. For each temperature, there is little to no trend in active site RMSD based on the polymer composition. In the main text we describe how this likely due to an inability of composition to control contacts specifically on the active site. However, above 298K the active site RMSD tends to be lower than the PETase alone across compositions. This is the same data that is displayed in Figure 4.10(a). 131

4.12 Whole protein RMSD as a function of polymer composition at 298K (a), 320K (b), and 350K (c). For each temperature the colorbar is adjusted to be centered at the whole protein RMSD value of the GoMartini PETase active site with no polymers bound. This is done to provide a direct comparison between the bound states and unbound state of PETase. For each temperature, there appears to be a slight trend where compositions with more charge and hydrophobicity have lower RMSD values. This is the same data that is displayed in Figure 4.10(b). 132

4.13 Active site conformation is affected by polymer contacts. In Figure 4.6 in the main text, we show how compositions with less perturbed active sites have more contacts very near the active site than those with more perturbed active sites. Here, we show the raw data for that figure. In (a) and (b) we see the subtraction of the average number of contacts for compositions with more perturbed active sites from those with less

perturbed active sites. In (c) and (d) we apply the threshold absolute value of one and graph the beads which meet those values in spherical coordinates. The yellow dots are positive meaning that these spots have at least one more polymer contact on average for less perturbed active sites. The purple dots are negative meaning that they have at least one more contact when the active site is more perturbed. The positive - negative convention is just based on how the subtraction was done. In (e) and (f) these distributions are shown on the GoMartini model to get a fuller sense of where the contacts are relative to the active site. (g) is just the GoMartini model and is provided as a comparison. 133

5.1 The *pdu* operon in *Salmonella enterica* serovar Typhimurium LT2 contains the genes encoding proteins responsible for formation of the 1,2-propanediol utilization microcompartment (Pdu MCP). These include enzymes that perform both key pathway steps and cofactor recycling functions (orange) and shell proteins that encase these enzymes (bacterial microcompartment, BMC, domain-containing genes shown in blue, bacterial microcompartment vertex, BMV, domain-containing gene shown in green). Notably, only one shell protein in the *pdu* operon, PduN, contains a BMV domain. 139

5.2 (a) Depiction of different *pdu* operon genotypes used in this figure. (b) Coomassie-stained SDS-PAGE of purified Pdu MTs (Δ PduN) and Pdu MCPs (WT) comparing the protein content in these purified structures, where labels indicate the Pdu protein the band corresponds to (i.e.

C/G for PduC/PduG), except for Lys, which indicates lysozyme.

(c) Comparison of structures formed in Pdu MCP-forming strains (WT) and Pdu MT-forming strains (Δ PduN). Scale bars in optical and fluorescence micrographs are 5 μ m. (d) Phase contrast and GFP fluorescence micrographs showing the impact of increased PduN-FLAG expression on the formation of Pdu MT structures versus closed Pdu MCP structures, where increasing arabinose concentration correlates with increasing expression of the PduN-FLAG protein off the pBAD33 plasmid. Scale bars in optical and fluorescence micrographs are 5 μ m. (e) Coomassie-stained SDS-PAGE, anti-FLAG western blot, and negatively stained TEM on Pdu MCPs purified from a pduN knockout strain supplemented with PduN-FLAG off a plasmid. Source data for (b) and (e) are provided as a Source Data file in the original manuscript [1]. Similar results to those reported in (b–e) were observed across three independent biological replicates, except for TEM imaging of thin cell sections, which was performed on multiple cells in a given biological sample, but not with biological replicates.

141

5.3 (a) Schematic of the 1,2-propanediol utilization pathway encapsulated in Pdu microcompartments. (b) Strains containing different compartment geometries (MCPs in Wild Type, blue lines, MTs in Δ PduN, green lines), without compartment expression (Δ PocR, grey lines), and with broken compartments (Δ PduA PduJ, red lines) grown in minimal media (NCE) with 1,2-propanediol as the sole carbon source. Data are

presented as mean values \pm standard deviation over three biological replicates. (c) Concentration of key pathway metabolites over the course of the growth described in (b). Data are presented as mean values \pm standard deviation over three biological replicates. Source Data for panels (b) and (c) are provided as a Source Data file in the original manuscript [1].

144

5.4 (a) Schematic of the PduA-PduN interface used for these simulations, where PduA is shown in blue and PduN is shown in green. (b) Potential of mean force (PMF) calculated from AAMD simulations as a function of bending angle, θ_B , between PduA and PduN. $\Delta G_{0^\circ \rightarrow 40^\circ}$ is the difference in the PMF between the 0° and 40° bending angles. (c) PMF calculated from AAMD simulations as a function of the distance between PduA and PduN, used to calculate the total interaction energy (ΔG) between these two oligomers. (d) Schematic of the PduA-PduA interface used for these simulations. (e) PMF calculated from AAMD simulations as a function of bending angle, θ_B , between two PduA hexamers. $\Delta G_{0^\circ \rightarrow 34^\circ}$ is the difference in the PMF between the 0° and 34° bending angles. Calculations used calculate data points in (b), (c), and (e) are described in Method section. Error bars on plots in (b), (c), and (e) represent the sampling error on the calculated energies, estimated by splitting simulation data into different sections and observing the differences in the calculated potential as described in the

Methods section, Calculation specifics. Source data for plots (b), (c), and (e) are provided as a Source Data file in the original manuscript [1]. 149

- 5.5 (a) Representative snapshot from the all-atom molecular dynamics simulation at the lowest energy 40° bending angle. The inset shows a short range cation-pi interaction between Arginine 66 on the PduA hexamer and Phenylalanine 71 on the PduN pentamer.
- (b) Representative snapshot from the all-atom molecular dynamics simulation at a 70° bending angle. The inset shows hydrogen bonding between Asparagine 67 on each PduA hexamer. This is believed to explain the second well in the bending potential landscape in Fig. 5.4e. 150
- 5.6 (a) PMF calculated from AAMD simulations as a function of the distance between HO Hex and HO Pent, used to calculate the total interaction energy (ΔG) between these two oligomers. (b) Potential of mean force (PMF) calculated from AAMD simulations as a function of bending angle, θ_B , between HO Hex and Pent. $\Delta G_{0^\circ \rightarrow 30^\circ}$ is the difference in the PMF between the 0° and 30° bending angles. (c) Schematic of the HO Hex-HO Pent interface used for these simulations, where HO Hex is shown in blue and HO Pent is shown in orange. Hydrogen bonds between oppositely charged amino acids form at the bottom the interface. (d) PMF calculated from AAMD simulations as a function of the distance between HO Hex and HO Hex, used to calculate the total interaction energy (ΔG) between these two oligomers. (e) PMF calculated from AAMD simulations as a function of bending angle,

- θ_B , between two PduA hexamers. $\Delta G_{0^\circ \rightarrow 21^\circ}$ is the difference in the PMF between the 0° and 21° bending angles. (f) Schematic of the HO Hex-HO Hex interface showing the formation of Arginine pairs at the bottom interface. Calculations used calculate data points in (a), (b), (d), and (e) are described in the Methods section. Error bars on plots in (a), (b), (d), and (e) represent the sampling error on the calculated energies, estimated by splitting simulation data into different sections and observing the differences in the calculated potential as described in the Methods section, Calculation specifics. 152
- 5.7 Balance of forces in calculation of $F_{\theta_B}(z)$. 158
- 5.8 Mapping between bending angle, θ_B (degrees), and the distance between the centers of mass of the pentamer and hexamer, z (nm). 159
- 5.9 Histograms showing the parallel “windows” used to calculate the potential of mean force. The overlap of the windows is shown by the distance between the centers of mass of the pentamer and hexamer, z (nm). 160
- 6.1 (A) Images of VP1 based on protein data bank (PDB) entry 1SVA [2] and its coarse-grained equivalent. The pink beads are rigid and represent the globular portion of the protein. The modeled VP1 pentamer can be split into 5 identical units, just as the real VP1 pentamer contains 5 chains with the same primary structure. Each unit has a rigidly attached C-terminal ligand comprised of nine beads which

are only rigid in the middle (but are freely rotating) and terminate by a purple connector bead.(B) The connector beads hybridize with the cyan connection site with an energy given by the parameter ϵ . Each unit also contains an N-terminal ligand made of 5 flexibly connected beads. The first two are uncharged while the last 3 represent the positively charged residues found on this part of the VP1 protein.

167

6.2 The two general cases of the implicit ion model. (Left) The case of screened electrostatics where the salt provides uniform screening of the electrostatic interactions present in the system. The strength of this screening is given by the Debye screening length, λ_d , as shown in Equation 6.1. (Right) The case of template binding at high salt. We model this by putting the $q_{eff}=0$ and adding a short range attraction, α , that comes from depletion since Debye-Huckel is not valid (Equation 6.2).

168

6.3 The $T = 1$ capsid is made of $N=12$ VP1 pentamers. The relative potential energy per VP1 and pairwise distance distribution of the capsid as a function of the template diameter is measured by first allowing 12 VP1 pentamers to bind to a large template and then slowly reducing the size of the capsid such that no VP1 pentamers are released during this process illustrated at the top left of the figure. Results on the bottom left show that the minimum of the potential energy occurs at a template diameter of 9 nm. On the bottom right, we see 3 peaks matching the icosahedral symmetry (see main text) and large regions of

zero probability, indicating the static nature of the VP1 pentamers in this configuration. Since this static $T = 1$ configuration is stable over many template sizes we believe that it would also be robust against changes in template size based on changes in salt, pH, *etc.*, provided that these changes do not impact the VP1-VP1 or VP1-template interactions too much. This minimum energy structure also recovers the presence of a three helix triangle located at the three fold symmetry points of the icosahedron [3], shown at the top right. The image of the three-helix triangle with the scale bar was adapted with permission from Kler *et al.*, ACS Chemical Biology 2013, 8, 2753–2761. Copyright 2013 American Chemical Society.

170

6.4 (Left) The N=13 shows deviation from the icosahedral symmetry observed in the case of the N=12, $T = 1$ capsid. Instead, a six-coordinated VP1 pentamer is observed (top left), consistent with the Euler formula for closed shells made of regular polygons. The six-coordinated VP1 pentamer is also mobile as shown by the absence of zero-probability regions in the pairwise distribution function (which were observed for N=12 in Figure 6.3). (Right) This structure is only favored over the $T = 1$ for 12 nm and above templates where it has a lower energy per VP1 pentamer (U/N) than the $T = 1$ capsid. At smaller template diameters, one VP1 pentamer will be spontaneously released and the $T = 1$ capsid reforms.

172

- 6.5 The $N = 10$ capsid can adopt different configurations based on the curvature of the template. (Right) The 11 nm template shows icosahedral symmetry like the $N = 11$ and $T = 1$ case, just with two holes that are second nearest neighbors. (Left) For templates less than 11 nm, the capsid shows half of the template with a 5 fold symmetry, while the other 4 VP1 pentamers are found in an unstructured state. 174
- 6.6 Fraction of Templates with N bound pentamers at different time points along the simulations. Assembly of $T = 1$ capsids is sensitive to added salt concentration as shown by the fraction of templates having $N=12$ capsids at $t=5/5$. (Left) Assembly occurs only slightly faster when the Debye length is increased relative to biological salt conditions (Center). (Right) Having only short ranged attraction decreases the observed fraction of $T = 1$ capsids assembled even when the attraction is strong. 176
- 6.7 An example of binding of the final VP1 pentamer to an $N=11$ capsid during a simulation using $\epsilon = 8 k_B T$, $\lambda_d = 1.0$ nm, $q_{eff} = 0$, and a 10 nm template. The black arrows point forward in time, while the red arrow points to a decrease in the total free energy. The process requires cooperative interactions between the final VP1 pentamer, the bound VP1 pentamers, and the template. This process shows that the C-terminal ligands stabilize the 12th VP1 pentamer on the partially assembled capsid until it finds an orientation where it can bind to the surface of the template. To find this orientation, the VP1 pentamer

- searches a rough free energy landscape and thus it is important that the interactions are weak enough to be reversible. 178
- 6.8 An illustration of the shortcomings of the simplistic model to represent the ease of making the first contact. Considering the full extensibility of the connector domain, the first contact between a partial capsid and a VP1 pentamer can occur at a much longer interaction distance. 179
- 6.9 (Top) The difference between the simplistic and extended model is displayed both in an initial conformation and in a typical conformation of a free VP1 pentamer. (Bottom) Inclusion of the full length connector in the model increases the rate of assembly. More $T = 1$ capsids are present at every step of the simulations. 180
- 6.10 Assembly simulations around spherical templates with a diameter of 10 and 11 nm. The top figures show the fraction of templates with N bound VP1 pentamers as a function of the fractional simulation time, t . The bottom shows the pairwise distance distributions of VP1 pentamers in each case. A peak is observed at $N = 10$ pentamers on the 10 nm template owing to lack of available binding sites, as compared to the 11 nm case. The lack of available template binding sites is shown by the lack of icosahedral symmetry seen in the pair distribution (especially at $t = 2/5$). 182
- 6.11 The full proposed growth mechanism of VP1 pentamers to an incomplete or pseudo-closed template. The VP1 pentamer first contacts a stray C-terminal ligand of a VP1 pentamer already bound to the template.

This binding can happen at large distances due to the length of the connector domains. If the capsid is incomplete, cooperative interactions can then bring the VP1 pentamer and the incomplete capsid together, and align them such that the final VP1 pentamer perfectly fits into the icosahedral hole on the template. This process is referred to as the elongation growth mechanism. If the capsid is pseudo-closed, the connected VP1 will have to allow the capsid to dynamically rearrange in order to add a VP1 pentamer to the capsid in a much slower process. The images at the bottom show what appears to be C-terminal ligand connections being made by free VP1 pentamers on pseudo-closed particles larger than $T = 1$ and smaller than $T = 7$ using electron microscopy [4], republished with permission of the Microbiology Society, from Simian Virus 40 VP1 Capsid Protein Forms Polymorphic Assemblies *In Vitro*, Kaneshashi *et al.*, Journal of General Virology 2003, 84, 7, 2003; permission conveyed through Copyright Clearance Center, Inc. It should be noted that there is no template in this image and thus the VP1 pentamers are connected to the pseudo-closed capsid and to each other only by C-terminal ligand interactions.

184

6.12 Part A shows the coarse grained structure of the VP1 including the globular body, C-terminal ligands, and N-terminal ligands. A full description of the individual beads that make up these sections and the potentials and parameters, which control their functionality is described in detail in the main text. Part B shows the addition of the

full C-terminal ligand length upon updating the model. The parameters and potentials for the ‘S’ beads that make up this section of the VP1 are also described in detail in the main text. 191

6.13 Assembly products at different values of ϵ under biological salt conditions ($\lambda_d = 1$ nm) is shown through distributions of bound pentamers to the templates and snapshots of the simulation box. Maximal assembly is observed at $8 k_B T$ whereas no assembly is observed at $6 k_B T$. At $10 k_B T$ kinetically trapped, glassy states form. $\epsilon = 7$ and $9 k_B T$ show intermediate behaviors. 197

6.14 Distributions of the number of bound VP1 pentamers at different N-terminal ligand charge fractions. Simulations are run with $\epsilon = 8 k_B T$, $\lambda_d = 1.0$ nm, and an 11 nm template. The charge fraction of the N-terminal ligand is q_{eff}^2 . As the fraction of charge is decreased, initially little difference is observed, followed by an increase in the polydispersity of $N \geq 10$ particles increases. Nucleation of the particles then becomes limited and eventually, at N-terminal ligand charge fraction of 0.3, the nucleation of particles is completely blocked. 198

6.15 The N=11 capsid forms a nearly perfect icosahedron with one missing vertex. The top of the figure shows this visually whereas the bottom of the figure shows it through the preservation of the icosahedral pairwise distribution from the $T = 1$ case. The important difference between this pairwise distribution and that of the $T = 1$ is the presence of non-zero probability throughout the distribution in the N=11 case (bottom). It

reflects the dynamic movement of the hole whereas the $T = 1$ capsid is static (see main text). 199

7.1 (a) The 1,2-propanediol utilization (Pdu) microcompartment shell is comprised of many different types of alphabetically named shell proteins including hexamers, pseudohexameric trimers, and a pentamer. (b) Secondary structure representations of the shell proteins which are selected for all-atom simulations probing their interactions with one another. (c) An example of the final complete shell in the coarse-grained model containing all three different coarse-grained shell proteins named A, B, and Z (d). The colors in (b) and (d) represent similar interaction strengths in the coarse-grained and all-atom simulations. 207

7.2 Molecular dynamics models of the assembly. (a) Our model consists of cargo beads, C, and three different hexamers denoted by A, B, and Z, each comprised of different types of beads (i.e., S1-S6, where S=A, B, or Z) that are connected in a rigid body. All bead types are equivalent across species except that their interaction energies are determined by parameters named for the two types of hexamers interacting (i.e., U_{AB} , U_{BZ} , U_{AZ}). All possible interactions are shown in the matrices in (d), (e), and (f). This includes interactions with other hexamers of the same type (U_{AA} , U_{BB} , U_{AB}) and cargo interactions, U_{AC} , U_{BC} , and U_{ZC} , which are the last column of (f). In the single hexamer system, shell proteins are simply abbreviated as S and have interaction parameters U_{SS} and U_{SC} . The total shell protein-shell

protein interaction strength is $3*U_{SS}$ in units of thermal energy, $k_B T$, since it is comprised of multiple interactions on the interface. The shell protein-cargo interaction strength is simply U_{SC} in units of $k_B T$ since there is only one interaction as shown. A full list of the forms and of these potentials can be found in the SI. (b) The cargo beads attract each other such that they form a phase separated liquid droplet. This attractive cargo-cargo interaction is constant throughout. (c) An example of a complete shell. (d) Interaction matrices describe all shell-shell and shell-cargo interactions in the simulation for the case of one type of shell protein, (e) two types of shell proteins, and (f) three types of shell proteins. These parameters correspond to the interactions shown in (a).

209

7.3 General results for the single shell protein system. (a) The six different types of structures formed are shown along with which interactions lead to their formation in a phase diagram. The x and y axes are the shell-shell interaction, U_{SS} , and shell-cargo interaction, U_{SC} , respectively. (b) The numbers correspond to the shell quality, Q_{Shell} , which measures the amount of defects (other than 12 pentameric holes) present with 1.00 meaning only those 12 gaps are present. Shells with higher shell-cargo interaction, U_{SC} , tend to have less defects.

212

7.4 Details on one component assemblies which form shells where $Q_{Shell} \geq 0.94$. (a) The size of single component shells. All one

component shells have similar cargo-nucleated assembly dynamics which look like the example shown in (b) for $U_{SC}=7$ and $U_{SS}=3$ 213

7.5 Stochastic binding of sheets to the cargo globule. (a) A typical binding process is shown through simulation snapshots at different time points. At $T=10\tau$, small nucleates can be seen forming in the bulk. At $T=35\tau$, these nucleates are more than large enough to bind the cargo, but none will until $T=50\tau$ when a large sheet randomly diffuses to the cargo. (b) The critical number of shell proteins needed in a sheet, N_{crit} , to bind the globule for sheets with different interactions. This is measured in simulations as the smallest sheet which is every observed to bind a bare cargo globule. (c) The average size of sheets that bind the cargo globule, $\langle N_{Bind} \rangle$. These values tend to be much larger than (b) due to the continued growth of sheets as they randomly diffuse and eventually bind. (d) The average time it takes for the first sheet to bind the cargo, $\langle T_{Bind} \rangle$, as a function of the shell-shell and shell-cargo interaction strengths. 214

7.6 The complex growth kinetics of a two component shell. (a) The shell quality, Q_{Shell} as defined in Equation 7.1, and time to bind the cargo, T_{Bind} , as the shell-cargo interaction of the recruiter, U_{BC} , is varied. There are two distinct regimes. At higher U_{BC} , to the right of the dotted line, there is the cargo-nucleated assembly where high quality shells form by nucleating on the cargo and thus $T_{Bind} = 0$. At lower U_{BC} , low quality shells with many holes form by nucleating in the bulk

and then binding the cargo at finite T_{Bind} . (b) The composition of the growing shell for a representative simulation in the bulk-nucleated regime when $U_{BC}=4$. In the blue shaded region, growth only occurs in the bulk in the form of sheets. A sheet eventually binds (i) due to the strong interaction between the recruiter, B, and the cargo globule. Once this occurs, growth happens *via* elongation of the nucleus on the globule as well as the continued binding of more bulk-nucleated sheets (ii) and (iii). (c) The composition of the growing shell for a representative simulation in the cargo-nucleated regime when $U_{BC}=8$. Nucleation and growth occur immediately on the cargo. (d) The interaction matrix for all interactions in the system. U_{BC} is the free parameter that determines the different growth pathways in (b) and (c). 216

7.7 Kinetics and properties of two component shells with self-interactions shown in (a). The mixing interaction, U_{AB} , is found in the interaction matrix in (b) where U_{BC} is a free parameter. The two components are combined at a 1:1 ratio. (c) As the interaction between B and the cargo, U_{BC} , is increased, the average shell quality, $\langle Q_{Shell} \rangle$ increases. The average shell size and variance, $\langle N_{Shell} \rangle \pm \sigma_{N_{Shell}}$, average fraction of component A in the shell, $\langle X_{A,Shell} \rangle$, and time to bind the cargo, $\langle T_{Bind} \rangle$, all decrease. The components become more mixed as U_{BC} is increased as shown by $\langle \rho_{AB} \rangle$ defined in Eq. 7.4. 220

7.8 Assembly in a three component shell. (a) As U_{BC} is varied, the same bulk-nucleated and cargo-nucleated assembly pathways from the two

component shell are present as in the two component system. At higher U_{BC} ($U_{BC} \geq 8$), there is the cargo-nucleated assembly where high quality (high Q_{Shell}) shells form by nucleating on the cargo. At lower U_{BC} ($U_{BC} \leq 6$), shells nucleate in the bulk as shown by the non-zero time it takes to bind the cargo, $\langle T_{Bind} \rangle$. Thus, the bulk-nucleated regime no longer leads to many holes in the shell when the third component is added. (b) The composition of the growing shell shown for a typical simulation where $U_{BC}=4$. The kinetics can be broken into three steps: (i) nucleation and growth of sheets in the bulk, (ii) a sheet is recruited and binds the cargo, (iii) component Z can add to the stable nucleus on the cargo and help to eliminate any gaps. (C) A visual illustration of the bulk-nucleated pathway with three steps from (b), (i) Nucleation, (ii) Recruitment, and (iii) Completion. (d) The matrix of all interactions used in (a) where U_{BC} is a free parameter. 222

7.9 Assembly in a three component shell. (a) As U_{BC} is varied, the same bulk-nucleated and cargo-nucleated assembly pathways from the two component shell are present as in the two component system. At higher U_{BC} ($U_{BC} \geq 8$), there is the cargo-nucleated assembly where high quality (high Q_{Shell}) shells form by nucleating on the cargo. At lower U_{BC} ($U_{BC} \leq 6$), shells nucleate in the bulk as shown by the non-zero time it takes to bind the cargo, $\langle T_{Bind} \rangle$. Thus, the bulk-nucleated regime no longer leads to many holes in the shell when the third component is added. (b) The composition of the growing shell shown for a typical

simulation where $U_{BC}=4$. The kinetics can be broken into three steps: (i) nucleation and growth of sheets in the bulk, (ii) a sheet is recruited and binds the cargo, (iii) component Z can add to the stable nucleus on the cargo and help to eliminate any gaps. (C) A visual illustration of the bulk-nucleated pathway with three steps from (b), (i) Nucleation, (ii) Recruitment, and (iii) Completion. (d) The matrix of all interactions used in (a) where U_{BC} is a free parameter. 223

7.10 Kinetic phase separation in a three component system. (a) Phase separation of A and Z leads to reduced shell quality. This phase separation is caused by increasing U_{ZC} such that Z can nucleate on the globule and thus is a kinetic phase separation. This is shown in (b), where Z also separates from B and in (c) where B and Z are still mixed due to the increase in U_{BC} such that B quickly integrates into the growing Z nucleus. 225

7.11 A comparison of one, two, and three component shells. (a) The x and y axes are the shell-cargo and shell-shell interactions averaged over the shell as defined in Equations 7.7 and 7.8 respectively. This allows us to plot one, two, and three component shells in terms of only two parameters. We include the phase diagram from Figure 7.3(a) in the background for comparison of shell assembly with one component to assembly with multiple shell proteins. All successful shell assembly falls on one or near one line in the parameter space as shown by the dotted arrow. This arrow overlaps with the regime where high quality shells

assemble (Q_{Shell} approaches 1) in the single shell protein case, while also extending to lower values of $\langle U_{SC} \rangle_{Shell}$ when three hexamer types are present. In other words, multi-component, cargo-encapsulating shells can assemble with a wider range of possible average interactions. In (b), (c), and (d) we show many properties of these assemblies can be described as a continuum by following the dotted arrow. (b) Given a well-formed shell, the assembly pathway is determined by the average interactions of that shell regardless of the number of components. (c) This is shown to be a continuum as described by T_{Bind} , which is 0 for cargo-nucleation, finite for bulk-nucleation and infinite for an empty assembly (Sheets in Bulk, the yellow region). (d) The size of the shell, $\langle N_{Shell} \rangle$, is determined by the relative strength of shell-cargo and shell-shell interactions regardless of the number of shell components. 227

7.12 Atomistic simulations show PduB and PduB' interact differently with PduA and likely play different roles in Pdu microcompartment assembly. Secondary structure representations of the PduA hexamer (a), PduB trimer (b), and PduB' trimer (c) show PduB and PduB' both have the Pdu AB1 interface while the N-terminus region of PduB is lacking from PduB' creating two unique interfaces known as Pdu AB2 and Pdu AB' along with the shared Pdu AB1(d). (e) Potential of mean force calculations show the different interaction strengths of the three interfaces from (d). We show the strongest interaction with PduA comes from Pdu AB2, due to the presence of the highly dynamic N-terminus

region of PduB (red inset). (f) Potential of mean force calculations as a function of the bending angle, θ_B , show energetically preferred bending orientations of the interfaces. The bending interactions at these interfaces are quite similar with some added flexibility for Pdu AB2 due to the presence of the N-terminus region. 231

7.13 Representative micrographs of Pdu MCPs prepared using various techniques. (a) Pdu MCPs prepared using standard negative staining techniques allows for Pdu MCP collapse as indicated by pooled staining at their center. (b) Pdu MCPs buffer exchanged to prevent collapse appear more inflated by TEM as indicated by reduced internal pooled staining. (c) Micrographs of Pdu MCPs in solution prepared using cryogenic methods appear the most inflated. Scale bars = 100 nm in each case. 233

7.14 The geometry of the rigid body hexamers in distance units. The parameters given are consistent across all types of hexamers, A, B, and Z in the main text. The rigid body consists of volume excluding X beads in addition to other beads which handle attractive interactions. These beads are named as S1 - S6 where S is A, B, or Z when the hexamer is a specific type of hexamer. Here they are named S to show the generality. The effective radius of the hexamer, r_{hex} is the distance from the center of geometry to the center of the S6 bead. 236

7.15 The relative probability of observing connections between two different species, ρ_{ij} , is defined in Equation 7.4. Here we show that AB

connections are far more probable than random mixing due to the strong U_{AB} . This agrees with our regular solution analysis. 242

7.16 The relative probability of observing connections between two different species, ρ_{ij} , is defined in Equation 7.4. These values that correspond to mixing (ρ_{AB} , ρ_{AZ} , and ρ_{BZ}) mostly follow our regular solution analysis. The one exception is for $U_{BC}=4$ and $U_{ZC}=6$. Here B and Z separate despite our predictions made based on their interactions (see Figure 7.10). We also include one data point where we vary composition (keeping $U_{BC}=6$ and $U_{ZC}=6$) and see no change to our relative connection probability for different species, showing that the normalization is reasonable. See Figure 7.17 for further details. 243

7.17 Additional data that are not reported anywhere else and so we include them for completeness. 244

7.18 Potential of mean force for interface Pdu AB2. The difference between this and Figure 7.12(d) is that we restrain all of the PduB backbone including the N-terminal. This shows that the interaction is strong even with reduced degrees of freedom for the N-terminal. Restraining all of PduB as we do here is more similar to methods used for interface Pdu AB1 and in previous work [1, 5]. However, we consider Figure 7.12(d) to be more accurate since the additional degrees of freedom of the N-terminal should not be neglected. 245

CHAPTER 1

Introduction

Polymers, or macromolecules, are composed of long strands of smaller molecules called monomers held together by chemical bonds [6]. This thesis explores how structural properties of different macromolecules affect encapsulation or compartmentalization of other molecules to achieve a specific function or application. Thus, the introduction begins with a discussion of the structure of individual macromolecules in Chapter 1.1. Chapter 1.2 introduces disordered polyelectrolyte complexes of oppositely charged, flexible heteropolymers that absorb certain species due to locally ordered hydrophobic domains. These materials are the subject of Part 1: Chapters 2, 3, and 4. In contrast, assembly of proteins with defined single chain structure is introduced in Chapter 1.3. They form shells which encapsulate enzymatic reactions as discussed in Part 2: Chapters 5 and 7 or genetic information, which could potentially be used for gene therapy as discussed in Chapter 6.

1.1. Structure of Individual Macromolecules

We first consider an ideal, flexible polymer. The conformation of an ideal polymer chain is modeled by a random walk, which defines the size of the polymer as a probability distribution rather than describing the position of individual atoms or monomers relative to one another [7]. This is because there is a time scale at which the positions of monomers in one configuration are completely decorrelated from the positions in the

original configuration [8]. These models can break down when the behavior of the polymer is not purely entropic, meaning the monomer units comprising the polymers have interactions such as being charged or having heterogeneous solvent interactions. This can lead to structures such as rod-like polyelectrolyte chains [9] or other exotic structures [10] in polymers containing chemically different groups.

One prominent example of this breakdown is the case of proteins [11]. Proteins are biological polymers comprised of a sequence of 20 unique amino acid monomers with a common peptide backbone. They often fold into well-defined structures and carry out some function within a cell. The structure of folded proteins is hierarchical, with the lowest level of organization being the primary structure. This is the sequence of amino acids comprising the protein. The secondary structure of a protein includes locally ordered structures known as α -helices and β -sheets. The tertiary structure of a protein is the 3-dimensional arrangement of these local structures relative to one another. After the tertiary structure, the final level of structure that defines a protein is its quaternary structure. This refers to the organization of two or more (in this thesis identical) polypeptide chains which interact to form a stable dimer, trimer, or other oligomer. Since the structure of the protein is well defined, there is no relevant time scale over which monomer positions become decorrelated. Instead, one can measure the stability of a folded protein structure through the small deviations in protein structure relative to a given structure, which is often obtained *via* x-ray crystallography. These deviations are quantified by the root mean square deviation [12] (RMSD) of the protein which is lower when the structure is more conserved.

The way a protein folds is often thought to completely determine its function. Thus, the ability to predict and affect how an amino acid sequence will fold is a major goal in the field of biotechnology [13]. However, in many cases some regions or the entirety of a protein will lack a clearly defined structure and this lack of folding can also be important for protein function [14]. These proteins are known as intrinsically disordered proteins or proteins containing intrinsically disordered regions if the protein is partially folded. These intrinsically disordered proteins and regions can act like flexible polymers and often include groups that interact with other proteins or other biological macromolecules. A protein can contain many intrinsically disordered regions, especially when it is an oligomer made of multiple copies of the same peptide chain containing intrinsically disordered regions. In this work, examples of 6-sided pseudo-hexameric trimers and 5-sided pentamers with intrinsically disordered regions will be shown and the function of these flexible regions will be investigated with the goal of understanding the functionality provided by a lack of protein structure.

1.2. Assemblies I: Disordered Polyelectrolyte Complexes

In this section, the relevant background for Part 1 is introduced. It begins by discussing the relevant interactions and general behavior of flexible polymers in complex phases such as polyelectrolyte complexes before going on to discuss the specifics of how they are used in Chapters 2, 3, and 4.

Due to their previously mentioned flexibility, synthetic polymers tend to exist in disordered or semi-crystalline states [15]. This is partly due to slow reptation of flexible polymers in dense systems, which leads to kinetic trapping, and partly due to the entropic

cost of limiting polymer chain conformations when the polymers are flexible. However, even completely non-crystalline polymers can exhibit order when there is more than one type of monomer present. Block copolymers are a great example of this, as the tendency for the chemically different blocks to segregate combined with the constraint of chain connectivity can lead to the formation of a wide range of microphases [16]. When monomer sequences are random instead of blocky, these interactions are predicted to lead to the formation of infinitely many phases [17]. Polymer phase behavior can become even more complex when interactions are long-range, that is, with the addition of charged monomers and their counterions. In a dry system, charged clusters will form and their specific structure will be affected by monomer sequence, among many other factors [18, 19]. However, these materials are also especially good at absorbing water due to the solvation of the charged groups and are often used as swelling hydrogel materials in diapers [20]. Their ability to swell in acidic conditions highlights the importance of pH in systems containing weakly charged monomers [21]. One can also add an oppositely charged polyelectrolyte to create a polyelectrolyte complex [22]. Polyelectrolyte complexes form due to the electrostatic interactions between the oppositely charged polymer chains and the entropic gain of the polymers releasing their counterions to the solution as their charge is cancelled by the oppositely charged polymer. This process is affected by all of the factors mentioned thus far including ionic conditions [23, 24, 25], monomer sequence or “blockiness” [18], rigidity of the polymers [26], and solvent quality [27].

One can also include components in these polyelectrolyte complexes which are not classical synthetic polymers. Some examples include charged micelles made of many lipid molecules [28] and biological enzymes [29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39].

These components can add further complications as they have well-defined 3-dimensional structures in contrast to semi-flexible polymers. This means that the concept of sequence blockiness can be, to some extent, replaced by the idea of chemical domains that exist on the surface of a protein, especially charged domains [40]. Investigations of polyelectrolyte complex structure, have found that flexible polymers can “self-optimize” with the surface of folded proteins leading to more contacts between certain surface patches on the protein and certain groups in the polymer sequence [41]. Similarly, blocky charged sequences found in intrinsically disordered proteins and regions are quite relevant to polyelectrolyte complex formation like in the case of purely synthetic polymers [42] despite not being found in a well-defined surface region. In biology, intrinsically disordered proteins and other charged biomolecules including RNA form polyelectrolyte complexes known as membraneless organelles. These membraneless organelles are phase separated compartments [43, 44], which increase reaction rates and help to avoid promiscuous enzyme behavior [45, 46] by controlling the localization of different proteins with given substrates and cofactors [47]. Thus, the role of these membraneless organelles is to spatiotemporally organize certain biochemical components.

Part 1 of this work will investigate how flexible components lead to macroscopically disordered systems that still retain local structure which is critical to their ability to compartmentalize other molecules. The first chapter of Part 1 (Chapter 2) investigates the assembly of a polyelectrolyte complex which mimics a membraneless organelle by using cationic and anionic random copolymers with hydrophobic, hydrophilic, and charged monomers, similar to the chemical diversity of naturally occurring amino acids. The anionic copolymer by itself will form single micellar structures with a hydrophobic core and a

hydrophilic corona in solution, absorbing a wide range of charged and uncharged contaminants from water including PFAS, which can be harmful to human health [48]. Adding the cationic polymer aggregates these originally stable structures into a macroscopic complex, which simulations show still contains hydrophobic domains like the micellar structures in solution. Interestingly, complex formation requires the charge on the polymers in these complexes to be unbalanced with three times more charge on the cationic polymer. The same effect is observed in simulations which show that the hydrophobic interactions between oppositely charged polymers are a major contributor to the ability of the cationic polymer to aggregate the anionic copolymer. Simulations also show how the hydrophobicity and charge both contribute to the absorption of these contaminants, especially the heterogeneity of the charge on the polymers within the complex.

This dissertation then investigates whether these polyelectrolyte complexes could be used to absorb a wider variety of contaminants. Since nanoplastics are also of increasing concern as a water soluble contaminant [49], Chapter 3 details the design of a complex similar to Chapter 2, but optimized to remove polyethylene terephthalate (PET) nanoplastics from water. To this end, anionic random copolymers with different amounts of hydrophobic groups were tested. In simulations of a single polymer and a PET surface, the lowest and highest hydrophobicity polymers make the most contacts with the surface. This is due to conformational changes in the polymer itself and agrees well with experiment. However, when these anionic random copolymers are added to a polyelectrolyte complex as in Chapter 2, there is not much absorption of nanoplastics for low hydrophobicity polymers. Increasing hydrophobicity changes the structure of the complexes as the more hydrophobic groups become less shielded from the solvent by hydrophilic groups

and hydrophobic domains within the complexes grow. Because they are embedded in the complex, low hydrophobicity polymers are not found in the extended conformations that allow them to bind the PET surface in the single polymer case. This leads to the most hydrophobic polymers absorbing nearly 100% of the nanoplastics, while the lowest hydrophobicity polymers absorb only 50%.

The complexes in Chapters 2 and 3 did not include any catalytic enzymes, which are among the most important components of membraneless organelles. Thus, Chapter 4 investigates the interaction between an enzyme known as PETase [50] and our anionic random copolymer in a nanoscale aggregate [32] rather than a macroscopic complex. This includes the ability of the anionic copolymer to stabilize the conformation of the protein at higher temperatures, which is important because PETase breaks down polyethylene terephthalate (PET) [51], one of the world’s most common plastics. PET has a glass transition above the temperature where PETase unfolds and becomes inactive [52]. Thus, for the reaction to occur in an efficient manner, the PETase must be stable at a higher temperature. This could allow for large-scale upcycling [53] of PET, keeping it from becoming harmful waste [54]. Simulations of the polymers with PETase showed how charge and hydrophobicity influence how much contact there is between the enzyme and the random copolymer by varying the mean charged and hydrophobic monomer fractions. Including charged groups lead to more contact between the polymer and protein and that these extra contacts were almost exclusively located on the positively charged sections of the surface of PETase. This is the “self-optimization” of a flexible polymer on the surface of a folded protein. However, the contacts near the enzyme’s active site lead to the greatest stabilization of the active site of PETase, which should in turn stabilize

catalytic function at elevated temperatures. Experiments also investigated the activity of the PETase-polymer complex on small molecule substrates and PET films at elevated temperatures, finding increased activity relative to the enzyme by itself in both cases.

1.3. Assemblies II: Crystalline Protein Shells

In this section, the relevant background for Part 2 is introduced. The section begins by discussing the general requirements for shell closure in microcompartments and viruses, detailing how they can be met in different ways in Chapter 5 and Chapter 6 respectively. Then, the multi-component nature of microcompartment shells is addressed along with the approaches taken in Chapter 7 to understand its ramifications.

Part 1 investigates how enzymes and their substrates are organized into membraneless organelles, helping to drive cell metabolism by catalyzing biochemical reactions. This contrasts with classical organelles which are separated from the cytoplasm by lipid bilayer membranes. Interestingly, bacteria have evolved another sub-cellular structure known as the bacterial microcompartment, which also accomplishes this task of segregating enzymes with the relevant substrates and reaction cofactors [55]. Unlike lipid membranes, which are considered to be liquid membranes due to the two-dimensional liquid-like diffusion of individual lipids, microcompartments have a more rigid crystalline structure that does not allow for two-dimensional rearrangement. These microcompartments are made of folded shell protein hexamers and pentamers, which surround a core made of enzymes. The enzymes carry out reactions that create energy and other materials crucial to the survival of the bacteria including CO₂ fixation [56]. The shell is selectively permeable

and thus allows substrates and products to enter or leave the compartment while intermediate chemical species are kept inside [57]. This localization of all relevant components increases reaction rates and protects the rest of the bacteria from potentially toxic chemical intermediates. Since these shells provide a diffusive barrier, which is crucial to the multi-step enzymatic reactions performed in the compartment, engineers seek to repurpose microcompartments as modular nano-bioreactors. They could potentially be used to fixate CO₂ outside of cells or to provide a diffusive barrier for enzymatic reactions that they do not naturally encapsulate. However, redesigning microcompartments is difficult without a fundamental understanding of the physical rules that govern the kinetics of assembly and the final structures formed by the shell proteins.

The assembly of many copies of 6-sided hexamers and 5-sided pentamers into a closed shell is quite similar to viral assembly [58, 59] and thus concepts developed for viral assembly [60, 61, 62, 63] can also apply to microcompartments. One such concept is the need for 12 5-sided pentameric shell proteins in order to form a closed structure when only hexameric and pentameric subunits are present as shown by the Euler theorem [64]. This is important for a microcompartment shell to provide a tight seal on the enzymatic reactions taking place inside. Another is the role of mechanical properties in determining assembly pathways and equilibrium structures [65, 66, 67]. Thus, Part 2 begins by investigating how the bending potential, defined as the relative free energy between two shell proteins as a function of the angle formed at the interaction interface, impacts the closure of crystalline shells. The depth of this bending potential (or lack thereof) can be thought of as another type of protein flexibility. However, it is an emergent property of two proteins rather than a property of a single protein. The first chapter of Part 2

(Chapter 5) identifies how different bending interactions between pairs of hexamers and pentamers leads to the formation of different unclosed structures when the pentamers necessary for closure are not included in the assembly. It starts with the specific example of the 1,2 propanediol utilization (Pdu) microcompartment and its pentamer PduN, along with a major hexamer PduA. This chapter provides experimental evidence that in the absence of the pentamer, microtubes form instead of a polyhedral microcompartment. Morphological changes are also shown to impact the performance of the multi-step reaction pathway encapsulated by the compartment/tube. The transition from compartment to tube is related to the strong bending potential and intrinsic bending angle between the pentamer and hexamer, which seems to dominate the morphology of the shell. The total bending interaction between PduA and PduN is shown to be caused by one short range interaction between amino acids that is only possible when the interface is highly bent. The Pdu system is then compared to another microcompartment system which, in the absence of the pentamer, forms its original crystalline shell structure with vacancies at the pentamer sites. This system is dominated by hexamer-hexamer bending interactions, explaining why the pentamer is not needed to induce the highly bent microcompartment morphology.

There are also many differences between bacterial microcompartment and viral assembly. One difference is the typical cargo the assembly occurs around. In microcompartment assembly, the cargo is normally a collection of folded enzymes, which aggregate to form the globule [68] around which shell assembly occurs. There are some exceptions where these enzymes are flexible intrinsically disordered proteins [69]. In viruses, flexible cargo is the norm. The cargo is often a single nucleic acid polymer or other polyelectrolyte, and

its flexibility is seen as very important to the assembly rate. However, in some biomedical imaging applications, viral assembly is done on a rigid nanoparticle template [70] and some shell proteins actually have flexible intrinsically disordered proteins built in [60, 71]. Using coarse-grained molecular dynamics, Chapter 6 investigates this interesting case of assembling pentamers where the flexibility is found on the shell proteins instead of the cargo. In this case, the flexibility of the shell proteins can accelerate the dynamics and make up for the lack of flexible cargo. This flexibility allows for closed shells with more than 12 pentamers, as the excess pentamers can act as 6-coordinated hexamers due to their flexibility. Thus, these shells still satisfy the Euler theorem.

Another difference between microcompartment assembly and viral assembly is the number of types of shell proteins present. Most viral assembly only includes one or two types of shell proteins while some MCP systems have as many as 8 different shell protein types. This has been shown to be crucial to their function, as deletion of just one or two shell proteins [72, 73, 74] can lead to failed assembly of the shell. This design strategy, where multiple components with different interactions form larger structures is actually typical of biological systems [75, 76], especially membraneless organelles [77], which inspired the polyelectrolyte complexes in Part 1. However, since much of the previous modeling has been focused on shells featuring only one type of shell protein or one hexamer and one pentamer [69], understanding these deletion experiments requires updated models, which include the multi-component nature of the assembly. Chapter 7 addresses the multi-component assembly problem by considering a system of three hexamers with different interaction energies between all possible pairs of hexamer types, as well as with the cargo. This chapter will show how the regime of interaction strengths

that leads to assembly of the shell is widened when there are more components with a diverse array of interaction strengths. This is made possible by a bulk-nucleated assembly mechanism, which is only available to multi-component shells. It is normally thought that cargo-encapsulating shells must nucleate with the cargo [78] and that this requires fairly strong interactions between shell and cargo (shell-cargo interactions) [79, 61, 68] and fairly weak interactions between shell proteins (shell-shell interactions) [69]. However, in the bulk-nucleated pathway, nucleation of the shell and cargo globule occur separately as has been hypothesized, but not fully demonstrated in experiment [74]. The total interaction of the immature shell with the cargo globule continues to grow as more shell proteins are added until it eventually binds and encapsulates the cargo due to the incorporation of a second type of shell protein. This type of shell protein is known as the recruiter and it has stronger shell-cargo interactions. This mechanism also requires a third type of shell protein which has weaker shell-shell interactions and is thus able to avoid kinetic traps and fill in gaps in the shell. The varied interactions necessary for the bulk-nucleated pathway are then shown in the Pdu system using atomistic simulations on various Pdu shell protein interfaces. This includes a 6-sided pseudo-hexameric trimer with flexible intrinsically disordered regions whose dynamics are likely important to the assembly as in Chapter 6.

Part 1

Disordered Polyelectrolyte Complexes

CHAPTER 2

Heterogeneous Charged Complexes of Random Copolymers for the Segregation of Organic Molecules

This chapter is based on the published work [80] of Jeremy Wang, Curt Waltmann, Han Umaña-Kossio, Monica Olvera de la Cruz, and John M. Torkelson, **Heterogeneous Charged Complexes for the Segregation of Organic Molecules**, *ACS Cent. Sci.*, 2021, 7, 5, 882-891, with modified details.

2.1. Abstract

Nature harnesses the disorder of intrinsically disordered proteins to organize enzymes and biopolymers into membraneless organelles. The heterogeneous nature of synthetic random copolymers with charged, polar, and hydrophobic groups has been exploited to mimic intrinsically disordered proteins, forming complexes with enzymatically active proteins and delivering them into nonbiological environments. Here, the properties of polyelectrolyte complexes composed of two random copolymer polyelectrolytes are studied experimentally and *via* simulation with the aim of exploiting such complexes for segregating organic molecules from water. The anionic polyelectrolyte contains hydrophilic and hydrophobic side chains and forms self-assembled hydrophobic domains. The cationic polymer is a high-molecular-weight copolymer of hydrophilic and charged side groups and acts as a flocculant. We find that the polyelectrolyte complexes obtained with this anionic and cationic random copolymer system are capable of absorbing small cationic, anionic,

and hydrophobic organic molecules, including perfluorooctanoic acid, a compound of great environmental and toxicologic concern. Importantly, these macroscopic complexes can be easily removed from water, thereby providing a simple approach for organic contaminant removal in aqueous media. Martini and coarse-grained molecular dynamics simulations explore how the microscale heterogeneity of these random copolymer complexes relates to their segregation functionality.

2.2. Introduction

Random copolymers have a statistical distribution of two or more types of monomers, leading to spatial heterogeneity in local composition as different regions of a copolymer chain may have different average composition. This type of heterogeneity from disordered polymer sequences is thought to be important in achieving biomimetic functions such as molecular-scale pattern-matching[**81, 82, 83, 84**]. Membraneless organelles, which are spatiotemporal aggregates of nucleic acids, enzymes and their substrates, and oppositely charged, intrinsically disordered proteins with rather random sequences of amino acid monomers, [**45**] likely utilize such concepts of disorder and heterogeneity. Because these membraneless organelles are analogous to polyelectrolyte complexes of oppositely charged random copolymers, the behavior of such complexes could provide insight into the behavior of membraneless organelles. This analogy has inspired research into the use of synthetic random copolymers to interact with enzymes, forming what can be considered to be a type of polyelectrolyte complex. [**85, 86**] Concentrating small-molecule substrates is also a crucial function of membraneless organelles. Here, we explore the possibility of using polyelectrolyte complexes of random copolymers as mimics of disordered proteins in

membraneless organelles with an aim to segregate small organic molecules from aqueous solution.

Polyelectrolyte complexes are generally formed when oppositely charged polymers are mixed in aqueous solution. [87, 88] Depending on factors including charge ratio, degree of polymerization, monomer sequence, [18, 42] ionic solution conditions, [23, 24, 25, 89] or solvent quality, [27] a wide range of phase behaviors for the polyelectrolyte complex can be observed, including the formation of colloidal suspensions, liquid coacervates, and solid precipitates. [90, 91, 92, 93] Colloidal suspensions of polyelectrolyte complexes have been investigated for their ability to encapsulate bioactive molecules and deliver such molecules in a biological environment.[94, 95, 96] Liquid coacervates of polyelectrolyte complexes have been shown to encapsulate and concentrate enzymes from solution, [97, 98, 99] similar to the capabilities of membraneless organelles.

The formation of solid precipitates with polyelectrolyte complexes can be particularly useful in separating particles from aqueous solution and removing contaminants from water through a flocculation process. [100, 101] Traditionally, flocculation is used to remove negatively charged colloids such as fine clay particles from water *via* the addition of a single species of a high-molecular-weight cationic polymer, which neutralizes the surface charge of the particles, acts as a bridge between them, and coagulates the particles into macroscopic flocs. [102, 103] However, polyelectrolyte complexes have also been used for flocculation purposes, [100] and solid polyelectrolyte complexes can also be effective at removing ionic compounds such as metal ions or charged organic compounds from water. [104]

This flocculation behavior provides a relatively simple experimental approach to measure the segregation of organic molecules into polyelectrolyte complexes. There have been studies demonstrating the ability of polyelectrolyte complex coacervates to partition and segregate small organic molecules, [28, 105, 106, 107] and separating the coacervates from the supernatant generally requires centrifugation techniques. In contrast, solid polyelectrolyte complexes that segregate organic molecules can be removed from solution through simple filtration. Segregation efficiencies can then be obtained by measuring the concentration of organic molecules in the filtered solution.

Building upon a rational design principle outlined in earlier work involving complexes of random copolymers and enzymes, [85] we hypothesize that a random anionic copolymer with hydrophilic, hydrophobic, and anionic monomers will form micellar-like structures in aqueous solution and flocculate with a random cationic copolymer, with the heterogeneity of the resulting complex providing favorable interactions with a wide range of organic molecules. Previously, others have shown that random copolymers with hydrophilic and hydrophobic groups exhibit protein-like folding and form single-chain micelles while segregating dye molecules in aqueous solution. [108] It is reasonable to expect that the addition of an anionic component would allow the three-component copolymer to form similar structures and encapsulate organic molecules, with the additional benefit of being able to remove the copolymer and dye in a flocculation-like process, which may enable applications in industrial water remediation.

Here, we develop a method to segregate and remove organic molecules from water using two oppositely charged random copolymers through experiments, simulations, and analysis. The anionic copolymer is comprised of hydrophilic, hydrophobic, and anionic

methacrylate groups. (Figure 2.1a) The cationic copolymer is composed of hydrophilic and cationic methacrylate groups. (Figure 2.1b) These random copolymer polyelectrolytes were synthesized using free radical polymerization and form macroscopic complexes when mixed (Figure 2.1c), successfully encapsulating several organic dyes with varying degrees of effectiveness. Three dyes, crystal violet, methyl orange, and phenolphthalein, were chosen as model molecules for their respective cationic, anionic, and hydrophobic natures as well as ease of quantification *via* UV-vis spectroscopy. Perfluorooctanoic acid was also chosen to demonstrate the relevance of this system in filtering difficult-to-remove chemicals from aqueous systems (Figure 2.1d).

We use coarse-grained molecular dynamics at two different length scales to study how the heterogeneous nature of these random polyelectrolyte complexes affects their ability to flocculate dyes. The simulations show complexes that are highly heterogeneous in composition with hydrophobic domains as well as heterogeneities in the charge distribution throughout the complexes. We explain the origin of these heterogeneities using statistical analysis that has been used previously to explain compositional heterogeneities observed in strongly incompatible random copolymers [17] and in random ionomers, which are molten state (dry) systems. [109, 110] Models of the dyes are also included in the simulations (Figure 2.1e), and we analyze the roles that hydrophobicity and charge play in the removal of the dyes.

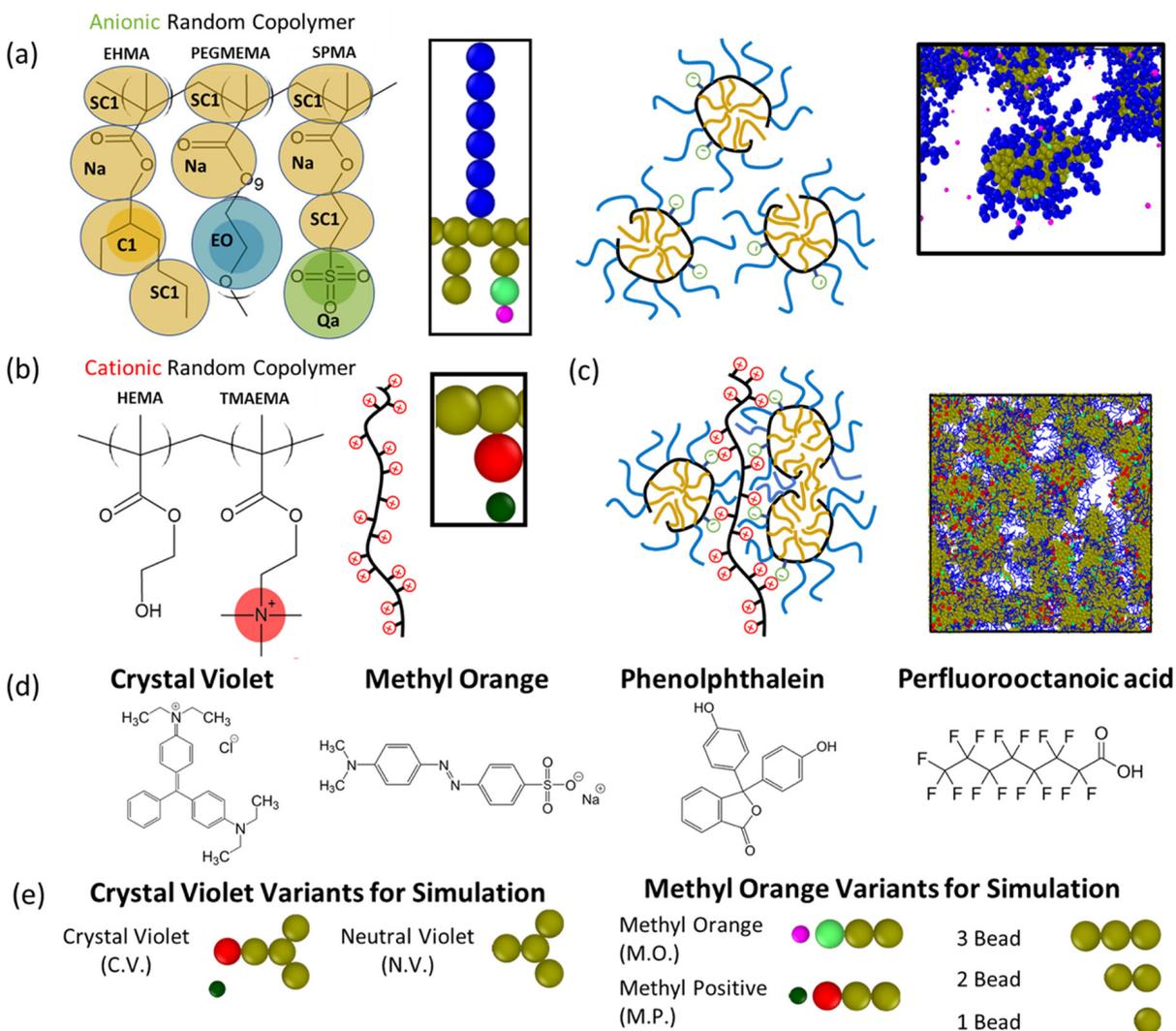


Figure 2.1. (a) Chemical structure, Martini parameterization, coarse grain model, illustration, and simulation snapshot of the anionic random copolymer. (b) Chemical structure, illustration, and coarse grain model of the cationic random copolymer. (c) Illustration and simulation snapshot of complexation between the anionic random copolymer and the cationic random copolymer. (d) Chemical structure of organic molecules used in experiments. (e) Coarse grain model of organic molecules used in simulations. Red and green beads correspond to positive and negative charges respectively, while tan beads are hydrophobic, and blue beads are hydrophilic.

2.3. Results and Discussion

2.3.1. Polymerization and Characterization of Random Copolymers

We synthesized the anionic and cationic copolymers *via* free radical polymerization. Aqueous size exclusion chromatography was used to determine apparent weight-average molecular weight (M_w), apparent number-average molecular weight (M_n), and apparent dispersity (M_w/M_n) values (Table 2.1). The values are apparent, as polymers that form hydrophobic domains can exhibit intermolecular aggregation in aqueous media *via* hydrophobic interactions. [111] This aggregation behavior can be confirmed for the anionic copolymer, as we obtained higher apparent M_w values with higher concentrations of polymer solution. (See the SI of the original manuscript [80], Table S4.) This interpolymer aggregation behavior likely explains the low measured dispersity of 1.1, which is significantly different from the dispersity of roughly 2 that we would expect for polymers produced by free radical polymerization of methacrylate monomers. [6] We note that such a very low apparent dispersity has also been observed in another charged polymer system with interchain aggregates that was synthesized by free radical polymerization. [112]

Table 2.1. Copolymer Characterization: Mole Fractions, Apparent Average Molecular Weights, Apparent Dispersity, and Apparent Average Degrees of Polymerization

	anionic random copolymer	cationic random copolymer
component 1 mol fraction	PEGMEMMA: 0.51	HEMA: 0.54-0.6
component 2 mol fraction	EHMA: 0.44	TMAEMA: 0.4-0.46
component 3 mol fraction	SPMA: 0.05	N/A
apparent M_w (g/mol)	290 000	10 000 000
apparent M_n (g/mol)	260 000	4 300 000
apparent dispersity	1.1	2.3
apparent DP_w	820	60 000 - 62 000
apparent DP_n	760	26 000 - 27 000

The cationic polymer also shows an anomalously high Mw value for a polymer synthesized by free radical polymerization, which is likely due to the fact that the hydroxyethyl methacrylate (HEMA) monomer used in the cationic polymer is susceptible to effects of chain transfer to the polymer and monomer and may act as a branching unit. [113] Thus, it is likely that the cationic polymer is highly branched in structure. This branching may be beneficial in the complexation process, as some studies suggest that highly branched flocculants exhibit better flocculation performance. [103] However, quantifying the degree of branching in polymers is not a simple process [114] and is not explored further in this study. We note that the dispersity of this polymer was measured to be 2.3, which is in line with expected values.

We analyzed the copolymer compositions *via* ^1H NMR spectroscopy. (Peak assignments are shown in SI of the original manuscript [80], Figures S3 and S4.) The anionic polymer has a molar composition of 51% polyethylene glycol methyl ether methacrylate (PEGMEMA), 44% (ethylhexyl methacrylate) (EHMA), and 5% sulfopropyl methacrylate (SPMA). From this information and the apparent Mw, we can calculate an apparent weight-average degree of polymerization (DPw) of 820. The peak assignments for the cationic copolymer cannot be exactly determined without knowing the branching ratio of the polymer, but upper and lower bounds can be determined for the strictly linear case and strictly branched case (one branch per HEMA monomer). Thus, a reasonable estimate for the molar composition of the cationic polymer is 54–60% HEMA and 40–46% TMAEMA, with an apparent DPw between 60 000 and 62 000. From this analysis, we can conclude that the cationic copolymer has a substantial charge fraction and is much

longer than the anionic copolymer, potentially making it an effective flocculant. The anionic copolymer has a significant hydrophobic composition while being slightly charged. See Table 2.1 for a summary of copolymer characterization.

2.3.2. Polyelectrolyte Complex Formation and Dye Filtration

We formed solid polyelectrolyte complexes by mixing 300 μL of 66 ± 2 mg/mL aqueous anionic copolymer solutions and 220 μL of 15.0 ± 1.5 mg/mL aqueous cationic copolymer solution in 10 mL of distilled water, which leads to a 6 to 1 ratio by weight of anionic copolymer to cationic copolymer. Initially, the mixture of copolymer solutions turns turbid and cloudy, indicating that polyelectrolyte complexes have grown to a size comparable to the wavelength of visible light. In less than a minute, macroscopic flocs can be observed, indicating the complexes have favorable interactions and a strong tendency to aggregate and coalesce into larger and larger structures. However, we also observe that the solution tends to be slightly turbid after macroscopic flocculation, indicating that there are colloidal polyelectrolyte complexes remain in solution. These are likely charge-stabilized colloids, as there is an excess of positive charge in the complexes. At this point, we add 20 μL of a 50 mg/mL magnesium sulfate solution, and this addition appears to coagulate the remaining polyelectrolyte complexes within a few minutes and leaves the solution clear. We believe this coagulation process is analogous to how multivalent cations are used to coagulate anionic colloids from solution. [115] The final aggregate sizes are usually on the order of millimeters and are robust to mechanical perturbation. When the mixture is stirred with a magnetic stir bar, the aggregates do not break apart even at stirring speeds exceeding 1000 rpm. After filtration through a 0.22 μm membrane filter, the measured

solid concentration in the filtered solution is 0.17 ± 0.02 mg/mL. This concentration corresponds to a polyelectrolyte complexation efficiency of 92%.

The amounts of copolymer solution that we mixed in the above description are determined by using a titration procedure. Starting with an initial mixture of distilled water and anionic copolymer solution, corresponding to a 10.3 mL solution containing 1.90 ± 0.05 mg/mL of anionic copolymer, we add the 15.0 ± 1.5 mg/mL cationic polymer solution in 20 μ L increments. We consistently find that macroscopic flocculation occurs at 220 mL of cationic solution added, which may correspond to a sort of equivalence point. However, this is not a traditional equivalence point for polyelectrolyte complexes, as the molar ratio of positive charges to negative charges of the complexes is not 1:1 but has a significant excess of positive charge with a ratio of 2.9–3.3:1. This amount of copolymer solution added to form macroscopic complexes does not change when adding dyes or contaminants at a concentration of 2 μ g/mL, except for the case of phenolphthalein, where 240 μ L of cationic solution was needed for flocculation. This difference is likely due to a slight salt concentration of about 1 mM NaCl from the preparation procedure and suggests that there is a salt concentration dependence on the formation of these complexes, which is normally observed in systems of aggregating polyelectrolyte complexes. [116] This effect may be explored further in a future study.

The removal efficiency for the dyes is determined by comparing the peak visible light absorption of the filtered samples with complexes removed to a calibration curve from stock solutions of the dye. We find that the removal of crystal violet, the cationic dye, is quantitative with a single filtration removing over 99.5% of the dye, reaching the detection limit of the instrument used. We obtain similar results for phenolphthalein, a

hydrophobic dye, with a removal efficiency of $> 98\%$. It should be noted that filtration experiments for phenolphthalein were performed in its colorless, neutral form, whereas quantification experiments were performed in its colored, charged form. The removal efficiency with a single filtration of methyl orange, the anionic dye, is $65 \pm 5\%$. We determine the removal efficiency of PFOA in a manner similar to the dyes, except using liquid chromatography with mass spectrometry using electrospray ionization. A value of $63.0 \pm 0.5\%$ is obtained for a single filtration. We also perform repeated filtrations for a sample of perfluorooctanoic acid, with the process of adding anionic copolymer solution, and then cationic copolymer solution and magnesium sulfate being repeated twice for a total of three filtrations. In total, $89.0 \pm 0.5\%$ of the perfluorooctanoic acid was removed in this experiment, demonstrating that this system can significantly reduce the concentration of environmentally relevant contaminants from aqueous systems. These results are shown in Figure 2.2(a).

2.3.3. Confirmation of Dye Encapsulation and Micelle Formation in an Anionic Copolymer

Crystal violet and methyl orange are solvatochromic dyes, exhibiting visible absorbance spectral shifts with changes in the hydrophobicity of the local environment. [117] We leverage this behavior to obtain information on the interactions of the dyes with the copolymer and resulting complex. Figure 2.2(b,d) shows that solutions of crystal violet exhibit a solvatochromic red shift when mixed with small amounts of anionic copolymer solution, with a peak absorbance shift from 593 to 598 nm. As the small amount of copolymer added does not change the overall polarity of the solvent, the crystal violet

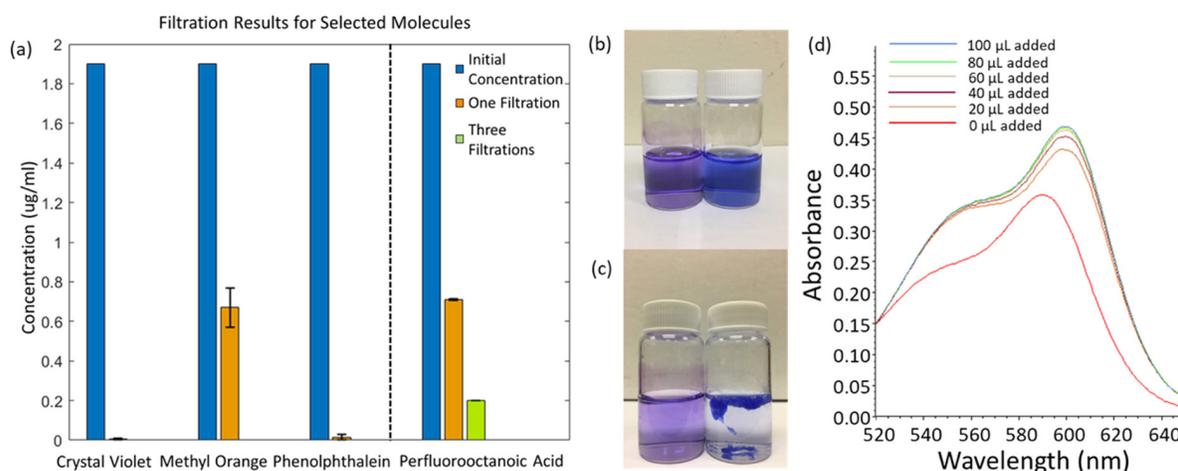


Figure 2.2. (a) Filtration results. For each of the three dyes, the results are averages from three separate filtration samples. Crystal violet and phenolphthalein are quantitatively removed. For perfluorooctanoic acid, results for one and three filtrations on a sample of perfluorooctanoic acid are shown. Error bars are standard deviations from three runs of a single sample. (b) Images of 2 $\mu\text{g/ml}$ aqueous solutions of crystal violet before (left vial) and after (right vial) addition of 100 μL of anionic copolymer solution (66 mg/mL). (c) Images of 2 $\mu\text{g/ml}$ aqueous solutions of crystal violet before (left vial) and after (right vial) encapsulation in a polyelectrolyte complex of anionic and cationic copolymer. (d) Visible absorbance spectra of a 2 $\mu\text{g/ml}$ solution of crystal violet in water as a function of added anionic copolymer solution (66 mg/mL). A significant solvatochromic shift is observed upon addition of trace levels of copolymer solution.

must be interacting strongly with the local hydrophobic domains of the anionic copolymer. This spectral shift is similar to the shift shown when anionic micelles of sodium dodecyl sulfate are formed in solution with crystal violet. [118] This shift is retained when complexes of the anionic and cationic copolymers are formed as shown in Figure 2.2(c), indicating that crystal violet is located near the hydrophobic pockets that exist within the polyelectrolyte complex. No solvatochromic shift is observed in absorbance spectrum when anionic copolymer is added to a solution of methyl orange nor does the resulting complex exhibit a visual color shift. These results indicate that methyl orange does not

interact strongly with the hydrophobic domains of the polyelectrolyte complex, possibly due to a weaker hydrophobic character and/or the same charge repulsion from the anionic copolymer. This may explain the lower removal efficiency of methyl orange compared to crystal violet. We turn to molecular dynamics simulations in order to differentiate more clearly the effects that charge or hydrophobicity have in the segregation and removal of these organic molecules and their molecular-scale interactions with the polyelectrolyte complexes.

2.3.4. Simulations of Polymers and Dyes

We use coarse-grained molecular dynamics at two different length scales to study the interactions of crystal violet, methyl orange, and variations of these molecules with the polymer complexes. The Martini model provides information on the conformation of the anionic random copolymer, while a more coarse, implicit solvent model is developed to study the formation of complexes and interactions with the dyes. Using the Martini model, we first perform simulations of only the anionic copolymers and their counterions without cationic copolymers or dyes (Figure 2.3(left)). The monomer fractions for the anionic copolymers match the fractions used in experiments (EHMA: 0.44, PEGMA: 0.51, SPMA: 0.05), and each copolymer has a degree of polymerization (DP) of 100. Figure 2.3(left) and (center) show that the anionic copolymers form micelles with both models, and the distribution of hydrophobic, hydrophilic, and negatively charged beads from the micelle center of mass. This distribution is different when cationic polymers are added to create the complexes as the anionic chains become more elongated and less micelle-like in Figure 2.3(right).

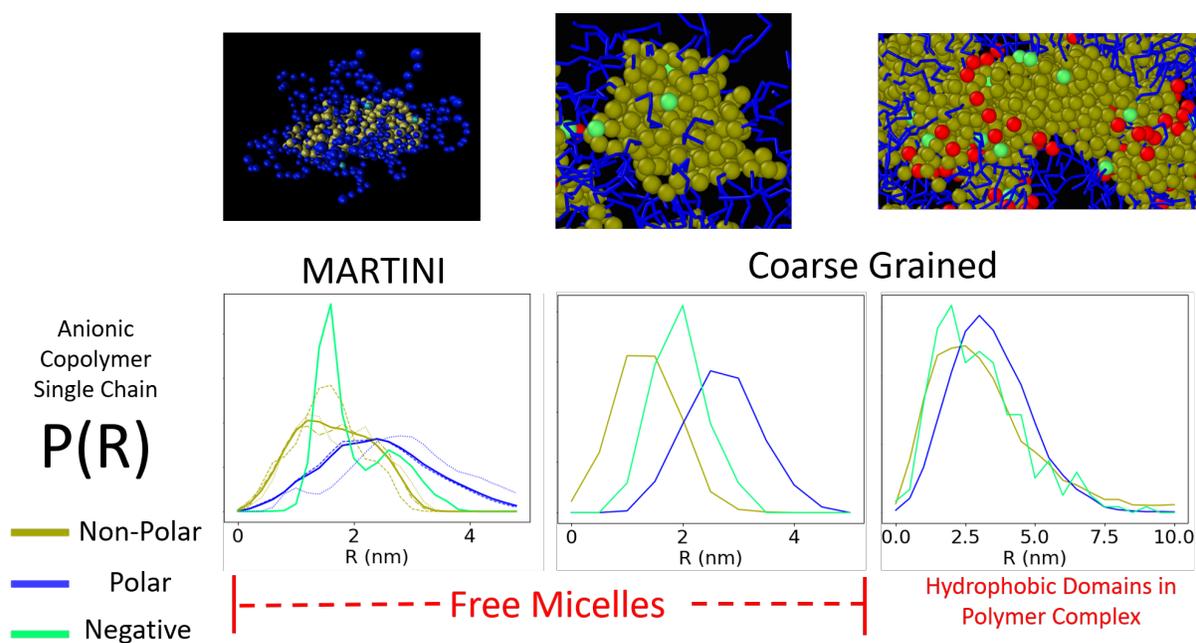


Figure 2.3. Distributions of different bead types in the Martini (right) and Coarse Grain (left) versions of the micelles. Both models show that the anionic polymers can form spherical micelles with dimensions of 4-5 nm. Distributions of beads which are hydrophobic are colored gold in both models and compose the hydrophobic cores of the micelles. Hydrophilic beads in both models are colored blue and form the corona of the micelles. The negatively charged beads in both models are colored teal and sit at the edge of the hydrophobic core about 2 nm from the center of mass. In the Coarse Grained model, more stretched conformations are observed when interacting with the cationic copolymers, similar to the Martini model.

Like the anionic copolymers, we use monomer fractions that correspond to experiments (HEMA: 0.54, TMAEMA: 0.46). In this case, $DP = 200$ is chosen in order to represent the larger molecular weight of the cationic polymer used in experiment. In both models, the anionic copolymers that are not interacting with the cationic copolymers take on micellar configurations due to the hydrophobic side chains and backbone. When interacting with the cationic polymers, the anionic copolymers take on much more stretched conformations that still feature hydrophobic domains.

The two models confirm that the anionic copolymer forms a hydrophobic core with a hydrophilic corona and charges sitting at the edge of the hydrophobic core. This demonstrates the ability of the coarse-grained model to capture the conformation of the methacrylate-based, random, charged copolymers. As was noted above, the experimental polymer charge ratio, i.e., the total ratio of positive charges on all of the copolymers to the total number of negative charges on all of the copolymers, was 2.9–3.3. (As in the simulations, counterions make the system charge neutral overall). Earlier experimental work done of complex coacervation has suggested that only polymer charge neutral systems form macroscopic phases whereas noncharge neutral systems should form smaller dispersions. [119] Some studies even presuppose that this condition should be met. [26] Even when studies have extended the modeling to include charge anisotropy and short-range attractions, the models include only fluctuations *via* linear response theory (or random phase approximation), and when ionic correlations are included, they are assumed to be local using a binding energy of ions to the chain backbone. [27, 120] The distribution of charge on the polymers, along with the formation of hydrophobic domains for the anionic copolymer, likely plays a role in the nonstoichiometric polymer charge ratio of the complexes. The polymers used in this study are nearly ideal random copolymers, as the reactivity ratios of the methacrylate monomers are nearly 1. [85] Thus, the charges are randomly distributed, and we can calculate the number fraction or probability of finding a sequence of N charged units on the polymer [121].

$$(2.1) \quad P(N) = f_c^N (1 - f_c)$$

where f_c is the charge fraction of the polymer. The cationic copolymer has a much higher charge fraction of 0.46, compared to the anionic copolymer charge fraction of 0.05. We can conclude that on average that the cationic copolymer has considerably longer and more frequent positive charge sequences than the anionic copolymer has negative charge sequences. We also note that the average distance between charge sequences is the reciprocal of the charge fraction, being 20 for the anionic copolymer and 2.2 for the cationic copolymer. Compounded with the fact that anionic charges are spread out over the surface of hydrophobic domains, regions of the cationic copolymer with longer charge sequences will require interactions with multiple hydrophobic domains to effectively compensate the charge. Steric effects will limit the number of hydrophobic domains that can aggregate in a local area, at times leading to uncompensated positive charges.

The coarse-grained simulations support this hypothesis, and a polymer charge ratio near 3 was required to create a percolated structure, in reasonable agreement with the experimental polymer charge ratio (2.9–3.3). In Figure 2.4, we explore the percolation (counting only the hydrophobic beads) of the system as more cationic polymer is added by examining the probability of finding a polymer in a cluster of a certain size as a function of the polymer charge ratio. This probability is a weight-average probability, as opposed to a number-average probability, meaning that the probabilities are normalized by the total number of polymers in a cluster of a certain size as opposed to the total number of clusters of a certain size. Thus, a delta function for a cluster size of 1 signifies a single cluster of all the components. At polymer charge ratios below 1, large clusters constituting up to 60% of all polymers are observed. In these clusters, cationic copolymers serve as high-valency cross-linkers, forming hydrophobic connections with on average eight

anionic copolymers in a “pearl-necklace”-like structure. [122] The ratio of charges on the average cationic polymer to the average anionic polymer is 9.2, meaning that the charge on the average cationic polymer is incompletely compensated by the 8 anionic copolymers on average to which it is connected. Consequently, free micelles and smaller clusters containing both cationic and anionic polymers are also observed. For the same reason, these smaller clusters always have a net positive charge even though the system is net negatively charged. As the charge ratio is increased above 1, more cationic polymer is added, and there is an electrostatic driving force for the free micelles to enter the densely connected phase. However, smaller dispersions are still observed, and only when the charge ratio continues to increase to the observed experimental ratio of 3 do we see the smaller dispersion completely incorporated into one large cluster.

The use of positively and negatively charged polymers combined with the statistical nature of copolymerization creates a system where individual polymers have a range of compositions in terms of charge sign, fraction, and hydrophobicity. It has been shown that amphiphilic copolymers with a distribution of compositions should phase separate into many phases with different compositions. [17, 123] However, we do not observe this in simulation or experiment due to the addition of the charged monomers and the energetic cost of creating many interfaces. Instead, we observe local charge heterogeneity as shown in Figure 2.5. This is shown by splitting the simulation box into many smaller cells of a certain size, L , and calculating the effective charge in those cells, Z_{eff}

$$(2.2) \quad Z_{eff} = N^+ + N^-$$

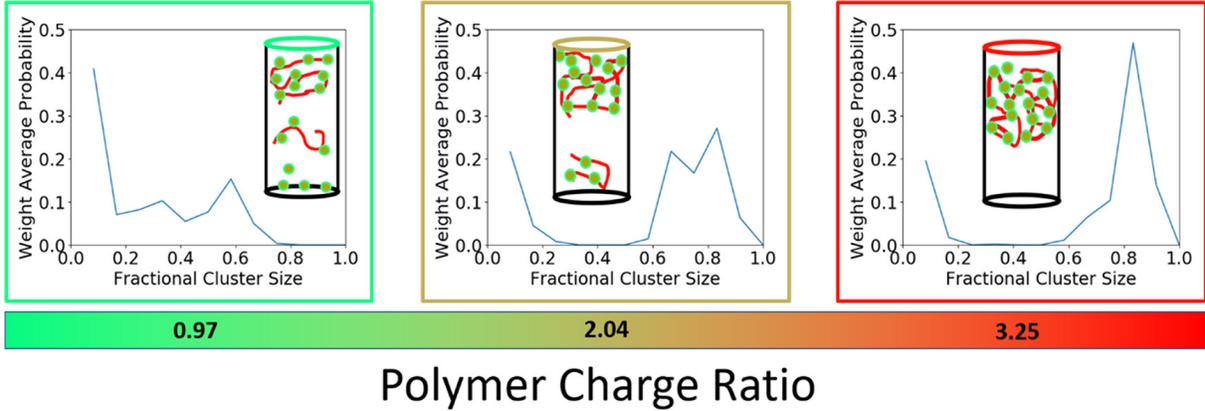


Figure 2.4. Probability of a chain being in a certain sized cluster, where the cluster size is measured as a fraction of the finite sized simulation, as a function of polymer charge ratio, which is defined as the ratio of the number of positive charges to the number of negative charges in polymer clusters (the system is electrically neutral due to counterions). At low charge ratio, free micelles, small clusters, medium-sized clusters, and large clusters comprised of nearly every chain are observed (left). As the polymer charge ratio is increased above 1, there is an electrostatic driving force for the free micelles to enter the dense phase; the free micelles should be incorporated in the dense phase even if it is not connected through hydrophobic interactions (center). As the charge ratio is further increased to 2 and above, the medium-sized clusters effectively disappear, leaving the polymers in one dense phase (red test tube). This agrees well with experiments where a polymer charge ratio above 3 is necessary to drive all the polymers into a macroscopic phase.

where N^+ is the number of positive charges in the cell and N^- is the number of negative charges in the box. In Figure 2.5(a), N^+ and N^- are restricted to be charges on the polymers; in Figure 4(b), they can be any charge including those from the counterions. At small cell sizes, we observe two peaks at ± 1 with and without the inclusion of counterions in the effective charge of the box. That is, the system develops domains with different fractions of charge. The energy penalty, F_c , associated with this charge heterogeneity is proportional to the square of the effective charge, divided by the cell size, L , in terms of the Bjerrum length, $l_B = e^2/(4\pi\epsilon_0\epsilon k_B T)$ with ϵ_0 being the permittivity of vacuum,

ϵ being the relative permittivity of the media, e being the elementary charge, k_B being Boltzmann's constant, and T being absolute temperature. Here, we use the Bjerrum length in water, 0.7 nm, which comes from its bulk dielectric constant, $\epsilon = 80$.

$$(2.3) \quad \frac{f_c}{k_B T} \propto \frac{Z_{eff}^2 l_B}{L} \propto l_B L^5$$

The compensation for this charge heterogeneity must come from the hydrophobic interactions of the anionic and cationic copolymers. The energy of these hydrophobic interactions, F_H , comes from the interface between the solvent and hydrophobic domains. It scales with the surface tension, γ , and, by dimensional analysis, the cell size squared.

$$(2.4) \quad \frac{f_H}{k_B T} \propto \gamma L^2$$

Thus, the charged term has a much stronger scaling with the cell size, and as a result, the two peaks at ± 1 are observed only at small length scales, obtained by minimizing the sum of eqs 2.3 and 2.4 giving $L_{hetero} \propto (\frac{\gamma}{l_B})^{1/3}$. Free counterions help to compensate the charge on the polymers, decreasing this length scale in Figure 2.5(b). Above this length scale, the population of cells with different numbers of charges tends toward a Gaussian distribution with zero average net charge, while the distribution width becomes broader as L increases. This is due to the stretched conformations of the copolymers in complexes as described in Figure 2.3. We note that when L increases beyond a critical value, the width of the charge distribution should shrink again, because there is no system with

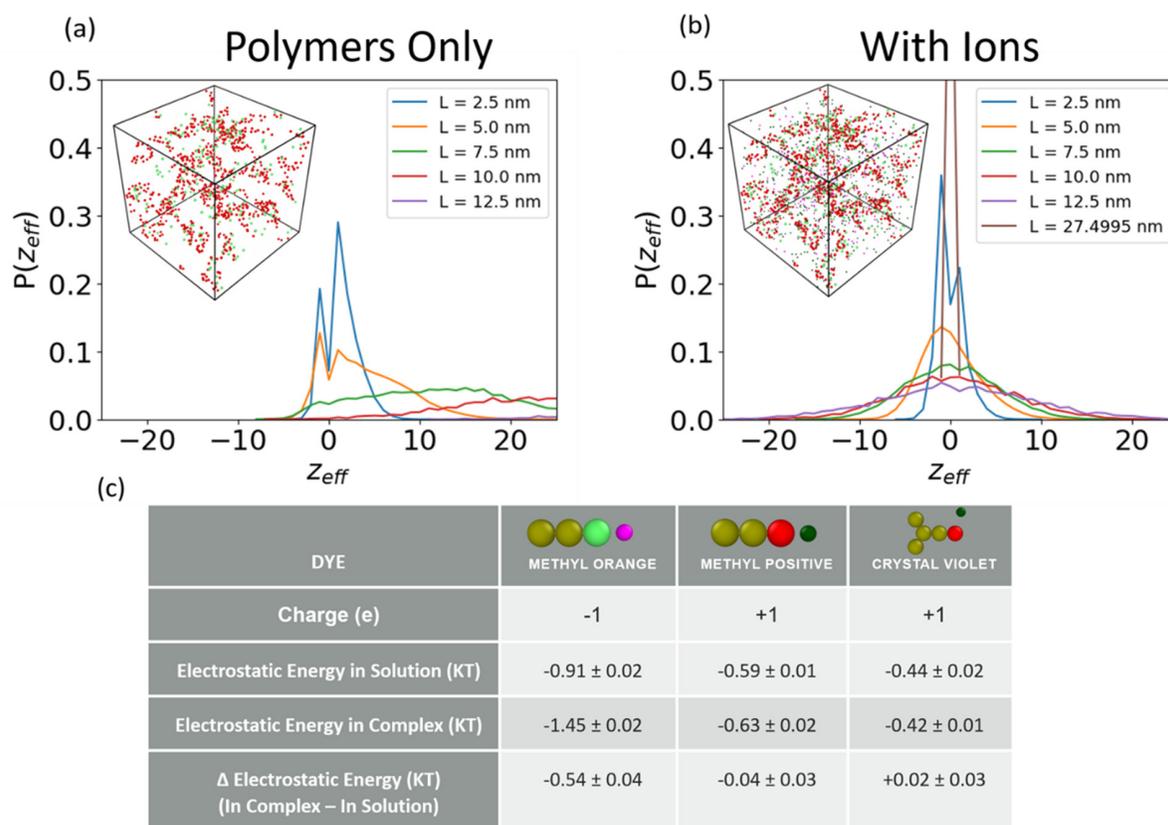


Figure 2.5. (a) Calculations of charge heterogeneity for charges on polymers. The simulation box is split into smaller cells of different lengths, L , and then the effective charge from the polymers in these boxes is calculated according to eq 2.2. Box sizes of 2.5 and 5 nm show two peaks where the effective charge is ± 1 . This shows how the hydrophobic energy of the polymers leads to local charge segregation in these polyelectrolyte complexes. (b) Calculations of charge heterogeneity for charges on polymers and counterions. The same calculation is performed as described above. Free counterions help to negate some of this charge segregation, but at small length scales, the two peaks are still observed. At larger length scales, with a Gaussian distribution centered at 0, effective charge is observed. (c) Electrostatic driving force for dye segregation. The electrostatic energy of different dyes when they are free in solution and segregated in the complexes as well as the energy difference between the states.

macroscopic excess charge. Finally, when L approaches the system box size, we observe a delta function at zero given the electroneutrality condition imposed in the simulations.

The charge heterogeneity also impacts the absorption of the dyes. In order to explore the generality of the method to segregate different molecules, we used seven variations of crystal violet and methyl orange coarse grain dyes shown in Figure 2.2. These molecules were added into the simulations with a polymer charge ratio of 3.25 at a ratio of 1 dye molecule to 130 polymer monomers. A dye is considered segregated if any of its hydrophobic beads is within a certain distance of a hydrophobic bead belonging to a polymer. Following experimental trends in removal rate, the crystal violet has a higher condensation rate than methyl orange. This makes sense given the additional hydrophobic benzene ring in the crystal violet that effectively increases its hydrophobic interaction with the complex. Overall, for purely hydrophobic dyes, the more hydrophobic beads it contains, the higher the percentage of segregated contaminants in the sample. This dependence of segregation to hydrophobicity is in line with experimental results in related polymer–micelle complexes. [28] We also see that adding a charged bead to a given hydrophobic structure decreases the percentage of molecules segregated. These segregation results are shown in Figure 2.6(a).

As expected for the net positively charged polymer complex, the negative methyl orange dye is more readily segregated than its positive counterpart. This is supported by Figure 2.5(c), which shows that the electrostatic driving force is stronger for methyl orange than its positive counterpart. That is, for the negative dye, the electrostatic energy decreases upon condensation into the polymer complex, whereas for the positive dyes, there is almost no difference in electrostatic energy despite the polymer charge ratio of 3.25. Figure 2.5(c) also shows that, due to the charge heterogeneity demonstrated in Figure 2.5(a), the absorption of the positive dyes is not adversely impacted by the net

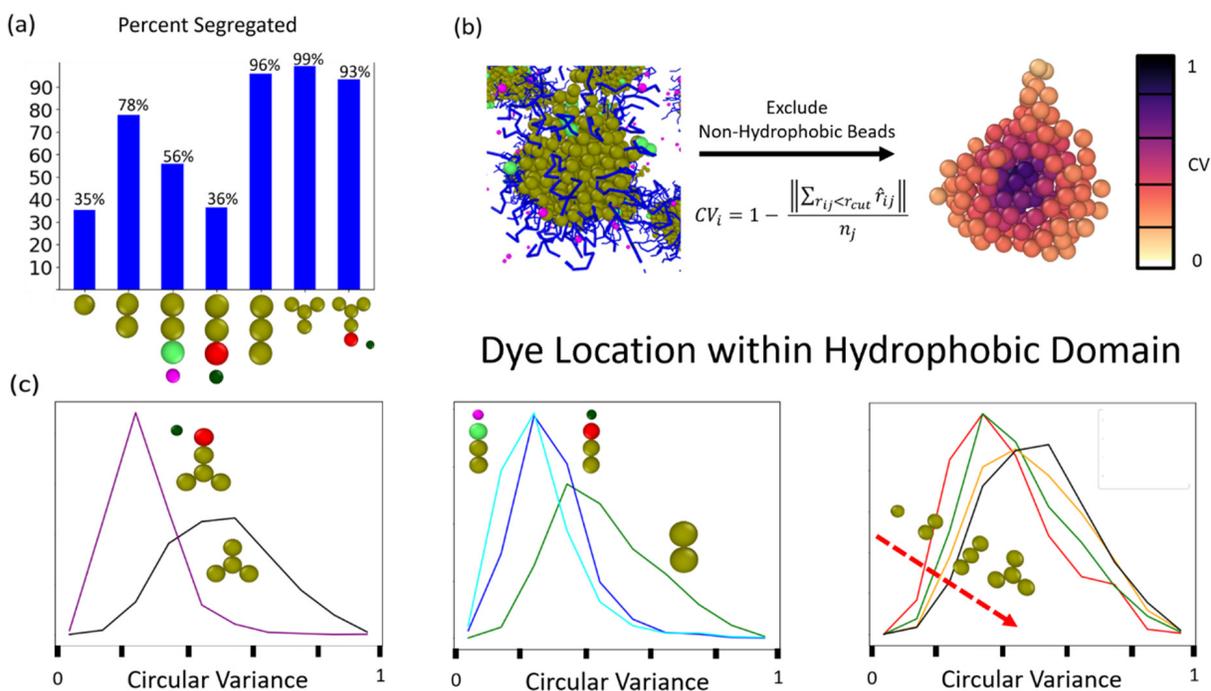


Figure 2.6. (a) Percentage segregation for the seven simulated dyes. The ratio of dyes to polymer monomers is 1:130, and the volume fraction of polymers is $\approx 10\%$. Segregation is defined as any hydrophobic dye bead being within a cutoff distance from a hydrophobic polymer bead. Almost all condensation occurs on the anionic polymers, which form stretched micelle-like structures with hydrophobic cores. The more hydrophobic a dye is, as measured by the number of hydrophobic beads, the higher the percentage segregated. Adding a charged bead to a hydrophobic dye always decreases the percentage segregated. (b) Circular variance as a measurement of dye location within hydrophobic domains. The circular variance is used to measure the degree of hydrophobic burial of a dye. It is calculated by taking the length of the vector sum of all the unit vectors from the dye to hydrophobic beads that are within the cutoff distance. This is then divided by the number of vectors and subtracted from 1. Thus, 1 is the maximum burial, and 0 is the minimum. (c) The position of dyes within hydrophobic domains as measured by circular variance. Left and center show specific comparisons for dyes with and without charges. The uncharged dyes are always much more buried. The right shows how the degree of burial continues to increase as the number of hydrophobic beads in the dye is increased.

positive charge on the complexes. The heterogeneity of the charges in the complex makes it possible for both negative dyes to reduce their energy upon condensation. In contrast, the positive dyes are relatively unaffected, because there are areas of net positive and net negative charge in the polymer complex, which is net positive. This encouraging generality of the method is not anticipated by simple intuition, which shows the importance of the heterogeneities in charge and composition caused by the random copolymers (some domains have positive charge and some negative) as shown in Figure 2.5(a,b).

This generality is explained by examining the location of condensed dyes within the hydrophobic cores present in the polymer complexes. To this end, we measure the hydrophobic circular variance. A full explanation of the circular variance is given in Figure 2.6(b), but it is used as a measure of the degree of burial of the dyes. Maximum burial by hydrophobic beads corresponds to a circular variance of 1, and minimum burial corresponds to a circular variance of 0. The distribution of circular variances for the different dyes is shown in Figure 2.6(c). There are two basic distribution shapes, one for charged and one for uncharged dyes. The distribution for charged dyes skews to lower values meaning that these dyes are restricted to be closer to the surface of the hydrophobic region due the charges preferring the ionic solvent environment. However, we notice a difference in the segregation behavior of oppositely charged dyes, with 10% of the segregated anionic methyl orange condensed to a hydrophobic bead on the cationic copolymer, compared to 1% for the segregated crystal violet. Despite the fact that both dyes reside at the interface of hydrophobic and hydrophilic regions of the polyelectrolyte complex, their hydrophobic interactions can be significantly different given that the methyl orange

is more likely to interact with the cationic copolymer than the crystal violet. This discrepancy may help explain why we do not observe a solvatochromic shift in experiments where methyl orange was segregated into complexes.

The distribution for the uncharged dyes tends toward higher values of burial, and the distributions indicate more burial as more hydrophobic beads are added. This burial means a stronger overall interaction between the hydrophobic portion of the dye and the hydrophobic domain of the complex, explaining the trend of lower percent segregated for charged dyes despite no adverse effects observed in the electrostatic potential (see Figure 2.5c). The stretched conformations of the anionic copolymers could be contributing to the removal of the charged dyes, because they have a higher surface area to volume ratio than spherical micelles and thus allow more low circular variance sites for the charged dyes to condense.

2.4. Future Outlook

We developed a method to segregate organic molecules from water into complexes formed by two oppositely charged, random copolymer species. We demonstrated that the heterogeneity of the complex plays an important role in providing favorable interactions to a wide variety of small molecules, as shown by our analysis of positively charged, negatively charged, and hydrophobic molecules. Hydrophobic interactions from hydrophobic cores play a dominant role for segregation into the complexes, and charged molecules undergo relatively favorable electrostatic interactions due to nanoscale charge segregation in the complexes. There are still many interesting aspects of this system to explore. The sizes and distribution of the hydrophobic domains, and whether such domains are formed

primarily through interactions between multiple polymer chains, are another aspect of heterogeneity in our system, which likely affects molecule uptake. We also claim that the hydrophobic domains, along with differences in charge distribution of the polyelectrolytes, leads to nonstoichiometric charge ratios for macroscopic complexes. It may then be possible to tune the charge distribution and hydrophobic content to change this charge ratio, which may also affect molecule uptake. Investigating how different copolymer compositions affect these various parameters and molecule uptake could be an active field of research.

Our work has implications for disordered systems such as membraneless organelles to concentrate small-molecule substrates necessary for enzymatic biological processes. Combining this result with the ability of synthetic random copolymers to form complexes with enzymes, [85] we expect that it may be possible to replicate the function of membraneless organelles in optimizing enzymatic activity by colocalizing an enzyme and its substrate, with potential industrial applications.

This method has the potential to provide an economical approach to remove a wide range of dye and contaminants from water on a large scale, as random copolymers can be synthesized inexpensively *via* free radical polymerization. The basis of favorable interactions between various molecules lies within the statistical distribution of monomers that leads to heterogeneity at the nanoscale, and controlling dispersity or other structural features of the polymer through more expensive techniques such as controlled radical polymerization is not necessary. This technique could be incorporated into existing water remediation processes *via* addition of a well-designed anionic random polyelectrolyte during a flocculation step. The flocculation behavior of these polyelectrolyte complexes

also has the potential to lead to the removal of hydrophobic particles from water such as enzymes and nanoparticles including nanoplastics, [49] something which traditional flocculants may have difficulty accomplishing, as they lack a significant hydrophobic interaction. Further studies on interactions between these heterogeneous polyelectrolyte complexes and bulk polymer surfaces are planned to explore such a possibility.

2.5. Methods

2.5.1. Molecular Dynamics Simulations

Molecular dynamic simulations were run at two different length scales. The finer simulations use the Martini forcefield with polarizable water. [124, 125] In these simulations, we study the distribution of different groups in micelles formed by the anionic copolymers. The model parameters used in these simulations were taken from a combination of sources studying similar polymers. [124, 126] The beads used are shown in Figure 2.1. The coarse-grained molecular dynamics simulations were performed using the HOOMD-blue package. [127, 128] An implicit water forcefield comprised of hydrophobic, hydrophilic, and charged beads was used to model both types of polymers and many different dyes. As shown in yellow in Figure 2.1, the backbone of all methacrylate polymers is composed of one hydrophobic bead per monomer, which is bonded to the next monomer *via* a harmonic bond. The side chains are unique to each monomer. The side chain of EHMA is modeled with two more hydrophobic monomers, PEGMEMA is composed of 6 hydrophilic beads (blue), and SPMA is a hydrophobic, followed by a negatively charged bead (green), which is accompanied by a positive counterion (pink). The cationic polymers are made of TMAEMA, which is modeled a positive bead (red) and accompanying counterion (green),

and HEMA, which has no side chain. Crystal violet, methyl orange, and variants of these dyes were also modeled using the same hydrophobic and charged beads as described in Figure 2.2. By utilizing additional angle potentials, their rigid shapes were conserved.

2.5.1.1. Coarse-Grained Model. Our forcefield features 7 types of beads. The “B” beads are used to represent hydrophobic subunits of polymers and contaminants, while the “L” beads represent the hydrophilic subunits. The rest of the beads represent positive and negatively charged subunits and ions. They are “QP”, “QM”, “QP_i”, “QM_i”, and “QM_{small}”. The beads ending in “i” are smaller ion beads. “QM_{small}” is the same size as an ion, but is used to represent the fact that we expect the negative charge on the SPMMA to be less shielded than the METAC.

Volume-excluding and hydrophobic pair interactions are handled by a Lennard-Jones interaction

$$(2.5) \quad V_{LJ}(r) = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^6\right]$$

For all pairs of beads, which are not both “B”, the potential is truncated at $2^{\frac{1}{6}}\sigma$, making it a purely repulsive WCA potential. [129] For pairs of “B” beads, the Lennard-Jones potential has a well depth, ϵ , of $1 k_B T$ and is truncated at 3σ . The mixing rule for σ is arithmetic.

$$(2.6) \quad \sigma = \sigma_{ij} = (\sigma_i + \sigma_j)/2$$

Explicit coulomb interactions are calculated using the PPPM method. [128] All bead types beginning with “Q” have a charge of 1.609×10^{-19} c. We use the bulk dielectric constant of water, 80. The charge, q , and diameter, σ , of each bead type is shown below in table 2.2.

Table 2.2. Non-bonded parameters for the Coarse-Grained Forcefield

Type	σ (nm)	q (e)
B	0.5	0
L	0.5	0
QP	0.5	1
QM	0.5	-1
QP _i	0.3	1
QM _i	0.3	-1
QM _{small}	0.3	-1

All bonds use a harmonic spring of the form

$$(2.7) \quad V_{bond}(r) = \frac{1}{2}k_{bond}(r - r_0)^2$$

where $k_{bond} = 120 k_B T / nm^2$ and $r_0 = 0.5nm$.

All angles used are harmonic angle potentials of the form

$$(2.8) \quad V_{angle}(r) = \frac{1}{2}k_{angle}(\theta - \theta_0)^2$$

Angle potentials used in contaminants use $k_{angle} = 120k_B T / rad^2$ and $\theta_0 = 120^\circ$ or 180° depending on the chemical shape of the contaminant. Angle potentials in the backbone of the polymers use $k_{angle} = 1k_B T / rad^2$ and $\theta_0 = 180^\circ$. These values were calibrated to observe gelation of the system.

2.5.1.2. Martini Model. We use Martini version 2.2 with polarizable water [124, 125]. The parameters for the random anionic copolymers come from a variety of sources [126, 130].

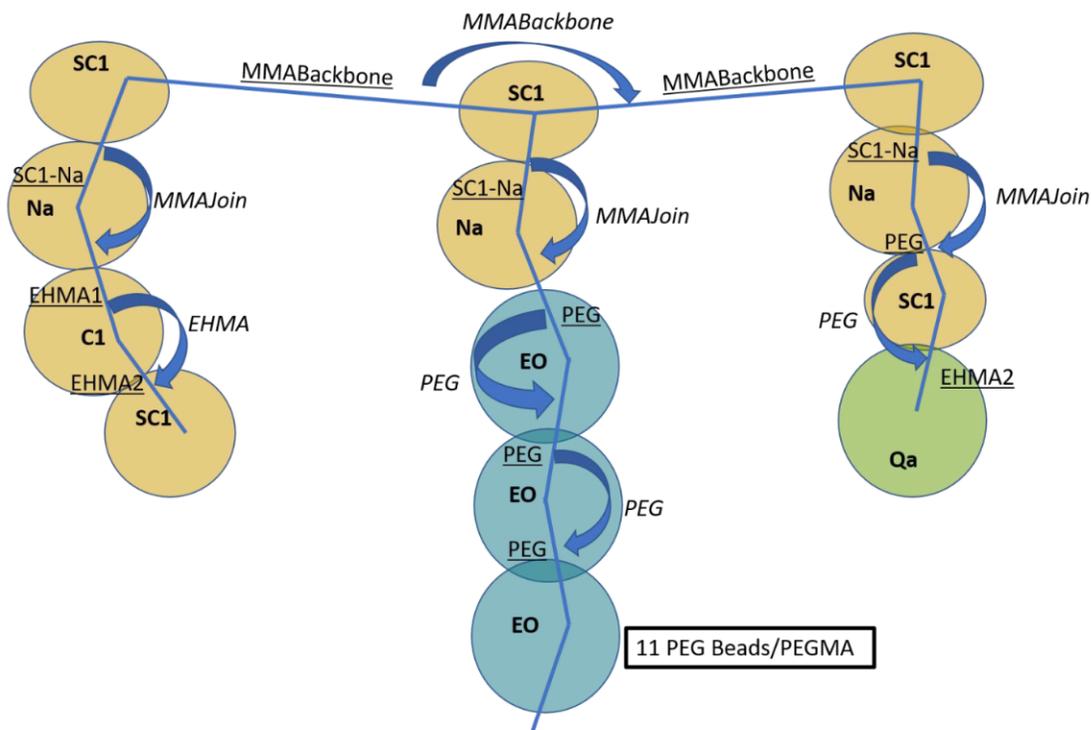


Figure 2.7. A full description of the bonded parameters for the random anionic copolymers. Angle names (see Table 2.4) are italicized and bond names (see Table 2.3) are underlined. Bead names control which the Lennard-Jones parameters all come from the martini forcefield.

All bonds use the same harmonic potential as in eq. 2.7 and the parameters are detailed in table 2.3.

All angles use the same harmonic potential as in eq. 2.8 and the parameters are detailed in table 2.4.

Table 2.3. Bond parameters for the Martini Forcefield

Type	$k_{bond}(kJ/mol/nm^2)$	$r_0(nm)$
MMABackbone	21 100	0.289
PEG	7 000	0.32
SC1-Na	17 000	0.282
EHMA1	17 000	0.54
EHMA2	1 250	0.425

Table 2.4. Angle parameters for the Martini Forcefield

Type	$k_{angle}(kJ/mol/rad^2)$	$\theta_0(^{\circ})$
MMABackbone	13	175
MMAJoin	67	144
PEG	400	122
EHMA	25	180

CHAPTER 3

Tailoring Interactions of Random Copolymer Polyelectrolyte Complexes to Remove Nanoplastic Contaminants from Water

This chapter is based on the work of Jeremy Wang, Curt Waltmann, Caroline Harms, Sumeng Hu, Jack Hegarty, Ben Shindel, Qifeng Wang, Vinayak Dravid, Kenneth Shull, John M. Torkelson, and Monica Olvera de la Cruz, **Tailoring interactions of random copolymer polyelectrolyte complexes to remove nanoplastic contaminants from water**, which has been submitted to Langmuir. Certain details have been modified.

3.1. Abstract

We investigate the usage of polyelectrolyte complex materials for water remediation purposes, specifically their ability to remove nanoplastics from water, on which there is currently little to no prior research. We demonstrate that oppositely charged random copolymers are effective at quantitatively removing nanoplastic contamination from aqueous solution. The mechanisms underlying this remediation ability are explored through computational simulations, with corroborating quartz crystal microbalance adsorption experiments. We find that hydrophobic nanostructures and interactions likely play an important role.

3.2. Introduction

Microplastic, and more recently, nanoplastic pollution are issues of emerging concern. A bulk of microplastics come from larger pieces of plastics exposed to the environment. The combination of UV light, which makes plastic particles brittle, and the constant abrasion of wave motion from bodies of water cause erosion of the macroplastics and the release of smaller particles.[131, 132, 133] Over time, these particles can become even smaller through other degradation processes which can be chemical or biological. [134, 135]

While the plastic particles themselves are thought to be relatively inert, they can still have adverse effects on environmental and human health. Microplastics can concentrate toxic chemicals or heavy metals already present in the environment. [136] Due to the hydrophobic nature of microplastics, hydrophobic chemicals such as bisphenols or other molecules of concern easily partition onto their surface,[137] while favorable electrostatic interactions between polymer surfaces and ions contribute to heavy metal adsorption.[138] Nanoplastics are thought to be considerably more potent at this partitioning and adsorption behavior due to their increased surface area to volume ratio.[139, 140] They also present additional concerns from their small size, as research has shown that these nanoplastics are much more biologically active and can penetrate cell membranes or even be cytotoxic. [141, 142] The combination of these aspects make nanoplastics potentially more hazardous than microplastics.

Currently, there are various techniques being investigated to remove microplastics and nanoplastics from water. Some recent methods include the use of advanced oxidation processes to decompose plastics through catalytic means.[143, 144] Others involve the

electrocoagulation of the plastic particles from water using charged electrodes.[145, 146] However, microplastics and nanoplastics tend to be quite dilute in bodies of water, leaving large volumes of water to be treated. [131] Thus, there is a need for economical and scalable processes to remove microplastics and nanoplastics from water. Industrial treatment processes such as sand filtration, activated carbon filters, and membrane filtration have been shown to be relatively effective at removing microplastics and nanoplastics, either by preventing particulate contaminants above a certain size from passing through the filter or by adsorption of the particles onto the filtration medium. [135, 139, 147] However, these techniques run into the issues of membrane fouling, pore clogging and adsorption capacity, which reduce the water flow and effectiveness of the filters over time. [135, 139]

The technique of flocculation, where negatively charged colloidal particles are aggregated through the use of oppositely charged flocculants, such as cationic polymers, has also been explored. [148, 149] Flocculation has particular advantages, as it is relatively simple to perform and is a widely adopted form of water treatment. It can also help reduce fouling of membranes used in subsequent membrane filtration.[150] Studies on the removal of nanoplastics through flocculation are somewhat sparse and the results can be quite variable, depending on factors such as size, solution pH, and flocculant dosage. Most studies have focused on polystyrene nanoplastics, and removal results typically range from 40% to 90+%, [151, 152, 153, 154] with one study reporting no removal.[147] One study used a methacrylic acid polymer as a surrogate for polyethylene terephthalate and found a removal rate of 88%. [155] Developing reliable and effective flocculation techniques to remove nanoplastics could go a long way in addressing nanoplastic pollution with established and accessible techniques.

In a previous paper [80], we used a flocculation-like technique to remove chemical contaminants from water. By combining a negatively charged random copolymer with a positively charged random copolymer, we were able to create macroscopic, solid complexes that could be removed from water through a simple filtration process. These polyelectrolyte complexes (PECs) have the ability to absorb various molecules due to favorable hydrophobic and charged interactions. In general, PECs are formed due to the presence of oppositely charged groups, but their interactions, properties, and phase behavior are highly tunable by a number of different factors. These can include solvent quality [27], ionic solution conditions [23, 156, 157, 89, 21], polymer rigidity [26], and monomer sequence (blockiness).[120, 42] One can even include other components in PECs including micelles[158] and bioactive proteins containing both positive and negative charges into these complexes.[159, 30, 40, 31]

The usage of random copolymers can provide additional tunability through the polymer composition, with the statistical distribution of the monomer sequence potentially enhancing polymer-surface interactions through a form of pattern recognition.[83, 82, 160, 84] Correlations in monomer sequence increase the affinity of polymer adsorption onto surfaces, particularly for heterogeneous surfaces. This concept was utilized to form nanoscale aggregates of oppositely charged random heteropolymers and single proteins, which have a heterogeneous surface, in organic solvents and polymer melts.[85, 161] We demonstrated complexation of random copolymers and the enzyme PETase in an aqueous solution,[162] which increased the activity of the PET-degrading enzyme on PET films. One potential mechanism for this enhancement is the ability of the copolymer-enzyme complex to stick and adsorb onto a PET surface due to a hydrophobic surface

affinity between the polymer and PET surface.[163] Here, we hypothesize that favorable interactions between macroscopic random copolymer PECs and PET can be utilized in flocculation to remove PET nanoplastics from solution.

In this study, we investigate the ability of random copolymer polyelectrolyte complexes to remove PET nanoplastics through a flocculation technique. As in our previous work, we use random methacrylate copolymers synthesized via free radical polymerization. The anionic copolymer is composed of hydrophobic ethyl-hexyl methacrylate (EHMA), hydrophilic polyethylene glycol methyl ether methacrylate (PEGMEMA)(500 Mn), and negatively charged sulfopropyl methacrylate (SPMA); the cationic copolymer is composed of the hydrophilic hydroxyethyl methacrylate (HEMA) and the positively charged trimethylaminoethyl methacrylate (TMAEMA). PET nanoplastics were prepared through nanoprecipitation techniques and loaded with Nile red fluorescent dye to enable the quantification of removal via fluorescence measurements. We observe that increasing the content of hydrophobic monomer EHMA in the anionic copolymer increases the amount of nanoplastics removed from solution, with the most hydrophobic polymer showing quantitative removal. Our hypothesis is that these PECs can adsorb and stick onto PET nanoplastics, encapsulating them into macroscopic flocs which can be removed relatively easily through coarse filtration. We investigate these polymer-surface interactions using quartz crystal microbalance (QCM) experiments to gauge the adsorption behavior of the anionic random copolymers on PET, along with corresponding Martini [164] coarse-grained molecular dynamics simulations. We show that hydrophobicity can non-monotonically impact polymer-surface interactions due to changes in polymer conformation. Finally,

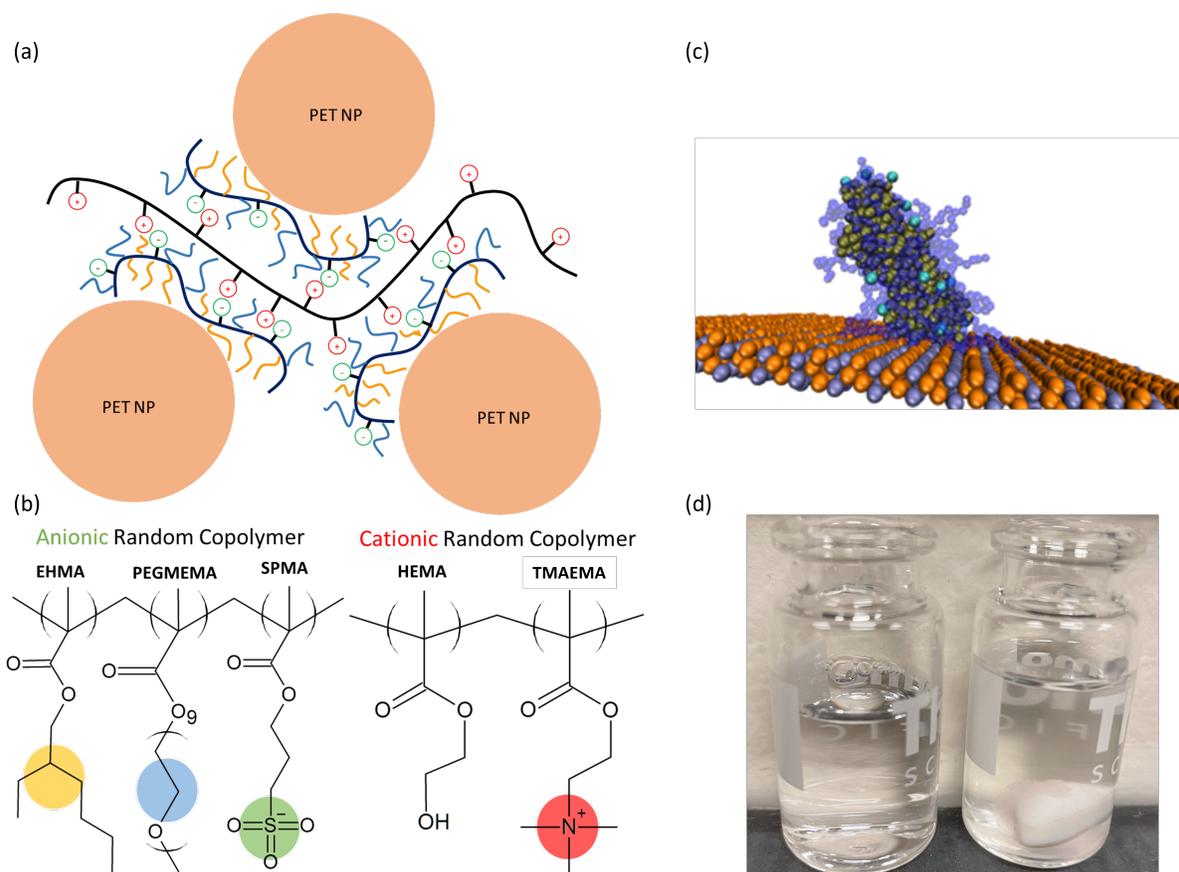


Figure 3.1. Schematic of polymer complexes and polymer adsorption onto PET. (a) Diagram of polyelectrolyte complex of random copolymers with adsorption of PET nanoparticles. (b) Chemical composition of the anionic and cationic copolymers. (c) Coarse-grained molecular dynamics simulation of anionic copolymer adsorbing onto PET. (d) Picture of nanoplastic spiked solution (left) and the polyelectrolyte complex formed after addition of random copolymer solutions, which has accumulated on a stir bar. (right)

implicit solvent coarse-grained molecular dynamics simulations of the copolymer complexes are used to understand how the nanostructure of the complexes may play a role in adsorption, again due to the arrangement of hydrophobic and hydrophilic groups within the complex.

Table 3.1. Anionic Copolymer Compositions

Samples	EHMA mol%	PEGMEMA mol%	SPMA mol%	M_n (g/mol)
Sample 1	54	36	10	268 000
Sample 2	49	42	9	183 000
Sample 3	33	59	8	248 000
Sample 4	25	67	8	304 000
Sample 5	51	33	16	274 000
Sample 6	39	26	35	201 000

3.3. Results and Discussion

Table 1 shows the various compositions of the anionic copolymers used in the study, which were chosen to investigate how relative hydrophobicity and charge fraction affect nanoplastic uptake. Samples 1-4 are in descending hydrophobic content, with Sample 1 being the most hydrophobic. Samples 5 and 6 are hydrophobic polymer compositions but with increased charge content. The cationic polymer was characterized with a 28% charge fraction (mol % TMAEMA) and an apparent number average molecular weight of 4.3 million g/mol.

The nanoplastic removal procedure is similar to that described previously.^[80] In short, anionic copolymer is added to an initial aqueous solution with our nanoplastic contamination. Under stirring, the cationic copolymer is then added until flocculation occurs and the solution becomes clear. A small amount of magnesium sulfate is then added to help coagulate the remaining flocs. The remaining solution is then filtered through a coarse 1 μm syringe filter, and fluorescence measurements are performed on the filtrate to determine the amount of nanoplastic removed.

The nanoplastics (Figure 3.2(a)) were characterized to have a number average size of 80 nm with a dispersity of 1.3. We find that we remove significant quantities of nanoplastics

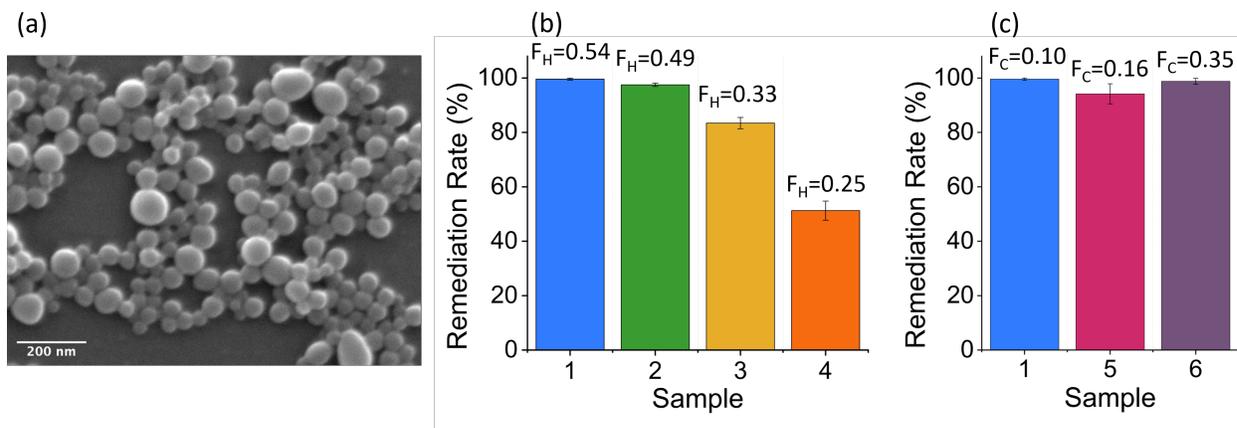


Figure 3.2. Image of PET nanoplastics and remediation rates from water. (a) SEM image of PET nanoplastics synthesized through a nanoprecipitation technique. (b) Remediation rates of nanoplastics for complexes that include polymer samples 1-4, which have a similar charge fraction. F_H denotes the fraction of hydrophobic monomer of the polymer sample. (c) Remediation rates of nanoplastics for complexes that include polymer samples 1, 5, and 6, which have roughly 50% more hydrophobic monomer than hydrophilic monomer. F_C denotes the fraction of charged monomer of the polymer sample

from our solution, with the complexes formed with our most hydrophobic sample (Sample 1) quantitatively removing nanoplastics below the detection limit of our fluorescence plate reader. We see in Figure 3.2(b) that there is a correlation between the hydrophobicity of the polymer composition and amount of nanoplastics removed. This would indicate that hydrophobic interactions play a significant role in the removal process. We find that charge does not have nearly as significant an impact on removal rate when comparing results of Samples 1, 5 and 6 in Figure 3.2(c), all with removal rates exceeding 90%.

To understand better the adsorption behavior of the polymers and the role it plays in nanoplastic removal, we perform quartz crystal microbalance experiments to measure directly the amount of anionic polymer adsorption on a PET surface.[165] We report

the mass of polymer adsorbed in Figure 3.3(a) as obtained by the Sauerbrey equation. Our results indicate that the adsorption is non-monotonic with hydrophobicity, with high adsorption for the most hydrophilic and most hydrophobic samples. These results appear to be counter-intuitive, especially given our nanoplastic removal results.

A similar trend is observed in Martini simulations (see Methods for details) of the anionic random copolymers on a PET surface in Figure 3.3. Hydrophobic interactions are quantified by counting the number of contacts between hydrophobic beads of the anionic copolymer and the PET surface, and we find that polymers with the lowest and highest hydrophobic monomer fractions, F_H , have highest numbers of hydrophobic contacts (see Figure 3.3). The simulations reveal a potential explanation for this behavior through polymer conformation. The highly hydrophilic copolymers form more caterpillar/train-like structures which can lie flat on the polymer surface. However, with increasing hydrophobic content, the polymers take more of a micellar conformation, with the formation of hydrophobic cores. This can lead to lower surface area of interaction between the polymers and the PET surface. This trend with hydrophobicity is similar to what was observed when the polymers interacted with a protein instead of a flat surface.^[162] At the highest hydrophobic monomer fraction, there is an insufficient amount of hydrophilic monomer to cover the core of the micelle, and as a result the polymer will wet the surface and create more hydrophobic contacts.

The discrepancy between the anionic polymer adsorption and the ability of the complex to remove nanoplastic particles is likely due to conformational and structural differences between the single polymers and polymer complexes. We used an implicit solvent,

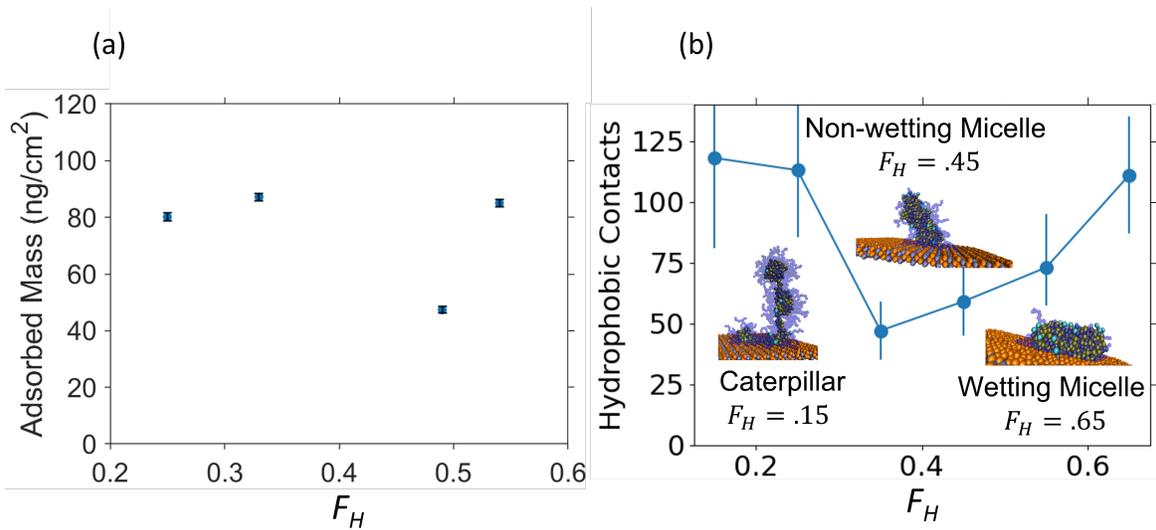


Figure 3.3. Polymer adsorption onto PET vs hydrophobic fraction F_H with experiments and simulation. (a) Mass of polymer adsorbed by samples 1-4 onto a film of PET as measured by QCM. The error bars are approximately the size of the markers used in the plot. (b) Simulation results of hydrophobic contacts between anionic copolymers and PET surface.

coarse-grained model to perform simulations of the polymer complexes with variable hydrophobic fractions of the anionic copolymer, $F_{H,anionic}$. We examined the local hydrophobic distribution of the complexes [17] at length scales smaller than the interdomain spacing of the complexes as calculated using the scattering function [109](Figure 3.4a). This is done by splitting the simulation box into many cells of size, L , and then measuring the fraction of hydrophobic monomers from both polymer species, F_H . The least hydrophobic complexes ($F_{H,anionic} = 0.25$) show a strong peak to the left of the mean composition suggesting many cells are more hydrophilic than the mean (Figure 3.4(c)). In Figure 3.4(e), there is also a weaker peak to the right of the mean, which suggests some local separation is present. For the most hydrophobic complexes, we instead see one peak near the mean, although the distribution is skewed hydrophilic in Figure 3.4(e). When the cell size,

L , is greater than the interdomain spacing, gaussian distributions around the mean are observed for all $F_{H,anionic}$ as expected. Snapshots of the complexes show much stronger segregation of polymer components into hydrophobic domains in Figure 3.4(d) than in Figure 3.4(b) due to the higher fraction of hydrophobic monomers, in agreement with the peak heights in Figure 3.4(a).

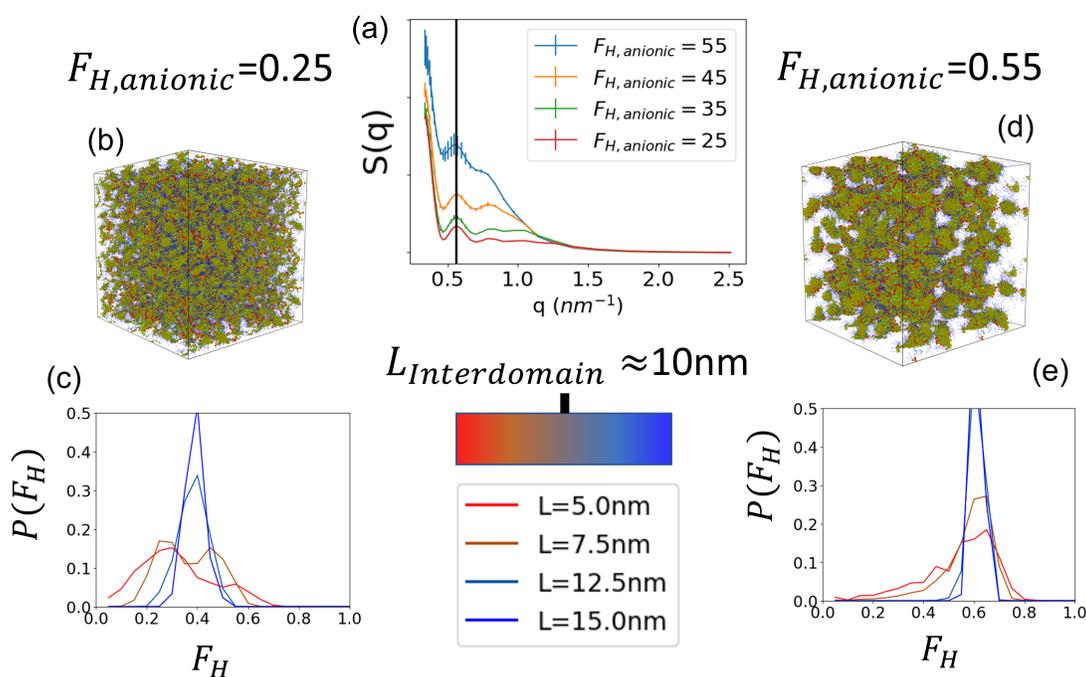


Figure 3.4. Simulation results for polyelectrolyte complexes with varying hydrophobic fraction of the anionic copolymer, $F_{H,anionic}$. (a) Scattering functions for $F_{H,anionic} = 25, 35, 45,$ and 55 all show peaks at $q = 0.56 \text{ nm}^{-1}$ and some show a smaller peak at $q = 0.78 \text{ nm}^{-1}$. These peaks correspond to the spacing between domains, which $\approx 10 \text{ nm}$. (b, d) Snapshots of the most and least hydrophobic complexes. (c,e) Local hydrophobic monomer density distributions show multiple peaks for the least hydrophobic complexes at length scales smaller than the interdomain spacing.

These domains may be similar to the “wetting micelle” conformations that were able to create more hydrophobic contacts with the surface in the single polymer simulations.[166,

167] In the single polymer simulations we also observed high absorption with the low hydrophobicity polymers due to their ability to lie flat on the surface in “caterpillar” conformations. However, in the complex they likely lack the necessary degrees of freedom for this to occur. Taken together, these results suggest that the properties and functions of these random copolymer polyelectrolyte complexes depend greatly on the nanoscopic details of polymer composition and conformation. For the particular case of removing nanoplastics, hydrophobicity appears to play a key role but is intertwined with the interactions between the two polymers in the complex system. The distribution of hydrophobic regions in the polyelectrolyte complex, as well as the PET surface, could be important.**[168]** A similar case was observed previously in terms of charge distribution and the ability to encapsulate molecular contaminants in aqueous solution.**[80]**

3.4. Conclusions

The ability to tune the interactions of the random copolymer polyelectrolyte complexes by adjusting hydrophobicity and charge clearly lends great utility and versatility to this remediation technique. While we have focused on the removal of PET nanoplastics in relatively pristine condition in this study and found hydrophobicity to be the dominating factor, the potential still remains to expand the type of contaminants that can be removed. Nanoplastics composed of different polymeric materials will likely have different surface properties and interactions, with a different composition that may be tailored to optimally remove said material. Environmental conditions can also potentially alter nanoplastic surface characteristics through the adsorption of biomolecules and the formation of a corona layer.**[169]** The removal of microplastic contamination may have entirely different

considerations. Biofilms, which tend to be a layer of naturally growing bacteria or other microorganism, readily accumulate onto microplastic surfaces and can present drastically altered surface properties.[134] The type of random copolymers that we used in this study have already been shown to interact well with biological molecules such as proteins,[162, 85] and thus may be well suited for interacting with such biological surfaces.

Thus, these random copolymer polyelectrolyte complexes may prove to be a potent water remediation technique due to the wide range of potential favorable interactions they can provide, with a simple mechanism of tuning these interactions through the adjustment of the polymer compositions. This in turn can provide a level of robustness in operating conditions and use cases that may very well extend beyond the case of addressing nanoplastic pollution. For example, in a preliminary experiment, we were able to quantitatively remove lead contaminated nanoplastics and lead ions from a sample solution with our flocculation technique. The abundance of electrostatic interactions in the polyelectrolyte complexes likely promote heavy ion adsorption, which has been demonstrated in other cases.[104, 149] While further understanding of the nanoscale interactions of these materials is needed to improve these functionalities, it is clear there is great potential in using this class of materials to address relevant issues in water remediation.

3.5. Methods

3.5.1. Martini Simulations of Random Copolymer Adsorption

3.5.1.1. Forcefield. We use the Martini [164] 2.2 [170] forcefield with polarizable water [171]. The potentials for the methacrylate-based random copolymers come from a few sources [126, 130] and are fully described in the Methods section of Chapter 2 and

used again in Chapter 4. For the polyethylene surface, we convert from the atomic crystal structure [172] using 3 STY [173, 174, 159] beads for the benzene ring and 2 more “Na” beads for the ester groups. The “Na” beads are also used in the random copolymer backbone for an equivalently bonded group of 4 atoms. During the simulations the PET surface is held using positional restraints. In the production simulations, we use the following parameters. The short-range Coulomb interactions were calculated up to 1.35 nm with the 3dc Ewald summation approach (relative permittivity of 2.5 for Martini with polarizable water) for the long-range electrostatic interactions in order to mimic 2D periodic boundary conditions in the x and y direction.[175] The box length was 20 nm x 20 nm x 20 nm with an extra 60 nm of vacuum in the z direction to avoid errors in the ewald sum coming from the periodic boundary. The LJ 12-6 potential interactions were truncated at 1.35 nm. The NVT ensemble was applied. The temperature was coupled at 300 K using the velocity rescaling method. The leapfrog integration time step of 10 fs was employed.

3.5.1.2. Protocol. For each hydrophobic fraction the polymer sequence is randomly generated under the constraint of the mean composition. The degree of polymerization is 200 monomers in order to sample a large number of different sequences within the chain. The mean composition features 3 different types of monomers. The charged monomer, SPMA, is always present at 10% while the hydrophilic monomer, OEGMA-9, varies based on the given hydrophobic fraction, F_H which determines the amount of the hydrophobic monomer, EHMA. We build the random copolymer with only the 20 positive counterions necessary to ensure electroneutrality. For each composition we run steered and non-steered binding simulations. To avoid the stochasticity of initial binding between the

random copolymer and the surface, we start with a steered MD run to initiate contact and then allow the polymer to react in a longer, completely unsteered simulation. In the steered MD simulations the first bead on the backbone of the polymer is pulled towards the PET monolayer at a rate of 1.0 nm/ns until it is 0.5 nm away from the monolayer surface using a harmonic spring with force constant 1000 kJ/mol. This bead is then held at this distance from the surface for 40 ns using the same spring with a pull rate of 0. For the final 180 ns of the simulation the polymer is completely unrestrained. This entire binding process is repeated 3 times for each composition.

3.5.1.3. Analysis. Hydrophobic PET-polymer contacts are defined as two hydrophobic beads, one in the polymer and one in the PET surface, within a cutoff of 0.53 nm from one another. This is approximately the lowest energy distance for the largest Martini beads. All bead types in the PET surface (STY and Na) as well as SC1 and C1 are considered hydrophobic according to the Martini forcefield.[164] Error bars on the total number of hydrophobic contacts are the standard error of 3 replicas which are run for each composition.

3.5.2. Simulations of Polyelectrolyte Complexes

3.5.2.1. Forcefield. Coarse grained molecular dynamics simulations of the polyelectrolyte complexes were performed using HOOMD.[176] We use an implicit solvent model for the polymer chains as described in the Methods [80] of Chapter 2. This model differentiates between hydrophobic, hydrophilic, and charged beads allowing use to include the basic chemical properties of the different monomers used in the experiments including the hydrophobic methyl methacrylate backbone of all monomers. Charged interactions are

calculated using a bulk dielectric constant of 80, which corresponds to a Bjerrum length, l_B of 0.7 nm. The calculation is performed using the PPPM method.[128] We use the NVT ensemble and the Langevin thermostat.

3.5.2.2. Protocol. We initialize the box by randomly placing 60 total polymer chains, 9 cationic and 51 anionic. We attempt to replicate the experimental parameters as all cationic polymers have a degree of polymerization of 200 monomers, and the mean charge fraction of those monomers is 30%, while anionic polymers have a degree of polymerization of 100 and the mean charge fraction is 10%. Counterions are included for all charged monomers, but no additional salt is added. All sequences are randomly generated assuming that all reactivity ratios are 1 [85] and under the constraint that the mean compositions of the anionic and cationic copolymers must be met exactly. Simulations are run with the timestep, $dt=0.001 \tau$ in LJ units which corresponds to $dt \approx 80$ fs in real units using the formula $\tau = \sqrt{\frac{md^2}{\epsilon}}$, where ϵ is the energy unit, $1 k_B T$ at 300K, d is the distance unit, 0.5nm and m is the mass unit 100 g/mol. We initially generate the polymers in an 40 nm x 40 nm x 40 nm box. The polymers are then annealed by raising the reduced temperature from $T=1$ to $T=2$ in LJ units over the course of 2 million timesteps, holding there for 2 million timesteps and then going back down to $T=1$ over 2 million timesteps. The box is then compressed to its final dimensions, 27.5 nm x 27.5 nm x 27.5 nm before repeating the annealing cycle. The final simulation is then run for 50 million timesteps which corresponds to ≈ 4 microseconds. This entire process is done 4 times for each composition and data is averaged over many frames per simulation and the 4 independent simulations.

3.5.2.3. Analysis. The scattering function $S(q)$ is computed according to

$$(3.1) \quad S(q) = \frac{1}{N_{pairs}} \sum_{pairs} \frac{\sin(qr)}{qr}$$

For computational efficiency we use an equivalent form where we create a histogram for the probability of a certain pair distance, r :

$$(3.2) \quad p(r_i) = \frac{N(r_i)}{N_{bins}}$$

$$(3.3) \quad S(q) = \sum_{bins,i} p(r_i) \frac{\sin(qr_i)}{qr_i}$$

This is much faster as it is over the number of bins which is far less than the number of pairs for the large amount of data we use.

$$(3.4) \quad S(q) = \sum_{bins,i} p(r_i) \frac{\sin(qr_i)}{qr_i} = \frac{1}{N_{bins}} \sum_{bins,i} N(r_i) \frac{\sin(qr_i)}{qr_i} = \frac{1}{N_{bins}} \sum_{pairs} \frac{\sin(qr)}{qr}$$

There is a different normalization factor but the relative peak heights are the same.

For the purpose of defining F_H in Figures 3.4(c) and (d), we use only the polymer beads ignoring the counterions. We use only the backbone bead of each monomer in computing the location of different monomers when we split the box into smaller boxes and calculate the hydrophobic fraction in each box. Each monomer is classified as hydrophobic if it is EHMA on the anionic copolymer or TMAEMA on the cationic copolymer.

CHAPTER 4

Functional Enzyme-polymer Complexes

This chapter is based on the published work [162] of Curt Waltmann, Carolyn E. Mills, Jeremy Wang, Baofu Qiao, John M. Torkelson, Danielle Tullman-Ercek, and Monica Olvera de la Cruz, **Functional enzyme-polymer complexes**, *Proc. Nat. Acad. Sci. U.S.A.*, 2022, 119, 13, e2119509119, with modified details.

4.1. Abstract

Engineered and native enzymes are poised to solve challenges in medicine, bioremediation, and biotechnology. One important goal is the possibility of upcycling polymers using enzymes. However, enzymes are often inactive in industrial, nonbiological conditions. It is particularly difficult to protect water soluble enzymes at elevated temperatures by methods that preserve their functionality. Through atomistic and coarse-grained molecular dynamics simulations that capture protein conformational change, we show that an enzyme, PETase, can be stabilized at elevated temperatures by complexation with random copolymers into nanoscale aggregates that do not precipitate into macroscopic phases. We demonstrated the efficiency of the method by simulating complexes of random copolymers and the enzyme PETase, which depolymerizes polyethyleneterephthalate (PET), a highly used polymer. These polymers are more industrially viable than peptides and can target specific domains on an enzyme. We design the mean composition of the random

copolymers to control the polymer-enzyme surface contacts and the polymer conformation. When positioned on or near the active site, these polymer contacts can further stabilize the conformation of the active site at elevated temperatures. We explore the experimental implications of this active site stabilization method and show that PETase-random copolymer complexes have enhanced activity on both small molecule substrates and solid PET films. These results provide guidelines for engineering enzyme-polymer complexes with enhanced enzyme functionality in non-biological environments.

4.2. Introduction

Enzymes have the potential to tremendously impact the fields of pharmacology [177, 178], biotechnology [33], and bioremediation [161]. They are especially useful for up-cycling plastics [53], which are currently polluting oceans [179, 180, 181] and freshwater supplies harming both humans [182] and animals [183, 54]. Enzymes such as lipases [184], cutinases [185, 186], hydrolases [31], and cytochrome P450 [187, 188] can catalyze a growing number of reactions due to advances in enzyme engineering [188, 189]. In addition to engineering new catalytic functions, attempts have been made to increase enzyme efficiency by creating multi-enzyme complexes [190, 191], immobilizing them on two-dimensional surfaces [192, 193, 187, 33, 194], embedding them in plastics [161], and modifying the amino acid composition of enzymes to increase their thermal stability [52, 195]. Stabilizing these enzymes is crucial for applications in non-biological conditions such as elevated temperatures and pressures. Here, we investigate how complexation with random copolymers can enhance the high temperature stability of the enzyme PETase [50], which degrades polyethyleneterephthalate (PET). PET is a glassy

polymer at room temperature. Therefore, for PETase to function effectively, the temperature should be raised above T_g (70°C) [172], which would normally denature the enzyme. Our simulations demonstrate the relationship between polymer composition and protein-polymer structure as well as the spatial correlations between chemically different monomers and the heterogeneous protein surface. We then show how the structure of the protein-polymer complex can impact the catalytic function of the enzyme, especially at elevated temperatures. Finally, we provide experimental verification of this enhanced catalytic activity of PETase in the presence of random copolymers.

Charged polymers have been used to encapsulate and stabilize proteins through either macroscopic segregation [29, 30] or microphase separation into nanoscale aggregates that do not precipitate into macroscopic phases [32, 31, 40]. This approach works on a wide variety of proteins including engineered proteins with non-biological functions [188, 85], because protein-polymer interactions can be altered in many ways to enhance complexation [38, 34] including by modifying the charge of the proteins themselves [36, 35]. Phase separation through engineering protein charge distribution has also been demonstrated *in vivo* [37] in cellular bodies known as membraneless organelles [45] (MLOs). MLOs are composed of proteins, nucleic acids, and small molecule metabolites [14] and are often formed from intrinsically disordered proteins (IDPs) or proteins with intrinsically disordered regions (IDRs). IDPs control the MLO composition based on specific interactions that originate from the IDP sequence [196] and this spatial organization helps to regulate the internal biochemistry of cells. Inspired by IDPs, we design a microphase separated protein-polymer complex using random copolymers. These random copolymers contain

monomers with a common backbone but different side chains that can be polar, nonpolar, or charged. The monomers, shown in Figure 7.2(d) and (e) are oligo-ethyl glycol methacrylate with a length of nine ethyl glycol units (OEGMA-9), ethyl hexyl methacrylate (EHMA), and sulfo-propyl methacrylate (SPMA). Due to their methyl methacrylate backbone, they are inexpensive to randomly polymerize in large-scale, industrial processes. They have also been used in previous computational and experimental studies of polymer complexes with proteins including horseradish peroxidase, glucose oxidase, and organophosphorus hydrolase as well as small organic molecules [85, 80]. Further, amphiphilic random copolymers are well-suited to complex with the heterogeneous polar and nonpolar surfaces of proteins [41] including PETase. Thus, in addition to the electrostatic attraction between the polymer and protein, the polar and nonpolar groups of the random copolymers can self-optimize [197], maximizing their interactions with protein surface domains making this complexation quite general [41, 85, 198].

In the present study, we explore how complexation with random copolymers can affect the conformation, and thus the function, of PETase especially at temperatures the enzyme does not experience *in vivo*. We achieve this by varying the mean polymer composition, which controls the polymer-protein surface interactions. This, in turn, affects complex formation and the spatial distribution of contacts on the surface of the enzyme. While there have been previous studies of protein-polymer complexation [38, 199], experimental studies are inherently limited in the direct observation of surface correlations [200] and previous computational studies have used coarse models that did not include charges [41] or vary the mean composition [85]. Other models [159] were used to study PETase and cytochrome P450 in the context of macroscopic complexes instead of nanoscale complexes

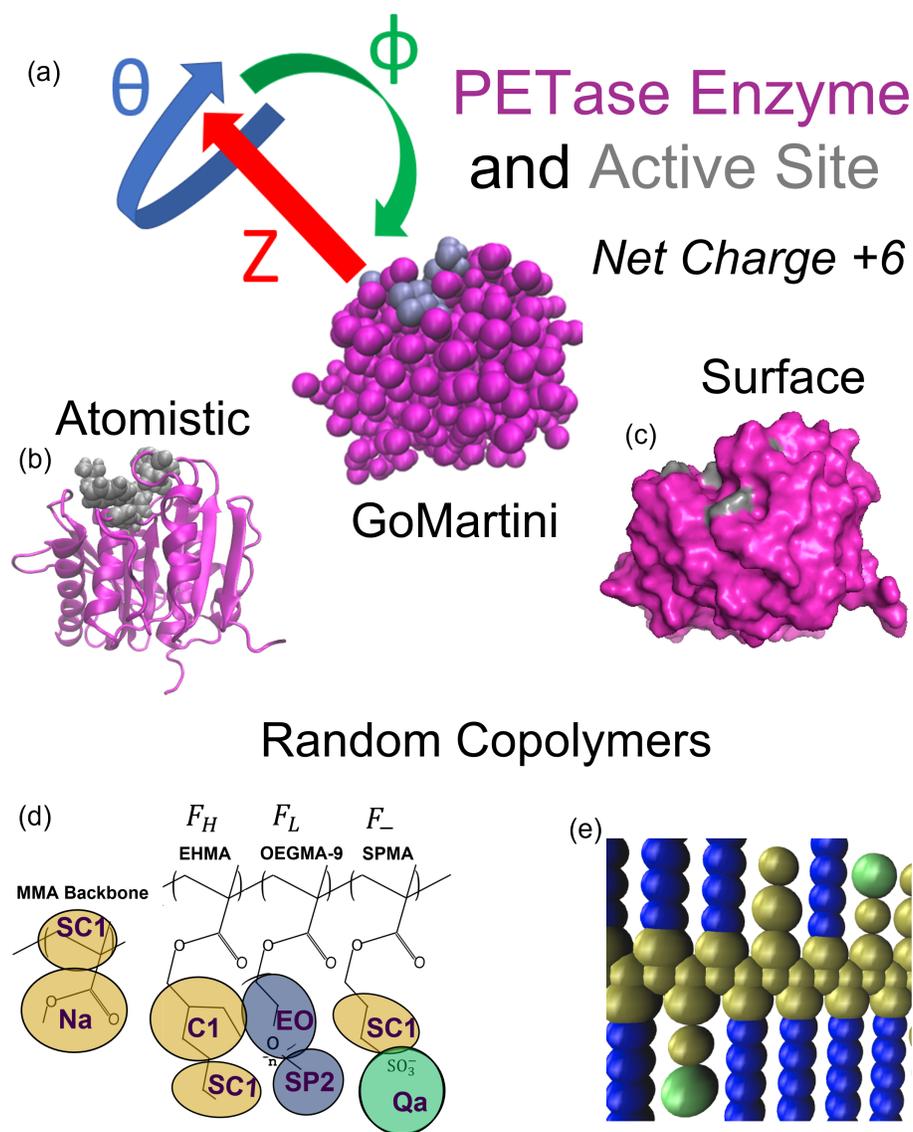


Figure 4.1. Description and models of PETase and the random copolymers. (a) GoMartini model of PETase in magenta with the active site in gray. (b) The secondary structure of PETase is shown with the same color scheme as in (a). The active site is shown using the van der Waals representation to highlight the cleft-like binding pocket for PET. (c) Surface representation of PETase. (d) Chemical and Martini description of the methacrylate-based random copolymers. Hydrophobic beads are tan, while hydrophilic beads are blue, and negatively charged beads are cyan. F_H , F_L , and F_- refer to the percentage of the respective monomer in the random copolymers. (e) Snapshot of the Martini random copolymer model with the colors corresponding to (d).

and they were also unable to study the effect of the polymers on the conformation of the enzyme active site, which we show is affected by the spatial distribution of protein-polymer contacts both at room and elevated temperatures. Specifically, we demonstrate that when these PETase-random copolymer complexes have an abundance of polymer-active site contacts, they can have less perturbed active sites than the protein by itself and remain stable as the temperature is increased.

4.3. Results and Discussion

4.3.1. PETase Structure at Elevated Temperature

PETase denatures as temperature is increased, leading to detectable decreases in activity above room temperature [52, 195]. In Figure 4.2(a), all-atom molecular dynamics simulations demonstrate the deformation of the active site at elevated temperatures. We use protein RMSD [12] to measure the conformation of the protein relative to the energy minimized crystal structure, not including rotational or translational diffusion. We measure the conformation of the whole protein as well as the active site using a previously published definition of the seven active site residues [50]. This provides metrics for the behavior of the whole protein and the active site, whose conformation relative to the crystal structure correlates with enzymatic activity [51].

As shown in Figure 4.2(a), there is an upward trend for both the whole protein and active site RMSD values as the temperature increases and PETase activity decreases. The active site RMSD at 330K is an outlier - despite decreasing from 320K, it is still higher than 298 or 310K. In agreement with studies highlighting the flexibility of the active site [51], we see that the active site is more perturbed by increases in temperature (especially at

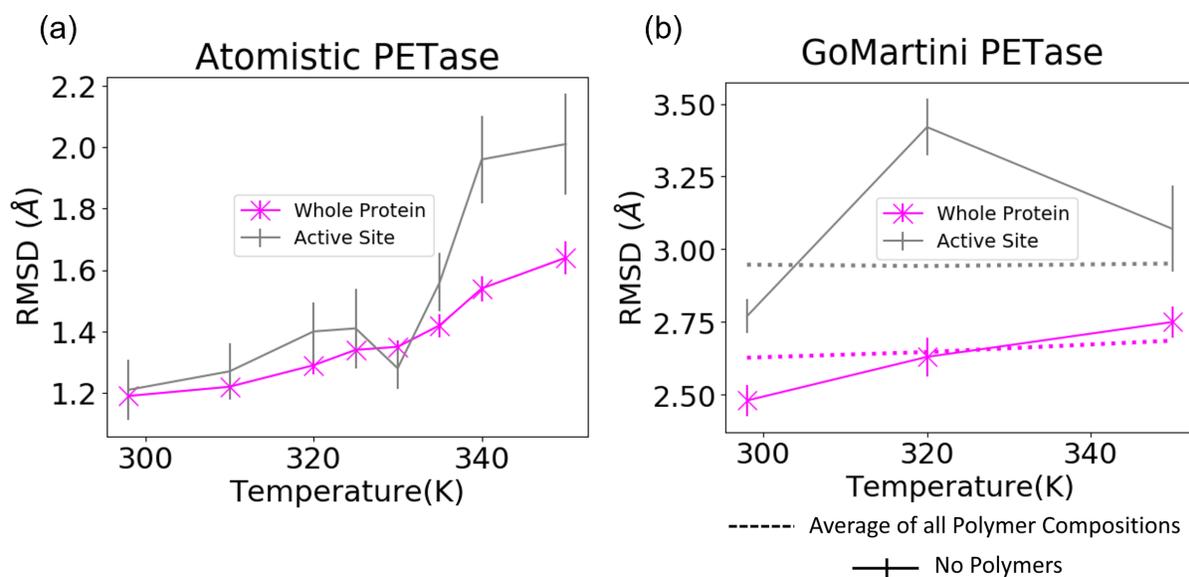


Figure 4.2. Comparison of the GoMartini PETase (b) alone to atomistic simulations (a) and to the GoMartini PETase complexed with random copolymers as the temperature is increased (b). RMSD is used to measure the conformation relative to the crystal structure with higher values signifying more deformation. (a) Results for the atomistic model show a general increase in the RMSD of the protein as well as the active site as is expected based on known decreases in activity with increasing temperature. (b) The GoMartini model shows an increase in RMSD for the whole protein and the active site at 320K, but at 350K the active site RMSD decreases, unlike the entire protein backbone. The active site RMSD may be inaccurate at high temperatures, but can still be used as a baseline to measure the thermal stability of the PETase-polymer complexes. In these complexes, temperature dependence of the whole protein and active site conformation is nearly eliminated.

the highest temperatures) when compared to the whole protein. In the following sections where random copolymers complex with PETase, we use the GoMartini [201] model of PETase, which agrees reasonably well with the temperature-dependent conformation of atomistic PETase (Figure 4.2(a)). As shown in Figure 4.2(b), the GoMartini PETase also shows an upward trend for both RMSD values with temperature as well as larger increases

for the active site than the whole protein. Although the RMSD values are different between the atomistic and GoMartini models, they are also measured on different time scales, 10 ns for the atomistic results and 300 ns for the GoMartini results (see Methods for more details), with longer times potentially leading to further perturbation of the protein structure. The ability to access these longer time scales is a great advantage of the GoMartini model. In agreement with atomistic simulations, the GoMartini model does have a noticeable increase at 320K. However, at 350K, the active site RMSD decreases instead of increasing. Here, the model is failing to accurately predict the active site behavior of PETase, although the whole protein RMSD displays the correct trend. Despite this issue, the GoMartini model still provides a baseline that can be compared to the case of polymer complexation to evaluate the ability of the polymers to stabilize the active site. This issue is also mitigated by restricting our analysis of the active site to 320K in some cases such as Figure 4.6(c).

4.3.2. Enzyme-Polymer Complexation

Since the stabilization of enzymes through microphase complexation depends on the structure of the adsorbed random copolymers, we seek to understand how to control the adsorption through mean composition of the polymer. We measure the adsorption *via* the number of contacts between the polymer and the enzyme at three set temperatures. The results at 298K are shown in Figure 4.3. The trends observed at this temperature continue at higher temperatures (see Appendix Figure 4.8).

As shown in Figure 4.3(a), the mean compositions with the most contacts between polymers and protein tend to be those with a relatively high percentage of negatively

(a) Total Normalized Polymer-PETase Contacts

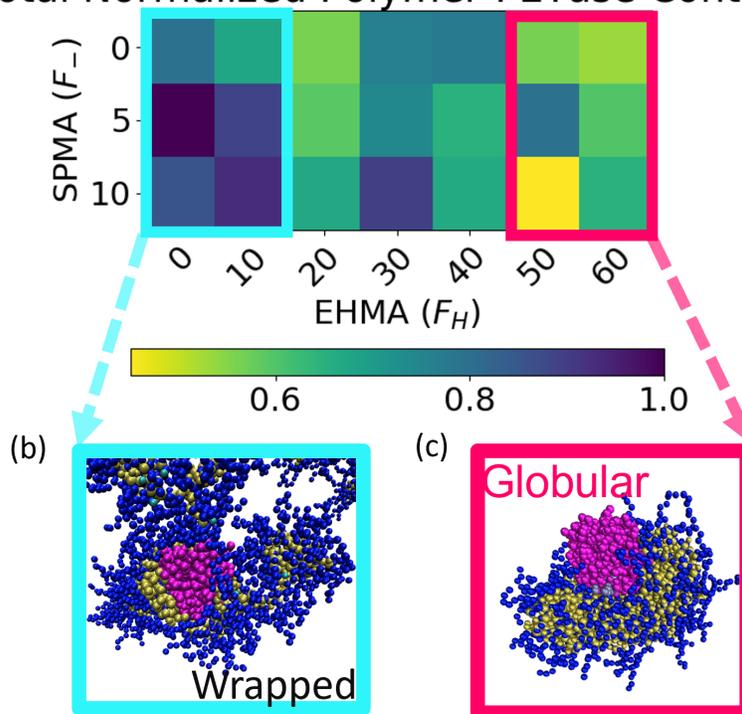


Figure 4.3. Random copolymer complexation with the PETase protein for different mean polymer compositions. F_H is the percentage of hydrophobic EHMA, F_- is the percentage of negatively charged SPMA and F_L is the percentage of hydrophilic OEGMA-9. These percentages sum to 100 and thus F_L , which is not displayed, is $100 - F_H - F_-$. (a) The number of contacts is shown as a function of composition at 298K. A maximum is observed at very low percentages of EHMA and higher percentages of SPMA, while a local minimum is observed at $F_H=20\%$. (b) Simulation snapshot of a wrapped polymer conformation, which occurs at low values of F_H and is characterized by a high percentage of contacts between the enzyme and polymer backbone. (c) Simulation snapshot of a globular polymer conformation, which occurs at high values of F_H and is characterized by micelle-like behavior of the amphiphilic polymers.

charged SPMA, F_- , and a lower percentage of hydrophobic EHMA, F_H . Thus, additional hydrophobic EHMA monomers are unnecessary for the polymer to complex with

the hydrophobic domains of the protein surface since the polymer backbone is already hydrophobic.

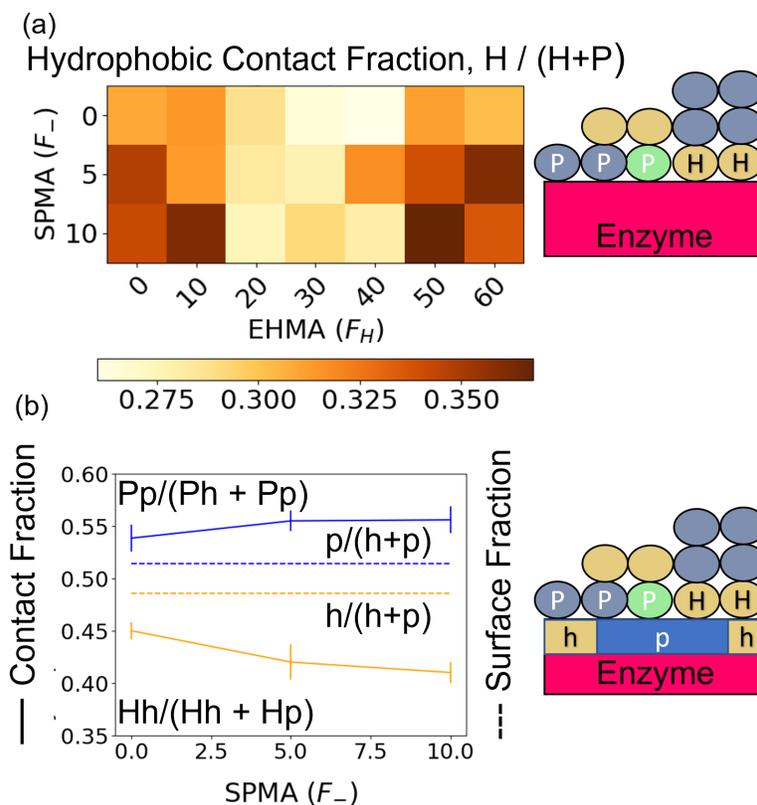


Figure 4.4. Hydrophobic interactions affect PETase-random copolymer complexation. (a) The fraction of contacts that involve hydrophobic polymer beads. There are two local maxima, one at very low F_H where the polymer backbone wraps around the protein surface and one at high F_H where the EHMA increases the baseline hydrophobic fraction of the polymer. (b) The fraction of these hydrophobic contacts that occur on the hydrophobic surface of PETase is lower than the hydrophobic surface fraction of PETase. This struggle to optimize the interaction could be related to ill-defined hydrophobic domains due to partially hydrophobic amino acids in the Martini model. However, polar-polar interactions seem to be optimized and this increases as charge is added, while the opposite occurs for hydrophobic-hydrophobic interactions.

This is further illustrated in Figure 4.4(a) by the fraction of protein-polymer contacts that involve a hydrophobic polymer bead either from the backbone of the polymer or the side chains. Local maxima of the fraction of hydrophobic contacts occur at both the highest and the lowest values of F_H . These low F_H conformations wrap around the PETase surface (see Figure 4.3(b)), as opposed to higher F_H compositions that lead to globular polymer conformations (see Figure 4.3(c)). Thus, the wrapped conformation of the low F_H random copolymers allow the hydrophobic backbone to more easily access the protein surface leading to more hydrophobic and total protein-polymer contacts. The connectivity of the hydrophobic backbone and hydrophobic globular conformations seems to lead to the creation of contacts with both the polar and hydrophobic parts of the protein as shown by the weaker hydrophobic correlations between polymer and protein (Figure 4.4(b)). The polar correlations are stronger, possibly due to reduced connectivity of the polar side chains and the ability to interact with the aqueous solvent instead of the protein surface. We note this analysis is highly sensitive to how the hydrophobicity of different beads is defined (see Methods).

The number of polymer-enzyme contacts grew as F_- increased, especially for wrapped conformations. This increased attraction between the enzyme and the random copolymers is intuitive since PETase has a +6 net charge. The increase in contacts occurs on the positively charged part of the surface as shown in Figure 4.5. Thus, the charge of the polymer influences the spatial distribution of the polymer on the dipolar PETase surface [50] and the overall number of contacts. There is also some competition between optimizing charged and hydrophobic interactions as the addition of charge slightly weakens the hydrophobic correlations (Figure 4.4(b)). These contacts were biased to a specific

Additional Contacts (ACs) of Negatively Charged Polymers Are Located on Positive Protein Domains

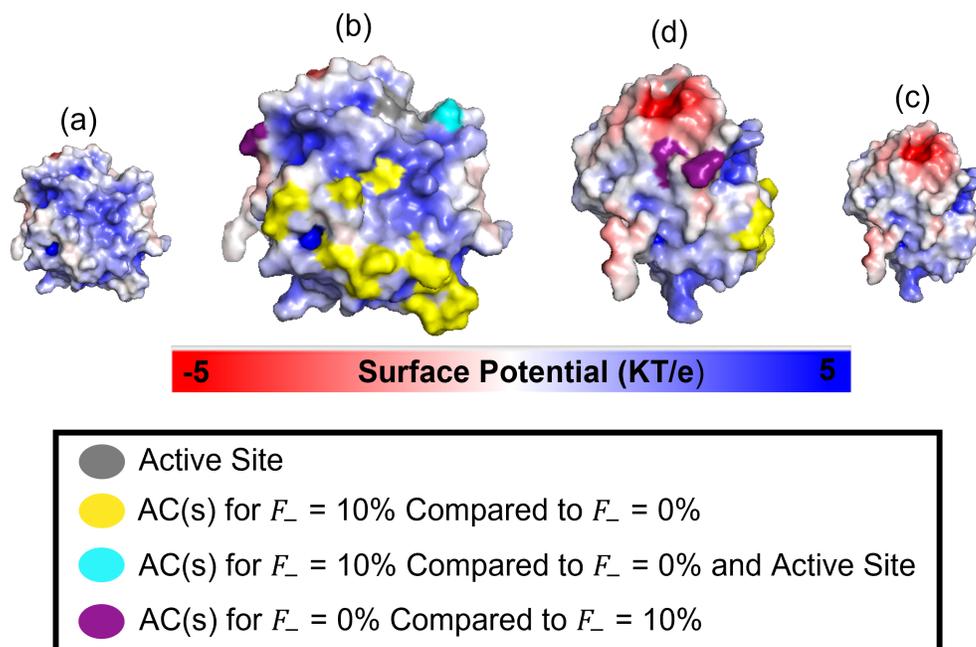


Figure 4.5. Including negatively charged monomers increases contact with positive surface domains. (a) and (b) show PETase with only the surface potential and with additional contacts (ACs) overlaid respectively. One can see that the yellow and cyan sites preferred when $F_- = 10%$ are overwhelmingly on the positive part of the protein due to the addition of negative charge to the polymer. (c) and (d) show the a rotated orientation of the protein. The full data set for this figure, more method description, and higher temperature results can be found in the Appendix Figure 4.9.

protein domain despite the randomness of the polymer sequence, suggesting that using more controlled polymerization or peptide engineering is not necessary to intentionally contact these charged domains. It also suggests that the inverse can be achieved, i.e., a certain protein domain could be targeted by engineering the surface potential. Experiments have also demonstrated that protein charge domains are crucial in macrophase behavior of protein-polyelectrolyte complexes [30, 199]. This is also true of polymer

blockiness [120, 42, 22], or the tendency of like monomers to be grouped together in the polymer sequence, in phase separation of polyelectrolyte complexes especially as ionic conditions are varied [25, 23, 89].

4.3.3. Stability of the Complexed Enzyme

In Figure 4.2(b), we measure how complexation affects the conformation of the protein at room and elevated temperatures, using the RMSD values for the protein and the active site. These RMSD values are a measure of the protein conformation relative to the PETase crystal structure with lower values representing less perturbed active sites and thus a more active enzyme.

We find that the random copolymers are very effective at preventing PETase from deforming as the temperature is increased. This is shown by the RMSD values for the active site and the whole protein averaged over all polymer compositions and comparing with the results for PETase without polymers (Figure 4.2(b)). The RMSD values as a function of temperature with polymers bound is almost flat for the whole protein and the active site; the actual values for the active site and the whole protein are only slightly greater than the protein alone at 298K. The perturbations caused by polymer binding are less than the perturbations seen at elevated temperatures as at 320K. Thus, the stability of the active site is much improved with the addition of polymers. The distribution of active site RMSD values at 320K shows that for nearly every polymer composition the active site RMSD is lower than without polymers (Appendix Figure 4.10(a)). The RMSD values for the active site also have a much broader range when compared to the values of the whole protein (Appendix Figure 4.10(b)). Further, this range includes

some active site RMSD values that are lower at room temperature with polymers than without. This wide range of RMSD values for the active site is not well explained by trends in mean polymer composition (Appendix Figure 4.11). To explain the wide range of active site behavior we look at the location of the contacts for those compositions where active sites are less perturbed than the PETase alone at room temperature and compare to compositions where active sites are more perturbed than the PETase alone at room temperature (see Figure 4.6). We find that less perturbed compositions have significantly more contacts near or on the active site of the enzyme at both 298K (Figure 4.6(c)) and 320K (Figure 4.6(d)). This suggests that contacts on the active site stabilize instead of perturb the active site. Thus, although we found a correlation between polymer contacts on the positive section of the enzyme and mean polymer composition in Figure 4.5, these contacts were not concentrated on the active site, which explains the absence of correlation between mean polymer composition and active site conformation (Appendix Figure 4.11). In other words, the composition of the polymers (especially charge) biases the spatial distribution of contacts, but it does not bias these contacts to the active site, where they provide stability. This does not necessarily have to be the case for PETase or any given enzyme, since previous work has shown that complexation can be influenced by engineering the net charge on various proteins [35, 36]. A similar strategy could be used to change the charge distribution near the active site. Thus, we suggest that engineering the spatial distribution of charges near the active site could increase the activity of an enzyme-polymer complex by biasing the charged polymers near the active site and stabilizing the active site. While active site stabilization due to local contacts may not be completely general, previous studies have illustrated that the addition of

polymers to enzymes can enhance the activity at elevated temperatures in water [202, 31], suggesting the phenomenon extends beyond PETase. Moreover, most active sites are partially hydrophobic [203, 204, 205, 206], suggesting that polymers with hydrophobic groups can stabilize the active site with local contacts as shown here.

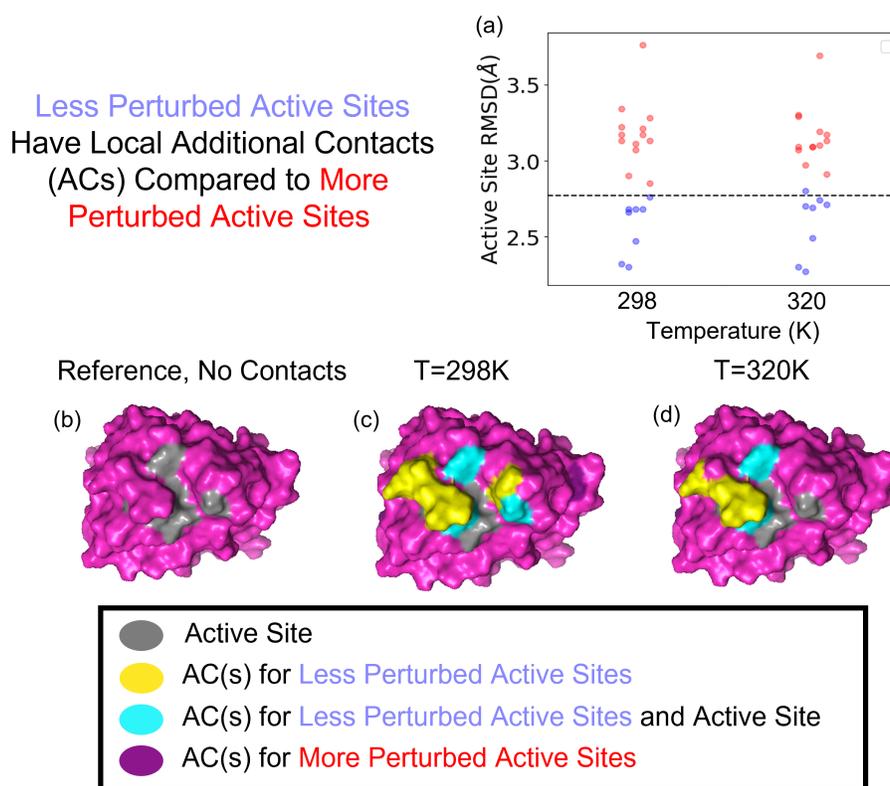


Figure 4.6. Less perturbed active sites are stabilized by additional contacts near the active site. (a) The distribution of active site RMSDs at 298K and 320K. The less perturbed active sites are colored in blue while the more perturbed active sites are colored in red. Each point refers to a different polymer composition and the line refers to PETase alone at 298K. These are the groups being compared in (c) and (d). (b) Surface representation of PETase with no excess contacts shown for comparison. (c) Comparison at 298K shows many additional contacts (ACs) around the active sites for less perturbed compositions. These contacts stabilize the active site instead of further perturbing it at 320K (d) as well. The full data set for this figure and more method description can be found in the Appendix, Figure 4.13.

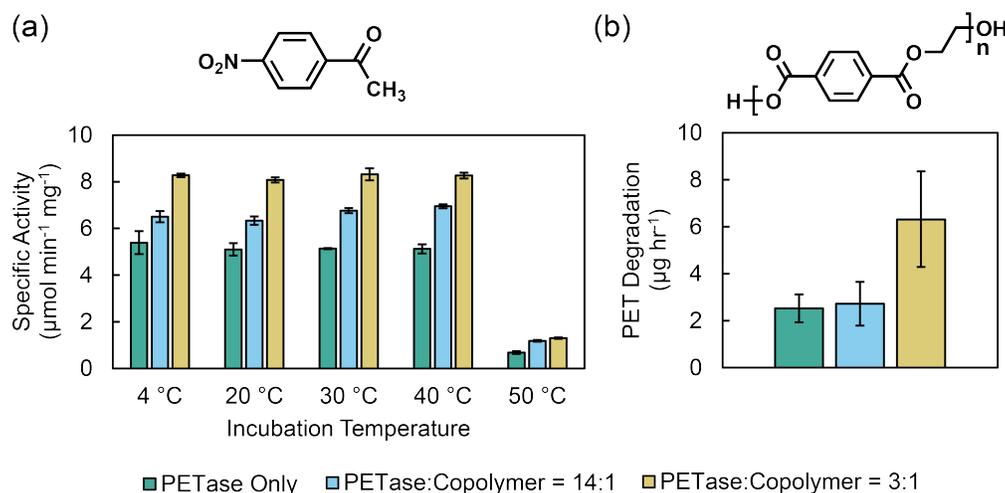


Figure 4.7. Activity of PETase and PETase/copolymer complexes at two different PETase:copolymer molar ratios. (a) Specific activity against small molecule substrate p-nitrophenyl acetate after one hour incubation at various temperatures. Error bars represent standard deviation over 3 replicate experiments (b) PET degradation activity over five hours at 35 °C. Error bars represent the 95% confidence interval on activity values.

4.3.4. Activity of Enzyme-Polymer Complexes on Small Molecule and Solid Substrates

We performed experiments to assess the impact of active site stabilization on enzyme activity. We first examined the temperatures at which our PETase/copolymer complexes are stable. We incubated PETase or PETase/copolymer mixtures at a range of temperatures (4-50 °C) for 1 hour, and assayed esterase activity at room temperature against the small molecule substrate, p-nitrophenyl acetate (Figure 4.7(a)). Here, we used a small molecule substrate to avoid confounding effects of temperature on PETase's PET degrading activity (an established phenomenon [52, 195]). The copolymer used in these studies has a mean composition with $F_H = 43\%$ and $F_- = 12\%$, within the normal range of F_H

in the simulations and a reasonably similar value of F_- (see Materials and Methods section for further detail). Interestingly, we found that for all incubation temperatures, the specific esterase activity of PETase was enhanced in the presence of copolymer, and that this activity enhancement increased at higher copolymer concentrations (Figure 4.7(a)). Our data also indicate that PETase activity in PETase/copolymer complexes is stable at temperatures up to 40 °C. Notably, this is less than the glass transition temperature of PET (70 °C). While the activity enhancement with the addition of copolymer could improve the utility of the enzyme, further thermal stabilization of the PETase enzyme is likely needed to realize the full potential of this system. For example, polymers with side groups that penetrate PET could be used to decrease its T_g .

Next, we examined the activity of PETase/copolymer complexes on solid PET as a substrate over five hours at 35 °C (Figure 4.7(b)). The presence of copolymer enhances the activity of PETase towards PET, similar to the results obtained with our small molecule activity assay. We confirmed that none of the formulations tested exhibited any decrease in esterase activity after incubation at 35 °C for five hours (Figure S8 in the original manuscript [162]), suggesting that the enhanced enzyme activity with copolymer is not due to changes in the temporal stability of the enzyme. Thus, our experimental data suggest that PETase activity is improved upon addition of copolymer regardless of the substrate. These results, in conjunction with our simulation studies, suggest that copolymer binding impacts active site conformation, thus altering PETase activity. We observe a wide distribution of active site conformations in simulations in the presence of copolymer compared to the naked enzyme (Figure 4.2(c)). While some of these active site conformations may be less active than the native enzyme, we hypothesize that there may

be a few active conformations present at any given time that confer substantially higher enzyme activity. An alternative hypothesis is that the random copolymers interact favorably with the substrate leading to an increase in local substrate concentration. Indeed, it has been shown that these random copolymers can bind hydrophobic small molecules [80], and that IDRs in proteins, similar to the random copolymers in Figure 4.3(b), promote higher order assembly [207]. However, because the increase in PETase activity appears to be independent of substrate (solid PET or aqueous p-nitrophenyl acetate), it is more likely that the interactions responsible for this activity enhancement are between the PETase and the random copolymer rather than the random copolymer and the substrate.

4.4. Conclusions and Outlook

The ability to use various types of enzymes in industrial conditions could dramatically impact biotechnology, pharmacology, and bioremediation. Here, we show that (i) PETase is functionally stabilized by complexation with industrially scalable random copolymers. We demonstrate this through simulations of whole protein and active site conformations and using experiments measuring enzyme activity in response to thermal challenge; (ii) this effect is further enhanced when random copolymers form more contacts with the enzyme active site; and (iii) polymer composition biases the conformation and location of the random copolymers on a protein despite the randomness of the polymer sequences. Thus, engineering the surface potential of the protein, using standard protein mutation or modification techniques, could bias polymers to bind near the active site, further increasing enzyme activity. This approach may be compatible with a variety of enzymes as long as their surfaces can be engineered without misfolding. However, the use of copolymer

complexation to modify active site conformation does not necessitate further protein mutation or modification as the diverse functional groups on the copolymer should permit binding to a variety of protein surfaces. Further, this approach can also be used in the absence of information on the sequence and structure of an enzyme. Thus, this strategy has the potential to be a widely accessible route to increasing enzyme functionality, especially given that it requires only a random polymerization. Future experiments using substrates with various physical properties could elucidate the role of interactions between substrates and random copolymers, opening a new avenue for engineering enzyme activity via complexation with copolymers. These interactions could be controlled by using monomers with specific affinities and promote assembly on solid substrates like PET films.

Complexation with random copolymers is a welcome addition to the enzyme engineering toolbox, offering an orthogonal, versatile strategy for increasing functionality either in conjunction with or independent of other protein engineering techniques.

4.5. Methods

4.5.1. Atomistic Simulations

We performed all atom molecular dynamics simulations to study PETase in aqueous solution using the package GROMACS (version 2016.3) [208]. We used the most recent CHARMM [209] forcefield. The recommended CHARMM TIP3P water model [210] was applied with the structures constrained via the SETTLE algorithm [211]. For simulations at different temperatures only the temperature was varied and all other parameters remained the same. The periodic boundary conditions were employed in all

dimensions. The neighbor searching was calculated up to 12 Å using the Verlet particle-based method and was updated every 20-time steps. The Lennard-Jones (LJ) 12-6 interactions were switched off from 10 to 12 Å via the potential-switch method in GROMACS. The short-range Coulomb interactions were truncated at the cut-off distance of 12 Å, and the long-range interactions were calculated using the Smooth Particle Mesh Ewald (PME) algorithm [212, 213]. The NPT ensemble (constant number of particles, pressure, and temperature) was employed. The temperature was coupled using the Nosé-Hover algorithm (reference temperature was varied, characteristic time 1 ps). The isotropic Parrinello-Rahman barostat was employed with the reference pressure of 1 bar, the characteristic time was 4 ps, and the compressibility of $4.5 \times 10^{-5} \text{ bar}^{-1}$. All the covalent bonds were constrained, which supported an integration time step of 1 fs. These parameters were recommended for the accurate reproduction of the original CHARMM simulation on lipid membranes [214], and have been verified in our simulations on proteins [85, 215, 198, 216] and lipid membranes [217].

Ten different 10 nanosecond simulations, containing only the protein, water, and the counterions necessary for electroneutrality, are carried out at each temperature (298K, 310K, 320K, 325K, 330K, 335K, 340K, and 350K) and data is collected for the last 5 nanoseconds. Using GROMACS[12], the root mean square deviation (RMSD) is all atoms relative to that atom in the energy minimized crystal structure with rigid body translation and rotation accounted for. Thus, this is not a measure of diffusion, but a measure of how conserved the PETase conformation is with lower values being more conserved. The value for each atom is then averaged over the entire protein for the “Whole Protein RMSD” (ignoring the first 20 and last five residues, which are less structured and

thus have high fluctuations leading too large errors). For the “active site RMSD” we include only the atoms in the active site, according to previous definitions of the active site [50]. The RMSDs of the PETase active site are of particular interest since they correlate with observed catalytically inactivity of PETase at temperatures above room temperature. Error bars are the standard error based on the number of independent runs, $N=10$.

4.5.2. GoMartini PETase Only Simulations

To model the complexation of the random copolymers with the PETase (PDB: 6EQE) [50] surface, we use the Martini [164, 170] forcefield with polarizable water [171]. The temperature dependant melting behavior of PETase is captured by the GoMartini [201] protein model. This model builds on network-based models [218] by replacing harmonic bonds between non-covalently bonded residues with Lennard-Jones interactions where epsilon is 12 kJ/mol [219], increasing the ability of the protein to denature including at elevated temperatures. Remarkably, it has been very recently demonstrated that the GoMartini model with these parameters is capable of protein configuration change for several mutations of copper, zinc, superoxide dismutase, a protein associated with neurodegenerative disorder amyotrophic lateral sclerosis. The topology files for the GoMartini 2.2 model were generated using the program `go_martinize.py` and the epsilon value of 12 kJ/mol. The program `go_martinize.py` was modified such that the neighboring contact map was built up to a cut-off distance of 1.1 nm [219, 220, 39]. In the production simulations, the recommended parameters [221] for the Martini 2.2 potential were employed, which are summarized here. The short-range Coulomb interactions were calculated up to 1.1 nm

with the particle mesh ewald summation (relative permittivity of 2.5) for the long-range electrostatic interactions. The LJ 12-6 potential interactions were truncated at 1.1 nm. The NPT ensemble was applied. The temperature was coupled at 298, 320, or 350K using the velocity rescaling method. The isotropic pressure coupling (reference pressure 1 bar, time constant 5.0 ps, compressibility $3 \times 10^{-4} \text{ bar}^{-1}$) was employed using the Parrinello-Rahman algorithm. The leapfrog integration time step of 10 fs was employed.

We use a similar methodology for the GoMartini as for the atomistic simulations. We simulate only the protein, polarizable water, and the counterions necessary for electroneutrality. Five independent simulations are run for 300 nanoseconds, and data is collected for the second half of the simulation. We calculate the root mean square deviation (RMSD) between the beads which comprise the PETase protein in the molecular dynamics and the respective energy minimized crystal structures using GROMACS[**12**]. Thus, the RMSD is a measure of how conserved the PETase conformation is with lower values being more conserved. We then average this quantity over the beads in the GoMartini PETase model (once again ignoring the first 20 and last 5 residues) and all beads in the active site providing us with two different RMSD values corresponding to different parts of the PETase. The active site group uses the same definition of the 7 active site residues (Thr88, Trp159, Ser160, Trp185, Asp206, His237, Ser238) as the all-atom RMSD calculations. Error bars are the standard error based on the number of independent runs, $N=5$.

4.5.3. Martini Simulations of PETase and Random Copolymers

The Martini model for the monomers comes from a variety of sources: one for the methacrylate backbone [126], one for the PEO oligomers [130], and another for the rational approach to coarse graining organic molecules including those with charge [164].

Each simulation involves PETase and four copolymers with a degree of polymerization (DP) of 100 monomers. Sequences are built at random under the constraint of a given mean composition. This is accomplished by creating a computational pool of 400 monomers ($N=100 * 4$ polymers) with the given ratio of monomer types. Monomers are randomly chosen one at a time and added to the growing chains until no monomers are left. The result is 4 chains with completely random sequences that likely do not have the same composition, but have the given mean composition as a group. The counterions included are those necessary to neutralize the positively charged protein and one counterion for each negatively charged polymer monomer. The compositions constrain the total percentages of monomers in the box while individual copolymers have gaussian fluctuations in composition. This also means that the volume fraction of polymers plus PETase fluctuates, but is around 5 percent. For all 21 compositions, the simulation is initialized with only the PETase and polymers allowing the copolymers to quickly contact PETase. This box is then solvated and cyclically annealed from 298K to 350K and back down three times at a rate of one cycle per 80ns. Twelve configurations are taken from the last half of the 240ns annealing process. For each of these configurations, the simulation is run for an additional 50ns at each temperature (298K, 320K, and 350K), a total of 1800ns per composition. From these simulations, all relevant values are measured and averaged by temperature.

4.5.3.1. Definition of Polymer-Protein Contacts. Like in the calculation of the hydrophobic surface fraction, a contact occurs between two beads when they are less than $.53 \text{ nm}$ from each other. These contacts can either be hydrophobic-hydrophobic (Hh), hydrophobic-polar (Hp), polar-polar (Pp), or polar-hydrophobic (Ph). This is determined by bead type, not monomer or amino acid type, consistent with the calculation of the hydrophobic surface fraction. Again we use the definition that “C” or “N” beads are hydrophobic and all others are polar [164]. Thus most polymer monomers and many amino acids contain both polar and hydrophobic beads [218].

4.5.3.2. Calculation of Hydrophobic Surface Fraction. In Figure 4.4(b) in the main text we calculate the hydrophobic fraction of the surface of PETase as follows. First we define which beads are in the surface. These are beads which make, over a time average, one contact with a water bead. A contact means that the PETase bead is within the $.53 \text{ nm}$, the equilibrium separation distance, of a water bead. Then, all beads defined to be in the surface are labeled as either polar or hydrophobic. There are 4 general types of beads in the Martini forcefield: “Q”, “P”, “C”, and “N”. We consider “C” and “N” to be hydrophobic and others to be polar [164]. The hydrophobicity of the polymer beads is defined in the same way. The fraction is just the number of beads considered to be hydrophobic divided by the total number of beads. The fraction calculated, $.49$, is significantly higher than $.28$ [198], which has been reported previously in atomistic simulation for the hydrophobic surface fraction of PET. This is based on a different definition of hydrophobic which labels each bead based on the identity of the amino acid they are a part of. In the Martini model, amino acids often contain both hydrophobic and polar beads making the definitions non-equivalent. These definitions accounts for the vast

majority of the difference since, by using the other definition of hydrophobicity on the martini model, we get a surface fraction of .32, in reasonable agreement with atomistic result. We think the bead-by-bead definition makes more physical sense for the Martini simulations, especially since our polymer monomers also contain both hydrophobic and polar beads, which we show is very important to explain the behavior of the system. In other words, we perform the analysis at the length scale of the coarse graining.

4.6. Appendix

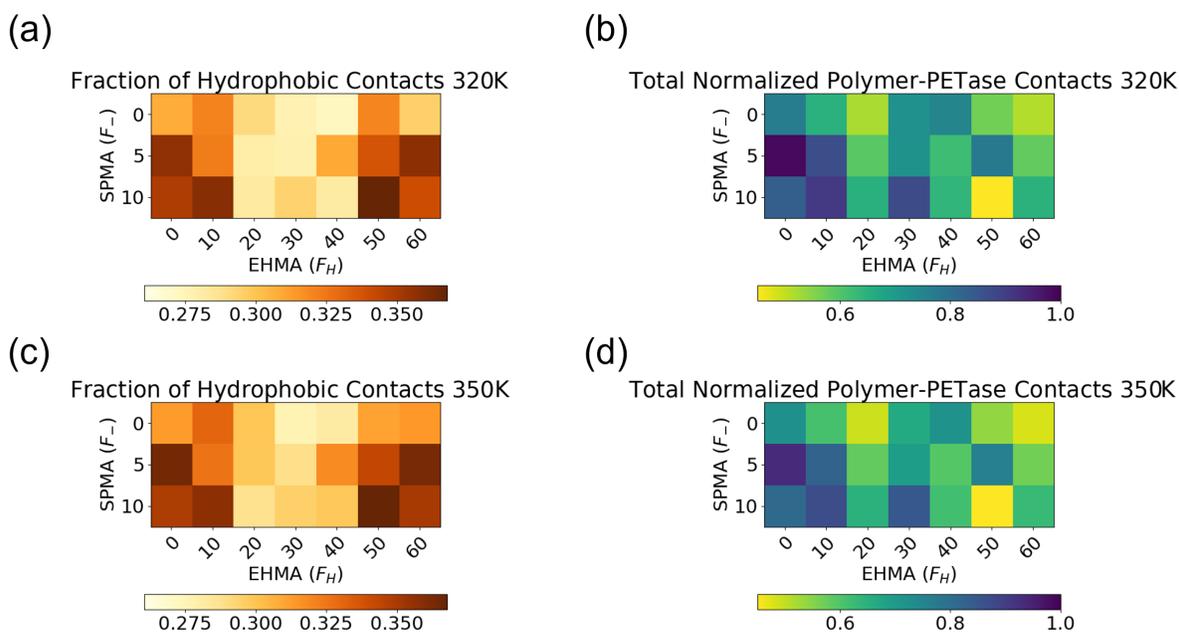


Figure 4.8. Contacts at Elevated Temperature. At 320K and 350K the same trends are observed in terms of which compositions lead to the most contacts and the fraction of those contacts that are hydrophobic. In general, the total number of contacts does decrease for every composition as temperature is increased, while the fraction of those contacts that are hydrophobic increases.

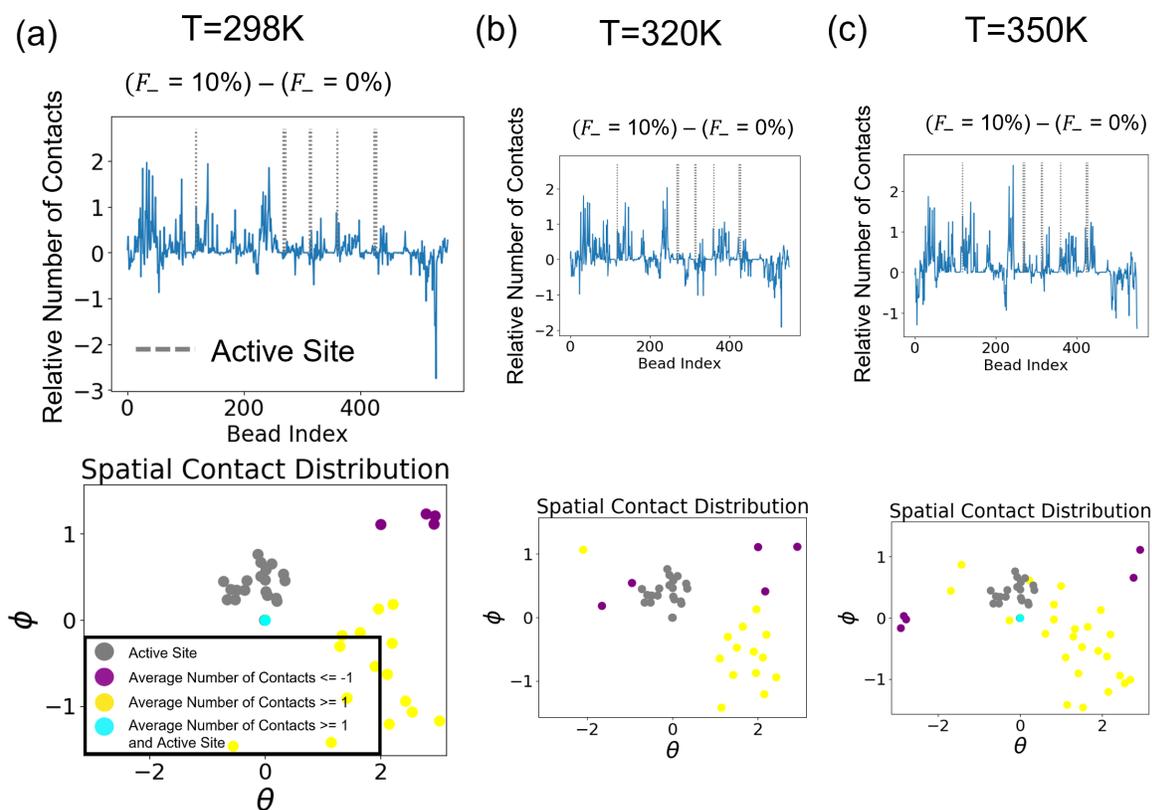


Figure 4.9. Charged Polymers Affect Spatial Distribution of Contacts. In Figure 4.5 in the main text, we show that the positive section of the protein is more contacted when negative charges are present on the polymers. Here, we explain the full methodology. The first step is calculate the average number of contacts between the random copolymers and every single bead that makes up the GoMartini PETase. Then we average that over all compositions with $F_- = 0\%$ and do a separate average over all compositions with $F_- = 10\%$. Then we subtract the $F_- = 0\%$ data from the $F_- = 10\%$ data. That is what is shown at the top of (a), (b), and (c) with the active site shown in dashed gray lines. At the bottom of (a), (b), and (c) We then take that data and graph the spatial position of those bead in spherical coordinates if they meet the cut off threshold of having an absolute value equal to or greater than 1 as described in the legend at the bottom of (a). In Figure 4.5 in the main text, any residue featuring a bead that reaches a threshold value at the bottom of (a) at 298K is colored over the electrostatic surface potential of the protein. This shows that these excess contacts for $F_- = 10\%$ occur on the positive part of the protein. In (b) and (c), we also show this spatial distribution for 320K and 350K. This shows that although the exact beads that break the threshold value change, the preference for the same region of the protein in spherical coordinates is clear.

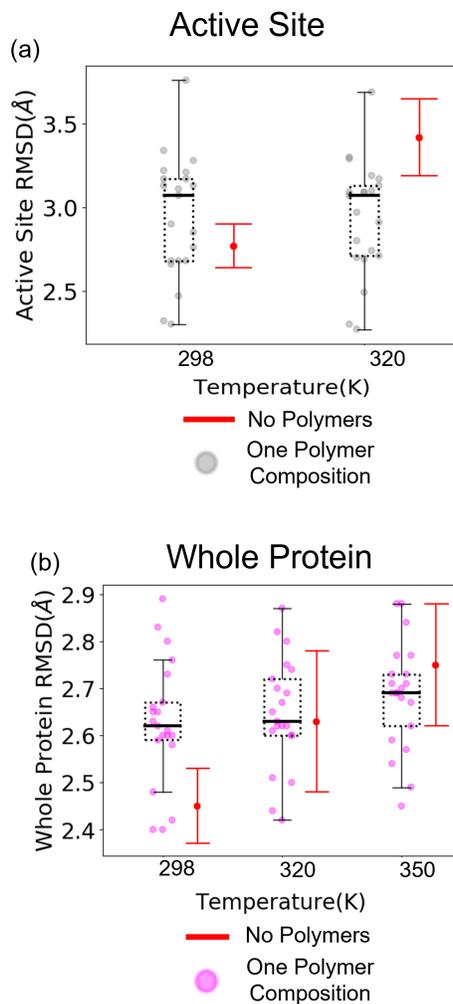


Figure 4.10. The distribution of RMSD values when polymers are attached. The data for attached polymers is shown as a boxplot with individual data points corresponding to different polymer compositions. The box extends from the first quartile of the data to the third quartile meaning half of the distribution is inside the box. A line is drawn at the median. The protein by itself is shown in red as a single average point and the distribution represented by the standard deviation and thus 66% of the data is inside the red bars. These data as a function of composition can be found in the Appendix Figures 4.11 and 4.12 respectively. (a) The distribution of the active site RMSD values at 298K and 320K. 350K is not shown due to inconsistencies in the active site behavior of the atomistic and GoMartini models at 350K. The distribution is much wider when polymers attached suggesting the active site conformation is modified and not just “trapped” by the presence of the polymers. (b) The distribution of the whole protein RMSD values at 298K, 320K, and 350K. This distribution when polymers are attached is narrower than the active site RMSD in (a).

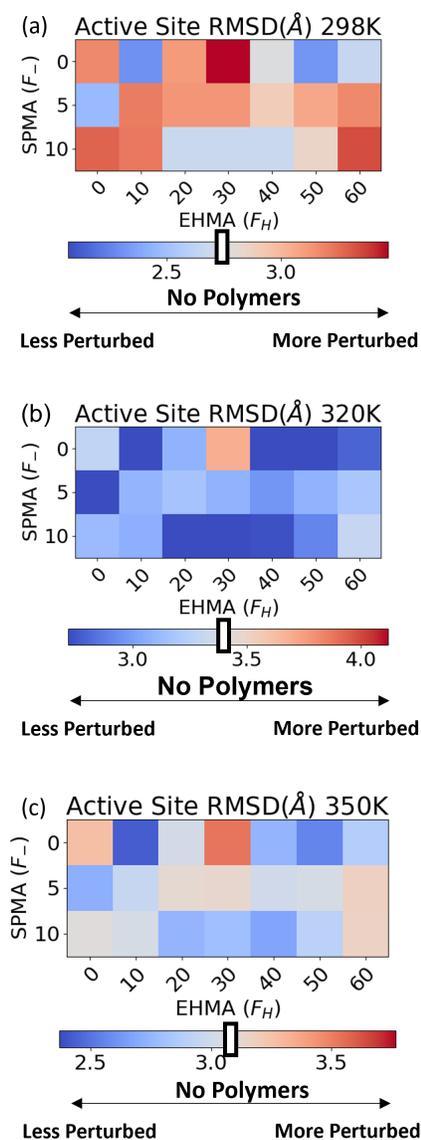


Figure 4.11. Active site RMSD as a function of polymer composition at 298K (a), 320K (b), and 350K (c). For each temperature the colorbar is adjusted to be centered at the active site RMSD value of the GoMartini PETase active site with no polymers bound. This is done to provide a direct comparison between the bound states and unbound state of PETase. For each temperature, there is little to no trend in active site RMSD based on the polymer composition. In the main text we describe how this is likely due to an inability of composition to control contacts specifically on the active site. However, above 298K the active site RMSD tends to be lower than the PETase alone across compositions. This is the same data that is displayed in Figure 4.10(a).

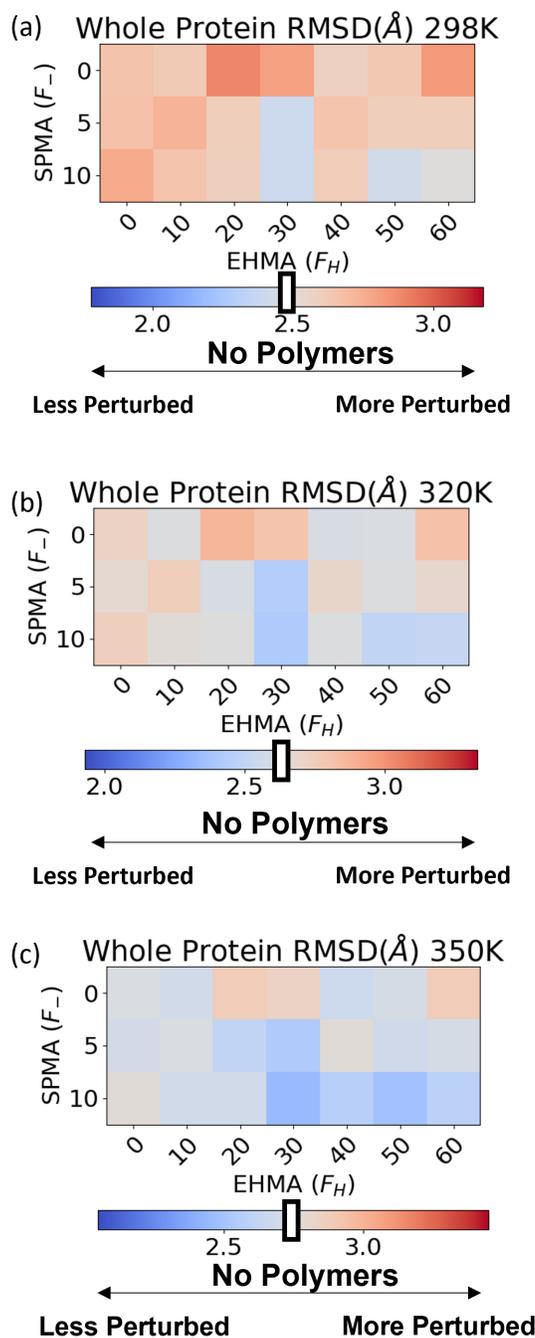


Figure 4.12. Whole protein RMSD as a function of polymer composition at 298K (a), 320K (b), and 350K (c). For each temperature the colorbar is adjusted to be centered at the whole protein RMSD value of the GoMartini PETase active site with no polymers bound. This is done to provide a direct comparison between the bound states and unbound state of PETase. For each temperature, there appears to be a slight trend where compositions with more charge and hydrophobicity have lower RMSD values. This is the same data that is displayed in Figure 4.10(b).

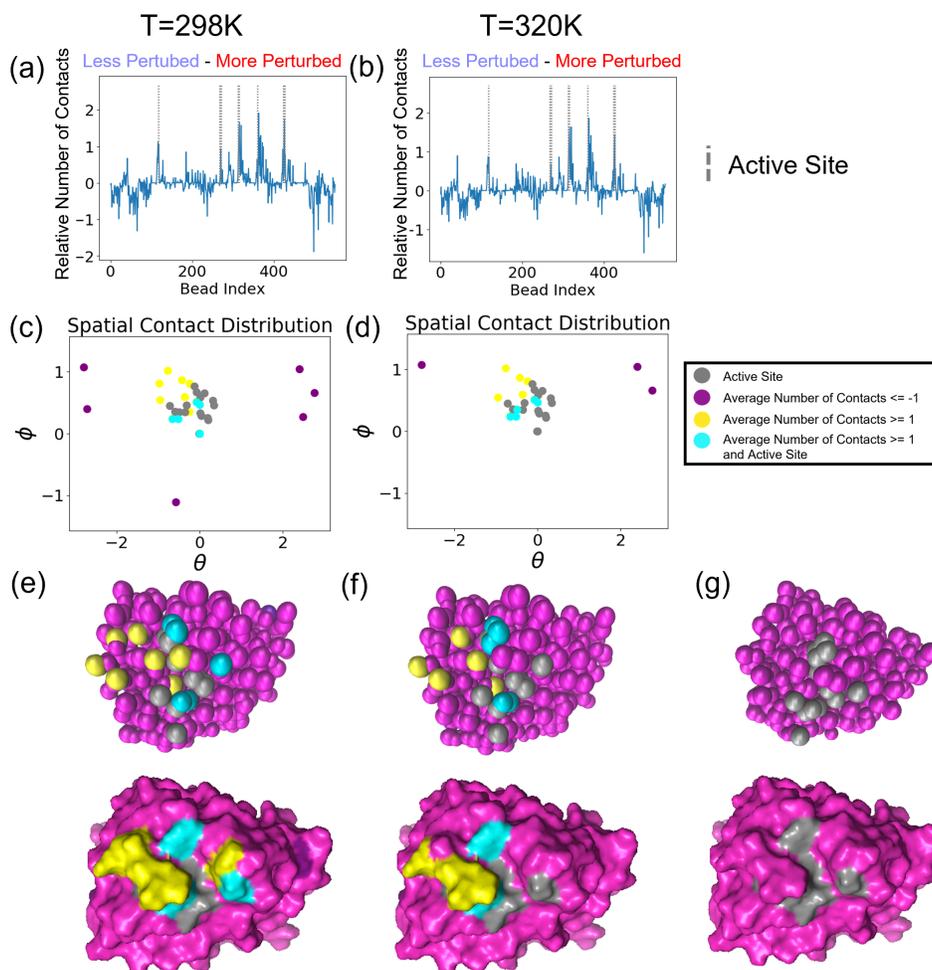


Figure 4.13. Active site conformation is affected by polymer contacts. In Figure 4.6 in the main text, we show how compositions with less perturbed active sites have more contacts very near the active site than those with more perturbed active sites. Here, we show the raw data for that figure. In (a) and (b) we see the subtraction of the average number of contacts for compositions with more perturbed active sites from those with less perturbed active sites. In (c) and (d) we apply the threshold absolute value of one and graph the beads which meet those values in spherical coordinates. The yellow dots are positive meaning that these spots have at least one more polymer contact on average for less perturbed active sites. The purple dots are negative meaning that they have at least one more contact when the active site is more perturbed. The positive - negative convention is just based on how the subtraction was done. In (e) and (f) these distributions are shown on the GoMartini model to get a fuller sense of where the contacts are relative to the active site. (g) is just the GoMartini model and is provided as a comparison.

Part 2

Crystalline Protein Shells

CHAPTER 5

Vertex Protein PduN Tunes Encapsulated Pathway Performance by Dictating Bacterial Metabolosome Morphology

This chapter is based on the published work [1] of Carolyn E. Mills, Curt Waltmann, Andre G. Archer, Nolan W. Kennedy, Charlotte H. Abrahamson, Alexander D. Jackson, Eric W. Roth, Sasha Shirman, Michael C. Jewett, Niall M. Mangan, Monica Olvera de la Cruz, and Danielle Tullman-Ercek, **Vertex protein pduN tunes encapsulated pathway performance by dictating bacterial metabolosome morphology**, *Nat. Comm.*, 2022, 13, 3941, with modified details and additional atomistic free energy calculations.

5.1. Abstract

Engineering subcellular organization in microbes shows great promise in addressing bottlenecks in metabolic engineering efforts; however, rules guiding selection of an organization strategy or platform are lacking. Here, we study compartment morphology as a factor in mediating encapsulated pathway performance. Using the 1,2-propanediol utilization microcompartment (Pdu MCP) system from *Salmonella enterica* serovar Typhimurium LT2, we find that we can shift the morphology of this protein nanoreactor from polyhedral to tubular by removing vertex protein PduN. Analysis of the metabolic function between these Pdu microtubules (MTs) shows that they provide a diffusional barrier capable of shielding the cytosol from a toxic pathway intermediate, similar to native

MCPs. However, kinetic modeling suggests that the different surface area to volume ratios of MCP and MT structures alters encapsulated pathway performance. Finally, we report a microscopy-based assay that permits rapid assessment of Pdu MT formation to enable future engineering efforts on these structures.

5.2. Introduction

Spatial organization of biological processes is essential to life across many organisms, from multicellular eukaryotes to unicellular prokaryotes. Once thought to lack subcellular organization, bacteria utilize an array of strategies for segregating specific processes within the cell. One such example is bacterial microcompartments (MCPs), which are organelles that encase specific sets of enzymes in a protein shell [222, 55]. Genes associated with MCPs are found in 45 bacterial phyla [223, 224], and are classified by the metabolic pathway segments they encapsulate. At the highest level, MCPs are classified as either carboxysomes or metabolosomes based on whether they encase pathways involved in anabolic or catabolic processes, respectively. Carboxysomes aid many carbon-fixing bacteria by increasing CO_2 concentration in the vicinity of the carboxylating enzyme ribulose biphosphate carboxylase/oxygenase (RuBisCO) [225, 226]. Metabolosomes, on the other hand, aid in metabolism of a broad array of substrates and thus encapsulate many different pathway chemistries; however, these pathways typically share a unifying feature of passing through a toxic aldehyde intermediate [227, 228]. Sequestration of this toxic intermediate is thought to aid in metabolism of niche carbon sources such as 1,2-propanediol and ethanolamine, providing a competitive growth advantage to the enteric pathogens that often harbor metabolosomes [229, 230].

MCPs represent attractive engineering targets in a variety of applications, from bio-production, where heterologous enzyme encapsulation could improve pathway performance [231], to antibiotic development, where disruption of these MCP structures could eliminate a competitive growth advantage [229]. However, metabolosomes in particular exhibit diversity in shape and size, and it is not well-understood how these features relate to function [224, 232, 233, 234, 235]. A variety of engineering fields, from catalysis [236] to drug delivery [237], have illustrated the importance of shape and size on nanomaterial performance. The relevance of these features has yet to be meaningfully investigated in MCP systems.

The 1,2-propanediol utilization (Pdu) MCP is a model metabolosome that aids in breakdown of 1,2-propanediol [238]. Pdu MCPs exist in a variety of bacteria [223, 224, 230] and both the encapsulated pathway [230, 239, 238] and the structure [240] of these metabolosomes have been investigated. The pdu operon contains 21 genes encoding for the proteins that make up the Pdu MCP shell as well as the main pathway and cofactor recycling enzymes (Fig. 5.1). Eight proteins compose the Pdu microcompartment (MCP) shell—PduA, PduB, PduB', PduJ, PduK, PduN, PduT, and PduU [241, 242]. Of these eight proteins, seven (PduABB'JKTU) contain one or more bacterial microcompartment (BMC) pfam00936 domains, and, as such, form the hexagonal multimers that assemble into the facets and edges of the microcompartment [242, 73, 243, 244, 245, 246, 247]. PduN is the sole bacterial microcompartment vertex (BMV) pfam03319 gene in the pdu operon and is thus expected to form pentamers that cap the vertices of the Pdu MCP [235, 248, 249, 250, 251, 252]. PduN is a low abundance component of the MCP shell, but it is essential for the formation of well-formed compartment structures [241, 242, 253].

While prior studies have illustrated that aberrant structures form in the absence of PduN, the functionality and nature of these structures have yet to be explored in any detail. Further, studies on both alpha- and beta-carboxysomes showed that strict closure of the shell is required for these microcompartments to confer their biologically relevant growth benefits and that this cannot be achieved in the absence of pentameric vertex shell proteins like PduN [254, 255]. It is unclear how important this strict closure is for metabolosome systems like the Pdu MCP, as modeling studies have shown that a moderate diffusional barrier between the cytosol and an enzyme core is sufficient for mediating toxic intermediate buildup [256]. Previous work has suggested the differing importance of various shell proteins, including PduN, in Pdu MCP function [242]; but questions remain about precisely how MCP morphology controls Pdu pathway performance.

Here, we describe our detailed characterization of an MCP-related structure that we call Pdu microtubes (Pdu MTs) that form when vertex protein PduN cannot incorporate into the Pdu MCP shell, and use molecular dynamics modeling to understand the molecular interactions responsible for this morphology shift. This is accomplished through a comparison of Pdu and another MCP system that, in contrast, forms spherical compartments with pentameric gaps in the absence of the pentamer. Together, these results represent a key step towards understanding the complex interplay between shell protein interactions, compartment morphology, and encapsulated pathway performance.

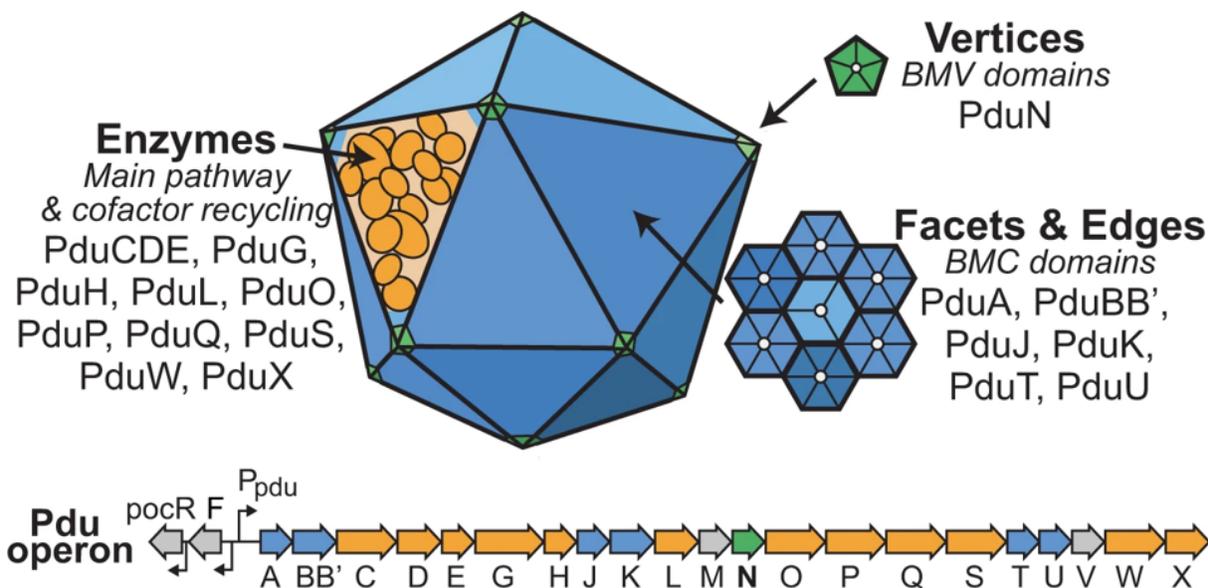


Figure 5.1. The *pdu* operon in *Salmonella enterica* serovar Typhimurium LT2 contains the genes encoding proteins responsible for formation of the 1,2-propanediol utilization microcompartment (Pdu MCP). These include enzymes that perform both key pathway steps and cofactor recycling functions (orange) and shell proteins that encase these enzymes (bacterial microcompartment, BMC, domain-containing genes shown in blue, bacterial microcompartment vertex, BMV, domain-containing gene shown in green). Notably, only one shell protein in the *pdu* operon, PduN, contains a BMV domain.

5.3. Results

5.3.1. PduN mediates the morphology of Pdu compartment structures

We first explored the impact of PduN on in vivo assembly of Pdu MCPs using a combination of fluorescence microscopy and transmission electron microscopy (TEM) on thin cell sections of both wild type (WT, PduN-containing) and *pduN* knockout strains (Δ PduN). Our fluorescence microscopy assay uses a green fluorescent protein (GFP) reporter fused to an encapsulation peptide, herein referred to as ssD for signal sequence from PduD, that is sufficient for encapsulation of heterologous proteins in Pdu MCPs [257]. Thus,

compartment distribution throughout the cell is indicated by the presence of the green fluorescence associated with the ssD-GFP reporter encapsulated within the MCP lumen. As in previous studies, expression of Pdu MCPs in the wild type (PduN-containing) background results in punctate fluorescence throughout the cell (Fig. 5.2c), suggesting that well-formed compartments are distributed within the cell [258, 259, 260]. In contrast, when the pduN gene is knocked out, expression of the pdu operon results in lines of fluorescence, typically aligned with the long axis of the cell (Fig. 5.2c). These lines of fluorescence indicate the formation of elongated structures within the cell capable of recruiting ssD-tagged GFP. Indeed, thin cell section TEM on cells expressing the pdu operon in the pduN knockout strain confirms the presence of tube structures, henceforth referred to as Pdu MTs (Fig. 5.2c). Interestingly, both fluorescence microscopy and thin cell section TEM show that these Pdu MTs appear to inhibit cell division, as the structures traverse multiple cleavage furrows (Fig. 5.2c, Supplementary Fig. 1 in the original manuscript [1]). While striking, such elongated structures are not unprecedented in the MCP literature—similar extended structures have also been observed in cells expressing pentamer-deficient carboxysomes, for example [254]. However, little is known about the structure or protein content of these tube structures.

We thus sought to examine, in detail, the structure of the Pdu MTs formed by expression of the pdu operon in our pduN knockout strain. These Pdu MTs are comprised of many of the same shell proteins as Pdu MCPs, evidenced by the presence of PduA, PduB, PduB', PduJ, and PduU bands by SDS-PAGE in both samples (Fig. 5.2b). Notably, bands associated with enzymatic cargo (PduCDE, PduG, PduP, PduQ, PduS) are also present in the purified Pdu MT sample. TEM analysis of purified Pdu MTs (Fig. 5.2c,

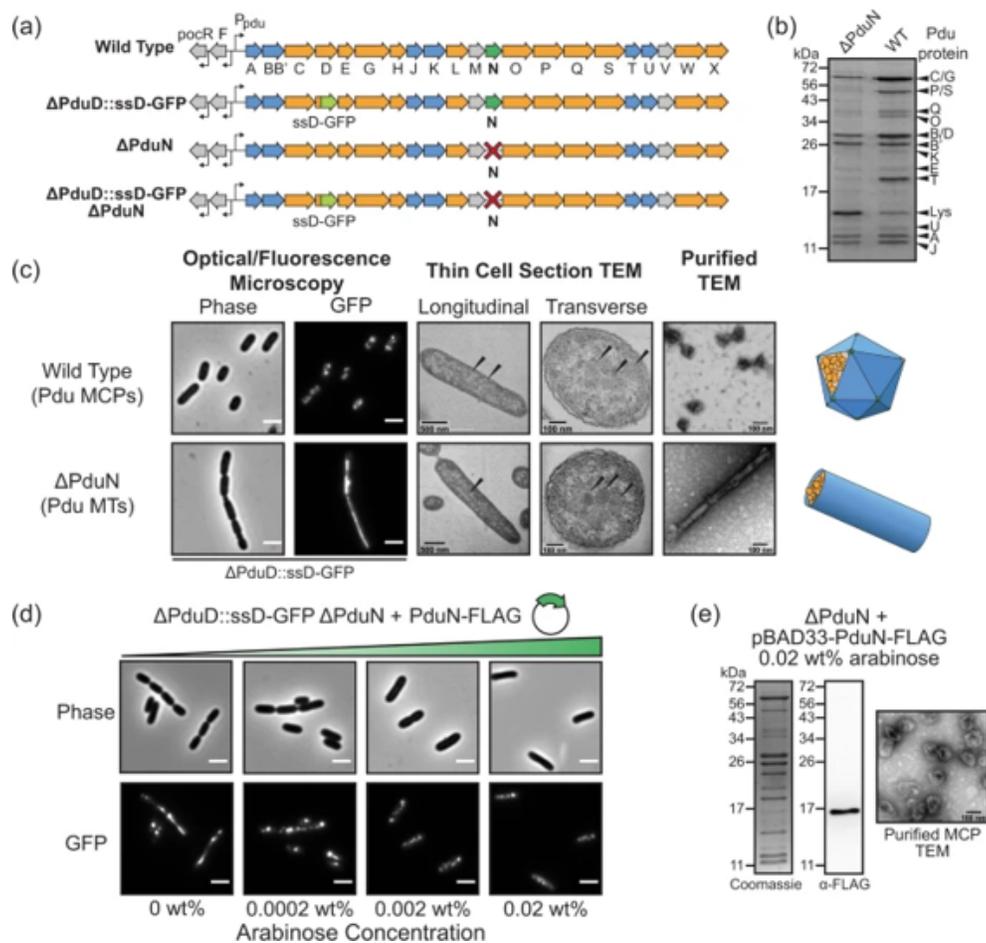


Figure 5.2. (a) Depiction of different *pdu* operon genotypes used in this figure. (b) Coomassie-stained SDS-PAGE of purified Pdu MTs ($\Delta PduN$) and Pdu MCPs (WT) comparing the protein content in these purified structures, where labels indicate the Pdu protein the band corresponds to (i.e. C/G for PduC/PduG), except for Lys, which indicates lysozyme. (c) Comparison of structures formed in Pdu MCP-forming strains (WT) and Pdu MT-forming strains ($\Delta PduN$). Scale bars in optical and fluorescence micrographs are 5 μ m. (d) Phase contrast and GFP fluorescence micrographs showing the impact of increased PduN-FLAG expression on the formation of Pdu MT structures versus closed Pdu MCP structures, where increasing arabinose concentration correlates with increasing expression of the PduN-FLAG protein off the pBAD33 plasmid. Scale bars in optical and fluorescence micrographs are 5 μ m. (e) Coomassie-stained SDS-PAGE, anti-FLAG western blot, and negatively stained TEM on Pdu MCPs purified from a *pduN* knockout strain supplemented with PduN-FLAG off a plasmid. Source data for (b) and (e) are provided as a Source Data file in the original manuscript [1]. Similar results to those reported in (b–e) were observed across three independent biological replicates, except for TEM imaging of thin cell sections, which was performed on multiple cells in a given biological sample, but not with biological replicates.

Supplementary Fig. 2 in the original manuscript [1]) shows that these tubes are 50 ± 10 nm in diameter, in agreement with diameters observed in cell sections. This dimension is distinct from the 20 nm diameter of rods self-assembled from PduA and PduJ shell proteins alone [73, 261], indicating that some combination of the other shell proteins present and the encapsulated cargo mediates the size and curvature of these Pdu MTs [261, 262]. These results suggest that the Pdu MTs formed by our pduN knockout strain are complex multi-protein assemblies, similar to Pdu MCPs.

5.3.2. Pdu microtubes control the metabolic flux of the 1,2-propanediol utilization pathway

Having shown that elongated Pdu MTs form in the absence of PduN, we next sought to probe the metabolic functionality of these tubes, and how organization into MTs versus MCPs impacts pathway performance. We hypothesized that the morphological shift from Pdu MCPs to MTs may negatively impact pathway performance, as we expect the Pdu MTs to have open ends that would increase exchange between the enzymatic core and the cytosol.

We explored the impact of compartment geometry on Pdu pathway performance by comparing the growth and external Pdu metabolite profiles (1,2-propanediol, propionaldehyde, 1-propanol, and propionate, Fig. 5.3a) of four strains—wild type (MCP-forming), Δ PduN (MT-forming), Δ PduA PduJ (broken compartment control [73]) and Δ PocR (no pdu operon expression control [263, 264, 265]). We grew these strains on 1,2-propanediol with excess adenosylcobalamin (adoB12), a condition that permits distinction of compartment-forming conditions that successfully sequester the toxic propionaldehyde

intermediate away from the cytosol [242, 57, 266]. Cell growth and metabolite profiles (Fig. 5.3b, c) show that control strains, Δ PocR and Δ PduA PduJ, grow as expected. When there is no expression of the pdu operon (Δ PocR), there is no cell growth over time, as none of the enzymes capable of 1,2-propanediol metabolism are present (Fig. 5.3b). Metabolite tracking confirms that no 1,2-propanediol is consumed (Fig. 5.3c). When the operon is expressed, but compartments cannot properly form (Δ PduA PduJ), cell growth and 1,2-propanediol consumption initially occur rapidly (Fig. 5.3b, c; doubling time of 2.338 ± 0.003 h between 3 and 9h), as there is no shell protein barrier preventing enzymes access to 1,2-propanediol. Consequently, this strain exhibits the most rapid initial generation of propionaldehyde, propionate and 1-propanol (Fig. 5.3c). However, after 12 h, a lag in growth begins to occur as propionaldehyde buildup exceeds a threshold value (doubling time of 62 ± 18 h between 12 and 18 h). This stalls propionate uptake into central metabolism, explaining both the observed growth lag and the delayed propionate consumption in this strain between 12 and 30 h (Fig. 5.3b). Several groups have reported this in strains with a broken compartment phenotype [239, 242, 73, 266] where it was hypothesized that propionaldehyde inhibits the methylcitrate pathway [267].

In contrast, strains containing Pdu MCPs (wild type) and Pdu MTs (Δ PduN) exhibit growth profiles consistent with a well-encapsulated Pdu pathway [242]. Initial growth and 1,2-propanediol consumption are slightly slower than the broken compartment control (Δ PduA PduJ), corresponding to doubling times of 3.17 ± 0.08 h and 2.64 ± 0.13 for wild type and Δ PduN strains between 3 and 9 h, respectively. However, growth of WT and Δ PduN strains eventually surpass the Δ PduA PduJ strain at later time points as propionaldehyde buildup begins to impact growth, evidenced by doubling times of 9.2

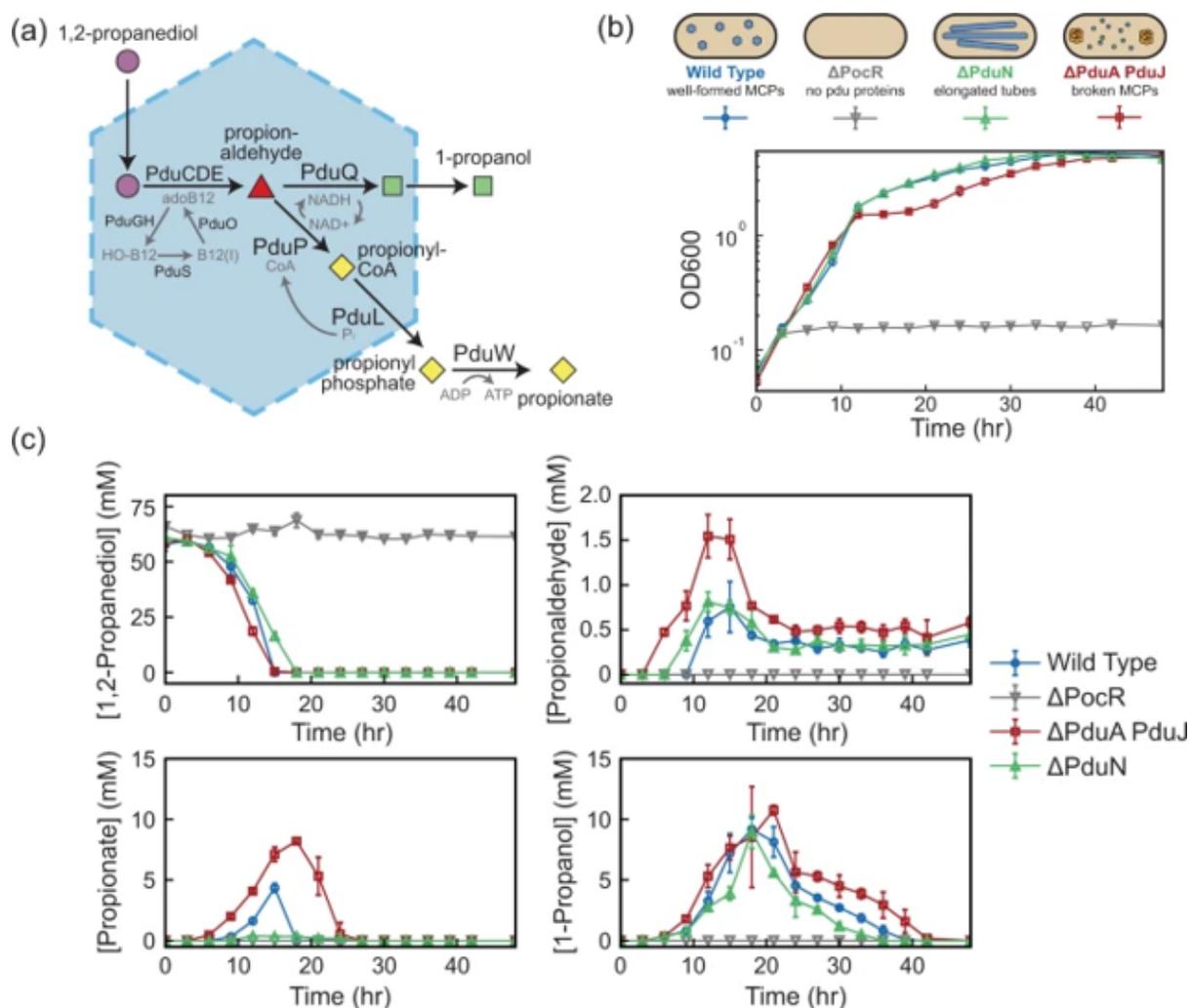


Figure 5.3. (a) Schematic of the 1,2-propanediol utilization pathway encapsulated in Pdu microcompartments. (b) Strains containing different compartment geometries (MCPs in Wild Type, blue lines, MTs in Δ PduN, green lines), without compartment expression (Δ PocR, grey lines), and with broken compartments (Δ PduA PduJ, red lines) grown in minimal media (NCE) with 1,2-propanediol as the sole carbon source. Data are presented as mean values \pm standard deviation over three biological replicates. (c) Concentration of key pathway metabolites over the course of the growth described in (b). Data are presented as mean values \pm standard deviation over three biological replicates. Source Data for panels (b) and (c) are provided as a Source Data file in the original manuscript [1].

Δ 1.6 h for the WT strain and 9.2 ± 1.9 h for the Δ PduN strain between 12 and 18 h. Strains containing Pdu MCPs (WT) and Pdu MTs (Δ PduN) both exhibit a lower peak concentration of propionaldehyde than the broken compartment strain (Δ PduA PduJ); however, the buildup of propionaldehyde is slightly more rapid in the Pdu MT strain than the Pdu MCP strain, where there are detectable propionaldehyde levels at 9 h of growth (Fig. 5.3c). This suggests that the change in geometry from MCP to MT subtly alters passive substrate transport in and out of the compartment, impacting the accessibility of substrates to the enzymatic core. This could either be due to changes in compartment surface area or potential open ends of Pdu MTs. Significantly, compared to the Pdu MCP strain (WT), the Pdu MT strain (Δ PduN) exhibits lower peak propionate concentrations and more rapid consumption of 1-propanol (Fig. 5.3c). This suggests that in these growth conditions, the Pdu MT geometry favors more rapid uptake of propionate into central metabolism, again, possibly due to changes in average substrate transport in and out of Pdu MTs versus Pdu MCPs. Taken together, these results indicate that the diffusional barrier provided by the Pdu MT protein shell is sufficient to prevent toxic propionaldehyde buildup in the cytosol.

Observing that knocking out pduN caused the formation of Pdu MTs instead of Pdu MCPs, we hypothesized that PduN is directly responsible for mediating the morphology of Pdu microcompartments. To test this hypothesis, we supplemented our pduN knockout strain with a plasmid containing FLAG-tagged PduN and observed changes in compartment morphologies at varying inducer levels using fluorescence microscopy (Fig. 5.2d). We find that increasing PduN-FLAG expression decreases the formation of elongated structures (Pdu MTs), and increases the observation of punctate fluorescence (Pdu MCPs)

(Fig. 5.2d). Interestingly, even with no inducer present (0 wt% arabinose), a decrease in the percent of cells with elongated structures is observed (Supplementary Fig. 3 in the original manuscript [1]). This is likely a result of leaky PduN-FLAG expression; because PduN constitutes only 0.6% of the total shell protein content, it is not surprising that even very low levels of PduN would impact shell closure [268]. We validated these microscopy results by purifying compartments from a pduN knockout strain supplemented with PduN-FLAG off a plasmid (0.02 wt% arabinose). These compartments exhibit the characteristic polyhedral geometry of Pdu MCPs by TEM and the characteristic banding pattern of well-formed Pdu MCPs by SDS-PAGE (Fig. 5.2e). Further, anti-FLAG western blotting on these same purified compartments confirmed the presence of PduN-FLAG in these well-formed structures (Fig. 5.2e). We conclude that PduN plays a direct role in the formation of Pdu MCPs, likely by facilitating capping of MCP vertices.

5.3.3. All-atom simulations of PduN and PduA

Next, we examined the molecular underpinnings of how PduN facilitates MCP closure by investigating the interaction interface responsible for PduN incorporation using all-atom molecular dynamics (AAMD) simulations. Previous work modeling the interface between two PduA hexamers revealed that preferred interaction angles between hexamers play a key role in higher order assembly of these proteins [5]. We hypothesized that similar studies comparing the PduN interaction interface to the PduA/PduA interface could yield insight into the specific, unique features that allow PduN to initiate Pdu MCP vertex capping. We selected PduA as the interacting partner for PduN based on previous studies showing that PduA and PduN interact *ex vivo* [241]. We built an estimated

model of the PduA/PduN and PduA/PduA interfaces using a homology-based approach that leveraged the solved crystal structure of the HO MCP (PDB: 5V74). This structure provides exquisite molecular detail of how homologous shell proteins assemble to form an MCP shell (see Methods for details) [235, 262]. Using this model as a starting point, we performed AAMD simulations of this interface to examine the energetics associated with various bending angles between PduA and PduN as well as between PduA and PduA (Fig. 5.4). Specifically, we calculated the potential of mean force (PMF) between each pair of protein oligomers as a function of the bending angle between the two components (Fig. 5.4b, e) and the distance between their centers of mass in the y-direction (Fig. 5.4c). More details on the calculation can be found in the Methods section. The resulting PduA/PduN bending energy landscape revealed a strong preference for a 40° bending angle between PduA and PduN (Fig. 5.4b) with the bending energy ($\Delta G_{0^\circ \rightarrow 40^\circ} = 6 \pm 2$ kcal/mol) comprising over half of the total interaction strength ($\Delta G_{PduN/PduA} = 10 \pm 2$ kcal/mol, Fig. 5.4c). Notably, this is higher than bending angles (30°) between hexamer/pentamer components reported in the crystal structure of an MCP shell from *Haliangium ochraceum* [235]. This preference for a bent interaction is distinct from the bending energy landscape of the PduA/PduA interface, which has only shallow minima ($\Delta G_{0^\circ \rightarrow 40^\circ} = 1.2 \pm 0.3$ kcal/mol) that constitute less than a quarter of the total PduA/PduA interaction energy ($\Delta G_{PduA/PduA} = 11 \pm 2$ kcal/mol [5]). We note two things about this bending energy landscape. First, that the energy minimum at 34° is consistent with previous models investigating PduA/PduA bending interactions [261]. Second, we note that a second minimum exists at a PduA/PduA interaction angle of 70°; however, given that this bending angle would not permit assembly of larger icosahedra or

polyhedra like those formed in the Pdu MCP system, we do not believe it is physically relevant to the discussion here. Interestingly, while the bending angle preference is dramatically different between these two interfaces, the magnitude of the PduA/PduN and PduA/PduA interaction is similar [5] (Fig. 5.4c). Together, this suggests that PduN could provide an energetically favorable bending point that allows for the closure of the shell without requiring less favorable bending of the PduA/PduA interface. Since this bending is intrinsic to the PduA/PduN interaction, even dimers, trimers, or any other PduN-featuring oligomers would also be highly bent. Thus, their incorporation would quickly disrupt the formation of any smaller Pdu MTs or flat sheets that are likely present early in the assembly process due to the low concentration of PduN.

In addition to measuring the bending energy landscapes, we were also able to track the conformation of the proteins as the bending angle, θ_B , was varied. At the minimum energy bending angles we observed the formation of strong short-range interactions at the interface between the PduN and PduA as well the interface between PduA and PduA (Fig. 5.5). In the case of PduN and PduA, we show a cation-pi interaction between ARG66 on PduA and PHE71 on PduN (Fig. 5.5a) and in the case of PduA/PduA interaction, a hydrogen bond forms between ASN67 on each of the PduA hexamers (Fig. 5.5b). These interactions provide a chemical understanding of the energy landscapes from Figure 5.4b, e. The ability to relate the energy landscape to the actual amino acids involved provides the opportunity to engineer the interactions between hexamers by changing the amino acid sequence. Ideally one could translate this physical knowledge of the interactions into the ability to change MCP size and shape, which, provides the opportunity to alter reaction kinetics. One way to see how these landscapes affect the assembled morphology

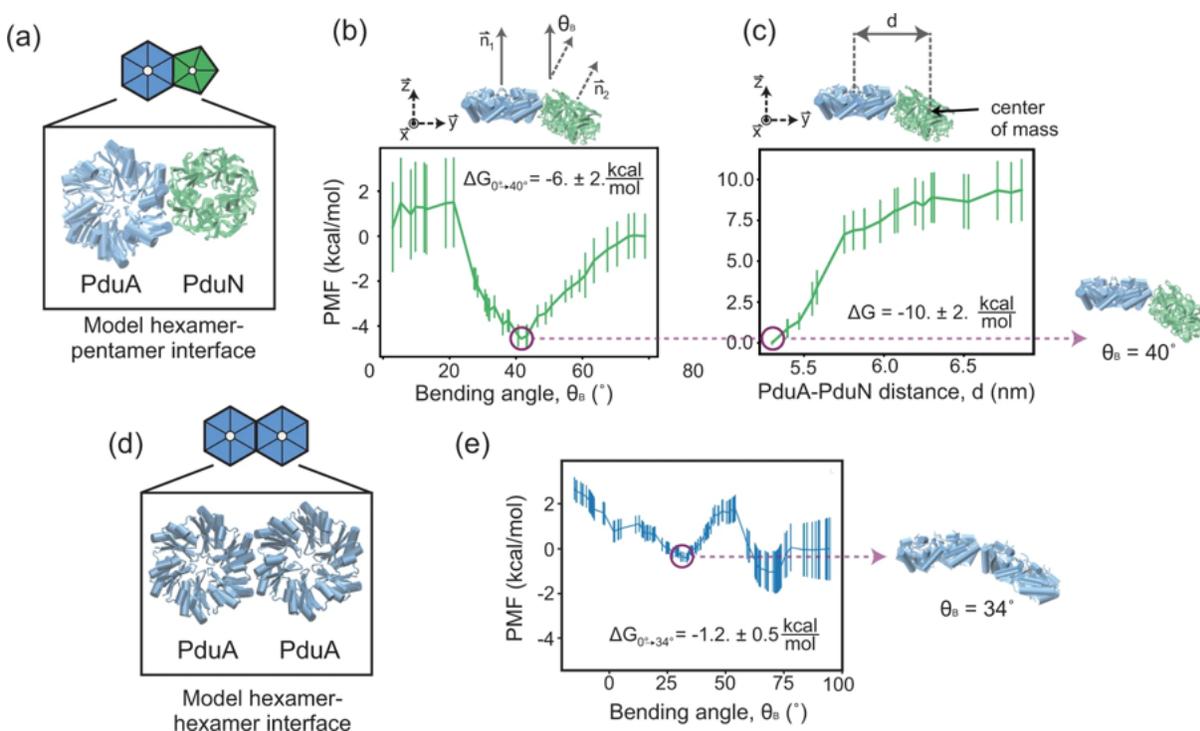


Figure 5.4. (a) Schematic of the PduA-PduN interface used for these simulations, where PduA is shown in blue and PduN is shown in green. (b) Potential of mean force (PMF) calculated from AAMD simulations as a function of bending angle, θ_B , between PduA and PduN. $\Delta G_{0^\circ \rightarrow 40^\circ}$ is the difference in the PMF between the 0° and 40° bending angles. (c) PMF calculated from AAMD simulations as a function of the distance between PduA and PduN, used to calculate the total interaction energy (ΔG) between these two oligomers. (d) Schematic of the PduA-PduA interface used for these simulations. (e) PMF calculated from AAMD simulations as a function of bending angle, θ_B , between two PduA hexamers. $\Delta G_{0^\circ \rightarrow 34^\circ}$ is the difference in the PMF between the 0° and 34° bending angles. Calculations used calculate data points in (b), (c), and (e) are described in Method section. Error bars on plots in (b), (c), and (e) represent the sampling error on the calculated energies, estimated by splitting simulation data into different sections and observing the differences in the calculated potential as described in the Methods section, Calculation specifics. Source data for plots (b), (c), and (e) are provided as a Source Data file in the original manuscript [1].

is to compare these energy landscapes with those of another MCP system with different morphologies.

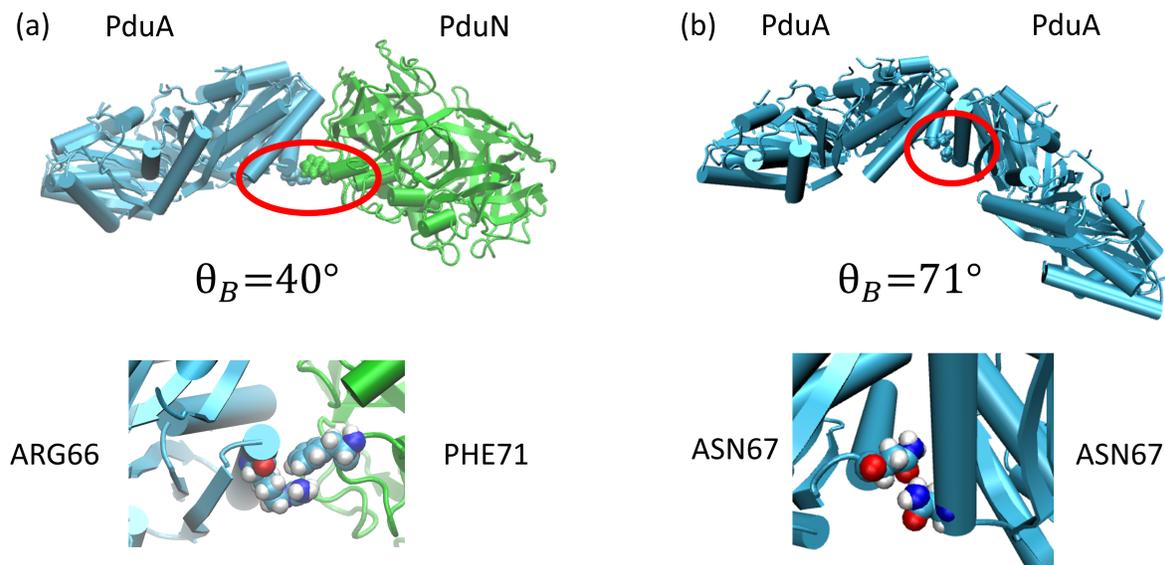


Figure 5.5. (a) Representative snapshot from the all-atom molecular dynamics simulation at the lowest energy 40° bending angle. The inset shows a short range cation-pi interaction between Arginine 66 on the PduA hexamer and Phenylalanine 71 on the PduN pentamer. (b) Representative snapshot from the all-atom molecular dynamics simulation at a 70° bending angle. The inset shows hydrogen bonding between Asparagine 67 on each PduA hexamer. This is believed to explain the second well in the bending potential landscape in Fig. 5.4e.

5.3.4. Comparison to a pentamer hole forming microcompartment system

We have shown that, in the absence of the PduN pentamer, Pdu MTs will form. However, this behavior is non-trivial as, in other MCP systems a compartment of the same size will form, but with gaps at the pentameric vertices [269, 270, 271]. We investigate one such system, using AAMD to measure the total interaction energy and bending energy landscape for hexameric and pentameric shell proteins from *Haliangium ochraceum* [235].

We will refer to these proteins as HO Hex and HO Pent respectively. This system is known to form a T=9 icosahedral shell featuring 12 5-sided HO Pent proteins, 60 6-sided HO Hex proteins, and 20 other 6-sided pseudo-hexameric trimer proteins. We will ignore the trimers as the pentamer holes have been in structures featuring only HO Hex [269]. Our hypothesis is that the formation of assemblies with pentameric gaps should be reflected in the bending energy landscapes of the constituent proteins. Furthermore, these bending energy landscapes should be related to the amino acid sequences of the proteins just as in the case of PduN and PduA as demonstrated in Fig 5.4 and 5.5. Thus, we should be able to relate differences in the assembly behavior of the HO system and the Pdu system directly to the amino acid sequences of constituent shell proteins.

AAMD calculations show that the bending energies (Fig. 5.6b, e) in this system are stronger than even the PduN-PduA interaction. The energy difference between the flat (0°) and lowest energy bending angle for the hexamer-hexamer interaction is an order magnitude larger in the HO system (Fig. 5.6e) than in the Pdu system (Fig. 5.4e). This helps to explain why structures with missing pentamers are assembled in the HO system, while tubes assemble in the Pdu system. Tubes require flat bending angles in one direction, while the shells with pentameric holes have all 30° bending angles [270]. The energetic cost to create these flat interfaces seems to be too high in the HO system, but not in the Pdu system. The calculations also showed that the total HO Hex-HO Hex (Fig. 5.6d) interaction is much stronger than the HO Hex-HO Pent interaction (Fig. 5.6d). Thus, it is also likely that the hexamers assemble much faster than the pentamers in the HO system. This may lead the pentamers to fill in the gaps that naturally form as defects, which is

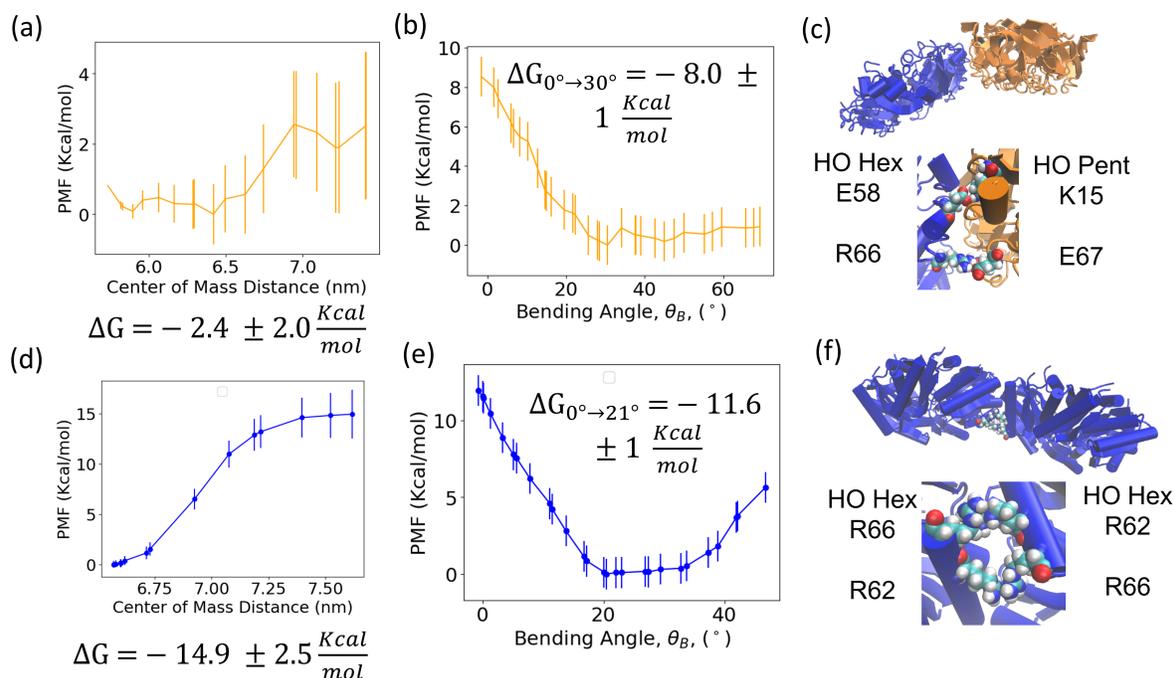


Figure 5.6. (a) PMF calculated from AAMD simulations as a function of the distance between HO Hex and HO Pent, used to calculate the total interaction energy (ΔG) between these two oligomers. (b) Potential of mean force (PMF) calculated from AAMD simulations as a function of bending angle, θ_B , between HO Hex and Pent. $\Delta G_{0^\circ \rightarrow 30^\circ}$ is the difference in the PMF between the 0° and 30° bending angles. (c) Schematic of the HO Hex-HO Pent interface used for these simulations, where HO Hex is shown in blue and HO Pent is shown in orange. Hydrogen bonds between oppositely charged amino acids form at the bottom the interface. (d) PMF calculated from AAMD simulations as a function of the distance between HO Hex and HO Hex, used to calculate the total interaction energy (ΔG) between these two oligomers. (e) PMF calculated from AAMD simulations as a function of bending angle, θ_B , between two PduA hexamers. $\Delta G_{0^\circ \rightarrow 21^\circ}$ is the difference in the PMF between the 0° and 21° bending angles. (f) Schematic of the HO Hex-HO Hex interface showing the formation of Arginine pairs at the bottom interface. Calculations used calculate data points in (a), (b), (d), and (e) are described in the Methods section. Error bars on plots in (a), (b), (d), and (e) represent the sampling error on the calculated energies, estimated by splitting simulation data into different sections and observing the differences in the calculated potential as described in the Methods section, Calculation specifics.

consistent with the observation that without the pentamers these gaps are simply not filled, with no other obvious change to the assembly [269].

As we hypothesized, the differences in interactions can be related to the amino acid sequence of the proteins. The strong bending potential of the HO Hex-HO Hex interaction seems to come from two arginine pairs [272] found at the bottom of interface involving arginines at amino acid position 62 and 66 (Fig. 5.6f, labeled as R62 and R66). The arginine at position 66 is conserved from the sequence of PduA and is heavily involved in a cation- π interaction with a phenylalanine of PduN (Fig. 5.5a), however, it has no partner in the interaction with the other PduA. The arginine at position 62 in HO Hex aligns very well with R66 across the interface and creates the totally different bending potential observed in the HO Hex-HO Hex interaction. The HO Pent-HO Hex interaction has a similarly strong bending interaction (Fig. 5.6b) as the PduA-PduN bending interaction (Fig. 5.4b) once again involving R66 on the hexamer (Fig. 5.6c). The HO Hex-HO Pent bending energy (Fig. 5.6b) appears higher than the total energy (Fig. 5.6b), but this apparent contradiction is resolved by the fact that in an unbiased simulation low bending angles are unstable for HO Pent-HO Hex and the two shell proteins repel each other.

5.4. Conclusions

There is great interest in repurposing MCPs for metabolic engineering applications, where they have the potential to alleviate bottlenecks such as slow pathway kinetics, toxic intermediate buildup, and cofactor competition [55, 231]. While strides have been made in loading non-native cargo into these systems in a controlled fashion [260, 273,

256, 270, 274, 275, 276], the selection criteria for an MCP system in any given engineering application is lacking. This includes MCP features such as size and morphology. Here, we report an in-depth characterization of an alternative Pdu compartment geometry, Pdu MTs, which form when vertex protein PduN cannot incorporate into the Pdu shell. Intriguingly, this shift in morphology upon loss of BMV-containing proteins is not universal across compartment systems—in the absence of vertex proteins, γ -carboxysomes form elongated structures similar to Pdu MTs [254], but α -carboxysomes predominantly form regular icosahedra [255]. Further, other metabolosomes can form closed icosahedra in the absence of pentamers [270, 274, 271, 269]. We show that in at least some cases, this is a consequence of the molecular interactions between shell proteins, specifically the preferred bending angle between these shell proteins. On this front, we anticipate that MD simulations can provide key insights towards understanding differences between a larger variety of compartment systems.

Comparison of growth and pathway performance in cells expressing Pdu MTs and Pdu MCPs showed that Pdu MTs prevent buildup of the toxic propionaldehyde intermediate in the native Pdu pathway. This result suggests that Pdu MTs provide a diffusive barrier between the cytosol and the encapsulated enzyme core. However, we note that the morphology change from spherical MCPs to cylindrical MTs necessarily changes the surface area to volume ratio of the compartment. We expect that the different surface area to volume ratio made available by these Pdu MT structures will prove beneficial to engineered encapsulated pathways with different kinetic profiles. Future analysis of different encapsulated pathways across different compartment geometries will provide valuable insight in this regard.

5.5. Methods

5.5.1. PduA/PduN Homology Modeling and Initial Relaxation

The initial structure for the atomistic model of the PduA/PduN interface was generated as follows. The PduA structure was taken from PDB 3NGK [244]. The structure of the PduN subunit was estimated using the Phyre2 web portal [277]. The pentamer structure was then generated by aligning five copies of this PduN subunit structure to the HO Pent structure extracted from PDB 5V74 [235] using the MatchMaker tool in UCSF Chimera [278, 279]. This structure was then minimized using default parameters in UCSF Chimera's Minimize Structure tool. To build the PduA/PduN interface, a HO Hex/HO Pent interface was extracted from PDB 5V74, which is a solved crystal structure of a full microcompartment from *Haliangium ochraceum* [235]. Chimera's MatchMaker tool was then used to align the PduA hexamer and PduN pentamer to the HO Hex and HO Pent structures, respectively. The PduA/PduN interface structure was then minimized again using the default parameters in Chimera's Minimize Structure tool [278]. The PduA/PduA interface was generated in the same way, except using a HO Hex/HO Hex interface from the PDB 5V74 structure. HO Hex/HO Hex and Ho Pent/HO Hex structures were taken directly from the PDB structures. Prior to running simulations, the all interface models were solvated in water containing 100 mM NaCl. Using the GROMACS molecular dynamics engine [280], the system was subject to a 100 ps constant pressure, temperature (NPT) equilibration with the protein backbones restrained. Steered MD simulations were then run to create configurations where the proteins adopt many different bending angles or distances (depending on the nature of the calculation).

Umbrella sampling was then performed in the z-direction and mapped back onto the θ_B direction, converting the forces in the z-direction to those in the θ_B direction as described in the next section, Calculation of the Bending Potential of Mean Force.

5.5.2. Calculation of the Bending Potential of Mean Force

The bending potential between the hexamer and pentamer, $V(\theta_B)$, is calculated from the forces on the pentamer at a given bending angle, according to the definition of the force,

F_{θ_B}

$$(5.1) \quad F_{\theta_B} = -\frac{\delta V}{\delta \theta_B}$$

Here V is the general interaction potential between the hexamer and pentamer, so it is necessary to measure only the forces in the θ_B -direction. Using the GROMACS simulation engine [280], we position restrain the backbone of the hexamer to not move in any cartesian direction while the pentamer is allowed to move in the yz-plane using harmonic restraints. We also restrain the center of mass distance of the pentamer and hexamer in the z-direction using a harmonic spring. We run simulations at many different angles, which correspond to different z-distances between the centers of mass of the protein. We are thus able to calculate a potential of mean force by performing a discrete summation over those angles

$$(5.2) \quad V(\theta_B) = -\sum_{i=0}^{n-1} \langle F_{\theta_B}(\cdot, z_i) \rangle (\langle z_{i+1} \rangle - \langle z_i \rangle)$$

$F_{\theta_B}(Z)$ is the component of the spring force as logged from the simulations, F_{spring} , in the θ_B -direction

$$(5.3) \quad F_{\theta_B}(z) = F_{spring} \cos(\theta_B)$$

Components of F_{spring} perpendicular to the θ_B -direction point directly from the center of the pentamer to the pentamer-hexamer interface and are balanced by the constraints on the hexamer, F_{hex} (Fig. 5.7). Since we can measure the bending angle, θ_B , in the simulation, we create a one-to-one map between z and θ_B , where z is the distance between the centers of mass of the pentamer and hexamer in the z -direction(Fig. 5.8).

We are careful to use many different “windows” (i.e. make n large and $(z_{i+1} - z_i)$ small) to calculate the mean forces at mean positions in a pseudo-continuous manner with overlap between states, especially near the minimum. In Figure 5.9, each color represents a different “window” and there is significant overlap. The same method applies in the case of two hexamers. The initial configuration for the different windows are created by starting with the protein interface, which was created and relaxed as described in the main text. Then, with position restraints on the hexamer and pentamer as described above, but no z center of mass constraint, the z center of mass of the pentamer is pulled at a rate of 1 Å/ns to create configurations with all the necessary z values for all the different windows. These simulations take about 20-30 nanoseconds, with independent simulations pulling up and down in the z direction. The pentamer-hexamer interface acts like a hinge and does not detach in this process. Each window is then run for at least

10 nanoseconds (details in the Calculation Specifics Section) to compute the mean force, $\langle F_{\theta_B}(\langle z_i \rangle) \rangle$.

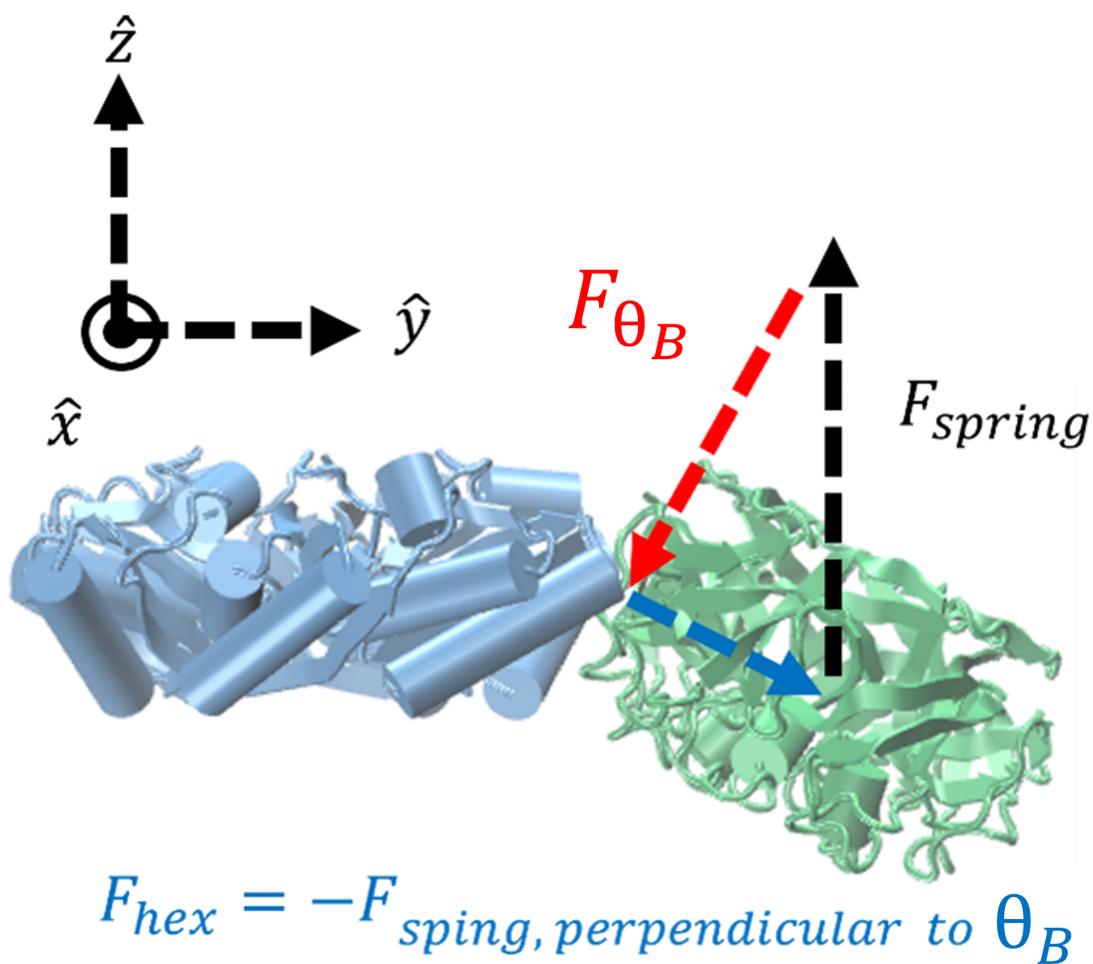


Figure 5.7. Balance of forces in calculation of $F_{\theta_B}(z)$.

5.5.2.1. Simulation Details. We performed all atom molecular dynamics simulations using the package GROMACS (version 2016.3) [280]. We used the most recent CHARMM [209] forcefield. The recommended CHARMM TIP3P water model [210] was applied with the

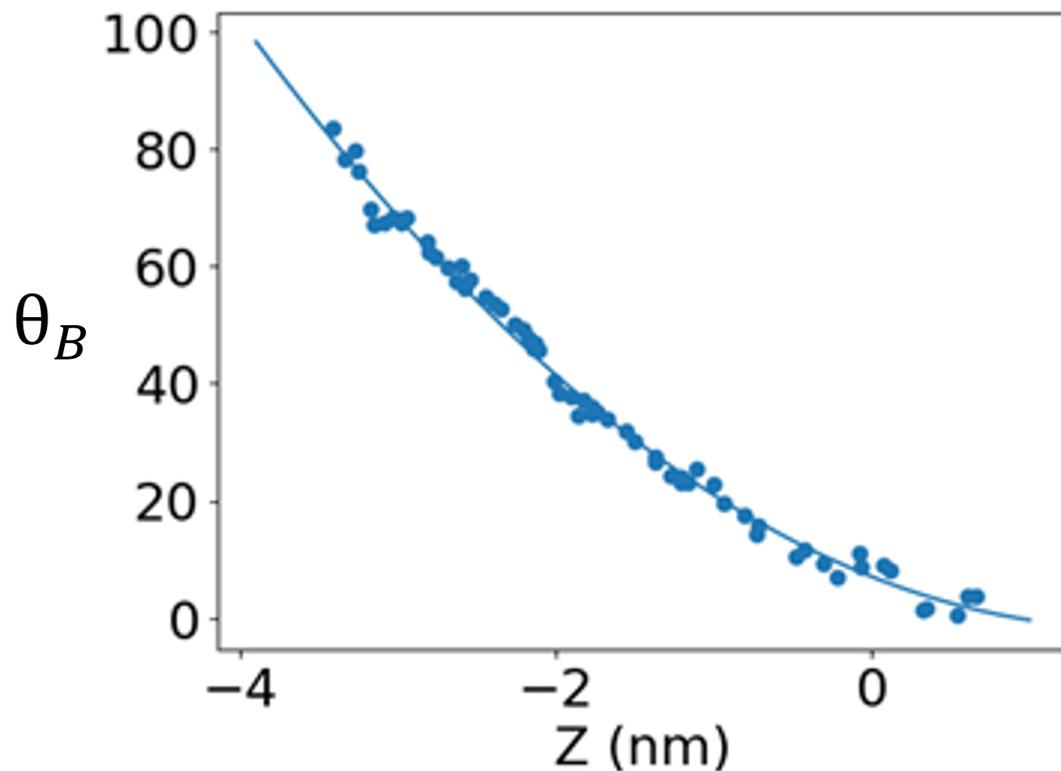


Figure 5.8. Mapping between bending angle, θ_B (degrees), and the distance between the centers of mass of the pentamer and hexamer, z (nm).

structures constrained via the SETTLE algorithm [211]. The periodic boundary conditions were employed in all dimensions. The neighbor searching was calculated up to 12 Å using the Verlet particle-based method and was updated every 20-time steps. The Lennard-Jones (LJ) 12-6 interactions were switched off from 10 to 12 Å via the potential-switch method in GROMACS. The short-range Coulomb interactions were truncated at the cut-off distance of 12 Å, and the long-range interactions were calculated using the Smooth Particle Mesh Ewald (PME) algorithm [212, 213]. The NPT ensemble (constant

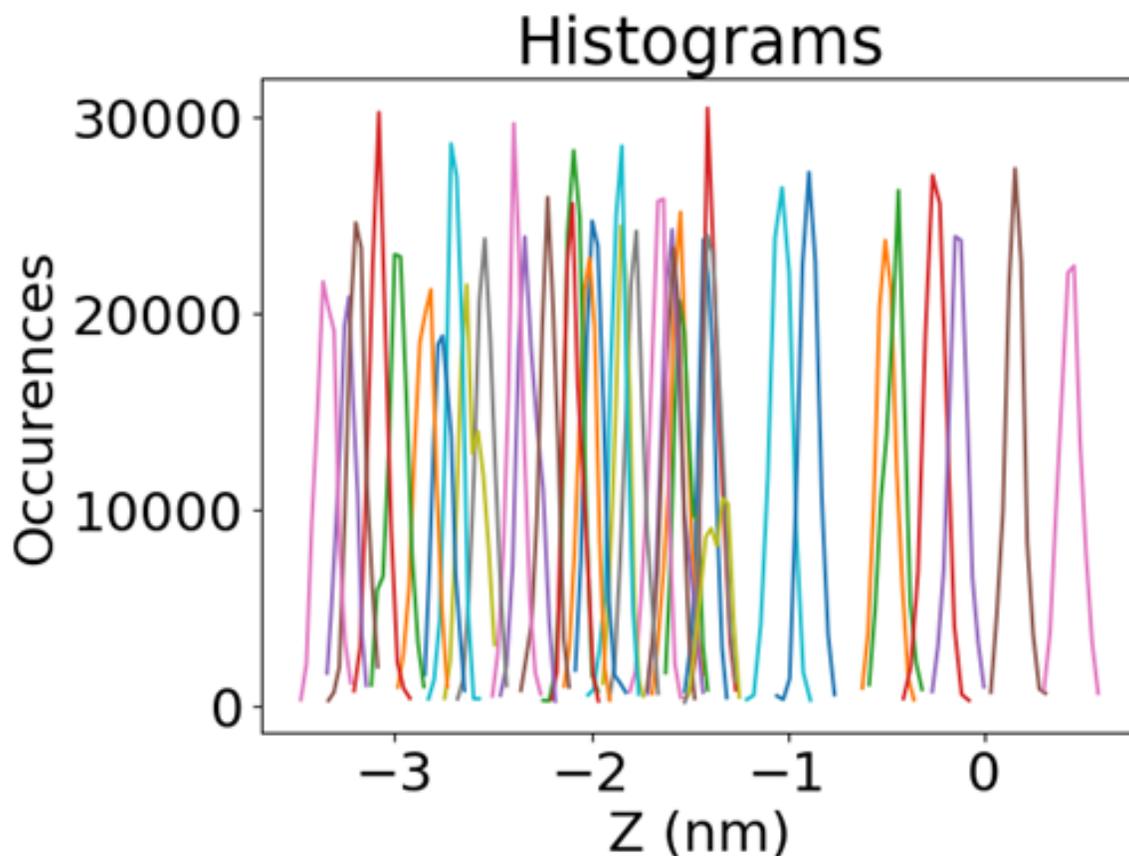


Figure 5.9. Histograms showing the parallel “windows” used to calculate the potential of mean force. The overlap of the windows is shown by the distance between the centers of mass of the pentamer and hexamer, z (nm).

number of particles, pressure, and temperature) was employed. The temperature was coupled using the Nosé-Hover algorithm (characteristic time 1 ps and reference temperature 298K). The isotropic Parrinello-Rahman barostat was employed with the reference pressure of 1 bar, the characteristic time was 4 ps, and the compressibility of $4.510^{-5} \text{bar}^{-1}$. All the covalent bonds were constrained, which supported an integration time step of 1 fs. These parameters were recommended for the accurate reproduction of the original

CHARMM simulation on lipid membranes [214], and have been verified in simulations on proteins [85, 215, 281, 216] and lipid membranes [217].

5.5.2.2. Calculation Specifics. The total interaction of the PduN-PduA interface utilizes 24 windows run for 5 nanoseconds each.

The bending of the PduN-PduA interface utilizes 37 windows. Each window is run for 10 nanoseconds.

The bending of the PduA-PduA interface utilizes 51 windows run for 15 nanoseconds each.

The total interaction of the HO Pent-HO Hex interface utilizes 20 windows run for 15 nanoseconds each.

The bending interaction of the HO Pent-HO Hex interface utilizes 32 windows. Each window is run for 15 nanoseconds.

The total interaction of the HO Hex-HO Hex interface utilizes 14 windows run for 15 nanoseconds each.

The bending interaction of the HO Hex-HO Hex interface utilizes 28 windows run for 15 nanoseconds each.

Differences in number of windows and run time reflect the complexities of the energy landscapes and an effort to reduce error bars relative to the magnitude of the energies. Error bars are based on sampling error and estimated by splitting the data in different sections (first half vs. second half, even data points v. odd data points) and observing the differences in the calculated potential.

CHAPTER 6

Assembly and Stability of Simian Virus 40 Polymorphs

This chapter is based on the published work [207] of Curt Waltmann, Roi Asor, Uri Raviv, and Monica Olvera de la Cruz, **Assembly and Stability of Simian Virus 40 Polymorphs**, *ACS Nano*, 2020, 14, 4, 4430-4443, with modified details.

6.1. Abstract

Understanding viral assembly pathways is of critical importance to biology, medicine and nano-technology. Here, we study the assembly path of a system with various structures, the simian vacuolating virus 40 (SV40) polymorphs. We simulate the templated assembly process of VP1 pentamers, which are the constituents of SV40, into icosahedral shells made of $N=12$ pentamers ($T = 1$). The simulations include connections formed between pentamers by C-terminal flexible lateral units, termed here “C-terminal ligands”, which are shown to control assembly behavior and shell dynamics. The model also incorporates electrostatic attractions between the N-terminal peptide strands (ligands) and the negatively charged cargo, allowing for agreement with experiments of RNA templated assembly at various pH and ionic conditions. During viral assembly, pentamers bound to any template increase its effective size due to the length and flexibility of the C-terminal ligands, which can connect to other VP1 pentamers and recruit them to a partially completed capsid. All closed shells formed other than the $T = 1$ feature the ability to dynamically rearrange and are thus termed “pseudo-closed”. The $N=13$ shell can even spontaneously

“self-correct” by losing a pentamer and become a $T = 1$ capsid when the template size fluctuates. Bound pentamers recruiting additional pentamers to dynamically rearranging capsids allows closed shells to continue growing *via* the pseudo-closed growth mechanism for which experimental evidence already exists. Overall, we show that the C-terminal ligands control the dynamic assembly paths of SV40 polymorphs.

6.2. Introduction

Understanding viral assembly pathways is critical to biology [282] and biotechnology [283, 284, 285, 286, 287, 288]. In particular, virus-like particles (VLPs) have become attractive candidates for many applications in nano- and bio-technology, such as catalysis [289], gene therapy [290] and vaccination [291]. One extensively studied virus for biological applications is the simian vacuolating virus 40 (SV40), a member of the polyomavirus family. Discovered in 1960 [292], SV40 is an enveloped virus that, *in vivo*, is made of up of 3 unique proteins, VP1, VP2, and VP3. VP2 and VP3 are located on the inner part of the viral capsid, while the outer part is made up of 72 VP1 pentamers [2]. Together these proteins form a 48 nm, $T = 7$ [64] icosahedral capsid around a double stranded DNA (dsDNA) genome that is wrapped around ~ 20 histone octamers. In this conformation, as in many viruses, the dsDNA is strongly compacted by oppositely charged proteins including polyamines [293]. *In vitro*, it has been shown that a full capsid shell made of only 72 VP1 pentamers will assemble around the bare dsDNA to form the 48 nm $T = 7$ capsid [3, 294]. Other structures will also form around a variety of other templates including RNA [295, 60, 4], nanoparticles [70, 296, 297, 298], and even micron sized particles [299]. The simplest structure is a $T = 1$ icosahedral capsid which is made up

of only 12 VP1 pentamers [300]. The ability to form multiple structures is part of what makes SV40 VP1 an attractive candidate for technological applications [291]. In particular, it is possible for VP1 to achieve $T=1$, $T=7$, and intermediate capsid sizes by being 5 or 6 coordinated (a pentamer or a hexamer respectively). This offers an opportunity to study viral assembly pathways and explore templates to modify assembly mechanisms.

The VP1 pentamer is made up of 5 identical polypeptide chains, each around 360 amino acids in length. The binding between these chains is very strong [301], and thus the VP1 pentamer can be considered a single unit that has five-fold symmetry. Each chain can be functionally divided into 3 parts: the C-terminal ligands, the N-terminal ligands, which function as the template binding domain, and the globular body. The middle sections of the chains form the globular section of the protein and their structure has been described in great crystallographic detail [2]. The final 12 residues on the N-terminal ligands are unstructured but contain positively charged residues. These residues bind negatively charged templates through a charged interaction and have been included in previous models of capsid assembly [61, 302] where the cargo is assumed to be a flexible polyelectrolyte that is essential for driving the assembly. This interaction has been shown to be independent of the C-terminal ligands [303]. In previous models of viral assembly [61, 302], the N-terminal peptides are termed arms while here they are termed N-terminal ligands, and in those models the C-terminal ligands are not included. The C-terminal ligands interdigitate with the body to form the connections between VP1 pentamers. These connections are essential to the formation of the capsid as their removal makes capsid assembly impossible [304]. The exact connection mechanism and topology vary based on the symmetry of the capsid [3], the presence of multivalent ions

[294], and local coordination [2]. Cryo-EM reconstructions of the $T = 1$ and $T = 7$ capsids have shown that the C-terminal ligands have a very specific topology [2], while attempts to reconstruct the C-terminal ligand topology of intermediate-sized particles were unsuccessful [3] suggesting these intermediates may have no well defined topology or arrangement of 5 and 6-coordinated VP1 pentamers.

The actual assembly process of viruses is not fully understood [59, 305]. Many models have been used to understand equilibrium phase behavior [306] and equilibrium capsid shapes [307]. However, they can lack dynamic assembly information. Modeling these dynamic pathways is difficult as due to the large size of the completed capsids, atomistic detail becomes intractable and as such more coarse-grained approaches have been taken. There are multiple coarse-grained models, some of which observe polymorphism [308], based on the assembly of regular polygons [309, 310, 311, 61, 308], so-called shape-based models. These shapes tile together to form static capsids. However, there is experimental evidence that the VP1 pentamer assembly is a highly dynamic process and models are required to explain the growth mechanism. As Donald Caspar stated “they [VP1 pentamers] will behave like an animate creature ... erratically flexing its donor organ near the end of each tentacle and grasping with its acceptor organ near the base of each face of its five-sided head.” [312] Thus, we propose a model for the assembly of SV40 polymorphs, not based on the assembly of rigid shapes, but one that allows the C-terminal ligands to bind to a specific point on the body of the VP1 pentamer, which is now a globular cylinder that preserves the 5-fold symmetry of the VP1 protein (see Figure 6.1). The proposed model also implicitly mimics the interactions between the N-terminal ligands and a negatively charged template in various pH and ionic conditions (see Figure 6.2), in

agreement with our experimental results on the assembly of the VP1 pentamers on 524 nucleotide RNA.

The inclusion of the C-terminal ligands in the simulations allows us to show the following “animate behavior” [312], (1) the $T = 1$ capsid is a static structure while other incomplete and pseudo-closed structures (*i.e.*, 13 VP1 pentamers and slightly smaller structures which may or may not be closed) are dynamic. (2) Fluctuations in template size can cause this 13 VP1 pentamer structure to spontaneously release a VP1 pentamer and reform the $T = 1$ capsid in what appears to be a self-correcting mechanism. (3) During assembly, C-terminal ligands are able to first connect to a partially completed capsid and then the N-terminal ligands bind to the spherical template.(4) This increases the effective size and flexibility of any partial capsid regardless of the template, making it more effective in recruiting additional VP1 pentamers. (5) This mechanism leads to pseudo-closed structures that continue to grow by forming connections with additional VP1 pentamers *via* the free C-terminal ligands and connection sites on the structure. The structure can then dynamically rearrange, allowing the N-terminal ligands of the additional VP1 pentamer to bind the template.

6.3. Results and Discussion

Simulations of the behavior of a single capsid are first performed in order to understand the products which may be present during assembly. Here, all simulations are performed with $\epsilon = 8k_B T$ because it gives the necessary strength to assemble the VP1 pentamers into icosahedral shells (lower values do not lead to this assembly and larger values lead to aggregation of the VP1 pentamers in the bulk as shown in Figure 6.13 in the Appendix).

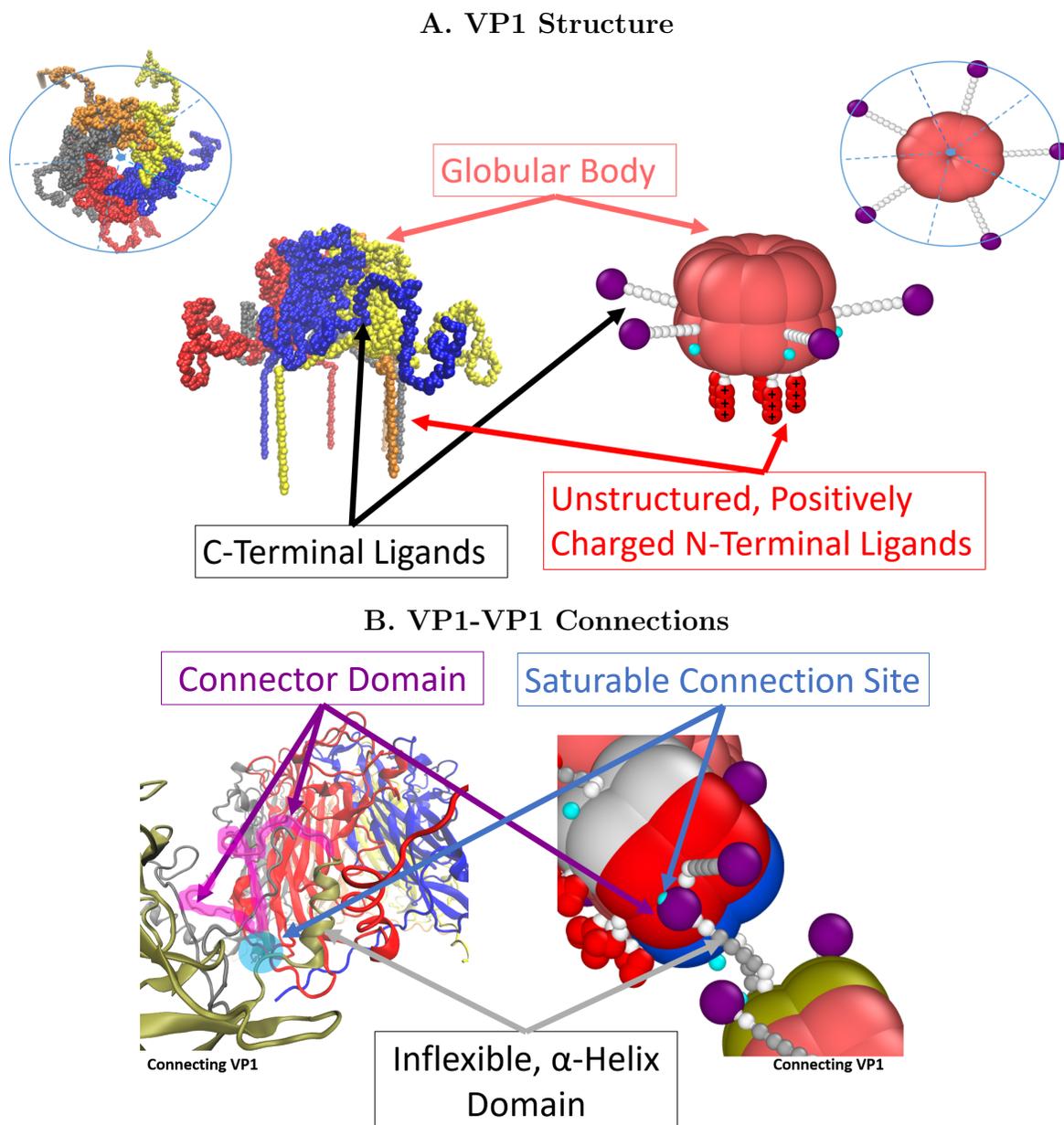


Figure 6.1. (A) Images of VP1 based on protein data bank (PDB) entry 1SVA [2] and its coarse-grained equivalent. The pink beads are rigid and represent the globular portion of the protein. The modeled VP1 pentamer can be split into 5 identical units, just as the real VP1 pentamer contains 5 chains with the same primary structure. Each unit has a rigidly attached C-terminal ligand comprised of nine beads which are only rigid in the middle (but are freely rotating) and terminate by a purple connector bead. (B) The connector beads hybridize with the cyan connection site with an energy given by the parameter ϵ . Each unit also contains an N-terminal ligand made of 5 flexibly connected beads. The first two are uncharged while the last 3 represent the positively charged residues found on this part of the VP1 protein.

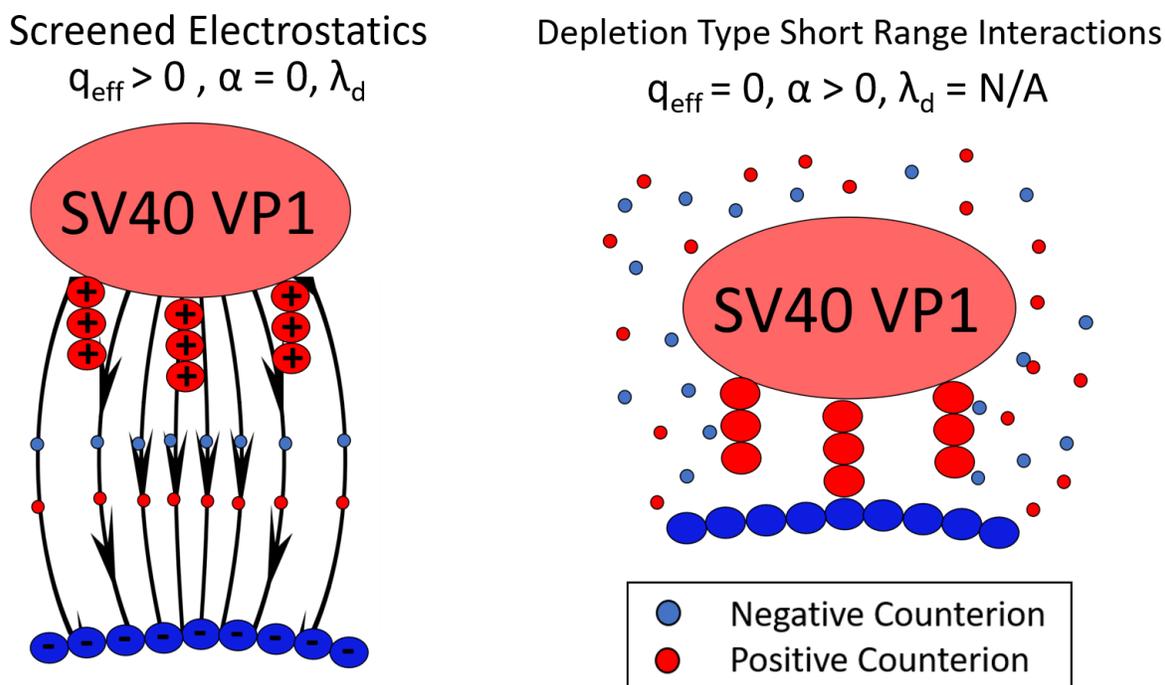


Figure 6.2. The two general cases of the implicit ion model. (Left) The case of screened electrostatics where the salt provides uniform screening of the electrostatic interactions present in the system. The strength of this screening is given by the Debye screening length, λ_d , as shown in Equation 6.1. (Right) The case of template binding at high salt. We model this by putting the $q_{\text{eff}}=0$ and adding a short range attraction, α , that comes from depletion since Debye-Huckel is not valid (Equation 6.2).

6.3.1. Icosahedral Capsid ($T=1$, $N=12$)

Simulations of the $T = 1$ capsid (*i.e.*, a rigid template with $N=12$ bound VP1 pentamers) are performed at different curvatures using $\epsilon = 8 k_B T$, $\lambda_d = 1.0 \text{ nm}$, and $q_{\text{eff}} = 1$. Icosahedral symmetry is observed in agreement with experiments and the definition of a $T=1$ icosahedral lattice. This is measured by the distribution of pairwise distances between the centers of VP1 pentamers. Figure 6.3 shows that at the lowest free energy, obtained for a radius of 9 nm, three pair correlation peaks are observed. The first peak is

the nearest neighbor peak, followed by the second and third characteristic distances peaks. Using these peaks, $8.2 \left(\frac{16.4}{2}\right)$ is the center to vertex distance and 8.6 is the edge length. For a regular icosahedron the ratio of these distances, $\frac{d_{\text{center,vertex}}}{\text{edge length}}$ is 0.95 which is equal to $\frac{8.2}{8.6}$. The last peak can also be used to estimate the total capsid diameter by adding twice the height of a pentamer above the center of mass (6.6 nm) to the third peak giving a total equilibrium capsid size of ≈ 23 nm. This is in good agreement with the reported values of 24.5 nm from solution X-ray scattering experiments [60]. Also in good agreement with experiments, is the existence of 3 helix triangles (see top right of Figure 6.3), which have been observed in cryo-electron microscopy (EM)[3]. This suggests the 3 helix structure is somewhat determined by capsid geometry since the orientational dependence of the hydrophobic interactions, credited for stabilizing the 3 helix structure, does not exist in this model. In addition to the 3 peaks, the pair distribution of the VP1 pentamers shows large regions of zero probability indicating that VP1 pentamers are vibrating only about the $T = 1$ icosahedral lattice points, but not sliding past each other. This is true of the $T = 1$ capsid over many template sizes and implies that it could withstand changes in solution conditions that may swell [294, 313] or shrink the template. In the experimental section, we will show evidence this may be occurring as pH is increased.

6.3.2. Non-Icosahedral Shell: N=13

The stability of closed shells geometries that can not contain icosahedral symmetry, but may be an “off-path” intermediate are also important in understanding viral assembly. This section focuses on one example of this class of shells, the N=13 capsid. As shown in Figure 6.4, icosahedral symmetry is disrupted by the 13th VP1 pentamer, which occupies

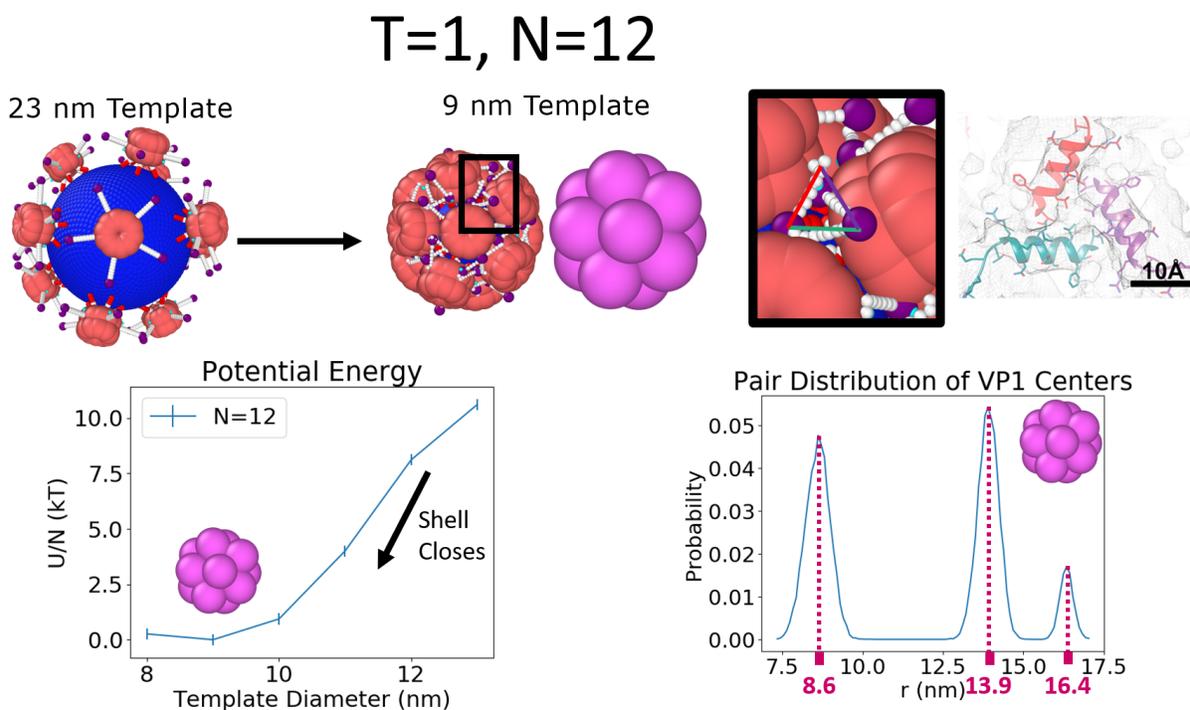


Figure 6.3. The $T = 1$ capsid is made of $N=12$ VP1 pentamers. The relative potential energy per VP1 and pairwise distance distribution of the capsid as a function of the template diameter is measured by first allowing 12 VP1 pentamers to bind to a large template and then slowly reducing the size of the capsid such that no VP1 pentamers are released during this process illustrated at the top left of the figure. Results on the bottom left show that the minimum of the potential energy occurs at a template diameter of 9 nm. On the bottom right, we see 3 peaks matching the icosahedral symmetry (see main text) and large regions of zero probability, indicating the static nature of the VP1 pentamers in this configuration. Since this static $T = 1$ configuration is stable over many template sizes we believe that it would also be robust against changes in template size based on changes in salt, pH, *etc.*, provided that these changes do not impact the VP1-VP1 or VP1-template interactions too much. This minimum energy structure also recovers the presence of a three helix triangle located at the three fold symmetry points of the icosahedron [3], shown at the top right. The image of the three-helix triangle with the scale bar was adapted with permission from Kler *et al.*, ACS Chemical Biology 2013, 8, 2753–2761. Copyright 2013 American Chemical Society.

a 6 coordinated position. A mix of 6 and 5 coordinated VP1 pentamers are also observed in the $T = 7$ capsid [2], which is made up of 72 VP1 pentamers (12 five-coordinated pentamers and 60 six-coordinated pentamers). Both of these structures follow the Euler formula for closed shells of regular polygons, $V-E+F=2$, which predicts that the number of six-folds is smaller by 12 than the number of total units under the constraint that only hexagons and pentagons are allowed (see Appendix for derivation). It is important to note that the location of the six-fold is dynamic relative to the template, implying the VP1 pentamers neighboring the six-coordinated VP1 pentamer are constantly reorganizing relative to each other. This is shown by non-zero probability at all distances at the bottom left of Figure 6.4. Since it is a closed, non-icosahedral shell that has dynamic behavior it will be referred to as pseudo-closed. This is in contrast to the $T = 1$ capsid which displayed regions of zero-probability and thus no reorganization. The underlying reason for the dynamics of the shell seems to be topological frustration of trying to insert a six-coordinated VP1 pentamer with only five C-terminal ligands into the shell. Shrinking the template to be smaller than 12 nm in diameter causes one of the VP1 pentamers to spontaneously detach and the $T = 1$ forms again. This may function as a self-correcting mechanism for assembly with flexible, dynamic templates. The released VP1 pentamer is always a nearest neighbor to a six-coordinated pentamer and its release seems to alleviate the internal stress caused by the six-coordination of the VP1 pentamer. In the $T = 7$ capsid, there is no α - *helix* in the γ subunit of the six-coordinated VP1 pentamer and this may help to relieve this stress.

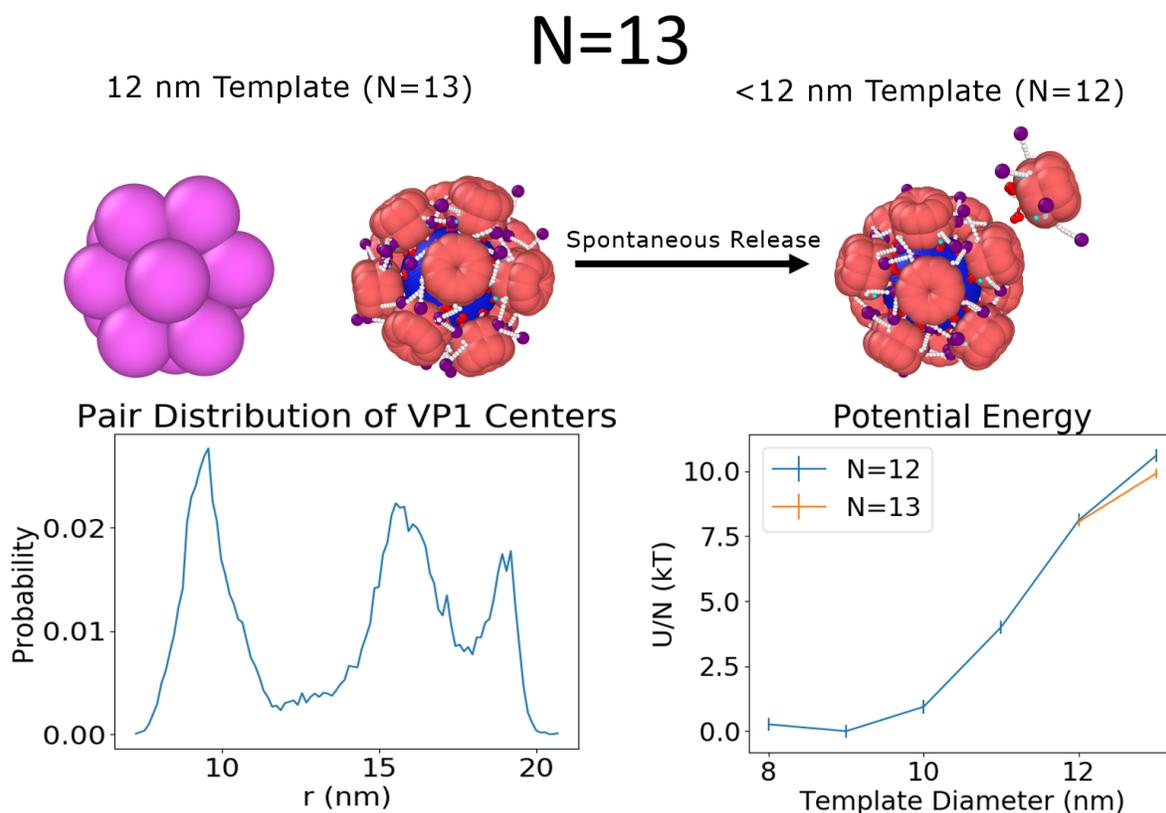


Figure 6.4. (Left) The $N=13$ shows deviation from the icosahedral symmetry observed in the case of the $N=12$, $T = 1$ capsid. Instead, a six-coordinated VP1 pentamer is observed (top left), consistent with the Euler formula for closed shells made of regular polygons. The six-coordinated VP1 pentamer is also mobile as shown by the absence of zero-probability regions in the pairwise distribution function (which were observed for $N=12$ in Figure 6.3). (Right) This structure is only favored over the $T = 1$ for 12 nm and above templates where it has a lower energy per VP1 pentamer (U/N) than the $T = 1$ capsid. At smaller template diameters, one VP1 pentamer will be spontaneously released and the $T = 1$ capsid reforms.

6.3.3. Assembly Paths: $N=10$ and $N=11$

The investigation of structures smaller than the $T = 1$ (*i.e.*, $N=10$ and $N=11$) provides the opportunity to understand possible intermediate states that may occur during assembly.

The $N=10$ on a 10 nm template is a dynamic, pseudo-closed shell like the $N=13$ and shows VP1 pentamers that have no exact position or specific coordination number (see Figure 6.5). While one half of the capsid seems to have five coordinated VP1 pentamers the other contains four VP1 pentamers that assume relatively larger fluctuations, which on average cover the template. When the template diameter is 11 nm, the pentamers spontaneously regain icosahedral symmetry with two mobile holes that are second nearest neighbors. The $N=11$ capsid (see Figure 6.15 in the Appendix) has only one hole, which does not close when the template size is varied. As mentioned earlier, these structures are intermediate states in the assembly of $T = 1$ capsids. The $N = 11$ and 11 nm $N = 10$ capsids show clear binding sites surrounded by free arms for the 11th and 12th VP1 to be added to the elongating capsid and this growth mechanism will be explored in the section: Elongation Mechanism on the Spherical Template. However, it is unclear how the 11th VP1 would be added to the pseudo-closed capsid. The pseudo-closed growth mechanism by which this is accomplished is also discussed in its own section Closed Shell Growth.

6.3.4. Assembly Factors

The assembly simulations (see details in the Methods section) can be performed using any value of the following parameters: q_{eff} , λ_d , α , and template diameter. Using these parameters, we show the effect of salt concentration, pH, and the importance of finely tuned C-terminal ligand interactions. The model is then updated to better represent the full length of the C-terminal ligands and the implications of this parameter are discussed. These results are compared with experiments and the underlying mechanisms of assembly

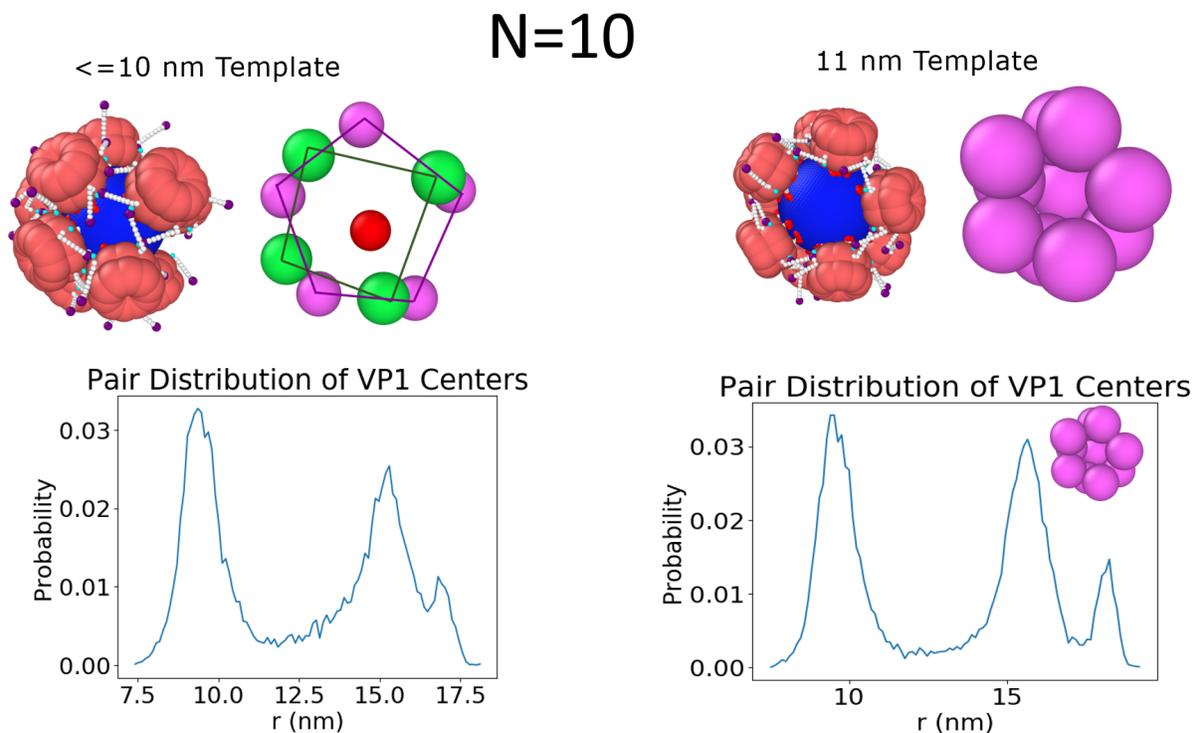


Figure 6.5. The $N = 10$ capsid can adopt different configurations based on the curvature of the template. (Right) The 11 nm template shows icosahedral symmetry like the $N = 11$ and $T = 1$ case, just with two holes that are second nearest neighbors. (Left) For templates less than 11 nm, the capsid shows half of the template with a 5 fold symmetry, while the other 4 VP1 pentamers are found in an unstructured state.

are examined in the simulations. Finally, by changing the template diameter from 11 to 10 nm, the growth of closed shells is evaluated.

6.3.5. Salt Concentration

The effect of salt is explored through three parameters, λ_d , q_{eff} , and α , which are functions of the ionic strength of the solution. Biological salt simulations were run using $\lambda_d = 1.0$ nm (or 92 mM added salt) and $q_{eff} = 1$. The effect of salt is explored through a lower salt

concentration, where λ_d is increased to 1.4 nm (or 47 mM added salt), and a higher salt concentration, where the charge is completely screened and depletion interactions in the form of Equation 6.2 are taken into account. The results of these studies are shown in Figure 6.6. The graphs at the bottom of Figure 6.6 show how the fraction of templates with N -bound VP1 pentamers, $F(N)$, progresses as a function of the fractional simulation time, t . $F(N)$ was sampled 20 times throughout the simulation with the average of the first four being reported as $t = 4/20 = 1/5$, the next four as $t = 2/5$, *etc.* The results show that VP1 pentamers bind very quickly at first and then the rate slows as available surface area is used while the amount of free VP1 pentamers decreases. Increasing the Debye screening length weakly increases the speed of assembly. At high salt concentrations, reliance only on short range interactions slightly decreases the formation of $T = 1$ capsids even when α is large enough to nucleate capsids. This is in part due to issues that arise when attempting to add the final VP1 pentamers as discussed in the next section. When the short range attraction is too weak ($\alpha = 1 k_B T$), no nucleation is observed at all.

6.3.6. Elongation Mechanism on the Spherical Template

The assembly simulations made it clear that the most difficult aspect of $T = 1$ assembly was adding the final VP1 pentamers to the incomplete capsid. This is somewhat different from the case of a flexible cargo like ssRNA which may act as an antenna, attract pentamers, and facilitate the assembly and the final closure of the capsid[60]. However, as was shown in our recent paper [314], few percent of particles with 11 (instead of 12) pentamers cannot be excluded at the signal-to-noise level of the experimental data.

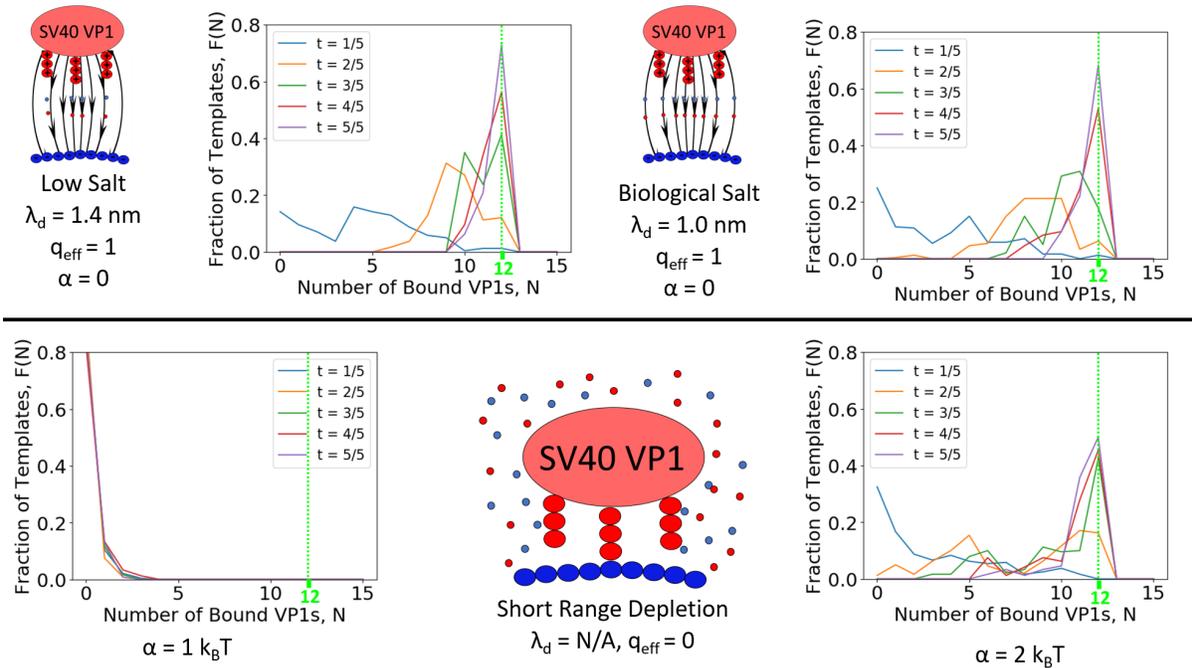


Figure 6.6. Fraction of Templates with N bound pentamers at different time points along the simulations. Assembly of $T = 1$ capsids is sensitive to added salt concentration as shown by the fraction of templates having $N=12$ capsids at $t=5/5$. (Left) Assembly occurs only slightly faster when the Debye length is increased relative to biological salt conditions (Center). (Right) Having only short ranged attraction decreases the observed fraction of $T = 1$ capsids assembled even when the attraction is strong.

In Figure 6.13 in the Appendix, even when $\epsilon = 7 k_B T$ there were a large fraction of $N = 10$ capsids, thus arriving at the $N = 10$ state was not the main issue. The inability to complete the capsid is due to the fact that these are the states where it is very hard to find free surface area on the spherical template to bind to. To understand how the binding of the final VP1 pentamers occurs, we observe successful binding of a 12th VP1 pentamer to an $N = 11$ capsid, which has icosahedral symmetry with one hole where binding can occur (see Figure 6.15 in the Appendix). The snapshots for this process are shown in Figure 6.7. At first, the VP1 pentamer made a single connection with the $N = 11$

capsid. The pentamer is then able to make more connections and stabilize itself on the capsid. This stabilization does not mean that the pentamer will bind, in fact the opposite occurs as the new connections form in such a way that it is orientationally impossible for the VP1 pentamer to bind the spherical template. To get out of this state, some of these connections have to dissociate and make the VP1 pentamer weakly attached, before creating new connections that orient the pentamer so that binding is possible. It is important to note that this process could not occur without the C-terminal ligands.

The character of this path makes it easy to understand why it was so difficult to attach the final VP1 to an $N = 11$ capsid. It also illustrates why the binding process is so sensitive to both the connection strength and electrostatics. If ϵ is too low the connections will not be able to stabilize the VP1 on the partial capsid whereas if ϵ is too high the connections may be too stable in the wrong orientation and the VP1 may eventually bind another template, leading to aggregation[**314**, **315**] as in Figure 6.13 in the Appendix. In this picture, the salt concentration controls how well oriented the capsid needs to be in order to bind the template. At higher λ_d (low salt) the final pentamer will be more strongly attracted even if it is farther away, whereas it will need to almost be in contact to bind *via* purely short range interactions. Some of these orientational effects could be mitigated by a flexible template that would grab the N-terminal ligands and held to orient the final VP1.

6.3.7. C-terminal Ligand Length

The full length of the connector domain is not included in the simplistic model depicted in Figure 6.1. This means the variety of possible attractive interactions involving connector

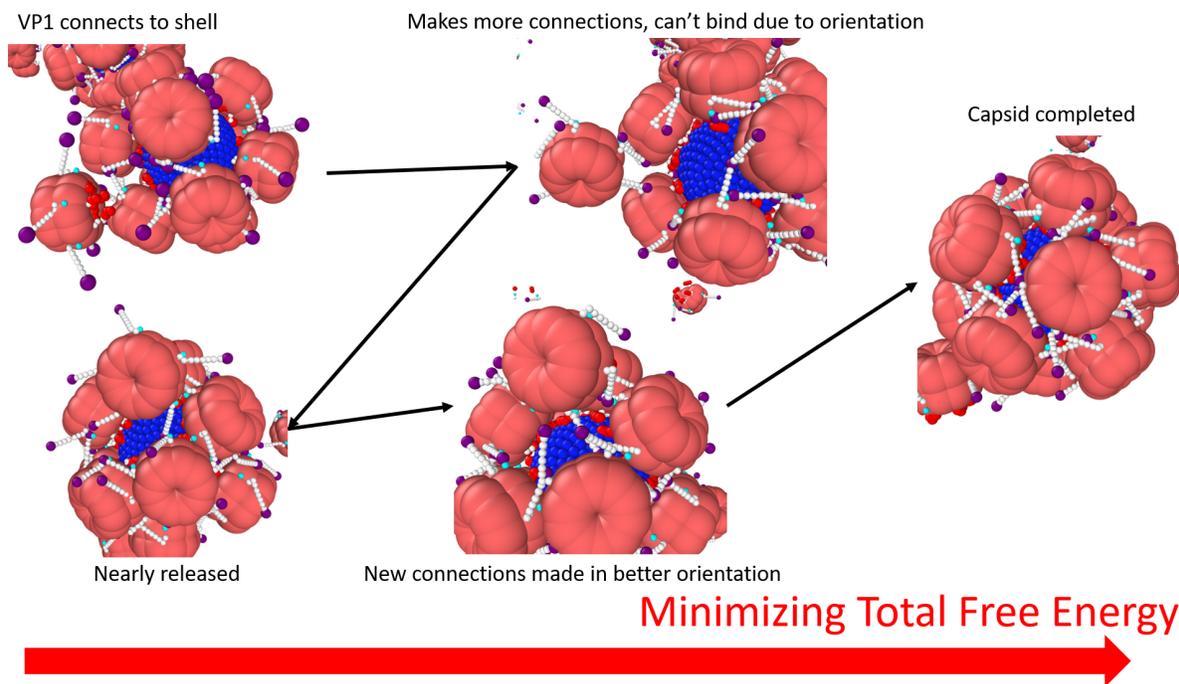


Figure 6.7. An example of binding of the final VP1 pentamer to an N=11 capsid during a simulation using $\epsilon = 8 k_B T$, $\lambda_d = 1.0 \text{ nm}$, $q_{eff} = 0$, and a 10 nm template. The black arrows point forward in time, while the red arrow points to a decrease in the total free energy. The process requires cooperative interactions between the final VP1 pentamer, the bound VP1 pentamers, and the template. This process shows that the C-terminal ligands stabilize the 12th VP1 pentamer on the partially assembled capsid until it finds an orientation where it can bind to the surface of the template. To find this orientation, the VP1 pentamer searches a rough free energy landscape and thus it is important that the interactions are weak enough to be reversible.

regimes and distance from the globular body at which they would occur are not present. This could be especially important in making first contact when attempting to add the final VP1 pentamer to an N=11 capsid, see Figure 6.8. To more accurately represent the 60 amino acid C-terminal ligand, 2 more beads were added to the end of the C-terminal ligand with equilibrium bond lengths of 2 nm, doubling the contour length (see Figure 6.9 in the Appendix). The potential between these beads is such that at a cost of $2 k_B T$

Distance at Point of First Contact

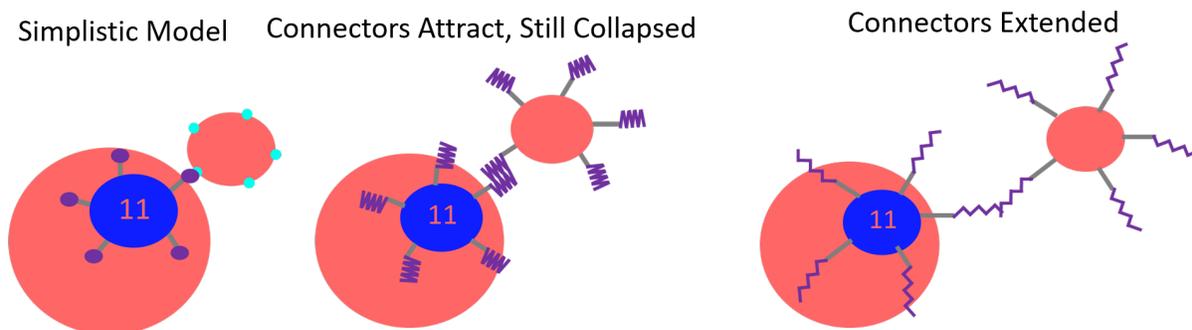


Figure 6.8. An illustration of the shortcomings of the simplistic model to represent the ease of making the first contact. Considering the full extensibility of the connector domain, the first contact between a partial capsid and a VP1 pentamer can occur at a much longer interaction distance.

the contour length of the new section can be 2 or 6 nm giving the C-terminal ligand a spring-like extensibility. This model accounts for the many possible configurations of the C-terminal ligand after the α -helix, which were “not well ordered” [3] in cryo-EM reconstructions of the $T = 1$ particle. The interactions of the simplistic model are kept completely identical whereas the additional two beads will have an interaction strength of $1 k_B T$ with all other types of beads in the system. The C-terminal ligand is composed of positively and negatively charged, hydrophobic, and hydrophilic residues so in some conformations it can have a small, but attractive interaction with any portion of a template or other VP1 pentamer. These sections of the C-terminal ligands will essentially behave as random copolymers, which have been shown to orient themselves to energetically favorable conformations [41].

As shown in Figure 6.9, the inclusion of the full C-terminal ligand increased the rate of assembly and yield of $T = 1$ capsids. This shows that the entire C-terminal ligand

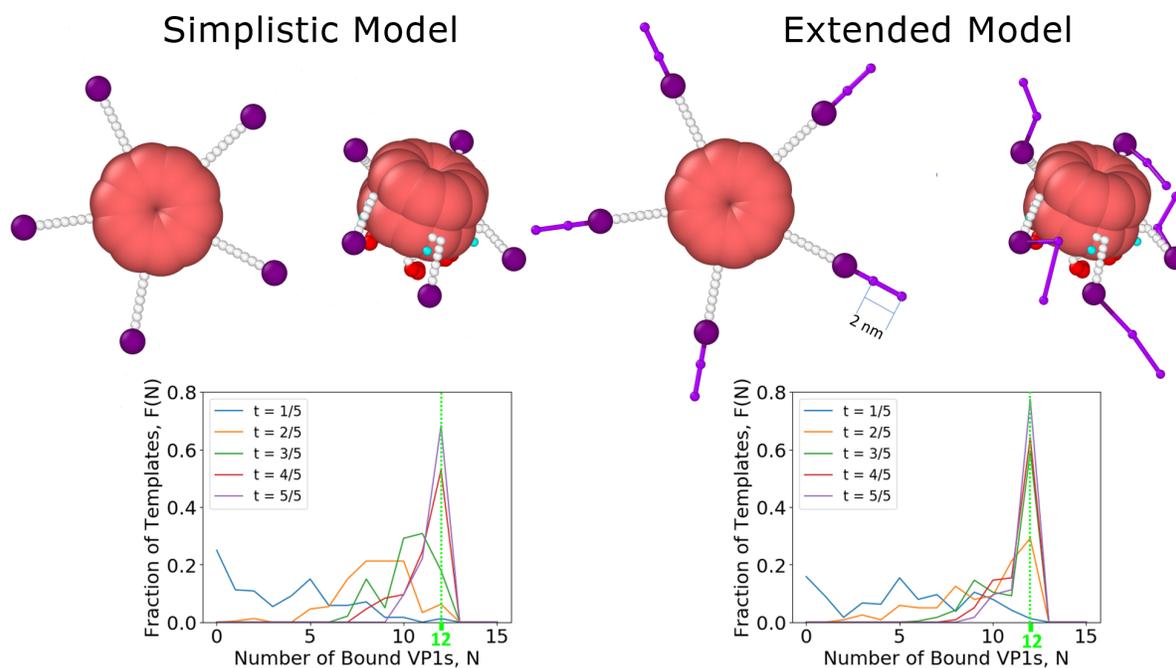


Figure 6.9. (Top) The difference between the simplistic and extended model is displayed both in an initial conformation and in a typical conformation of a free VP1 pentamer. (Bottom) Inclusion of the full length connector in the model increases the rate of assembly. More $T = 1$ capsids are present at every step of the simulations.

plays a role in assembly by increasing the range at which VP1 pentamers interact and helping to overcome the diffusion limit. This is often considered to be the role of a flexible polyelectrolyte template, but here we show the same effect can be obtained by including the extended C-terminal ligands of VP1 pentamers on a rigid template [70, 296, 297, 298, 299]. It follows that the effect does not actually require a template at all [314], although it would also serve to amplify the flexibility of a polyelectrolyte especially as it approaches full encapsidation.

6.3.8. Closed Shell Growth

Until now only the growth of incomplete capsids by elongation has been discussed, however, the growth of pseudo-closed shells (dynamic shells with no holes) can also occur. Figure 6.10 compares the assembly around spherical templates with diameters of 10 nm and 11 nm. As shown in Figure 6.5, the $N=10$ states on these templates were pseudo-closed and icosahedral with 2 holes respectively. Although in the end of the simulations both templates had some $T = 1$ capsids (*i.e.*, $N=12$ at $t=5/5$, where t is the fractional simulation time) there was a clear difference in the dynamics. The $N = 10$ state had a much longer lifetime on the 10 nm template than on the 11 nm template, as the 10 nm template has its largest peak at $N = 10$ at $t=3/5$. The higher stability can be attributed to the configurations assumed by the $N = 10$ particles at different curvatures, shown in Figure 6.5. Figure 6.10 shows the pairwise distance distributions of VP1 pentamers as a function of time. The presence of the two different $N = 10$ configurations is most clear for $t = 2/5$ at which $N = 10$ is the dominant structure on both templates. The 11 nm template displays the icosahedral pair distribution while the 10 nm template has a pseudo-closed shape similar to the $N = 10$ capsid on a 10 nm template in Figure 6.5. Thus the pseudo-closed state grows at a much slower rate, although it is surprising that it continues to grow at all.

The further growth of the $N = 10$ state on the 10 nm capsid is due to its dynamic character. Free VP1 pentamers first connect to a pseudo-closed shell and then dynamically rearrange to allow the connected VP1 pentamer to join the shell (Figure 6.11). The $N = 10$ state on a 10 nm capsid has nearly the same local packing density as the $T = 1$, $N = 12$ capsid. Yet, the results show that a $T = 1$ capsid does not continue to

Assembly on 10 nm Templates Assembly on 11 nm Templates

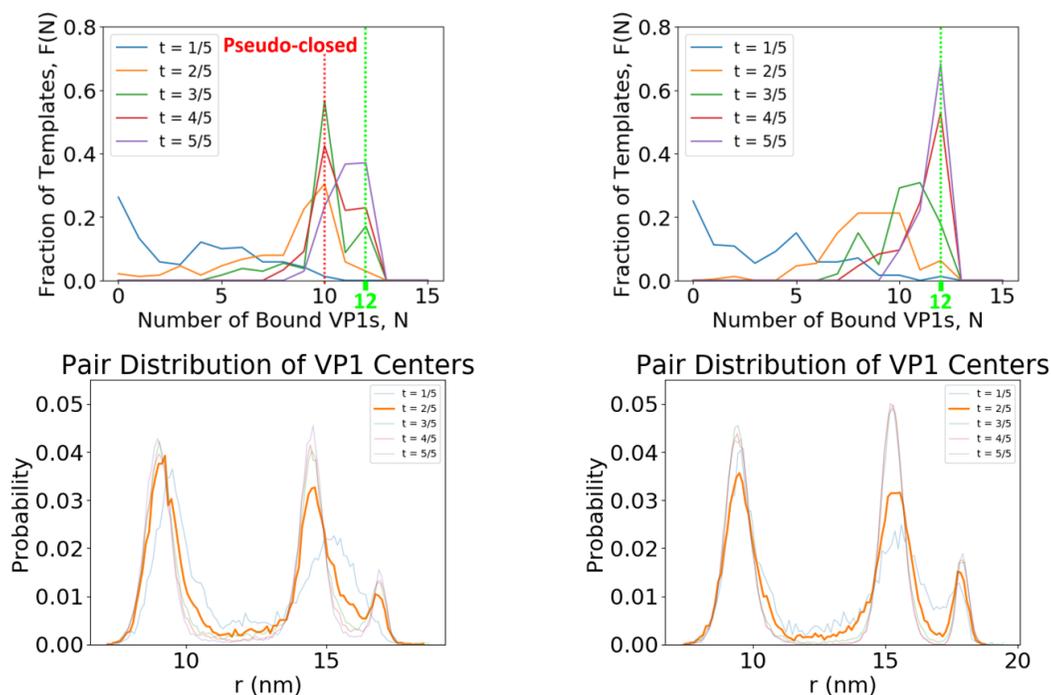


Figure 6.10. Assembly simulations around spherical templates with a diameter of 10 and 11 nm. The top figures show the fraction of templates with N bound VP1 pentamers as a function of the fractional simulation time, t . The bottom shows the pairwise distance distributions of VP1 pentamers in each case. A peak is observed at $N = 10$ pentamers on the 10 nm template owing to lack of available binding sites, as compared to the 11 nm case. The lack of available template binding sites is shown by the lack of icosahedral symmetry seen in the pair distribution (especially at $t = 2/5$).

add VP1 pentamers, because the capsid can not rearrange. The growth mechanism of pseudo-closed capsids was slower than the elongation stage, most likely, owing to the difficulty in connecting to a pseudo-closed shell and the longer rearrangement time scales. Previous experimental work [60] suggests that either these pseudo-closed $N=10$ particles

are not forming or that the time scale of the rearrangement is still untraceably fast. However, there is experimental evidence of this growth mechanism for closed shells of sizes between the $T = 1$ and $T = 7$ capsid. Attempts to solve the structure of these particles have failed.[3] We are suggesting here that they are dynamic/pseudo-closed shells and therefore they are able to grow *via* the pseudo-closed growth mechanism we describe in this paper. Our pseudo-closed growth mechanism explains this and other important experimental observations. Slow, stochastic growth *via* the pseudo-closed mechanism explains why particles often exist in heterogeneous size distributions [3, 4]. It also explains how Van Rosmalen, *et al.* observed ≈ 33 nm particles that become ≈ 45 nm particles after long incubation times [316]. Furthermore, it explains the EM images produced by Kaneshashi [4] of VP1 pentamers bound to 30 ± 5 nm particles without templates, shown in Figure 6.11.

6.4. Conclusions

We developed a model to study the assembly behavior of explicit VP1 pentamers with C-terminal ligands on rigid templates to form $T = 1$ capsids. The results of the simulations are consistent with assembly experiments of VP1 pentamers on 524 nt RNA, when the pH and salt concentration were varied. Therefore, the length and flexibility of C-terminal ligands can recruit additional pentamers to the growing capsid regardless of template flexibility. Simulations of single capsids show that the $T = 1$ capsid is static whereas all other structures are dynamic and VP1 pentamers are mobile relative to each other. Dynamic rearrangement of VP1 pentamers facilitates both the shrinking of the $N = 13$ back to a $T = 1$ capsid and the growth mechanism of pseudo-closed shells, where

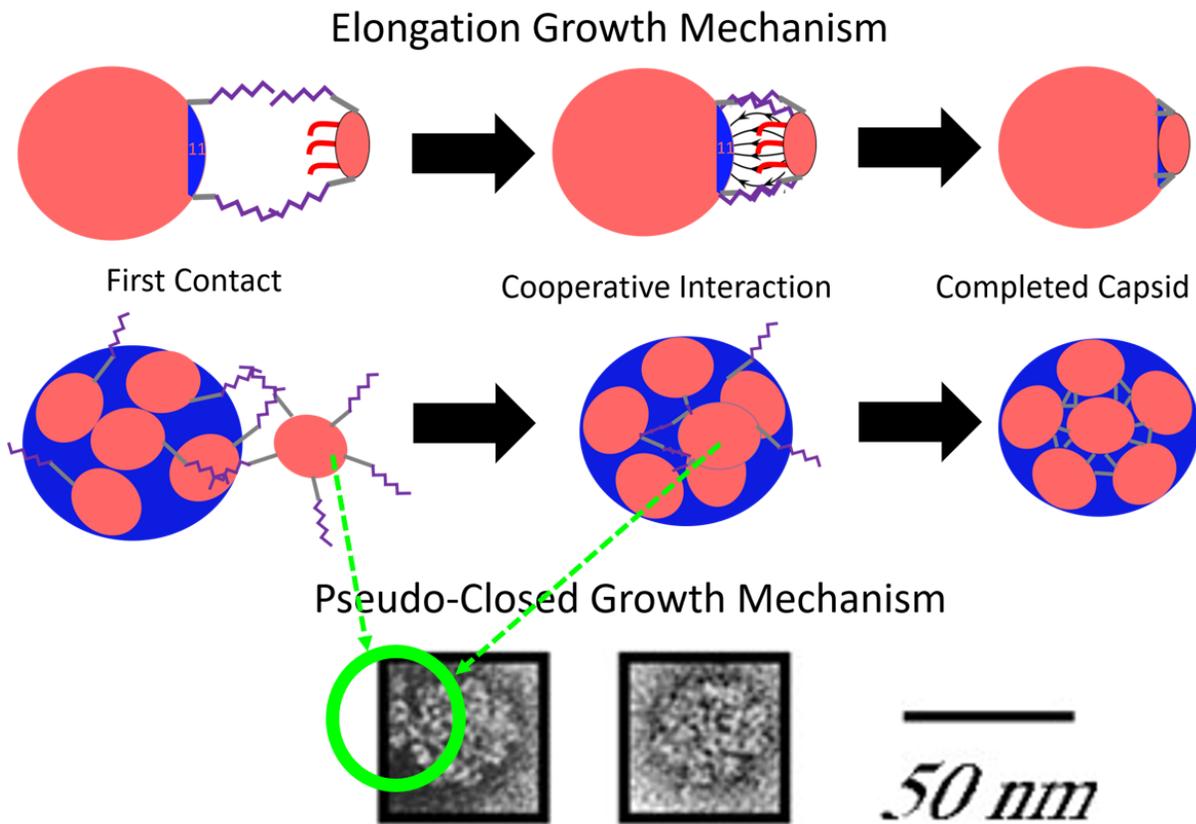


Figure 6.11. The full proposed growth mechanism of VP1 pentamers to an incomplete or pseudo-closed template. The VP1 pentamer first contacts a stray C-terminal ligand of a VP1 pentamer already bound to the template. This binding can happen at large distances due to the length of the connector domains. If the capsid is incomplete, cooperative interactions can then bring the VP1 pentamer and the incomplete capsid together, and align them such that the final VP1 pentamer perfectly fits into the icosahedral hole on the template. This process is referred to as the elongation growth mechanism. If the capsid is pseudo-closed, the connected VP1 will have to allow the capsid to dynamically rearrange in order to add a VP1 pentamer to the capsid in a much slower process. The images at the bottom show what appears to be C-terminal ligand connections being made by free VP1 pentamers on pseudo-closed particles larger than $T = 1$ and smaller than $T = 7$ using electron microscopy [4], republished with permission of the Microbiology Society, from Simian Virus 40 VP1 Capsid Protein Forms Polymorphic Assemblies *In Vitro*, Kaneshashi *et al.*, Journal of General Virology 2003, 84, 7, 2003; permission conveyed through Copyright Clearance Center, Inc. It should be noted that there is no template in this image and thus the VP1 pentamers are connected to the pseudo-closed capsid and to each other only by C-terminal ligand interactions.

dynamic shells can continue to add VP1 pentamers despite being already closed. This mechanism has been previously suggested based on experimental data. Our model shows that the C-terminal ligands are critical for the understanding of the kinetic growth of VP1 polymorphs, especially *via* the pseudo-closed growth mechanism.

6.5. Methods

6.5.1. VP1 Model

The coarse-grained VP1 model can be divided into 3 parts as shown in Figure 6.1. The first is the globular body made of 20 beads forming a rigid cylinder 8 nm in diameter and 6 nm in height. These dimensions correspond to the actual VP1 pentamer. 10 beads form a lower base ring whereas the other 10 are located 2 nm above the base ring. The body also contains 5 saturable connection sites, which allow interactions between VP1 pentamers. The second part of the structure comprises the C-terminal ligand of each VP1. Each VP1 pentamer contains 5 semiflexible C-terminal ligands, attached to the globular body. Each C-terminal ligand is made of 9 beads with a final connector bead, attracted to the connection sites *via* a Lennard-Jones potential whose strength is determined by a parameter ϵ . All of the interactions between the long connector domain, which has collapsed to fill the space between globular bodies, and the VP1 body in the final capsid state can be reduced to a single parameter using this approach. The connector beads also repel each other in order to avoid multiple C-terminal ligands connecting to a single connection site, thus making the connection sites saturable. Angle potentials along the C-terminal ligand create an inflexible domain, which represents an α - *helix* in the C-terminal ligand of the real VP1. The C-terminal ligand, though rigid, rotates freely. The

VP1 pentamer also contains 5 N-terminal ligands, made of a flexible chain containing 5 beads. The final 3 beads are positively charged to represent the charged residues in the unstructured N-terminals near neutral pH.

6.5.2. Template Model

The VP1 pentamers assemble around rigid, impenetrable spheres which have a uniform surface charge density (see the top of Figure 6.3). The total charge is $-524e$, which is the same as the total charge on the RNA used in the experiments. This choice of a rigid, spherical template is informed by the knowledge that the final conformation of polymers encapsidated by spherical viruses also tend to be spherically symmetrical [317, 318] or close to it [319] and thus will be accurate for completed capsids.

The rigid sphere excludes an *en masse* assembly path where a flexible polyelectrolyte first binds all viral capsomers in an unorganized fashion and then the capsid anneals in a slower step, which requires the template to change conformation as shown in other theoretical[315] and experimental work[79]. However, previous work with SV40 on short RNA's, the same RNA used in the experimental section of this paper, did not show an annealing step [320] in the assembly and instead proposed that it follows the nucleation - elongation mechanism, where incomplete capsids form before all VP1 pentamers are added. This same mechanism has also been shown in monte carlo simulations on cargo made of monomers, which are not covalently bonded to one another. [78] Although this work with SV40 [320] suggested the importance of the flexible template in binding VP1 pentamers to the incomplete capsid, here we show that the C-terminal ligands of the VP1 pentamers increase the effective size and flexibility of the incomplete capsid no matter the

template. To summarize, even though the rigid sphere does not capture the flexibility of the RNA, it is accurate to describe the conformation of the RNA inside of closed capsids. Moreover, once the partial capsid forms the C-terminal ligands determine the effective size of the aggregate. We expect this partial capsid will form *via* nucleation-elongation as previous experiments have ruled out *en masse* assembly on 524 nucleotide RNA.

6.5.3. Solution Salt Conditions

Electrostatics in these simulations are handled using an implicit ion model with two types of interactions. First, we extend the two state model of Alexander *et al.* for spherical symmetry [321] to the case of compacted flexible polymers[322, ?] to construct an electrostatic potential for the template. This model introduces nonlinear effects in a Yukawa potential due to ion condensation *via* an effective reduced charge, q_{eff} , which can be computed using the nonlinear Poisson-Boltzmann equation for ion penetrable spheres[323] as opposed to the dense colloids used by Alexander *et al.* [321]. We note that polymer entropic effects are negligible as well as the ionic correlations inside the collapsed polymer in monovalent salts[322] (only in the presence of multivalent ions are these correlations significant[324]). Therefore, we do not include short-range interactions explicitly but assume the template is spherical, as mentioned in the Template Model section.

With these assumptions, the electrostatic energy is represented by a Yukawa potential given by,

$$(6.1) \quad F_{ele} = \frac{-q_{eff}^2 l_b e^{-\frac{r}{\lambda_d}}}{r}$$

where q_{eff} is in units of the elementary charge e , $l_b = e^2/(4\pi\epsilon_0\epsilon_r k_B T)$ is the Bjerrum length (ϵ_0 and ϵ_r are the permittivity of vacuum and the relative dielectric constant of the medium, respectively), $\lambda_d = \frac{1}{\sqrt{4\pi l_B c s}}$ is the Debye length, which in NaCl can be approximated by $\lambda_d(nm) = \frac{0.304}{\sqrt{I(M)}}$, where I is the ionic strength of the solution. This form of the Debye length breaks down at salt concentrations above about 300 mM [325, 326]. Since the effective charge, q_{eff} quantifies the portion of the charge not compensated by counterion condensation, at low salt concentrations (low ionic strength), q_{eff} approaches 1 while at very high salt concentration q_{eff} approaches 0. This form of the potential also provides an opportunity to mimic the deprotonation of the N-terminal ligands when the pH is increased. The deprotonation of specific amino acids at different solution pH is a highly complex function on the local environment [327]. Instead of modeling this complexity, we simply decrease the N-terminal charge as this gives the correct general trend. Since only the N-terminal ligand charges are affected and not the template charges, the N-terminal ligand charge fraction can simply be considered as q_{eff}^2 . The second type of interaction takes place at high salt concentration (above 300 mM), where in the case of not strongly charged surfaces, as in our system, the interaction is given by a short range attraction driven by the electrolyte correlations that generate a depletion type attraction at short distances, as shown by Li *et al.* [328], who showed that at high NaCl concentrations, the collective interactions between the salt ions produce depletion interactions between nanoparticles, similar to those generated by oligomers (see Appendix

in [328]). Therefore, we added a short range attraction between the charged N-terminal ligand beads and the template of strength, α

$$(6.2) \quad F_{short}(r) = \begin{cases} 0 & r \geq R + 1 \\ -\alpha & r < R + 1 \end{cases}$$

where R is the radius of the template. In molecular dynamics simulations, this attraction is implemented as an interaction between N-terminal ligand beads and a particle, located at the center of mass of the rigid template. α is positive at high salts where depletion interactions dominate.

The total free energy contribution of the implicit salt is then given by

$$(6.3) \quad F_{total}(I) = F_{short}(\alpha(I)) + F_{ele}(q_{eff}(I), \lambda_d(I)).$$

6.5.4. Simulation Methods

Two types of simulations were run. In the first, one large rigid template was initialized and surrounded with a given number of VP1 pentamers which bind immediately to the surface. The diameter of the template is then reduced slowly such that no VP1 pentamers are released during this process until the template becomes too small for all of the VP1 pentamers to bind. Individual states are then run for longer times in order to study the capsid at different curvatures.

The second type is the assembly simulations. Simulations are run by initializing $6 \times 6 \times 6$ simple cubic unit cells with a spacing of 25 nm for a total of 216 VP1 pentamers. This concentration is much higher than what is used experimentally but is useful to accelerate the simulations. 8 templates of a given size are then added to the box in an evenly spaced fashion. Any clashes between VP1s and templates are solved by removing the involved VP1 pentamers. Simulations are run for 200 million timesteps at $dt = .004 \tau$.

All simulations are run in HOOMD [127] using Langevin integration.

6.5.5. Type Mapping

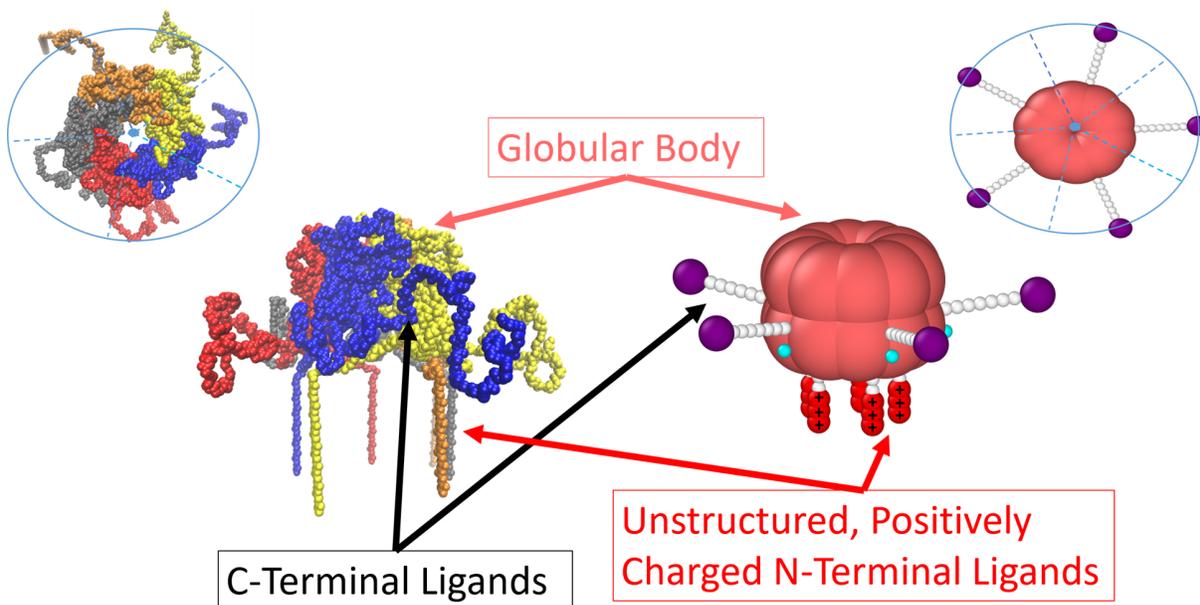
Energies are in units of $k_B T$ and distances are in units of nanometers.

Colors refer to Figure 6.12. The globular body is made of 21 pink ‘X’ beads. Two rings made of 10 beads located in a circle of radius 2 nm and one in the middle. The second ring is 4 nm above the first. All of these beads are part of one rigid body.

There are also 5 cyan ‘D’ beads representing the contact site which are located on the surface of the body 0.5 nm below the bottom ring.

The C-terminal ligands are made of 9 beads, 8 white ‘P’ beads and 1 purple ‘C’ bead, harmonically bonded to form a chain. The first bead is rigidly attached to the surface of the body 0.5 nm above the base ring. A series of angle potentials centered on beads 3-8 (the rigidly attached bead is number 1) keeps the rest of the beads nearly perfectly aligned as in an α – *helix*, while allowing it to leave the body without angular restriction. When the full length of the chain is included (see Part B of Figure 6.12), the chain is extended by two light purple ‘S’ beads which have an equilibrium an equilibrium bond

A. Original Model of VP1 Coarse Grained Structure



B. Updated Model Including Full C-terminal ligand Length

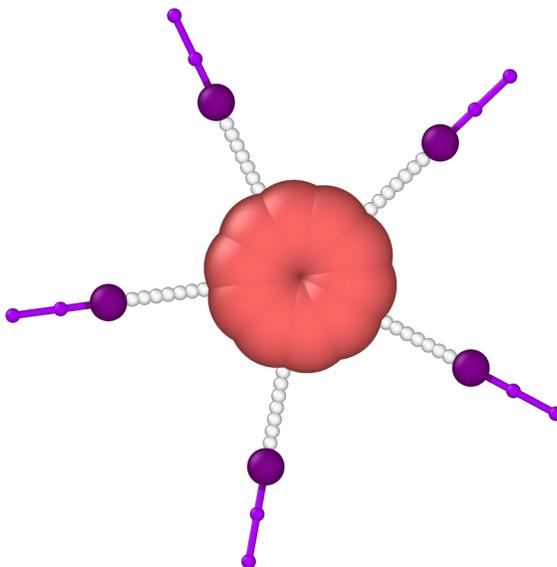


Figure 6.12. Part A shows the coarse grained structure of the VP1 including the globular body, C-terminal ligands, and N-terminal ligands. A full description of the individual beads that make up these sections and the potentials and parameters, which control their functionality is described in detail in the main text. Part B shows the addition of the full C-terminal ligand length upon updating the model. The parameters and potentials for the ‘S’ beads that make up this section of the VP1 are also described in detail in the main text.

length of 2 nm. These two beads model an extension of the chain by nearly 40 amino acids.

The N-terminal ligands are made of 5 beads, 2 white ‘P’ beads followed 3 red ‘qPp’ beads, all harmonically bonded to form a chain. The first bead is rigidly attached to the surface of the VP1 2 nm directly below the bottom ring.

Templates are made of ‘qPm’ beads, which are uniformly distributed on the radius of a sphere at a packing density which makes them impenetrable to other beads. They also contain a ‘center’ bead located at the exact center of the rigid template.

6.5.6. Volume Excluding Interactions

Volume Excluding interactions are handled through a purely repulsive Lennard-Jones interaction where $\epsilon=1 k_B T$ shown in Equation 6.4. Parameters are shown in Table 6.5.6.

$$(6.4) \quad U_{Exclusion}(r) = \begin{cases} 0 & r \geq \sigma \\ \frac{\sigma^{12}}{r} - 1 & r \leq \sigma \end{cases}$$

Type	$\sigma(\text{nm})$
X	2 no matter the other type (no mixing)
C	0.5
D	0.5
D-D	2.0
P	0.5
qPp	0.5
qPm	1.0

The mixing rule is simply arithmetic average with a few exceptions. ‘X’ beads where σ is always 2.0 nm to avoid contact site blocking. The σ value for D-D interactions is 2.0 nm in order to saturate the contact site.

Implicit Ions The effect of the ions is handled through two potentials as described in the main text. Both potentials are in units of KT. The first is the Debye-Huckel 6.5 between all beads whose names begin with ‘q’:

$$(6.5) \quad U_{Debye}(r, q_{eff}, \lambda_d) = \begin{cases} 0 & r \geq 3\lambda_d \\ \frac{-q_{eff}^2 l_b e^{-\frac{r}{\lambda_d}}}{r} & r \leq 3\lambda_d \end{cases}$$

where q_{eff} is in units of the elementary charge e , σ is the Lennard Jones σ , $l_b = e^2/(4\pi\epsilon_0\epsilon_r k_B T)$ is the Bjerrum length (ϵ_0 and ϵ_r are the permittivity of vacuum and the relative dielectric constant of the medium, respectively), $\lambda_d = \frac{1}{\sqrt{4\pi l_B c_s}}$ is the Debye length, which in NaCl can be approximated by $\lambda_d(\text{nm}) = \frac{0.3}{\sqrt{I(M)}}$, where I is the ionic strength of the solution.

The second potential is implemented as a potential between ‘qPp’ beads and ‘center’ beads at the center of the rigid templates. It is essentially a square well with a small linear portion over a range of 0.1 nm such that forces can be calculated in HOOMD. The full form is given in Equation 6.6, where R is the radius of the template and α is either 1 or $2 k_B T$.

$$(6.6) \quad F_{short}(r, q_{eff}) = \begin{cases} -\alpha & r \leq R + .9 \\ \alpha * r^{-(R+1)/.1} & R + .9 < r \leq R + 10 \\ r > R + 1 \end{cases}$$

6.5.7. Contact Site Interaction

The contact is a classic Lennard Jones Potential (Equation 6.7), where $\sigma = 0.5$ nm and ϵ is a parameter which represents VP1-VP1 contact strength as explained in the main text.

$$(6.7) \quad U_{LJ}(r, \epsilon) = 4\epsilon \left(\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right)$$

6.5.8. ‘S’ bead interactions

The two ‘S’ beads at the end of each C-terminal ligand represent a mostly unstructured 40 amino acid sequence containing positive, negative, hydrophobic, and hydrophilic residues. Thus it can likely find some confirmation where it has a weak positive interaction with any foreign object. To model this the ‘S’ beads are given a $1 k_B T$ attractive interaction (i.e., $\epsilon = 1 k_B T$) with all of the other beads in the system (except ‘qPm’) through a Lennard-Jones potential shown in Equation 6.7. The σ for ‘S’ beads is 0.5 nm and all

combinations are the same as in the repulsive Lennard Jones see Table 6.5.6. Since we want the ‘S’ bead interaction with the entire template object to be $1 k_B T$ and not the interaction with individual ‘qPm’ beads we borrow from the strategy used in the case of counterion release. Thus there is a Lennard Jones potential between ‘S’ and ‘center’ beads, where $\sigma = R+1$ and $\epsilon = 1 k_B T$.

6.5.9. Bonds

Nearly all bonds in the system are harmonic with a bond strength, $k = 300 \text{ nm/rad}^2$, and equilibrium distance, $r_0 = 0.5 \text{ nm}$. The harmonic bond potential is shown in Equation 6.8.

$$(6.8) \quad U_{Bond}(k, r) = \frac{k}{2}(r - r_0)^2$$

The only exception to this is for C-terminal ligand extending bonds containing the ‘S’ beads, which have use $k = 1 \text{ nm/rad}^2$ and $r_0 = 2.0 \text{ nm}$. As discussed, this is to model the configurations of a much longer section of the chain.

6.5.10. Angles

The only angle potential used is a harmonic angle potential with strength, $k = 900 \text{ nm/rad}^2$, and equilibrium angle, $\theta_0 = \pi$.

$$(6.9) \quad U_{Angle}(k, \theta) = \frac{k}{2}(\theta - \theta_0)^2$$

6.6. Appendix

6.6.1. Calibration of Connection Strength

The strength of VP1 connections, ϵ , is investigated using 6,7,8,9, and 10 $k_B T$. No assembly of $T = 1$ particles is observed at 6 $k_B T$, and very little at 7 $k_B T$. When we use 8 $k_B T$ the maximal yield of $T = 1$ particles are assembled. This is the connection strength that is used in all simulations in the main text. At 9 $k_B T$ the $T = 1$ particles are found alongside completed particles while at 10 $k_B T$ only aggregates are found. In Figure 6.13, these results are summarized by a combination of simulation snapshots and distributions of bound pentamers to the templates. While, this was intended to calibrate the model experimental results suggest that this strength can be changed in different solution conditions.

6.6.2. N-terminal Ligand Charge Deprotonation

In order to investigate the affect of pH we slowly reduce the fraction of charge on the N-terminal ligands mimicking the affect of deprotonating the charged amino acids as the pH is raised. At charge fractions near 1 no real difference is observed. Then we see a sharp decrease in the fraction of $T = 1$ particles at an N-terminal ligand charge fraction of .4, where all states $N \geq 12$ seem equally probable, as shown in Figure 6.14. This charge fraction is the only place in the simulations where we see particles with $N \geq 6$ losing VP1 pentamers during an assembly simulation although it was also observed when the N=13 template was decreased in size. When the charge fraction is further reduced nucleation is blocked entirely.

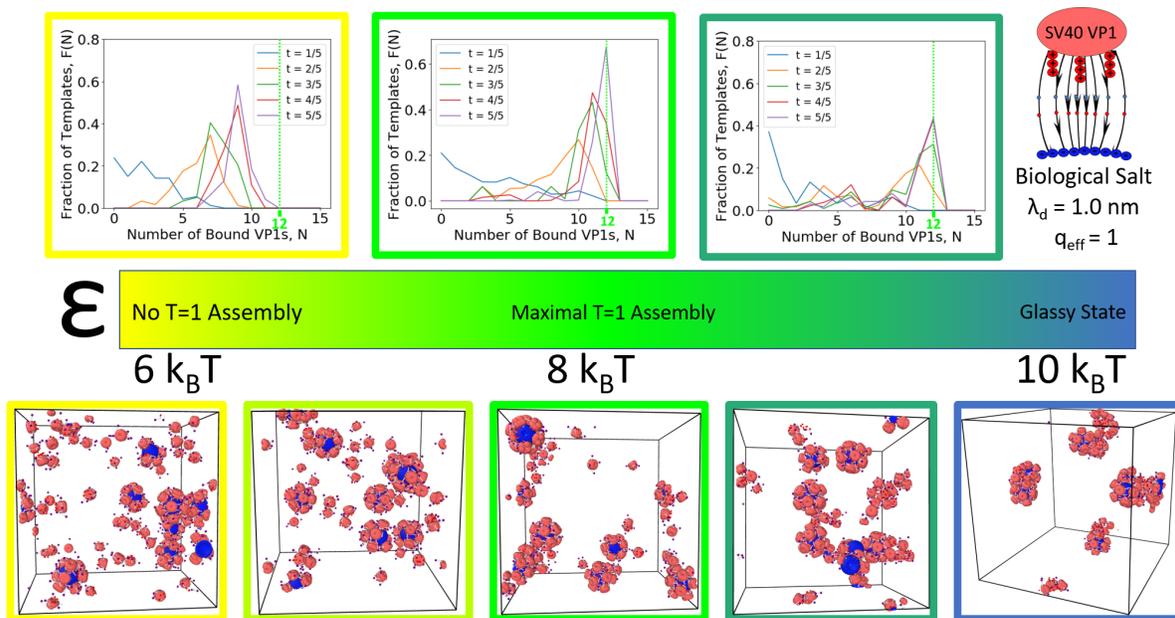


Figure 6.13. Assembly products at different values of ϵ under biological salt conditions ($\lambda_d = 1$ nm) is shown through distributions of bound pentamers to the templates and snapshots of the simulation box. Maximal assembly is observed at $8 k_B T$ whereas no assembly is observed at $6 k_B T$. At $10 k_B T$ kinetically trapped, glassy states form. $\epsilon = 7$ and $9 k_B T$ show intermediate behaviors.

6.6.3. N=11 Structure

The investigation of structures smaller than the $T = 1$ (i.e., $N=11$) provides the opportunity to understand possible intermediate states that may occur during assembly. In Figure 6.15, it is shown that the $N=11$ capsid preserves the icosahedral symmetry and shows only small deviations from the $T = 1$ capsid. Nearly the same peaks are observed although they have slightly shifted. In this case $\frac{d_{\text{center,vertex}}}{\text{edge length}} = \frac{8.1}{8.7} = 0.93$, an approximately 2% difference from the regular icosahedron. The presence of non-zero probability throughout the distribution along with observation of movement during the simulation shows that the hole is mobile just like the 6 coordinated VP1 pentamer in the $N=13$

N-terminal Ligand Charge Deprotonation

11 nm Template

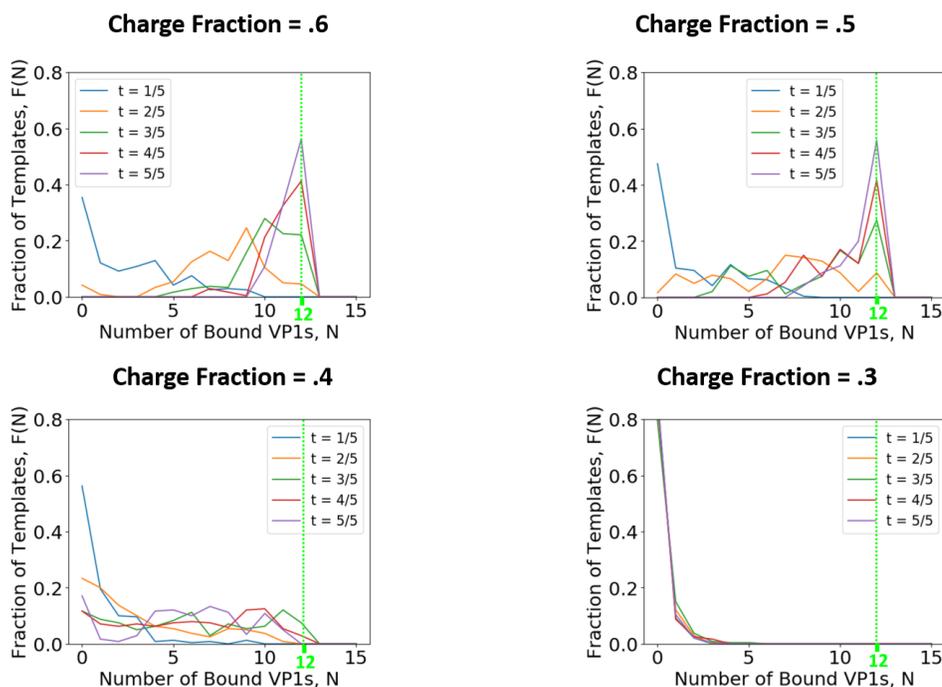
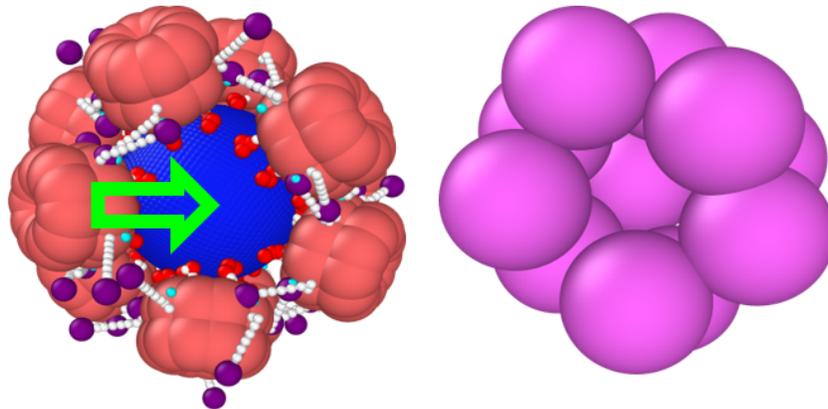


Figure 6.14. Distributions of the number of bound VP1 pentamers at different N-terminal ligand charge fractions. Simulations are run with $\epsilon = 8 k_B T$, $\lambda_d = 1.0$ nm, and an 11 nm template. The charge fraction of the N-terminal ligand is q_{eff}^2 . As the fraction of charge is decreased, initially little difference is observed, followed by an increase in the polydispersity of $N \geq 10$ particles increases. Nucleation of the particles then becomes limited and eventually, at N-terminal ligand charge fraction of 0.3, the nucleation of particles is completely blocked.

particle. This highlights the fact that the VP1-VP1 connections are able to break and reform, which is important for assembly behavior. Although it is mobile, the presence of the hole surrounded by many free C-terminal ligands and connection sites provides a consistent binding site for the 12th VP1 to complete the capsid.

N=11



Pair Distribution of VP1 Centers

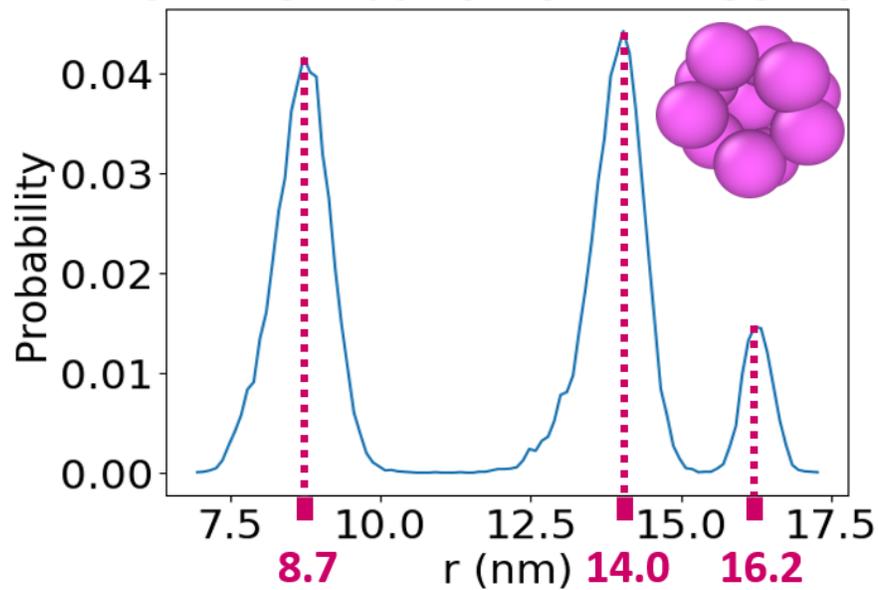


Figure 6.15. The $N=11$ capsid forms a nearly perfect icosahedron with one missing vertex. The top of the figure shows this visually whereas the bottom of the figure shows it through the preservation of the icosahedral pairwise distribution from the $T = 1$ case. The important difference between this pairwise distribution and that of the $T = 1$ is the presence of non-zero probability throughout the distribution in the $N=11$ case (bottom). It reflects the dynamic movement of the hole whereas the $T = 1$ capsid is static (see main text).

6.6.4. Euler Formula for an Icosahedral Shell of Pentagons and Hexagons

The basic Euler Theorem

$$(6.10) \quad F = E - V + 2$$

where F is the number of faces, V is the number of vertices, and E is the number of edges. Under the constraint that the shell is made only of 2 shapes, hexagons and pentagons we also have the following,

$$(6.11) \quad N = H + P$$

where N is the total number of shapes, H is the number of hexagons and P is the number of pentagons. All edges are shared between two faces.

$$(6.12) \quad E = \frac{5P + 6H}{2}$$

All vertices in our icosahedral shell are shared by 3 faces

$$(6.13) \quad V = \frac{5P + 6H}{3}$$

and the number of faces is equal to the number of shapes

$$(6.14) \quad F = N$$

By substituting N for F into Equation 6.10, using Equation 6.14 and then replacing $E - V$ with equations 6.12 and 6.13

$$(6.15) \quad N = \frac{5P + 6H}{6} + 2$$

And then replacing P with $N - H$ using equation 6.11 and rearranging it can be shown that

$$(6.16) \quad H = N - 12$$

implying that every shape added after 12 must be a hexagon to maintain the given constraints. In other words, all VP1 pentamers added to a structure where $N \geq 12$ must be 6 coordinated.

CHAPTER 7

Kinetic Growth of Multi-Component Microcompartment Shells

This chapter is based on the work of Curt Waltmann, Nolan W. Kennedy, Carolyn E. Mils, Eric W. Roth, Svetlana P. Ikononova, Danielle Tullman-Ercek, and Monica Olvera de la Cruz, **Kinetic Growth of Multi-component Microcompartment Shells**, which has been submitted to ACS Nano. Certain details have been modified.

7.1. Abstract

An important goal of systems and synthetic biology is to produce high value chemical species in large quantities. Microcompartments, which are protein nanoshells encapsulating catalytic enzyme cargo, could potentially function as tunable nano-bioreactors inside and outside of cells to generate these high value species. Modifying the morphology of microcompartments through genetic engineering of shell proteins has been shown to be a viable strategy to tune pathway performance. However, this is a difficult task without understanding how changing interactions between the many different types of shell proteins and enzymes affects microcompartment assembly and shape. Here, we use multi-scale molecular dynamics and experimental data to describe assembly pathways available to microcompartments comprised of multiple types of shell proteins with varied interactions. As the average interaction between the enzyme cargo and the multiple types of shell proteins is weakened, the shell assembly pathway transitions from (i) nucleating on the enzyme cargo to (ii) nucleating in the bulk and then binding the cargo as it grows to

(iii) an empty shell. Atomistic simulations and experiments using the 1,2-propanediol utilization microcompartment system demonstrate that shell protein interactions are highly varied and consistent with our multi-component, coarse-grained model. Furthermore, our results suggest that intrinsic bending angles control the size of these microcompartments. Overall, our simulations and experiments provide guidance to control microcompartment size and assembly by modulating the interactions between shell proteins.

7.2. Introduction

Microcompartments (MCPs) are proteinaceous shells that compartmentalize metabolic pathways in bacteria [256, 222]. The two main classes of these shells are carboxysomes, which encapsulate the enzyme Rubisco that is used for carbon fixation in autotrophic cyanobacteria [245, 329], and metabolosomes, which encapsulate catabolic pathways in different kinds of heterotrophs [330, 271]. Both carboxysomes and metabolosomes are polydisperse in size [240] ranging from 40-400 nm and composed of different types of shell proteins [268, 331, 332, 333]. These shell proteins include 5-sided pentamers, as well as 6-sided hexamers and pseudo-hexameric trimers, which have two unique types of sides and only 3-fold rotational symmetry. There are often many different kinds of hexameric shell proteins present in MCPs. For example, the 1,2-propanediol utilization (Pdu) metabolosome contains at least eight different types of shell proteins [268] as shown in Figure 7.1(a). The different types of shell proteins assemble together around enzymatic cargo, selectively controlling diffusion of substrates, products, and intermediate species in and out of the enzyme core [334, 57, 266]. These shells are polyhedra and studies suggest that surface patterns with different components are relevant to their biological

function [335, 268, 332, 336, 337, 338]. Since the behavior of MCPs is determined by its constituents, the engineering of shell proteins has been widely studied to create designer nanoreactors or “metabolic modules” where engineers control MCPs’ selective permeability, size, shape, and enzymatic cargo [5, 339, 329, 340, 56]. However, this is a difficult task without sufficient knowledge of the process by which these multi-component shells assemble. Here, we explore how the varying interactions of shell proteins affect their roles in shell assembly and lead to complex growth kinetics that expand the range of interactions for successful assembly. This is a specific example of a general principle in biology [75, 76, 77], where multiple components stabilize the formation of subcellular compartments. In this work, we extend this concept to MCPs using the example of the Pdu system.

There have been multiple studies of the growth of closed shells made of pentagonal and/or hexagonal components [341, 78], much of it done in the context of viruses and virus-like particles [61, 342, 60, 62, 343, 63]. Many viruses have only one type of shell protein oligomer [2, 344, 345] and it is thought that this is advantageous as it minimizes the length of the virus’s genetic material. Thus, previous models of shell assembly have also tended to consider only the minimal number of shell proteins necessary to form a closed shell [207, 68, 69]. However, the large number of shell proteins components in the Pdu MCP system (see Figure 7.1(a)) and others [235, 271] shows that this design strategy for biological shells is not universal and shells with multiple shell proteins should be considered. Recent experimental results for the Pdu MCP system highlight the need for this type of analysis. Specifically, the deletion of one minority shell protein, PduB (see Figure 7.1 (a) and (b)) causes the MCP to form empty shells that do not assemble around

enzymatic cargo [72, 346, 74]. To accurately describe how removal of specific hexamers changes the assembly pathway [73, 74], models must contain multiple different types of shell proteins. Here, we include three different types of coarse-grained hexameric shell proteins shown in Figure 7.1(c) and (d) and a phase separated cargo. These shell protein deletion experiments along with all-atom measurements of binding strengths between Pdu shell proteins (Figure 7.2(b)) motivate the choice of parameters in the coarse-grained model.

Previous models have illustrated the importance of multiple factors in shell growth, including the bending rigidity or flexibility of the components [65, 282, 207, 66] and especially the role of the cargo in driving the coalescence of shell proteins [3, 305, 68, 347]. For this reason, assembly mechanisms are characterized by how shell proteins interact with both one another (the shell-shell interaction, U_{SS} in Figure 7.2(d)) and the cargo (the shell-cargo interaction, U_{SC} in Figure 7.2(d)). There are two main assembly mechanisms that have been described for shells with encapsulated cargo, both of which require the cargo to nucleate formation of the shell [254, 341, 79]. The core-first assembly mechanism relies on phase separation of the enzymatic core followed by the nucleation of the shell on the previously assembled core. In concomitant assembly, the shell proteins and enzymatic core nucleate together simultaneously. In the context of viral assembly, the cargo is often a single nucleic acid [348, 3], charged polymer, or a nanoparticle [70, 296, 349] and thus all assembly is core-first (often referred to as templated assembly) unless there is no cargo at all [4] and the shell is “empty”. Core-first assembly can be further divided into *en masse* assembly, where many shell proteins bind the template and then anneal into a final structure, and assembly *via* elongation where a nucleus of shell proteins is

formed on the template followed by growth one shell protein at a time into a full, closed shell [79]. However, experiments [74, 72] have led to the idea that cargo-encapsulating shells may be nucleating separately from the enzyme cargo and thus the shell assembly is bulk-nucleated instead of cargo-nucleated. The coarse-grained model explored in this work with three shell protein components describes this scenario.

There are also experimental results on open [271, 350, 270, 269, 1] shells, which have gaps or holes. Interestingly, recent work has shown that open shells are still catalytically functional [1], and as such, important to systems biology. In some carboxysome systems, experiments suggest any defects allow leakage of CO₂ impairing carboxysome function [255], however this does not seem to be universal. Our three component model does not include a pentamer and as such the shells have holes or defects that lead to open shells, which we measure instead of classifying them as failed assemblies.

Here, we first show that shells with multiple types of shell protein interactions can access more assembly pathways than single component shells *via* coarse-grained molecular dynamics simulations. Specifically, we label the different shell proteins as the “nucleator” (A), “recruiter” (B), and the “completer” (Z) as shown in Figure 7.2(a) due to their roles in bulk-nucleated assembly. Then, atomistic simulations of Pdu MCP shell proteins demonstrate that the varied interaction strengths necessary for the nucleator, recruiter, and completer roles exist in the Pdu system. This is done by measuring interactions of PduA with both PduB and PduB', showing a large range of interactions (2-9 kcal/mol) including the weak interactions necessary for the “completer” role in the case of PduB'. Experimentally, electron microscopy is used here to image the partial collapse of wild type and empty Pdu microcompartments which lack PduB upon dehydration. This

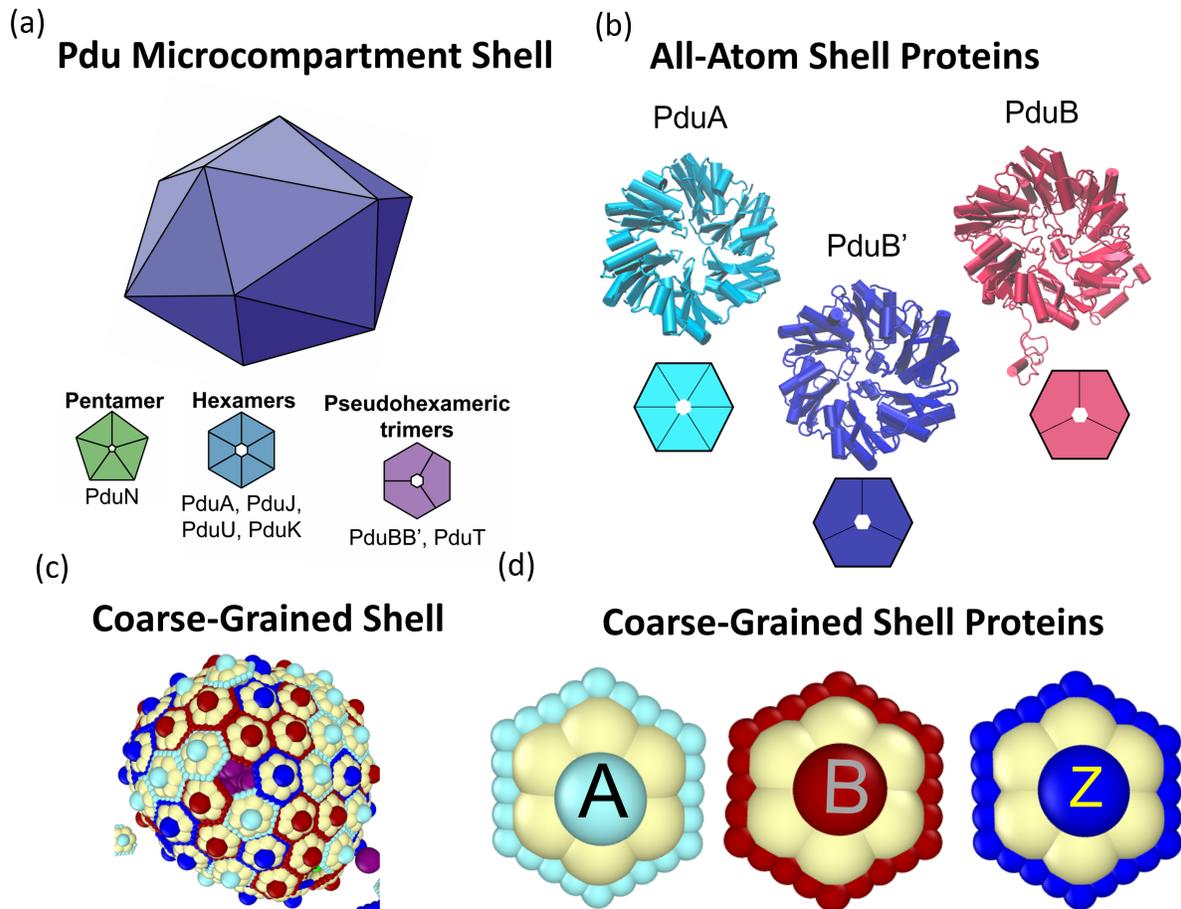


Figure 7.1. (a) The 1,2-propanediol utilization (Pdu) microcompartment shell is comprised of many different types of alphabetically named shell proteins including hexamers, pseudo-hexameric trimers, and a pentamer. (b) Secondary structure representations of the shell proteins which are selected for all-atom simulations probing their interactions with one another. (c) An example of the final complete shell in the coarse-grained model containing all three different coarse-grained shell proteins named A, B, and Z (d). The colors in (b) and (d) represent similar interaction strengths in the coarse-grained and all-atom simulations.

supports the coarse-grained prediction that water constitutes a significant portion of a bulk-nucleated shell's volume. We also observe that the shell is much smaller without PduB, which is consistent with our atomistic simulations measuring intrinsic curvature

at PduB interfaces. Overall, the work suggests the size of Pdu MCPs can be engineered *via* shell protein-shell protein interactions.

7.3. Coarse-Grained Simulation Model

We perform coarse-grained molecular dynamics simulations containing up to three different kinds of hexamer shell proteins and cargo beads shown in Figure 7.2(a). Each hexamer protein is comprised of a rigid body featuring many different bead types, which are named S_j , where $S = A, B, \text{ or } Z$ and $j = 1 - 6$. These S_j beads handle the specific interactions that align the hexamers and the form of their interactions is equivalent across species A, B, Z. The energies of these interactions, and thus the total interaction for different types of hexamers, are given by the parameters (examples in Figure 7.2(a)) in the matrices in Figure 7.2(d-f) for a one, two, or three shell protein system respectively. Hexamer-hexamer interactions are given by the parameter U_{SS} (shell-shell), where $S=A, B, \text{ or } Z$ for a system with more than one type of shell protein (Figure 7.2(e),(f)). The total interaction between the hexamers is $3 * U_{SS}$ in units of thermal energy, $k_B T$. This total interaction is varied between 3 and $15 k_B T$ by varying U_{SS} from 1 to 5. Shell protein A will have stronger interactions ($U_{AA} = 12 k_B T$) as has been shown experimentally and in atomistic simulations. Hexamer-cargo interactions are determined by the parameter U_{SC} , where $S=A, B, \text{ or } Z$ for system with multiple types of hexamers. U_{SC} is given units of thermal energy, $k_B T$ and is varied from 0-10 throughout this manuscript. Cargo-cargo interactions are always set to a value such that the cargo spontaneously phase separates (see Figure 7.2(b)) as has been shown in the Pdu microcompartment system [72].

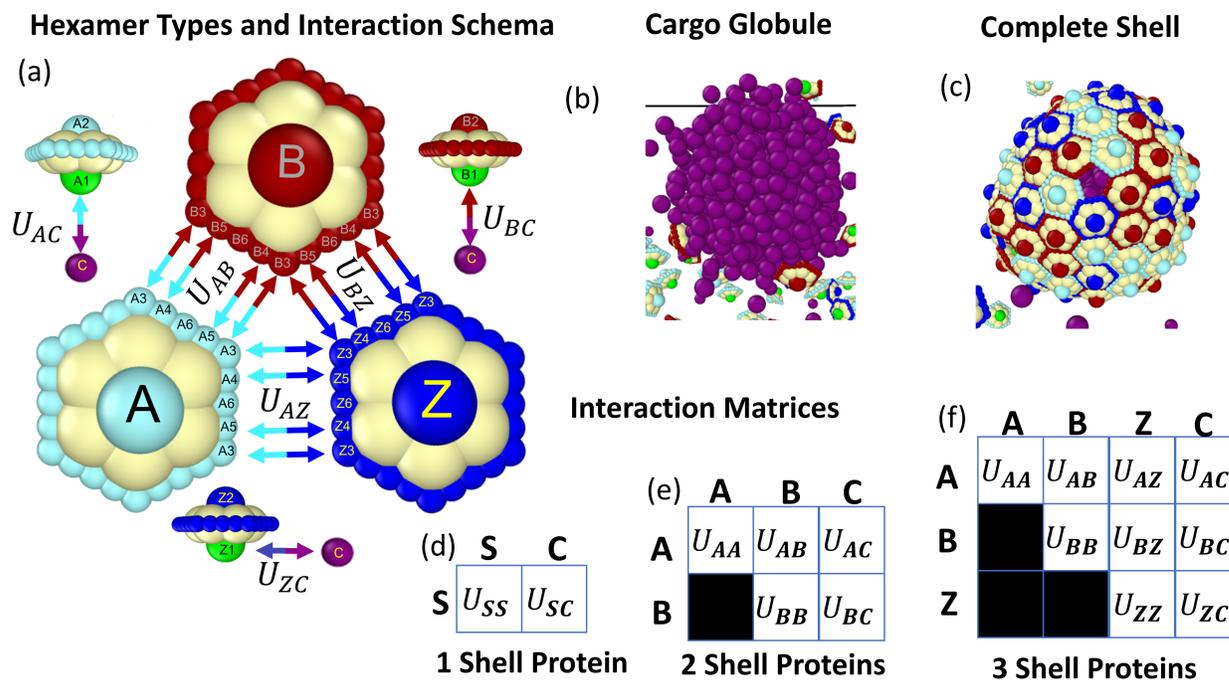


Figure 7.2. Molecular dynamics models of the assembly. (a) Our model consists of cargo beads, C, and three different hexamers denoted by A, B, and Z, each comprised of different types of beads (i.e., S1-S6, where S=A, B, or Z) that are connected in a rigid body. All bead types are equivalent across species except that their interaction energies are determined by parameters named for the two types of hexamers interacting (i.e., U_{AB} , U_{BZ} , U_{AZ}). All possible interactions are shown in the matrices in (d), (e), and (f). This includes interactions with other hexamers of the same type (U_{AA} , U_{BB} , U_{AB}) and cargo interactions, U_{AC} , U_{BC} , and U_{ZC} , which are the last column of (f). In the single hexamer system, shell proteins are simply abbreviated as S and have interaction parameters U_{SS} and U_{SC} . The total shell protein-shell protein interaction strength is $3*U_{SS}$ in units of thermal energy, $k_B T$, since it is comprised of multiple interactions on the interface. The shell protein-cargo interaction strength is simply U_{SC} in units of $k_B T$ since there is only one interaction as shown. A full list of the forms and of these potentials can be found in the SI. (b) The cargo beads attract each other such that they form a phase separated liquid droplet. This attractive cargo-cargo interaction is constant throughout. (c) An example of a complete shell. (d) Interaction matrices describe all shell-shell and shell-cargo interactions in the simulation for the case of one type of shell protein, (e) two types of shell proteins, and (f) three types of shell proteins. These parameters correspond to the interactions shown in (a).

We run our coarse-grained molecular dynamics simulations using the NVT ensemble with a Langevin thermostat. The simulations contain 148 hexamers and 600 cargo particles in a cubic box of length, 40 distance units. All simulations for a specific parameter set are averages of four simulations with unique, random initial placements of constituent particles. More details can be found in the Methods section.

7.4. Results and Discussion

7.4.1. Single Component Shells

In order to understand the behavior of multi-component shells we first studied the behavior of a system with only one type of shell protein. This establishes a baseline that we can then compare to the behavior of multi-component shells. This understanding also helps us to select components that replicate behavior seen in experiment such as PduB being responsible for assembly on a phase separated cargo globule [72]. Assembly of a single shell protein is characterized by only two parameters, the shell-cargo interaction, U_{SC} , and the shell-shell interaction, U_{SS} (see Figure 7.2(d)). Since there is only one type of shell protein, we do not specify a type for S. By varying these parameters in Figure 7.3(a), we characterize six different resulting structures. They include structures where “no nucleation” of hexamer assembly occurs, hexamers form “sheets in the bulk” that may attach to the cargo and become “sheets on cargo”, or hexamers assemble with the cargo to form either a shell or a “glassy state” when U_{SS} and U_{SC} are large. In Figure 7.3(b) and Figure 7.4 we characterize the complete shells that assemble. In Figure 7.3(b), we measure defects in these shells by the shell quality, Q_{Shell} . Q_{Shell} is defined as the number of hexamer-hexamer connections or interfaces present on the shell (6 per

shell protein, double counting), $N_{Connections}$, divided by the number of connections for a perfect shell, $N_{Perfect}^{Connections}$ of size N_{Shell} with twelve pentameric defects, which cause 60 unsatisfied connections between hexamers,

$$(7.1) \quad Q_{Shell} = \frac{N^{Connections}}{N_{Perfect}^{Connections}} = \frac{N^{Connections}}{6N_{Shell} - 60}$$

Thus, $Q_{shell} = 1$ means there are only the 12 pentameric defects while lower scores reflect the formation of more defects, approximately 1 hexameric hole for every .01 below $Q_{Shell} = 1$. This is how we define the difference between “Shells w/ Few Hexamer Holes” and “Shells w/ Hexamer Holes” in Figure 7.3(top) and values are shown in Figure 7.3(b).

Figure 7.3(a) shows perfect ($Q_{Shell} = 1$) or nearly perfect shells assemble consistently when the shell-shell interaction is relatively weak ($U_{SS} \leq 3, 9k_B T$) and the shell-cargo interaction is relatively strong ($U_{SC} \geq 5 k_B T$). Weaker shell-cargo interactions prevent nucleation, while stronger shell-shell interactions lead to highly defective states. Both perfect and defective shells tend to have around 90 hexamers depending on the exact parameters as shown in Figure 7.4(a) and there is little deviation in this size, $\sigma_{N_{Shell}}$, across different simulations with the same parameters. These shells assemble quickly (see Figure 7.4(b)), with nucleation and growth always occurring on the cargo globule when the shell-shell interaction is weak ($U_{SS} \leq 3$).

In the yellow region of Figure 7.3(a), when U_{SS} is 4 or greater, but the shell-cargo interaction is too weak to form a shell ($U_{SC} \leq 4$), sheets of hexamers nucleate in the bulk. However, when $U_{SC} \geq 1$ (the pink region of the phase diagram), the growing shell eventually binds to the cargo globule at a time which we define as T_{Bind} (see Figure 7.5(a)).

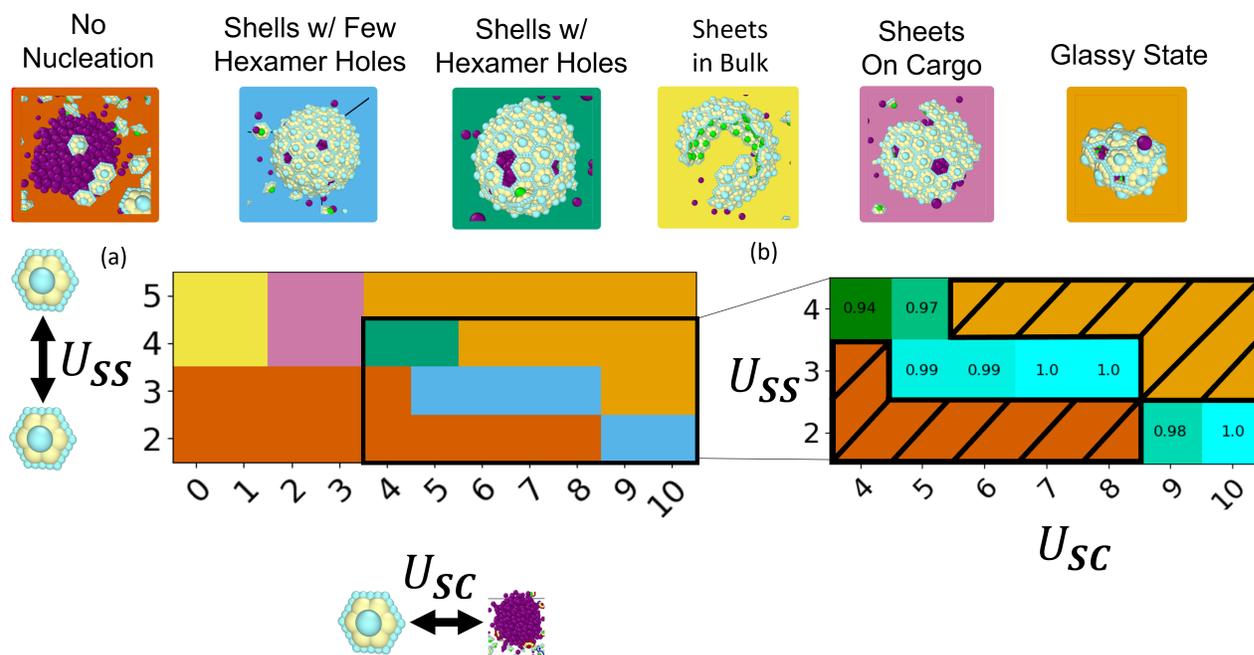


Figure 7.3. General results for the single shell protein system. (a) The six different types of structures formed are shown along with which interactions lead to their formation in a phase diagram. The x and y axes are the shell-shell interaction, U_{SS} , and shell-cargo interaction, U_{SC} , respectively. (b) The numbers correspond to the shell quality, Q_{Shell} , which measures the amount of defects (other than 12 pentameric holes) present with 1.00 meaning only those 12 gaps are present. Shells with higher shell-cargo interaction, U_{SC} , tend to have less defects.

T_{Bind} always refers to the first time a sheet binds to the cargo although it can occur many times. The binding occurs because the sheet's total interaction with the cargo is the sum of all the shell-cargo interactions of its constituent hexamers. Thus, as the sheet grows, the total interaction will reach some critical strength meaning that the shell has a critical size to bind the cargo, N_{Crit} . We attempt to measure N_{Crit} from simulation in Figure 7.5(b) as the size of the smallest sheet which binds the cargo. N_{Crit} is larger for smaller SC, but it is difficult to measure since the binding process is very stochastic. In Figure 7.5(c), we

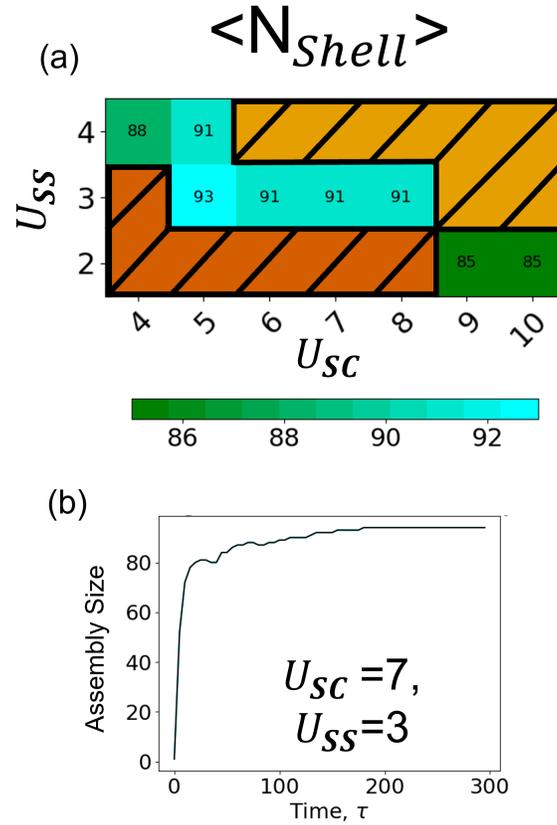


Figure 7.4. Details on one component assemblies which form shells where $Q_{Shell} \geq 0.94$. (a) The size of single component shells. All one component shells have similar cargo-nucleated assembly dynamics which look like the example shown in (b) for $U_{SC}=7$ and $U_{SS}=3$

show that the expected value for the size of a sheet when it binds the globule tends to be much larger than the critical size, confirming that the binding process is highly stochastic. Although the shell-shell interaction cannot affect the critical size, it can affect the kinetics of binding. Since it controls the rate of growth in the bulk, it also controls the timescale to reach N_{Crit} and thus the total time to for a sheet to bind the cargo, T_{Bind} , which is shown in Figure 7.5(d). This growth rate should also be affected by flexible regions on the shell proteins [207], which do not exist in the coarse-grained model but do exist in

the Pdu microcompartment system (see Figure 7.12). Experimentally, similar phenomena have been observed, as the assembly of sheet-like facets made of MCP shell proteins has been visualized using atomic force microscopy [351].

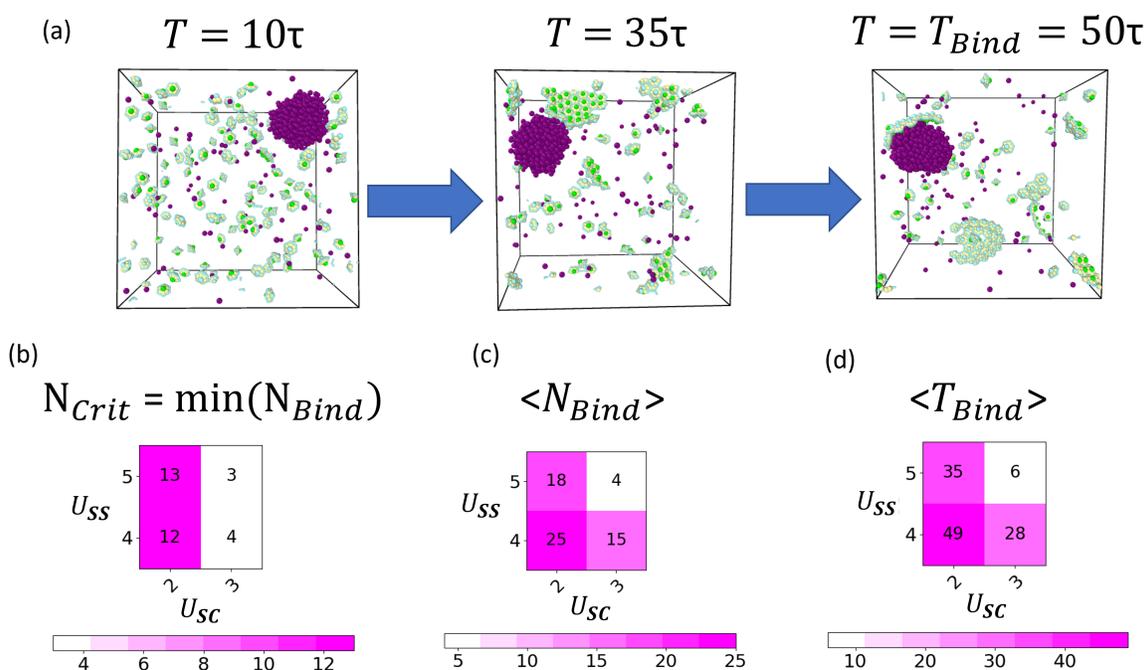


Figure 7.5. Stochastic binding of sheets to the cargo globule. (a) A typical binding process is shown through simulation snapshots at different time points. At $T=10\tau$, small nucleates can be seen forming in the bulk. At $T=35\tau$, these nucleates are more than large enough to bind the cargo, but none will until $T=50\tau$ when a large sheet randomly diffuses to the cargo. (b) The critical number of shell proteins needed in a sheet, N_{crit} , to bind the globule for sheets with different interactions. This is measured in simulations as the smallest sheet which is every observed to bind a bare cargo globule. (c) The average size of sheets that bind the cargo globule, $\langle N_{Bind} \rangle$. These values tend to be much larger than (b) due to the continued growth of sheets as they randomly diffuse and eventually bind. (d) The average time it takes for the first sheet to bind the cargo, $\langle T_{Bind} \rangle$, as a function of the shell-shell and shell-cargo interaction strengths.

7.4.2. Two Component Shells

Next, we applied our knowledge of the single shell protein system to understand how one “recruiter” shell protein can be responsible for cargo encapsulation as in the case of PduB [72]. We start with a shell component, A, that, like the shell without PduB, only assembles in the bulk. This is due to strong shell-shell interactions ($U_{AA}=4$) and weak shell cargo interactions ($U_{AC}=1$). This is shown in the yellow “sheets in bulk” region of Figure 7.3(a). For the second component, B, we choose a strong interaction with A ($U_{AB}=4$) such that it grows in the bulk with A, but a weak interaction with itself so that it does not nucleate by itself ($U_{BB}=2$).

We then vary, U_{BC} , the interaction strength between B and the cargo, C (all parameters are shown in Figure 7.6(d)). This gives two distinct regimes with different kinetics and shell quality, $\langle Q_{Shell} \rangle$, as shown in 7.6(a). When the interaction between B and the cargo is stronger ($U_{BC} \geq 8$) we see the cargo-nucleated regime, characterized by high shell quality and T_{Bind} of 0. This means that nucleation of hexamer assembly occurs on the cargo itself as in the single-component case as shown in Figure 7.6(c). In contrast, when $U_{BC} \leq 6$, $\langle T_{Bind} \rangle$ is non-zero, illustrating that nucleation of shell proteins is occurring in the bulk. This point is highlighted by one typical simulation for $U_{BC}=4$ in Figure 7.6(b). Similar to the one component case of Sheets on Cargo in Figure 7.5, shell proteins nucleate into sheets in the bulk before eventually binding the cargo as shown in Figure 7.6(b)(i). However, in the multi-component case, it is the integration of component B, with a relatively strong cargo interaction, $U_{BC} = 4$, that allows the sheet to bind the cargo globule. Once a nucleus has formed on the cargo globule, growth occurs *via* two different mechanisms. Sheets present in the bulk are still able to join the shell in a

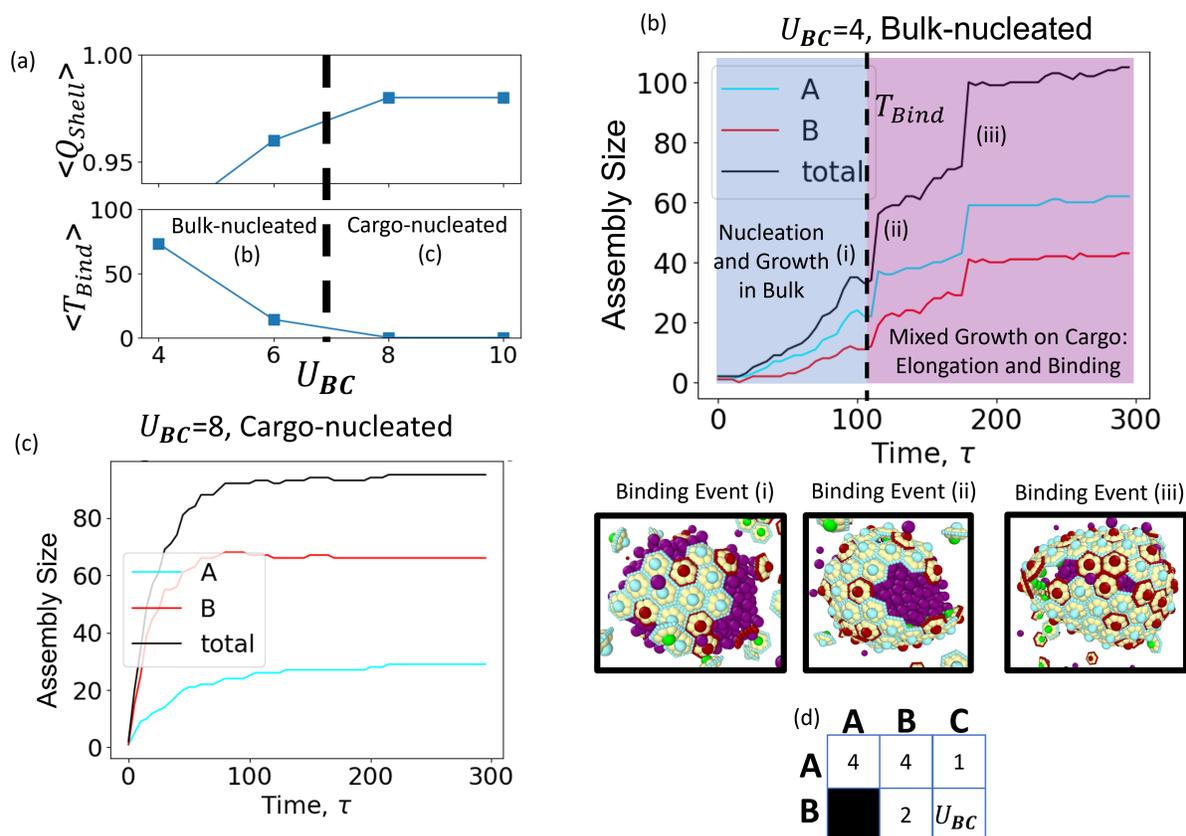


Figure 7.6. The complex growth kinetics of a two component shell. (a) The shell quality, Q_{Shell} as defined in Equation 7.1, and time to bind the cargo, T_{Bind} , as the shell-cargo interaction of the recruiter, U_{BC} , is varied. There are two distinct regimes. At higher U_{BC} , to the right of the dotted line, there is the cargo-nucleated assembly where high quality shells form by nucleating on the cargo and thus $T_{Bind} = 0$. At lower U_{BC} , low quality shells with many holes form by nucleating in the bulk and then binding the cargo at finite T_{Bind} . (b) The composition of the growing shell for a representative simulation in the bulk-nucleated regime when $U_{BC}=4$. In the blue shaded region, growth only occurs in the bulk in the form of sheets. A sheet eventually binds (i) due to the strong interaction between the recruiter, B, and the cargo globule. Once this occurs, growth happens *via* elongation of the nucleus on the globule as well as the continued binding of more bulk-nucleated sheets (ii) and (iii). (c) The composition of the growing shell for a representative simulation in the cargo-nucleated regime when $U_{BC}=8$. Nucleation and growth occur immediately on the cargo. (d) The interaction matrix for all interactions in the system. U_{BC} is the free parameter that determines the different growth pathways in (b) and (c).

stochastic manner (Figure 7.6(b)(ii) and (iii)) and thus growth post-binding can occur, while single shell proteins are also able to add to the edge of the nucleus. Notably, not all of the hexamers in the final shell are interacting with the cargo and thus there is visible empty space in contrast with the cargo-nucleated assembly (see Figure 7.6(b)(iii)). This is due to the weak overall shell-cargo interaction and is a behavior that was not seen in any single component shells. This also leads the final shell to be bigger than any one component shell with an average of 107 hexamers in the final structure (see Figure 7.7).

As the interaction strength between the recruiter and cargo, U_{BC} , is increased there are also a few other notable trends. The fraction of component A in the shell, $X_{A,Shell}$, defined as

$$(7.2) \quad X_{A,Shell} = \frac{N_{A,Shell}}{N_{Shell}}$$

decreases. This makes sense as a stronger interaction between B and the cargo leads to more B and thus less A in the shell. This implies one can engineer the composition and thus properties of MCPs by engineering shell-cargo interactions *via* encapsulation peptides [352]. For example, increasing diffusion of a certain substrate through the shell could be accomplished by increasing the shell cargo interaction of a shell protein which is responsible for the selective diffusion of that substrate. Increasing U_{BC} also leads T_{Bind} to decrease and then becomes irrelevant as nucleation occurs on the shell. Interestingly, this happens for $U_{BC}=8$, when B cannot nucleate by itself. Thus, it seems that A is solidifying a shell which is mostly made of B. To observe shell patterning that may be relevant to biological function, we quantified mixing of the components on the shell. This is measured by ρ_{ij} , probability of observing a connection between two components in the shell relative

to the probability of observing a connection between i and j if the species were mixed completely randomly. We consider the stoichiometry of the shell, not the whole system.

$$(7.3) \quad P_{ij} = \frac{N_{ij}^{Connections}}{(2 - \delta_{ij})N_{i,Shell}N_{j,Shell}}$$

$$(7.4) \quad \rho_{ij} = \frac{N_{Components} + (N_{Components})^2}{2} \frac{P_{ij}}{\sum_{i \geq j} P_{ij}}$$

We normalize such that the average value of the upper triangular ρ_{ij} matrix has an average value of 1. Thus, ρ_{ij} can be compared to the predictions regular solution model

$$(7.5) \quad F_{ij}^{mixing} = U_{ij}^{mixing} - TS_{ij}^{mixing}$$

by realizing that when $\rho_{ij} > 1$ the solution is more mixed than a random shell whose mixing is purely entropic. In this case the random solution model predicts $U_{ij}^{mixing} < 0$. For our model, the regular solution mixing parameter, χ_{ij} , is defined in terms of the shell-shell interactions of a given pair of species, U_{IJ} , from the interaction matrices, completely independent of shell-cargo interactions.

$$(7.6) \quad U_{ij}^{mixing} \propto \chi_{ij} \propto \frac{U_{II} + U_{JJ}}{2} - U_{IJ}$$

When $\chi_{ij} < 0$, the regular solution model predicts we should observe $\rho_{ij} > 1$ in the simulation and when $\chi_{ij} > 0$ it predicts $\rho_{ij} < 1$.

$$\begin{cases} \rho_{ij} > 1 & \chi_{ij} < 0 \\ \rho_{ij} < 1 & \chi_{ij} > 0 \end{cases}$$

For this two component system, independent of U_{BC} , $\chi_{AB} = -1$ and the simulations agree, with ρ_{AB} always 1.3 or greater. Despite this, ρ_{AB} is not independent of U_{BC} (see Figure 7.15) suggesting that changing the kinetics of assembly plays has an impact on shell patterning as well as stoichiometry. This is shown visually in Figure 7.7 for $U_{BC} = 10$, where the minority component A appears uniformly distributed on the shell. This is confirmed quantitatively as $\rho_{AA} = .2$, meaning A hexamers contacting A hexamers is 5 times less common than for a randomly mixed shell of the same stoichiometry. Biologically, this could allow for spatially uniform diffusion of a small molecule which displayed selective diffusion through a minority shell protein. We include all of the calculated ρ values in Figures 7.15 and 7.16.

N_{Shell} and $\sigma_{N_{Shell}}$ also decrease rapidly at higher U_{BC} , highlighting the importance of the weak cargo interactions and stochastic binding process in creating the larger, partially empty shells. The opposite trend is observed in the shell score which increases as U_{BC} increases, with the best shells occurring in the regime where B is not acting as a recruiter.

7.4.3. Three Component Shells

Although the two component shells were able to nucleate in the bulk, these bulk-nucleated shells were low quality as measured by Q_{Shell} in Figure 7.6(a). To create shells that nucleate in the bulk but also have few holes, we find it necessary add a third shell component to the system, Z. This component is analogous to other minority shell proteins that are

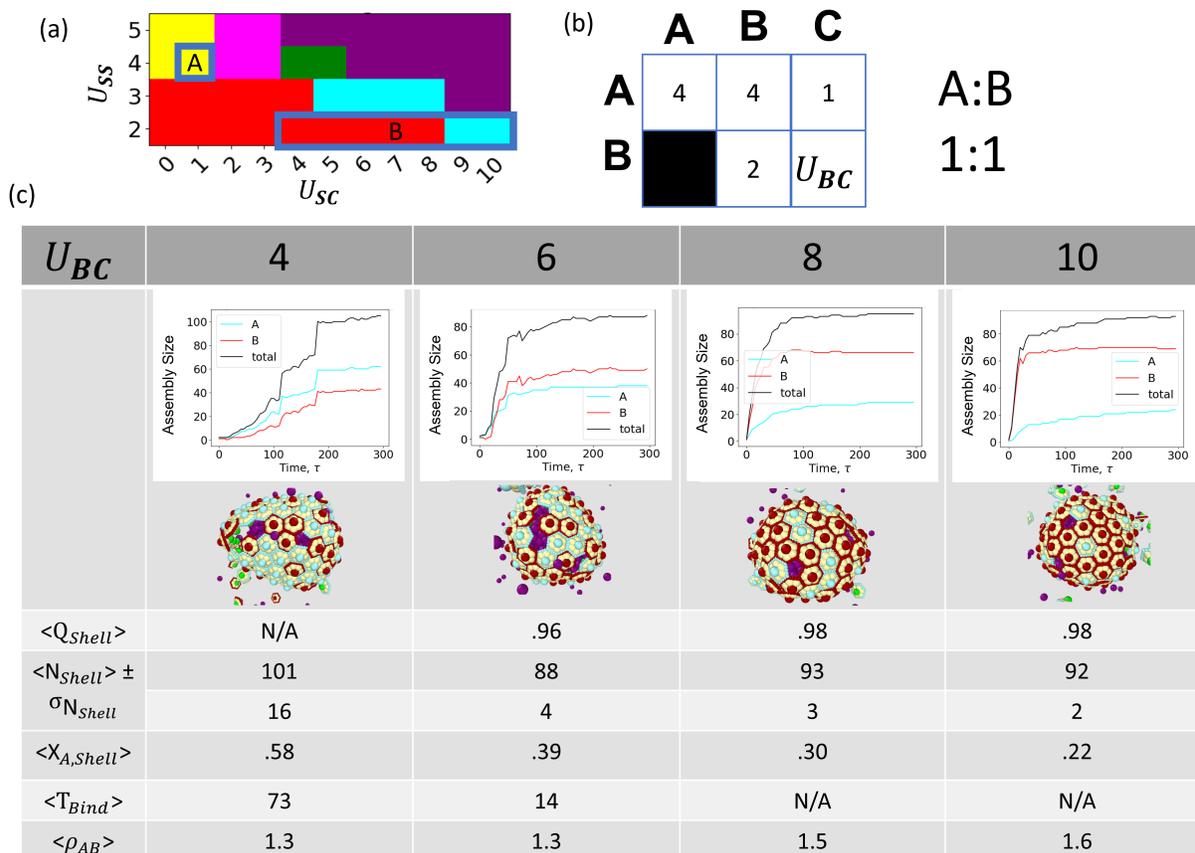


Figure 7.7. Kinetics and properties of two component shells with self-interactions shown in (a). The mixing interaction, U_{AB} , is found in the interaction matrix in (b) where U_{BC} is a free parameter. The two components are combined at a 1:1 ratio. (c) As the interaction between B and the cargo, U_{BC} , is increased, the average shell quality, $\langle Q_{Shell} \rangle$ increases. The average shell size and variance, $\langle N_{Shell} \rangle \pm \sigma_{N_{Shell}}$, average fraction of component A in the shell, $\langle X_{A,Shell} \rangle$, and time to bind the cargo, $\langle T_{Bind} \rangle$, all decrease. The components become more mixed as U_{BC} is increased as shown by $\langle \rho_{AB} \rangle$ defined in Eq. 7.4.

not involved in the process of binding the cargo. We use a shell-shell interaction of 3 for all pairs (U_{AZ} , U_{BZ} , U_{ZZ}) so that Z does not nucleate in the bulk with A and B, as well as a weaker interaction with the cargo, $U_{ZC}=4$, such that Z does not nucleate on the globule (all parameters shown in Figure 7.8(d)). The stoichiometric ratio is 2:1:1, A:B:Z unless

otherwise mentioned and the total number of shell proteins present in the simulation is the same as it was for the one and two shell protein systems. In Figure 7.8(a), we show that this leads to higher quality shells in the bulk-nucleated regime ($U_{BC} \leq 6$) in contrast to the two shell protein component system (see Figure 7.6(a)), where low-quality bulk-nucleated shells were observed. We also see similar trends in shell size, the deviations in that size, shell stoichiometry, and binding time as we did in the two component case (see Figure 7.9). An example of bulk-nucleated assembly is given in Figure 7.8(b), which shows the growth of the shell as a function of time when $U_{BC}=4$. Before the first binding event ($T < T_{Bind}$), nucleation and growth occur only in the bulk and mostly involve only shell proteins A and B as visually illustrated in Figure 7.8(c)(i). Then, at $T=T_{Bind}$, where recruitment of the shell to cargo globule occurs (Figure 7.8(c)(ii)), there is a sharp increase in the amount of B and Z that add to the shell while the rate of A incorporation decreases. This is because B and Z have stronger cargo interactions than A and can easily add to the shell growing on the cargo. In the bulk-nucleated regime, Z is acting as a “completer” of the shell, improving its overall quality, $\langle Q_{Shell} \rangle$ compared to when Z is not present (see the two component case, Figure 7.6(a)). We believe component Z is able to do this for two reasons that relate to its weaker shell-shell interactions. The first reason is that Z does not nucleate much in the bulk, leaving Z monomers available to fill in gaps left by the stochastic sheet binding process as shown for one representative case in Figure 7.8(b). The other reason is that Z can more easily rearrange on the surface of the shell due to its weaker interactions.

To demonstrate the importance of Z adding only after a nucleus has formed on the cargo, we increase U_{ZC} from 4 to 6 in Figure 7.10. Under these conditions Z can nucleate

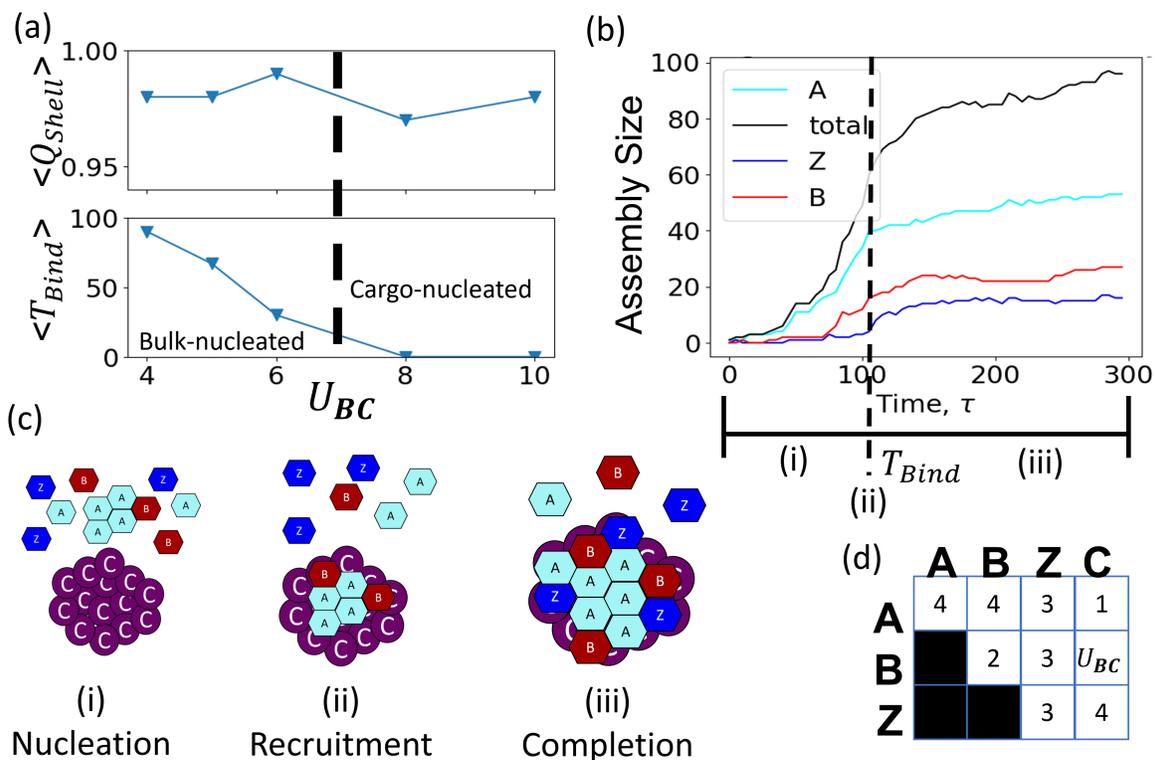


Figure 7.8. Assembly in a three component shell. (a) As U_{BC} is varied, the same bulk-nucleated and cargo-nucleated assembly pathways from the two component shell are present as in the two component system. At higher U_{BC} ($U_{BC} \geq 8$), there is the cargo-nucleated assembly where high quality (high Q_{Shell}) shells form by nucleating on the cargo. At lower U_{BC} ($U_{BC} \leq 6$), shells nucleate in the bulk as shown by the non-zero time it takes to bind the cargo, $\langle T_{Bind} \rangle$. Thus, the bulk-nucleated regime no longer leads to many holes in the shell when the third component is added. (b) The composition of the growing shell shown for a typical simulation where $U_{BC}=4$. The kinetics can be broken into three steps: (i) nucleation and growth of sheets in the bulk, (ii) a sheet is recruited and binds the cargo, (iii) component Z can add to the stable nucleus on the cargo and help to eliminate any gaps. (c) A visual illustration of the bulk-nucleated pathway with three steps from (b), (i) Nucleation, (ii) Recruitment, and (iii) Completion. (d) The matrix of all interactions used in (a) where U_{BC} is a free parameter.

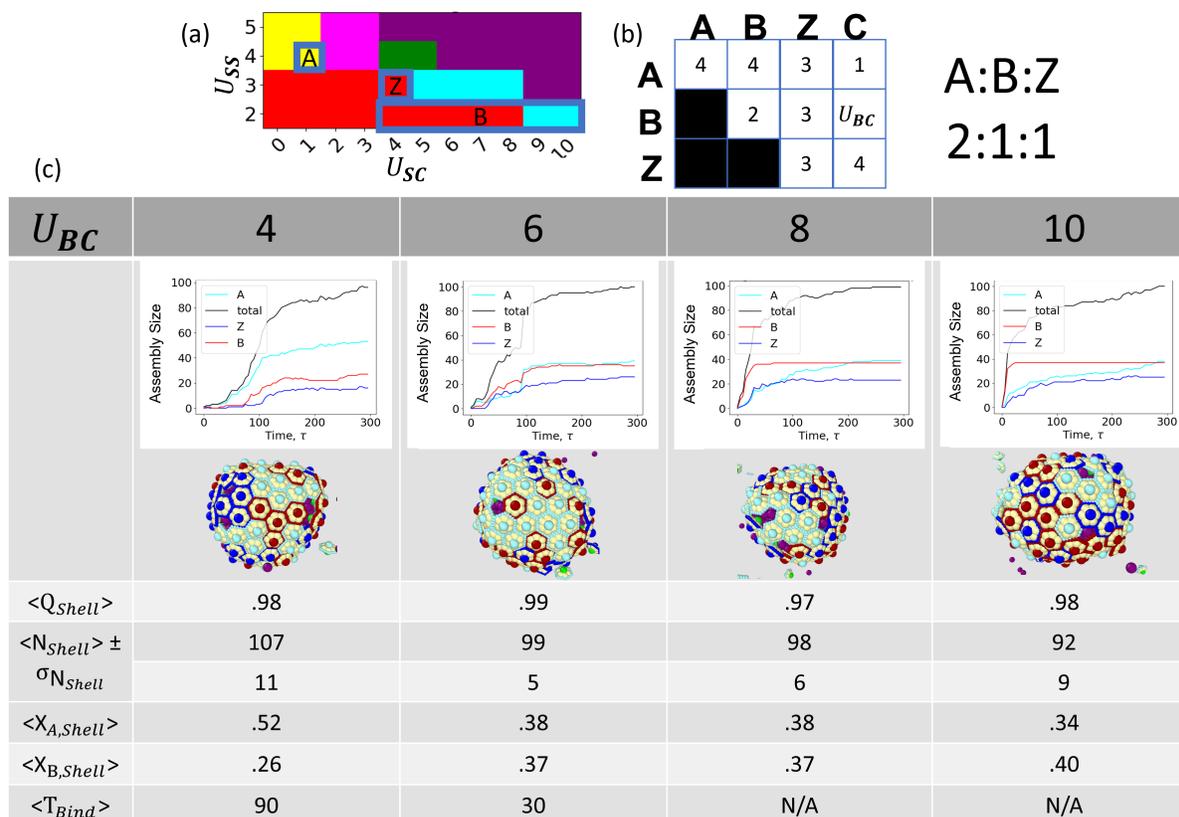


Figure 7.9. Assembly in a three component shell. (a) As U_{BC} is varied, the same bulk-nucleated and cargo-nucleated assembly pathways from the two component shell are present as in the two component system. At higher U_{BC} ($U_{BC} \geq 8$), there is the cargo-nucleated assembly where high quality (high Q_{Shell}) shells form by nucleating on the cargo. At lower U_{BC} ($U_{BC} \leq 6$), shells nucleate in the bulk as shown by the non-zero time it takes to bind the cargo, $\langle T_{Bind} \rangle$. Thus, the bulk-nucleated regime no longer leads to many holes in the shell when the third component is added. (b) The composition of the growing shell shown for a typical simulation where $U_{BC}=4$. The kinetics can be broken into three steps: (i) nucleation and growth of sheets in the bulk, (ii) a sheet is recruited and binds the cargo, (iii) component Z can add to the stable nucleus on the cargo and help to eliminate any gaps. (c) A visual illustration of the bulk-nucleated pathway with three steps from (b), (i) Nucleation, (ii) Recruitment, and (iii) Completion. (d) The matrix of all interactions used in (a) where U_{BC} is a free parameter.

on the globule immediately as shown in Figure 7.10(b) and (c). This leads to demixing of Z from either A and B or just A depending on U_{BC} . This is evident quantitatively through ρ_{BZ} and ρ_{AZ} , or visually in simulation snapshots. It is also evident in the stoichiometric growth curves, implying that this is a kinetic phase separation as opposed to a thermodynamic one. This must be true since we only changed U_{ZC} , which should not affect mixing according to the regular solution analysis. In Figure 7.10(a), we show that this kinetic phase separation is strongly correlated with shell quality, with the best shells having the most mixing between A and Z. Z cannot help to eliminate defects when it is phase separated from the rest of the shell. It should be noted that all ρ_{AZ} values are less than 1 consistent with the regular solution model analysis for $\chi_{AZ} = .5 > 0$. Allowing Z to nucleate on the globule did produce the only violation of this rule as for $U_{BC}=4$, $U_{ZC}=6$ $\chi_{BZ} = -.5 < 0$, but $\rho_{ij} = .8 < 1$. This implies that systems which rely on slow recruitment kinetics would not contain components that are able to nucleate on the cargo by themselves as this harms overall shell quality.

We now compare our results for multi-component shells to the single component shell results by taking an average of the interactions in our multi-component systems. To this end we define the shell-cargo interaction parameter averaged over all of the components in the shell, $\langle U_{SC} \rangle_{Shell}$

$$(7.7) \quad \langle U_{SC} \rangle_{Shell} = \frac{\sum_S^{Components} N_{S,Shell} * U_{SC}}{N_{Shell}}$$

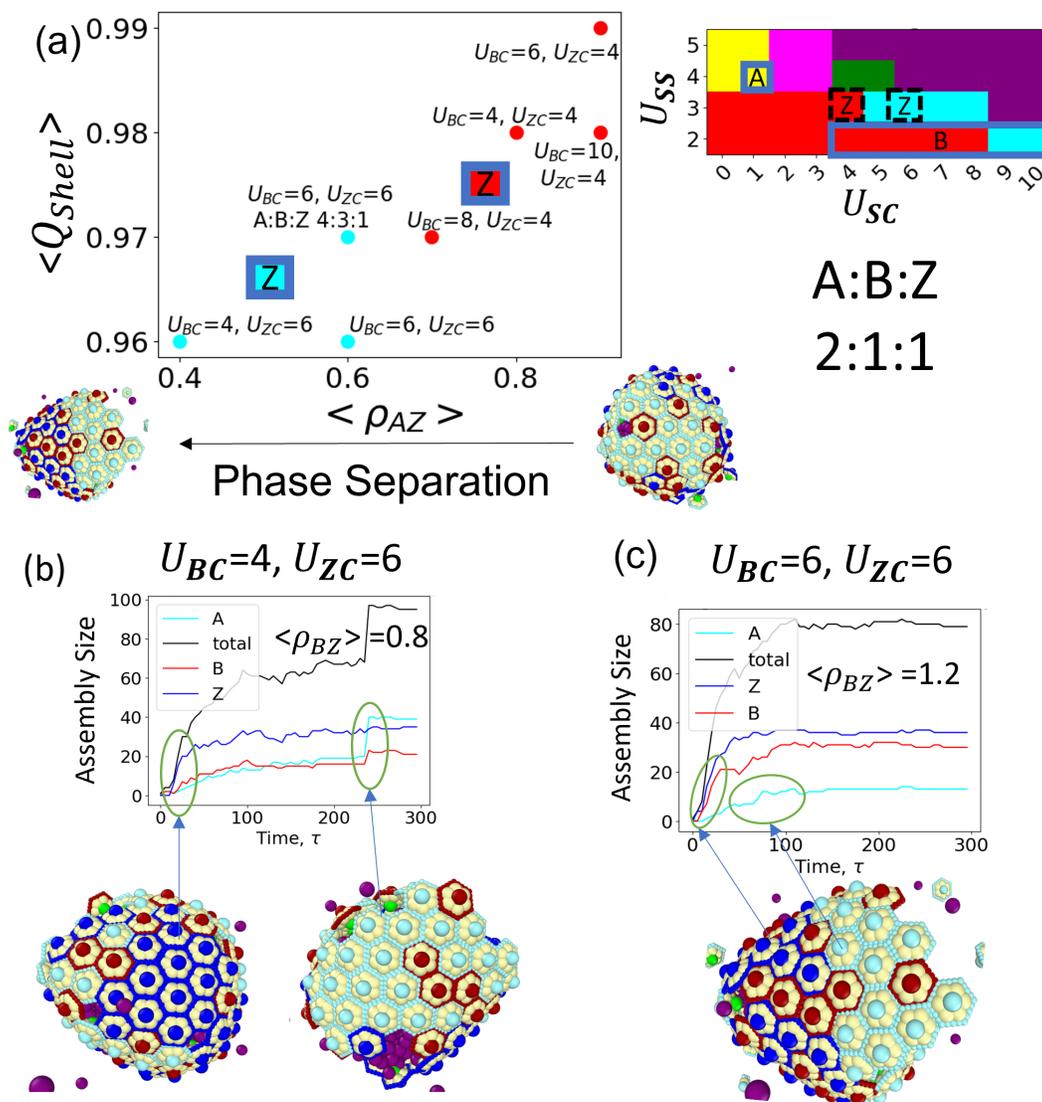


Figure 7.10. Kinetic phase separation in a three component system. (a) Phase separation of A and Z leads to reduced shell quality. This phase separation is caused by increasing U_{ZC} such that Z can nucleate on the globule and thus is a kinetic phase separation. This is shown in (b), where Z also separates from B and in (c) where B and Z are still mixed due to the increase in U_{BC} such that B quickly integrates into the growing Z nucleus.

where $N_{S,Shell}$ is the number of the shell proteins of type S, where S can be A, B, or Z, in the completed shell, U_{SC} is the shell-cargo interaction parameter for shell protein A, B, or Z and N_{Shell} is the total number of shell proteins in the completed shell.

We also define the average shell-shell interaction parameter averaged over all possible interacting pairs ($U_{AA}, U_{AB}, U_{AZ}, U_{BB}, U_{BZ}, U_{ZZ}$) in the shell, $\langle U_{SS} \rangle_{Shell}$

$$(7.8) \quad \langle U_{SS} \rangle_{Shell} = \frac{\sum_{SS}^{Pairs} N_{SS}^{Connections} * U_{SS}}{N_{Connections}}$$

where $N_{SS}^{Connections}$ is the number of the of connections between shell protein pairs AA, AB, AZ, BB, BZ, or ZZ. U_{SS} is the shell-shell interaction parameter for the pair and $N_{Connections}$ is the total number of connections between shell protein interfaces defined the same way as for Q_{Shell} in eq.7.1.

For single component shells $U_{SS} = \langle U_{SS} \rangle_{Shell}$ and $U_{SC} = \langle U_{SC} \rangle_{Shell}$. Thus, we can plot results for one, two, and three shell protein systems together in Figure 7.11(a). These results are plotted on top of the phase diagram for single component assembly from Figure 7.3(a) to compare single component shells to multi-component shells. We find that assembly of high quality shells occurs on or near a single line in the given parameter space as shown by the dotted arrow. In Figure 7.11(b), (c), and (d) we plot the assembly pathway, $\langle T_{Bind} \rangle$, and $\langle N_{Shell} \rangle$ all of which vary continuously along the dotted arrow as $\langle U_{SC} \rangle_{Shell}$ and $\langle U_{SS} \rangle_{Shell}$ are varied. This allows us to observe some striking similarities and differences between the single-component and multi-component shells. In the interaction regime where single component shells assemble, the multi-component shells assembled *via* the same cargo-nucleated assembly pathway (Figure 7.11(b)) where

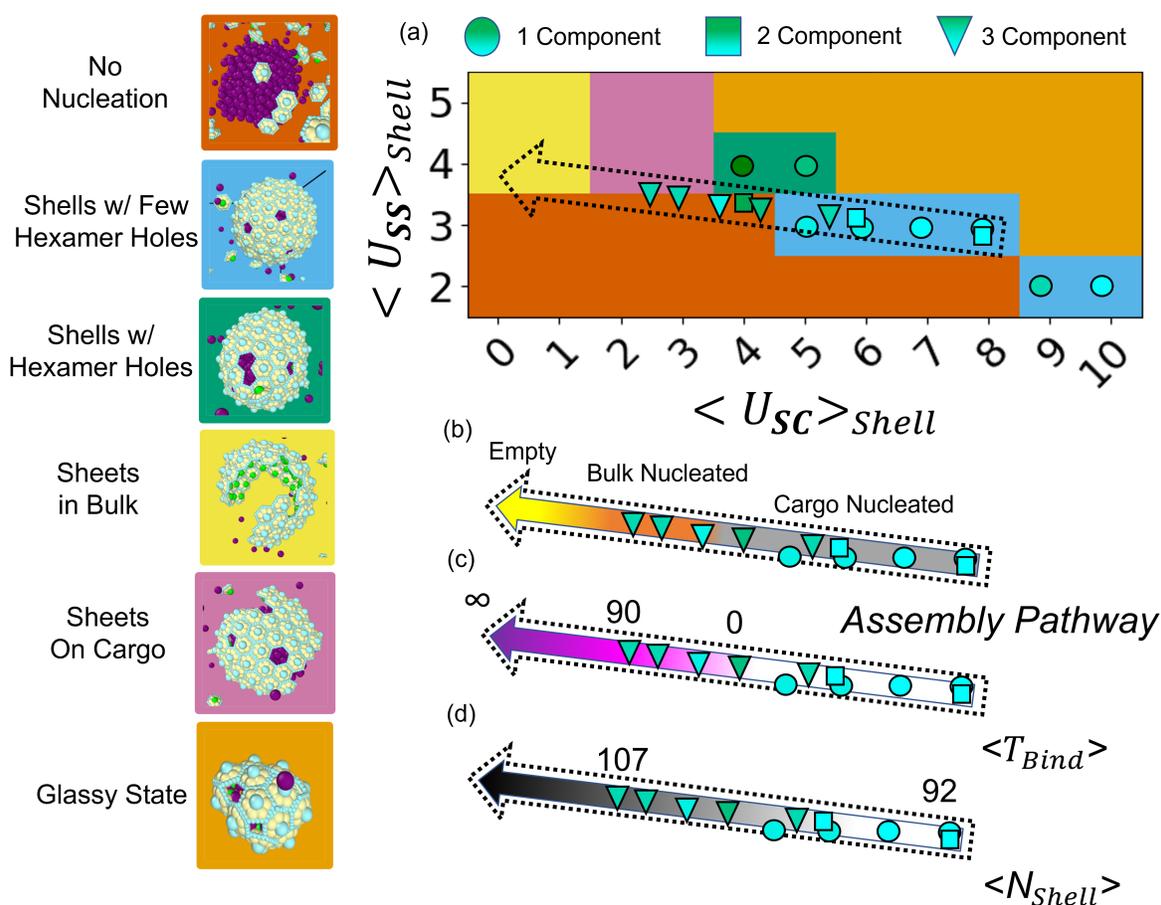


Figure 7.11. A comparison of one, two, and three component shells. (a) The x and y axes are the shell-cargo and shell-shell interactions averaged over the shell as defined in Equations 7.7 and 7.8 respectively. This allows us to plot one, two, and three component shells in terms of only two parameters. We include the phase diagram from Figure 7.3(a) in the background for comparison of shell assembly with one component to assembly with multiple shell proteins. All successful shell assembly falls on one or near one line in the parameter space as shown by the dotted arrow. This arrow overlaps with the regime where high quality shells assemble (Q_{Shell} approaches 1) in the single shell protein case, while also extending to lower values of $\langle U_{SC} \rangle_{shell}$ when three hexamer types are present. In other words, multi-component, cargo-encapsulating shells can assemble with a wider range of possible average interactions. In (b), (c), and (d) we show many properties of these assemblies can be described as a continuum by following the dotted arrow. (b) Given a well-formed shell, the assembly pathway is determined by the average interactions of that shell regardless of the number of components. (c) This is shown to be a continuum as described by T_{Bind} , which is 0 for cargo-nucleation, finite for bulk-nucleation and infinite for an empty assembly (Sheets in Bulk, the yellow region). (d) The size of the shell, $\langle N_{Shell} \rangle$, is determined by the relative strength of shell-cargo and shell-shell interactions regardless of the number of shell components.

T_{Bind} is 0 (Figure 7.11(c)) and grew to be the same size of roughly 91 shell proteins (Figure 7.11(d)). Thus, in the cargo-nucleated regime, it is accurate to describe the assembly of a multi-component shell as an assembly of a single shell component with the average interactions of all components in the shell. Outside of this regime, but still on the line, only the assembly of multi-component shells is observed. They have, on average, weaker shell-cargo interactions than is possible with only a single shell protein. Specifically, these are the bulk-nucleated shells (Figure 7.11(b)), which are larger (Figure 7.11(d)) and feature a non-zero T_{Bind} (Figure 7.11(c)). This leads to empty space inside the shell as observed in Figure 7.6(iii).

$\langle T_{Bind} \rangle$ quantifies the transition between the cargo-nucleated and bulk-nucleated pathways. In Figure 7.11(b) and (c) it is always zero for the cargo-nucleated regime and finite for the bulk-nucleated regime. By considering assemblies of sheets in the bulk to have a $\langle T_{Bind} \rangle$ of ∞ we can also describe empty assembly as part of this continuum quantified by $\langle T_{Bind} \rangle$ in Figure 7.11(c). This applies for N_{Shell} in Figure 7.11(d). Since the coarse-grained shell protein interactions have no intrinsic curvature (we show the interfaces between Pdu shell proteins are intrinsically curved in Figure 7.12(f)), the “sheets in bulk” represent the limit where N_{Shell} approaches ∞ . While we did not see assembly of high-quality shells all the way up to the regime of empty shell assembly (the yellow region in Figure 7.11(a)) it may be possible with further parameter-tuning and more shell protein components. Thus, for multi-component shells the three assembly mechanisms (cargo-nucleated, bulk-nucleated, and empty) may be continuous in the average interaction space shown in Figure 7.11(a). In contrast, cargo-nucleated and empty assembly are distinct regimes for assembly with a single shell protein as shown in Figure 7.3(a). Notably, this

does not mean that any combination of species with the correct average interactions leads to a high quality shell. We observed many low quality shells in Figures 7.6 and 7.10 when the kinetics were improper.

7.4.4. Atomistic Simulations of Pdu Shell Proteins: PduA with PduB and PduB'

To interrogate whether our coarse-grained model applies to the Pdu MCP system, which contains many different types of shell proteins, we run atomistic simulations of the PduA hexamer, PduB trimer, and PduB' trimer interfaces (see Figure 7.12(a-d)). We perform potential of mean force, PMF, calculations for to measure the interaction strengths for pairs of these shell proteins as has been done for other nanoscale buiding blocks [353, 354] including Pdu shell proteins [5, 1, 355]. The previous simulations measured the strength of interactions (including the effects of interface bending) between PduA hexamers and PduN pentamer, demonstrating strong interactions of 10-12 kcal/mol for PduA-PduN [1] and PduA-PduA [5, 355] interfaces. This and other experimental work [73] has shown that PduA leads microcompartment assembly like component A in our coarse-grained model and can nucleate shell protein assembly in the bulk. Thus, in order to show that PduB can play the role of the recruiter, B in the coarse-grained model, the PduA-PduB interaction must be strong enough for the two proteins to grow together in the bulk. In contrast, for PduB' to perform the role of the completer, Z in the coarse-grained model, it must have an interaction that is too weak for PduA and PduB' to grow together in the bulk as demonstrated in Figure 7.8. Results shown in Figure 7.12(e) confirm that this is the case for both pairs of shell proteins. Figure 7.12(b) shows that the PduA-PduB

interaction is actually two different interactions due to the 3-fold rotational symmetry of the 6-sided pseudohexameric trimer. The Pdu AB2 interaction is much stronger than Pdu AB1 (8.5 kcal/mol as opposed to 2 kcal/mol). This is consistent with experiments that have shown PduA has a strong preference for binding the Pdu AB2 interface over the Pdu AB1 interface [338]. The interactions thought to be necessary for assembly of nanoscale shells are around 5 kcal/mol or $8 k_B T$ [342, 207], greater than the Pdu AB1 interface, but less than the Pdu AB2 interface and on average quite similar to the expected value. Figure 7.12(e) also shows that the N-terminus is quite flexible and dynamic, which would increase the assembly rate of PduB in the bulk [207] and thus the rate at which it could recruit bulk-nucleated sheets to the cargo. This flexibility is also evident in the bending potential of the Pdu AB2 interface as shown in Figure 7.12(f). The PduA-PduB' also has two unique interfaces, one of which is also the weak Pdu AB1 interface shown in Figure 7.12(d). The second interface (Pdu AB') is different only due to the deletion of the N-terminus and this makes the interaction far weaker than Pdu AB2 as shown in Figure 7.12(e). The weaker interaction and loss of flexibility make it a poor recruiter and thus a better completer. We also measure the bending potential of mean force in Figure 7.12(f) and show that all three interfaces have similar lowest energy bending angles of slightly less than 20° . This angle is $15 - 20^\circ$ lower than what was reported for PduA-PduA and PduA-PduN interfaces [1], but consistent with experiments showing that PduB microtubules have a much smaller bending angle than PduA microtubules [261].

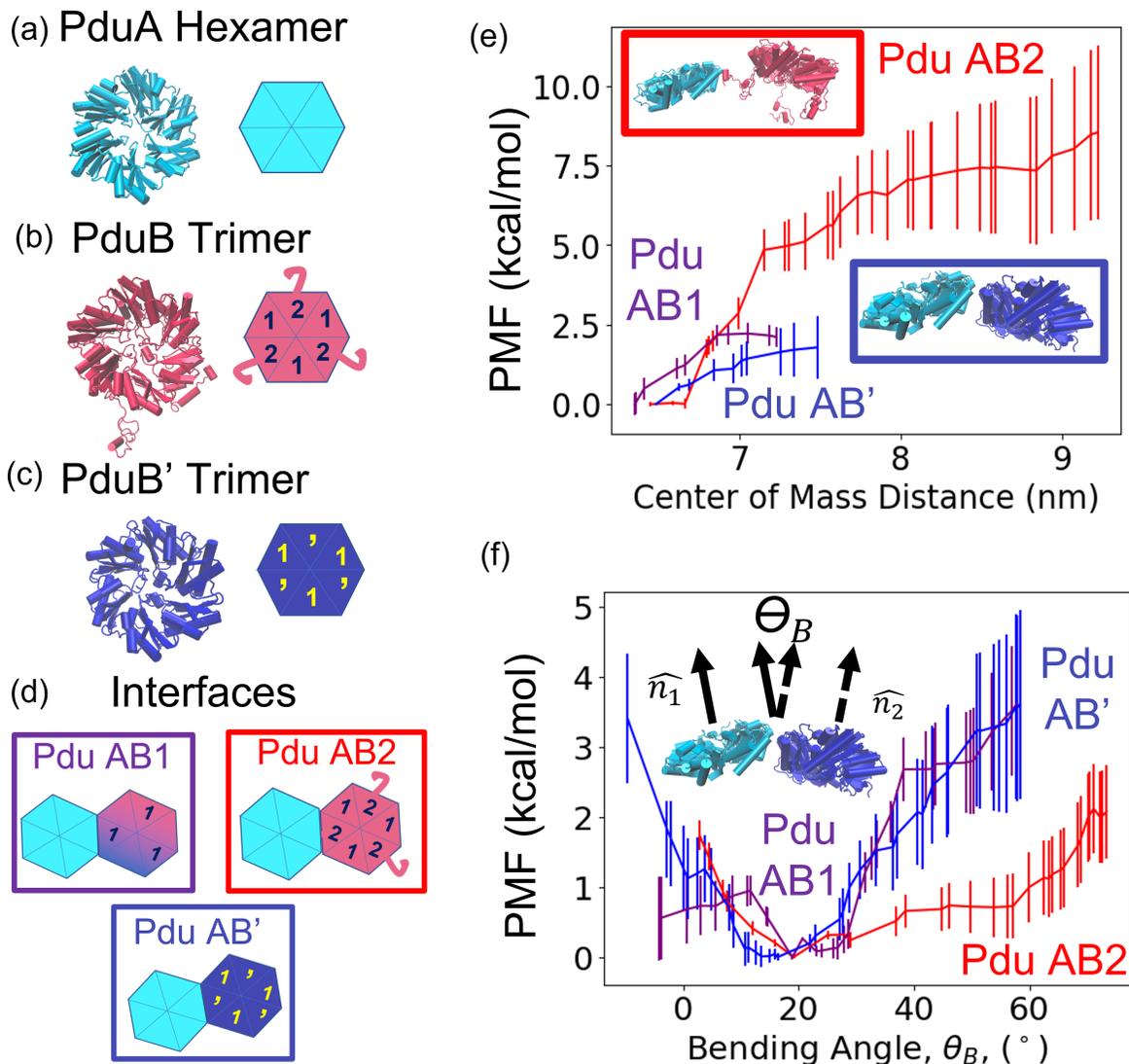


Figure 7.12. Atomistic simulations show PduB and PduB' interact differently with PduA and likely play different roles in Pdu microcompartment assembly. Secondary structure representations of the PduA hexamer (a), PduB trimer (b), and PduB' trimer (c) show PduB and PduB' both have the Pdu AB1 interface while the N-terminus region of PduB is lacking from PduB' creating two unique interfaces known as Pdu AB2 and Pdu AB' along with the shared Pdu AB1(d). (e) Potential of mean force calculations show the different interaction strengths of the three interfaces from (d). We show the strongest interaction with PduA comes from Pdu AB2, due to the presence of the highly dynamic N-terminus region of PduB (red inset). (f) Potential of mean force calculations as a function of the bending angle, θ_B , show energetically preferred bending orientations of the interfaces. The bending interactions at these interfaces are quite similar with some added flexibility for Pdu AB2 due to the presence of the N-terminus region.

7.4.5. Experimental

Transmission electron microscopy (TEM) of purified Pdu MCPs prepared for microscopy using various methods supports the hypothesis that Pdu MCPs contain a significant volume of water within their lumen as was shown for the bulk-nucleated regime. This is supported by the observation that under preparatory methods in which Pdu MCPs are allowed to collapse due to sample drying, Pdu MCPs appear smaller on average than Pdu MCPs prepared under conditions that limit collapse by keeping samples in aqueous solution. This phenomenon has been examined previously [240], but representative images are presented in Figure 7.13. To test this hypothesis further, we compared the amount of collapse between wild-type MCPs predicted to be full of enzymes with MCPs predicted to be empty. Empty MCPs were generated by knocking out a protein necessary for enzyme loading (PduB) [72]. We purified MCPs from the wild-type *Salmonella enterica subsp. Typhimurium* strain (WT MCPs) and the PduB knockout *S. Typhimurium* strain (Δ B MCPs). These purified MCPs were then imaged using cryo TEM and their diameters were measured. We compared this to the previously reported diameters in another study that used sample preparation for TEM in which compartments were allowed to collapse [72]. We found that WT MCPs collapsed by approximately 4% between techniques (126 nm vs 121 nm), whereas empty Δ B MCPs collapsed by approximately 14% (97 nm vs 83 nm), supporting the hypothesis that the collapse is related to dehydration and not the enzyme cargo. This also supports prior findings indicating that WT MCPs contain a significant volume of water leading to collapse upon drying [240]. The smaller size of the non-collapsed Δ B MCPs vs. non-collapsed WT MCPs is consistent with our atomistic

simulations in Figure 7.12 and previous work [1] showing that the PduB creates lower angle interfaces. This means that ΔB MCPs should have larger angles on average and thus a smaller diameter as we report here and in previous experiments [72]. Since the size of the shell seems to be controlled by the intrinsic curvature of these interfaces, engineering these interactions could be a viable route to controlling the size of MCPs.

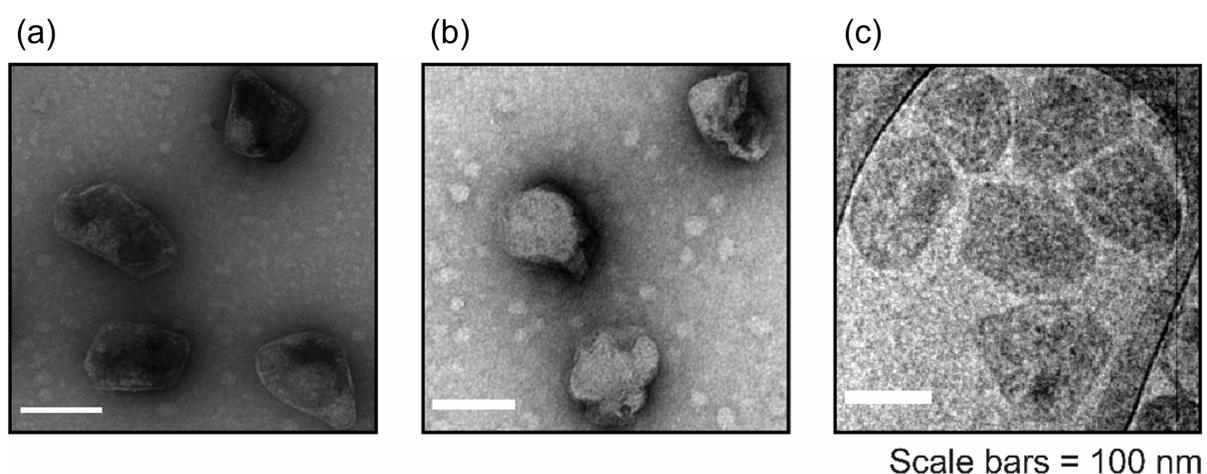


Figure 7.13. Representative micrographs of Pdu MCPs prepared using various techniques. (a) Pdu MCPs prepared using standard negative staining techniques allows for Pdu MCP collapse as indicated by pooled staining at their center. (b) Pdu MCPs buffer exchanged to prevent collapse appear more inflated by TEM as indicated by reduced internal pooled staining. (c) Micrographs of Pdu MCPs in solution prepared using cryogenic methods appear the most inflated. Scale bars = 100 nm in each case.

7.5. Conclusions and Outlook

Here, multi-scale molecular dynamics and genetic engineering elucidate the assembly of microcompartment shells with multiple types of shell proteins. In contrast to cargo-nucleated assembly, where shells and their kinetics can be described by single shell protein assembly with average interactions, distinct shell proteins play distinct roles in assembly

in the bulk-nucleated regime. Specifically, a nucleator with strong shell-shell interactions can grow sheets in the bulk, which are then bound to the cargo by a recruiter with a stronger shell-cargo interaction. The shells formed in this process have fewer defects when a third shell protein, termed the completer, adds to the growing shell only after the binding process is complete. This model suggests that at least three different types of shell proteins (plus a pentamer for shell closure) are likely necessary for these bulk-nucleated shells, and this may be highly relevant as engineers attempt to successfully assemble MCPs in cell-free environments. Having multiple components also expands the range of interactions where MCP assembly was successful, as has been observed in other multi-component biological assemblies.

In the Pdu MCP system, we find evidence for all three of these types of shell proteins using atomistic simulations that measured the binding strength of different pairs of Pdu shell proteins. These simulations suggest that there is a reason for the Pdu system to contain PduB and PduB', the latter of which lacks a flexible N-terminus. The lack of an N-terminus makes the PduB' interaction weaker, which our coarse-grained model shows can actually be helpful in reducing gaps in the shell. The N-terminus has also been linked to the interaction between PduB and the cargo and our results here do not contradict that hypothesis. There are still many more shell proteins in the Pdu system that were not mentioned here and further work is needed to understand how they impact the shell. It is possible that many of the shell proteins combine to play the different assembly roles we report. We also provide evidence that intrinsic curvature of the shell protein interfaces plays a large role in determining the size of these shells by combining measurements of bending energies in atomistic simulations with experiments removing PduB from the Pdu

MCP. This could be a promising route to engineering the size of these compartments in and out of cells and thus tuning enzymatic pathways. The goal of future investigations, both computational and experimental, should be to find modular design strategies for adjusting these binding angles and strengths by modifying the primary sequence of shell proteins in a way that can be applied to many different MCP systems.

7.6. Methods

7.6.1. Coarse Grain Simulations

7.6.1.1. Units. The distance unit, d , is ≈ 3 nm. The energy unit, $\epsilon = 1k_B T = 4.1 * 10^{-21}$ J at room temperature. The mass unit, m , is $\approx 320 \frac{g}{mol}$ or $\frac{.32}{6.022 * 10^{23}}$ kg. The derived time unit, $t^* = \sqrt{\frac{md^2}{\epsilon}}$, is $\approx 3.4 * 10^{-10}$ s. The timestep of the simulation, dt , is $.004 t^* = 1.4 * 10^{-12}$. τ , which is the unit of time given in the plots in the main text, is $.5 * 10^6 dt$ or ≈ 680 ns. Simulations run for at least 300τ , which is slightly over $200 \mu s$.

7.6.1.2. Hexamer Geometry.

7.6.1.3. Forcefield. The following interactions utilize a Weeks-Chandler-Andersen [129], WCA, potential. Defined as

$$(7.9) \quad U_{WCA}(r) = \left(\frac{\sigma}{r}\right)^{12}$$

when the interaction distance, r , is less than or equal to σ . And

$$(7.10) \quad U_{WCA,r>\sigma}(r) = 0$$

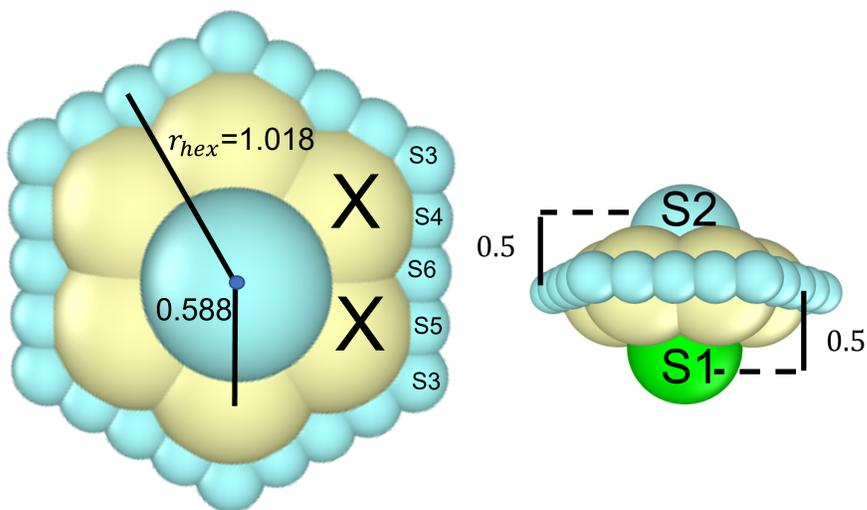


Figure 7.14. The geometry of the rigid body hexamers in distance units. The parameters given are consistent across all types of hexamers, A, B, and Z in the main text. The rigid body consists of volume excluding X beads in addition to other beads which handle attractive interactions. These beads are named as S1 - S6 where S is A, B, or Z when the hexamer is a specific type of hexamer. Here they are named S to show the generality. The effective radius of the hexamer, r_{hex} is the distance from the center of geometry to the center of the S6 bead.

when r is greater than σ . Below we show these parameters for pairs of beads types in Figure 7.14 and also the cargo beads, C.

S2-S2, $\sigma = 2.236$ - This is twice the radius of the hexamer, ($r_{hex} = 1.018$ as measured from the center of the hexamer to the center of the interface, which is also the center of the S6 bead) plus the equilibrium separation distance, $r_0 = .2$, of the morse potential used for all attractions between interfaces. Thus, this repulsion kicks in if two hexamers are ever connected at a concave angle.

S1-S2, $\sigma = 1.8$

X-C, $\sigma = 0.75$

X-S2, $\sigma = 0.75$

X-S3, $\sigma = 0.6$

C-S2, $\sigma = 0.5$

C-S3, $\sigma = 0.5$

C-S4, $\sigma = 0.5$

C-S5, $\sigma = 0.5$

C-S6, $\sigma = 0.5$

X-S3, $\sigma = 0.6$

S6-S4, $\sigma = 0.2$

S6-S5, $\sigma = 0.2$

S6-S6, $\sigma = 0.2$

S6-S3, $\sigma = 0.3$

For any unlisted possible pair there is no WCA interaction.

Morse potentials are used to model the attractive interactions along the interface of the hexamers as well as between the hexamers and the cargo

$$(7.11) \quad U_{Morse}(r) = D_0[\exp(-2\alpha(r - r_0)) - 2\exp(-\alpha(r - r_0))]$$

when r is less than or equal to r_{cut} and otherwise 0.

$$(7.12) \quad U_{Morse} = 0$$

D_0 is the well depth given in energy units, α controls the width of the interaction and is given in units of $\frac{1}{\text{distance}}$, and r_0 and r_{cut} have units of distance with r_0 being the lowest energy separation distance.

$$\text{S4-S5, } D_0 = U_{SS}, r_0 = .2, \alpha = 20, r_{cut} = 2$$

$$\text{S3-S3, } D_0 = .5U_{SS}, r_0 = .2, \alpha = 20, r_{cut} = 2$$

$$\text{S1-C, } D_0 = U_{SC}, r_0 = .5, \alpha = 5, r_{cut} = 3$$

U_{SS} and U_{SC} are the interaction parameters from the main text, which are named for the specific hexamers involved. For instance, U_{SS} between hexamers A and B is U_{AB} .

The Lennard-Jones potential is used only for the C-C interaction, and U_{CC} set to a value that ensures phase separation of the cargo, $1.8 k_B T$.

$$(7.13) \quad U_{LJ} = U_{CC} * 4 * \left(\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right)$$

when r is less than or equal to r_{cut} . It is 0 otherwise.

$$\text{C-C, } r_{cut}=3, \sigma=1$$

7.6.1.4. Protocol. All simulations contain 148 hexamers and 600 cargo beads in a cubic box with a length of 40 distance units or 120 nm. Initial conditions are randomly created. We use the NVT (constant number of particles, volume, and temperature) ensemble with a Langevin thermostat. Almost all simulations are run for at least 150×10^6 dt or $\approx 20 \mu s$. We make an exception when compartment growth has not plateaued as is the case when T_{Bind} becomes larger. In this case, the simulation is run until this plateau is reached. For a given parameter set, the simulation is run from random initial conditions four separate times and relevant quantities are averaged over the four independent runs.

7.6.2. Atomistic Simulations

7.6.2.1. Homology Modeling. The initial structure for the atomistic model of the PduA/PduB interfaces were generated as described previously [1]. The structure of PduA was taken from PDB 3NGK [244]. The structures of the PduB and PduB' subunits were estimated by the Phyre2 web portal [277]. Three copies of the estimated structures were aligned with the BMC-T structure extracted from PDB 6MZX [262] using the MatchMaker tool in UCSF Chimera [278, 279]. The trimer structure generated was then minimized using the default settings in UCSF Chimera's Minimize Structure tool. To build the PduA/PduB and PduA/PduB' interfaces, the two different BMC-H/BMC-T interfaces were first extracted from PDB 6MZX, a solved crystal structure of a fully intact microcompartment shell from *Haliangium ochraceum* [262]. We then used Chimera's MatchMaker tool to align PduA and PduB/PduB' structures with BMC-H and BMC-T structures, respectively.

7.6.2.2. Forcefield. We performed all-atom molecular dynamics simulations using GROMACS version 2016.3 [208] and the CHARMM36m [209] forcefield. The recommended CHARMM TIP3P water model [210] was applied with the structures constrained *via* the SETTLE algorithm [211]. The periodic boundary conditions were applied in 3 dimensions. The neighbor searching was calculated up to 12 Å using the Verlet particle-based method and was updated every 20 timesteps. The Lennard-Jones 12-6 interactions were switched off from 10 to 12 Å *via* the potential-switch method in GROMACS. The short range Coulomb interactions were truncated at the cutoff distance of 12 Å, and the long

range interactions were calculated using the Smooth Particle Mesh Ewald (PME) algorithm [212, 213]. The NPT ensemble (constant number of particles, pressure, and temperature) was employed. The temperature was coupled using the Nosé-Hover algorithm, characteristic time 1 ps. The isotropic Parrinello-Rahman barostat was employed with the reference pressure of 1 bar, the characteristic time was 4 ps, and the compressibility of $4.5 \times 10^{-5} \text{ bar}^{-1}$. All the covalent bonds were constrained, which supported an integration timestep of 1 fs. These are the recommended parameters for the accurate reproduction of the original CHARMM simulation on lipid membranes [214], and have been verified in further simulations of lipid membranes [217] as well as proteins [215, 198] including the PduA protein in this study [5, 1].

7.6.2.3. Protocol. We use the same basic protocol here as in our previous work where these calculations were performed on different shell proteins [1]. The PDB files of the homology modeling for the PduA-PduB interfaces are solvated in water containing 100 mM NaCl. Using the GROMACS molecular dynamics engine, the system undergoes a short constant pressure, temperature (NPT) equilibration 100 ps with the backbone restrained. For the potential of mean force (PMF) calculation, restraints are released for the PduA, while PduB still has a backbone restraint. The N-terminal is released from this constraint for the Pdu AB2 interface in the main text, but we keep it constrained in Figure 7.18 for comparison. Steered MD simulations are then run to create configurations where the proteins are at the many different distances sampled. This pulling step is done at a rate of $1 \frac{\text{Å}}{\text{ns}}$. These N configurations, or “windows”, are then run in parallel for 15 ns (3 independent 5 ns with simulations the same starting configuration) to gather the force data necessary for the PMF calculation. For interface Pdu AB1 (Figure 7.12(c)) we use

11 windows. For interface Pdu AB2 with all of PduB restrained (Figure 7.18) we use 20 windows. For interface Pdu AB2 with the PduB N-terminal unrestrained (Figure 7.12(d)) we use 34 windows. For interface Pdu AB' without the N-terminal we use 11 windows and run 5 independent 5 ns simulations for a total of 25 ns (Figure 7.12(e)). These are biased simulations where a harmonic bond is used to bias the proteins at the distance of the initial configuration. The average force exerted by the bond is measured during the simulation. The potential of mean force is then calculated using

$$(7.14) \quad V(x_N) - V(x_0) = - \sum_{i=0}^{N-1} \langle F_i(x) \rangle (\langle x_{i+1} \rangle - \langle x_i \rangle)$$

where V is the potential, $\langle x_i \rangle$ is the distance between the proteins averaged over the 15 or 25 ns and $\langle F_i(x_i) \rangle$ is the average force.

For the bending potential calculation, we use a similar method. The windows are generated in the same way except now we pull up or down in the z direction to create states at different bending angles, θ_B , which are mapped from the center of mass in the z dimension to give us $\theta_B(z)$. We can then measure the force in the z -direction and the z component of the center of mass and convert it to the force in the θ_B as

$$(7.15) \quad F_{\theta_B}(z) = F_z \cos(\theta_B(z))$$

For this calculation we run similar amounts of time and tend to use a higher number of windows to ensure overlap between states. For the Pdu AB1 interface we use 32 windows run for 15ns each. For the Pdu AB2 interface we use 35 windows run for 15 ns each. For the Pdu AB' interface use 40 windows run for 5 ns each.

Differences in number of windows and run time reflect the complexities of the energy landscapes and an effort to reduce error bars relative to the magnitude of the energies. Error bars are based on sampling error and estimated by splitting the data in different sections (i.e. first third vs. second third vs. last third) and observing the differences in the calculated potential, sometimes called block averaging.

The total calculation of the bending potential of mean force becomes

$$(7.16) \quad V(\theta_{B,n}(z_n)) - V(\theta_{B,0}(z_0)) = - \sum_{i=0}^{N-1} \langle F_{\theta_{B,i}}(z_i) \rangle (\langle z_{i+1} \rangle - \langle z_i \rangle)$$

7.7. Appendix

Relative Connection Probability, ρ

U_{BC}	4	6	8	10
	A B	A B	A B	A B
A	1.0	0.8	0.4	0.2
B	█	█	█	█
	1.3	1.3	1.5	1.6
B	0.7	0.9	1.1	1.2

Figure 7.15. The relative probability of observing connections between two different species, ρ_{ij} , is defined in Equation 7.4. Here we show that AB connections are far more probable than random mixing due to the strong U_{AB} . This agrees with our regular solution analysis.

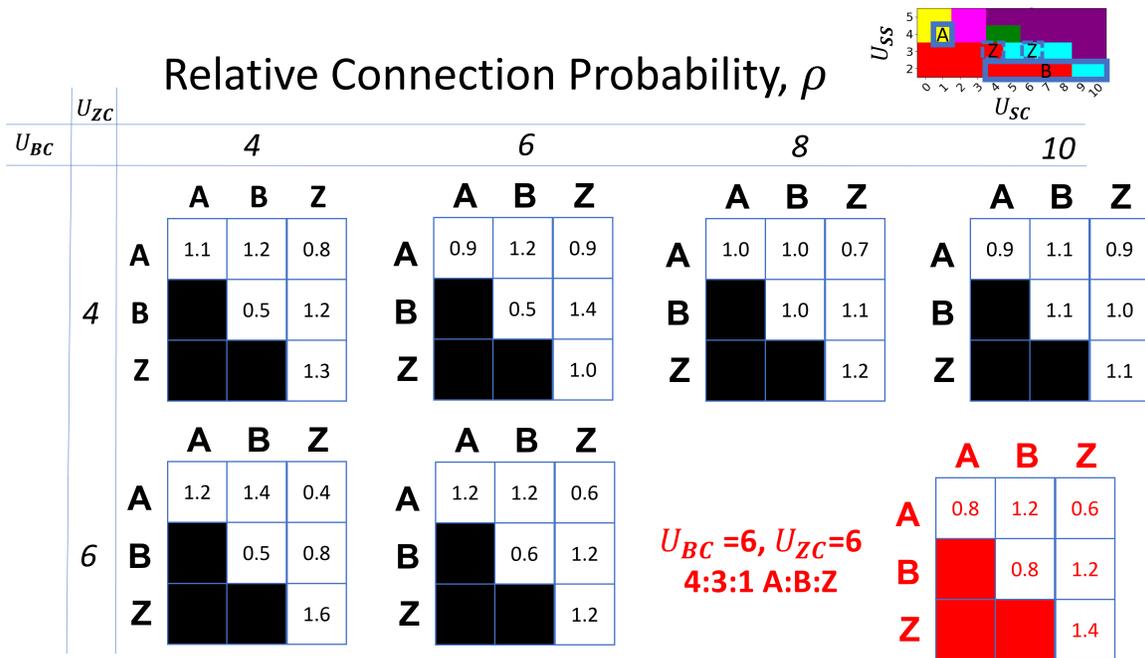


Figure 7.16. The relative probability of observing connections between two different species, ρ_{ij} , is defined in Equation 7.4. These values that correspond to mixing (ρ_{AB} , ρ_{AZ} , and ρ_{BZ}) mostly follow our regular solution analysis. The one exception is for $U_{BC}=4$ and $U_{ZC}=6$. Here B and Z separate despite our predictions made based on their interactions (see Figure 7.10). We also include one data point where we vary composition (keeping $U_{BC}=6$ and $U_{ZC}=6$) and see no change to our relative connection probability for different species, showing that the normalization is reasonable. See Figure 7.17 for further details.

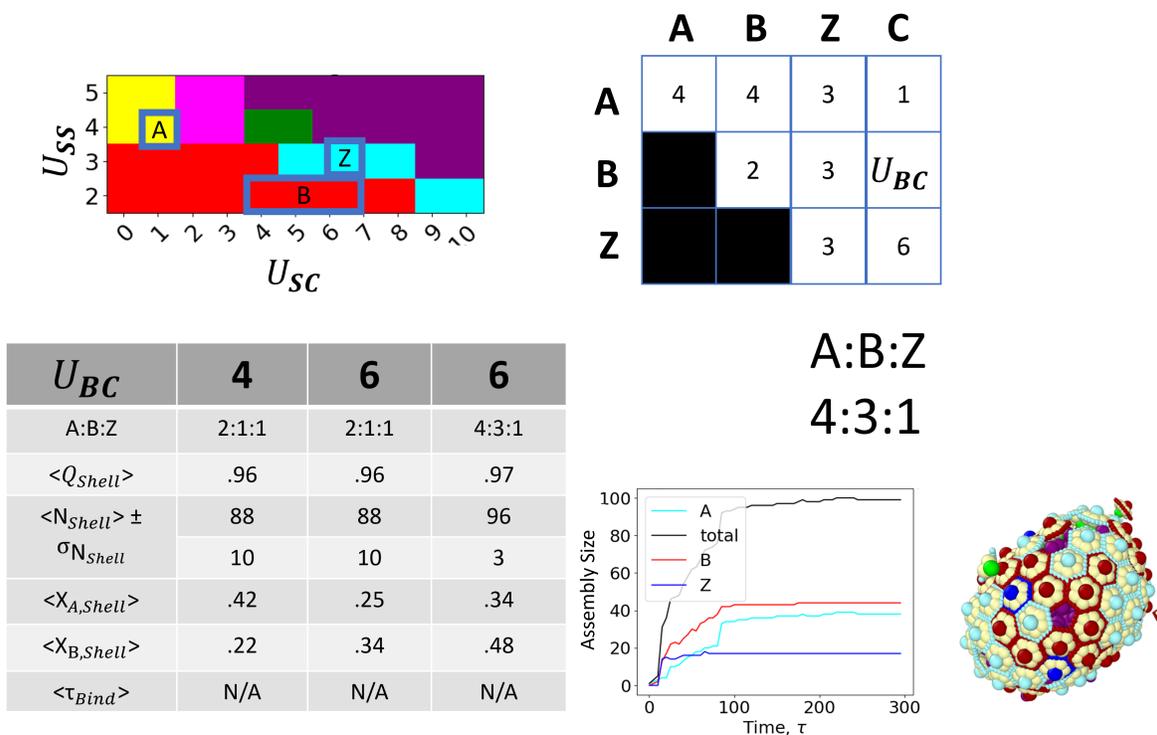
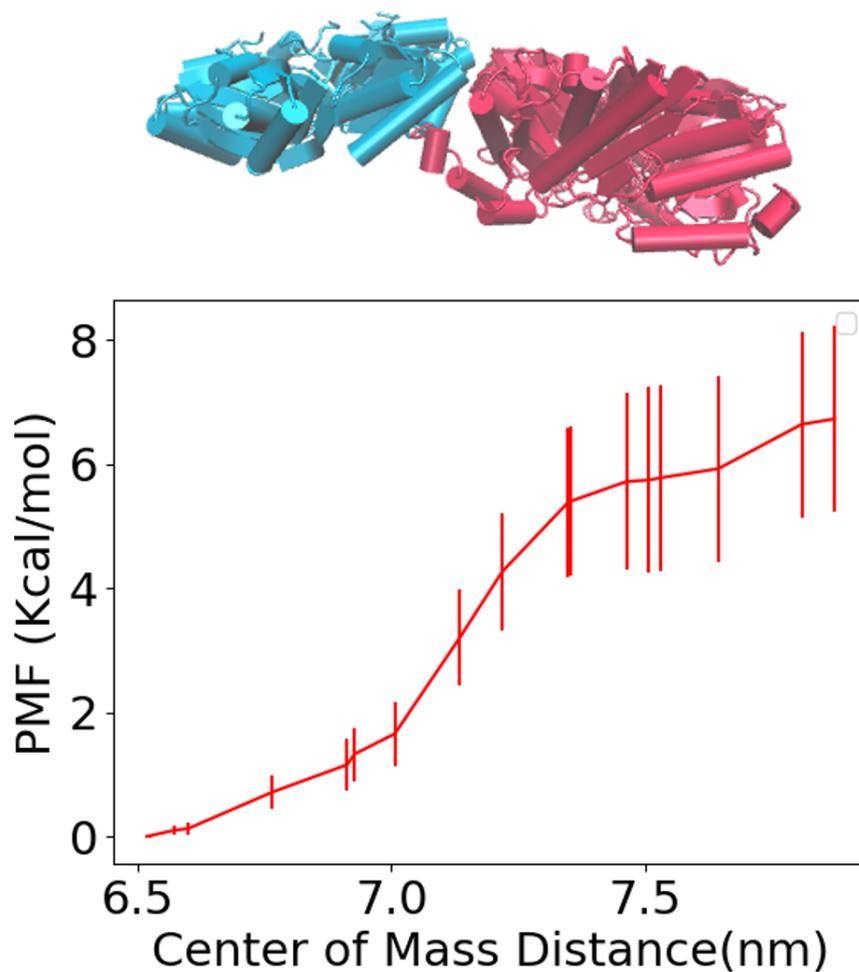


Figure 7.17. Additional data that are not reported anywhere else and so we include them for completeness.



$$\Delta G = -6.7 \pm 1.5 \frac{\text{Kcal}}{\text{mol}}$$

Figure 7.18. Potential of mean force for interface Pdu AB2. The difference between this and Figure 7.12(d) is that we restrain all of the PduB backbone including the N-terminal. This shows that the interaction is strong even with reduced degrees of freedom for the N-terminal. Restraining all of PduB as we do here is more similar to methods used for interface Pdu AB1 and in previous work [1, 5]. However, we consider Figure 7.12(d) to be more accurate since the additional degrees of freedom of the N-terminal should not be neglected.

CHAPTER 8

Conclusions

Starting with a few fundamental building blocks biology creates an incredibly diverse array of functional assemblies. These assemblies often include different types of macromolecules which each play an important functional role. These roles can be tuned by changing the chemical and structural properties of the given components. Part 1 demonstrated these concepts in synthetic polyelectrolyte complexes based on membraneless organelles where random copolymer polyelectrolytes served the function of intrinsically disordered proteins. The cationic polymer functioned as a crosslinker for the anionic polymers, which were able to segregate a wide array of small organic molecules based on the hydrophobic nanostructures they formed, demonstrating the different roles played by the two components. By changing the fractions of the monomers in our polymers in Chapter 3, we were able to adjust the nanostructure of the complexes and tailor them to specifically absorb nanoplastic contaminants from water. This selectivity has been seen in membraneless organelles in biological systems, which perform specific enzymatic functions by segregating a specific enzyme with the substrates it needs to be catalytically active. Inspired by this functionality, Chapter 4 explored the interactions between the anionic polyelectrolyte used in Chapters 2 and 3 and an enzyme known as PETase, which can depolymerize PET. We found that electrostatic interactions played a large role in the surface coverage of PETase by the polymer, which has a large positively charged surface domain. The polymer-enzyme interaction was also sensitive to the fraction of hydrophobic

monomers in the polymer. Having many of these monomers changed the polymer from, being wrapped around PETase to forming a globule, which attached to the surface, but covered less of it. This was very similar to what was observed in Chapter 3 when we analyzed the interaction of the anionic copolymer with a PET surface and showed that the least and most hydrophobic polymers adsorbed the best on PET surfaces due to the preferred conformations of the polymer. Experimentally, we showed that the presence of the polymer increased the activity of the PETase on small molecules and a PET film including at elevated temperature. This agreed with the simulations that showed increased stability of the enzyme at high temperature when the polymer was attached. Despite all of this work, we never investigated a polyelectrolyte complex that segregated the PETase and nanoplastic contaminants together. Such a system could remove nanoplastics from a volume of water and then also depolymerize them back into monomers, which also may be absorbed by the complex allowing for easy removal. This would require designing the complexes to simultaneously absorb the nanoplastics (which we did in Chapter 3) and the PETase. This would likely require more simulations to understand fully the best way to incorporate the PETase.

Like the copolymers in Part 1, the functional role of different shell proteins in microcompartments can also be tuned by changing their flexibility. We first investigated this in the case of pentameric shell proteins which are geometrically necessary to create a completely closed MCPs of hexagonal and pentagonal units. Chapters 5 and 6 showed that by changing this flexibility, the pentamer could play this role in three different ways. In the case of the Pdu MCP, we showed experimentally that the pentamer, PduN, dictates the morphology of the MCP. This system formed a tube-like assembly without the

pentamer, which we also showed impacts the encapsulated reaction pathway and thus the function of the MCP as a whole. Bending energy calculations allowed us to relate this morphology change to the rigid bending interactions between the pentamer and hexamer, PduA. We also explored an MCP system, which, in contrast, forms the MCP structure but with gaps on the pentamer sites when the pentamer is removed. Here, atomistic simulations showed that the dominant bending energy was coming from the hexamer-hexamer interactions instead. The interactions between different 6-sided shell proteins also had different intrinsic angles and we were able to show in Chapter 7 that removing PduB, a 6-sided pseudohexameric trimer, lead to a smaller shell as predicted by its preferred bending angles. Both of these systems contrast with the virus SV40, which forms various assemblies using only pentamers. This is due to intrinsically disordered regions, which act as flexible “arms” for the pentamer. Chapter 6 showed that this flexibility allows for the formation of non-icosahedral shells where the pentamer has either four or six nearest neighbors as well as modulation of the shell size based on the cargo instead of intrinsic bending angles. These intrinsically disordered regions were also shown to increase assembly rates in our coarse-grained model. Similar arm-like structures were also observed in the hexagonal trimer PduB of the Pdu MCP system. Interestingly, there is also a version of PduB, PduB', that lacks these intrinsically disordered arms, and this gives it much weaker interactions. In Chapter 7, we explored how having a range of interactions can lead to an assembly pathway for MCPs that is not feasible with less than 3 shell components. Here, the shell and the enzyme cargo nucleate separately and before a recruiter shell protein like PduB helps bind the two together. The third shell protein component

can help reduce defects in the shell, because it has weaker interactions like PduB'. Thus, changing the flexibility of PduB changes its role within this assembly pathway.

There are still many other shell proteins in the Pdu MCP system that have not been fully explored. Performing similar free energy calculations on Pdu shell proteins not studied here would provide more insight into the interactions present in the shell. The next extension would be to begin designing these interactions between shell proteins by mutating the amino acid primary sequence. Free energy calculations would be carried out in order to test how mutations affect bending angles and total interaction strengths. This would also help to build a general understanding of the role of different amino acid interactions in determining the overall interactions between two shell proteins as measured by the free energy calculations. The end goal would be to implement these mutations in a real system and show the effects of changing bending interactions on the overall structure.

References

- [1] Carolyn E. Mills, Curt Waltmann, Andre G. Archer, Nolan W. Kennedy, Charlotte H. Abrahamson, Alexander D. Jackson, Eric W. Roth, Sasha Shirman, Michael C. Jewett, Niall M. Mangan, Monica Olvera de la Cruz, and Danielle Tullman-Ercek. Vertex protein pdun tunes encapsulated pathway performance by dictating bacterial metabolosome morphology. *Nat. Commun.*, 13(1):3746, Dec 2022.
- [2] Thilo Stehle, Steven J Gamblin, Youwei Yan, and Stephen C Harrison. The structure of simian virus 40 refined at 3.1 Å resolution. *Structure*, 4(2):165 – 182, 1996.
- [3] Stanislav Kler, Joseph Che-Yen Wang, Mary Dhason, Ariella Oppenheim, and Adam Zlotnick. Scaffold properties are a key determinant of the size and shape of self-assembled virus-derived particles. *ACS Chem. Biol.*, 8(12):2753–2761, 2013. PMID: 24093474.
- [4] Shin-nosuke Kaneshashi, Ken-ichiro Ishizu, Masa-aki Kawano, Song-iee Han, Satoru Tomita, Hajime Watanabe, Kohsuke Kataoka, and Hiroshi Handa. Simian virus 40 vp1 capsid protein forms polymorphic assemblies *In Vitro. J. Gen. Virol.*, 84(7):1899–1905, 2003.
- [5] Yaohua Li, Nolan W. Kennedy, Siyu Li, Carolyn E. Mills, Danielle Tullman-Ercek, and Monica Olvera de la Cruz. Computational and experimental approaches to controlling bacterial microcompartment assembly. *ACS Cent. Sci.*, 7(4):658–670, 2021.
- [6] John Mackenzie Grant Cowie and Valeria Arrighi. *Polymers: chemistry and physics of modern materials*. CRC press, 2007.
- [7] Michael Rubinstein, Ralph H Colby, et al. *Polymer physics*, volume 23. Oxford university press New York, 2003.
- [8] P. E. Rouse and K. Sittel. Viscoelastic properties of dilute polymer solutions. *Journal of Applied Physics*, 24(6):690–696, Jun 1953.

- [9] B-Y Ha and Andrea J Liu. Counterion-mediated attraction between two like-charged rods. *Phys. Rev. Lett.*, 79(7):1289, 1997.
- [10] Jiahua Zhu, Shi yi Zhang, Kecheng Zhang, Xiang hui Wang, Jimmy W. Mays, Karen L Wooley, and Darrin J. Pochan. Disk-cylinder and disk-sphere nanoparticles via a block copolymer blend solution construction. *Nat. Commun.*, 4:2297, 2013.
- [11] Jeremy M. Tymoczko, John L. Berg and Lubert Streyer. *Biochemistry A Short Course*. WH Freeman, New York; New York, 3 edition, 2015.
- [12] Vladimir N. Maiorov and Gordon M. Crippen. Size-independent comparison of protein three-dimensional structures. *Proteins: Struct., Funct., Bioinf.*, 22(3):273–283, 1995.
- [13] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Zidek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, Aug 2021.
- [14] Timothy J. Nott, Evangelia Petsalaki, Patrick Farber, Dylan Jarvis, Eden Fussner, Anne Plochowitz, Timothy D. Craggs, David P. Bazett-Jones, Tony Pawson, Julie D. Forman-Kay, and Andrew J. Baldwin. Phase Transition of a Disordered Nucleage Protein Generates Environmentally Responsive Membraneless Organelles. *Mol. Cell*, 57(5):936–947, March 2015.
- [15] Yuji Higuchi and Momoji Kubo. Coarse-grained molecular dynamics simulation of the void growth process in the block structure of semicrystalline polymers. *Modelling and Simulation in Materials Science and Engineering*, 24(5):055006, may 2016.
- [16] Virginie Hugouvieux, Monique A. V. Axelos, and Max Kolb. Micelle Formation, Gelation and Phase Separation of Amphiphilic Multiblock Copolymers†. *Soft Matter*, 7(6):2580, January 2011.
- [17] B. W Swift and M. Olvera de la Cruz. Random copolymers in concentrated solutions. *European Physics Letters (EPL)*, 35(7):487–492, sep 1996.

- [18] Li-Wei Chang, Tyler K. Lytle, Mithun Radhakrishna, Jason J. Madinya, Jon Vélez, Charles E. Sing, and Sarah L. Perry. Sequence and entropy-based control of complex coacervates. *Nat. Commun.*, 8(1):1273, Dec 2017.
- [19] Han Umana-Kossio, Trung Dac Nguyen, Jeremy Wang, Monica Olvera de La Cruz, and John M. Torkelson. Unusual glass transition breadths of ionomers: Effects of thermal treatment and charge-carrying side chains. *Macromolecules*, 55(15):6536–6546, Aug 2022.
- [20] VV Vasilevskaya, AR Khokhlov, and K Yoshikawa. Single Polyelectrolyte Macromolecule in the Salt Solution: Effect of Escaped Counter Ions. *Macromol. Theory Simul.*, 9(8):600–607, NOV 14 2000.
- [21] D. Prusty, R. J. Nap, I. Szleifer, and M. Olvera de la Cruz. Charge regulation mechanism in end-tethered weak polyampholytes. *Soft Matter*, 16:8832–8847, 2020.
- [22] Charles E. Sing and Sarah L. Perry. Recent progress in the science of complex coacervation. *Soft Matter*, 16:2885–2914, 2020.
- [23] Mohsen Ghasemi and Ronald G. Larson. Role of electrostatic interactions in charge regulation of weakly dissociating polyacids. *Prog. in Polym. Sci.*, 112:101322, 2021.
- [24] Sarah Perry, Yue Li, Dimitrios Priftis, Lorraine Leon, and Matthew Tirrell. The effect of salt on the complex coacervation of vinyl polyelectrolytes. *Polymers*, 6(6):1756–1772, Jun 2014.
- [25] Mohsen Ghasemi, Sean Friedowitz, and Ronald G. Larson. Analysis of partitioning of salt through doping of polyelectrolyte complex coacervates. *Macromolecules*, 53(16):6928–6945, 2020.
- [26] Anisha Shakya, Martin Girard, John T. King, and Monica Olvera de la Cruz. Role of chain flexibility in asymmetric polyelectrolyte complexation in salt solutions. *Macromolecules*, 53(4):1258–1269, feb 2020.
- [27] Lu Li, Artem M. Rumyantsev, Samanvaya Srivastava, Siqi Meng, Juan J. de Pablo, and Matthew V. Tirrell. Effect of solvent quality on the phase behavior of polyelectrolyte complexes. *Macromolecules*, 54(1):105–114, 2021.
- [28] Mengmeng Zhao, Chao Wang, Haowei Jiang, Mahesh B. Dawadi, Bryan D. Vogt, David A. Modarelli, and Nicole S. Zacharia. Polyelectrolyte–micelle coacervates: intrapolymer-dominant vs. interpolymer-dominant association, solute uptake and rheological properties. *Soft Matter*, 15:3043–3054, 2019.

- [29] Xue Mi, Whitney C. Blocher McTigue, Pratik U. Joshi, Mallory K. Bunker, Caryn L. Heldt, and Sarah L. Perry. Thermostabilization of viruses via complex coacervation. *Biomater. Sci.*, 8:7082–7092, 2020.
- [30] Whitney C. Blocher McTigue and Sarah L. Perry. Design rules for encapsulating proteins into complex coacervates. *Soft Matter*, 15:3089–3103, 2019.
- [31] Carolyn E. Mills, Allie Obermeyer, Xuehui Dong, Jeremy Walker, and Bradley D. Olsen. Complex coacervate core micelles for the dispersion and stabilization of organophosphate hydrolase in organic solvents. *Langmuir*, 32(50):13367–13376, 2016. PMID: 27993031.
- [32] Allie C. Obermeyer, Carolyn E. Mills, Xue-Hui Dong, Romeo J. Flores, and Bradley D. Olsen. Complex coacervation of supercharged proteins with polyelectrolytes. *Soft Matter*, 12:3570–3581, 2016.
- [33] Hursh V. Sureka, Allie C. Obermeyer, Romeo J. Flores, and Bradley D. Olsen. Catalytic biosensors from complex coacervate core micelle (c3m) thin films. *ACS Appl. Mater. Interfaces*, 11(35):32354–32365, 2019. PMID: 31441305.
- [34] Aaron Huang, Justin M. Paloni, Amy Wang, Allie C. Obermeyer, Hursh V. Sureka, Helen Yao, and Bradley D. Olsen. Predicting protein–polymer block copolymer self-assembly from protein properties. *Biomacromolecules*, 20(10):3713–3723, 2019. PMID: 31502834.
- [35] Chad S. Cummings and Allie C. Obermeyer. Phase separation behavior of supercharged proteins and polyelectrolytes. *Biochemistry*, 57(3):314–323, 2018. PMID: 29210575.
- [36] Rachel A. Kapelner and Allie C. Obermeyer. Ionic polypeptide tags for protein phase separation. *Chem. Sci.*, 10:2700–2707, 2019.
- [37] Vivian Yeong, Emily G. Werth, Lewis M. Brown, and Allie C. Obermeyer. Formation of biomolecular condensates in bacteria by tuning protein electrostatics. *ACS Cent. Sci.*, 6(12):2301–2310, 2020.
- [38] Rachel A. Kapelner, Vivian Yeong, and Allie C. Obermeyer. Molecular determinants of protein-based coacervates. *Curr. Opin. Colloid Interface Sci.*, 52:101407, 2021.
- [39] Hao Sun, Baofu Qiao, Wonmin Choi, Nicholas Hampu, Naneki C. McCallum, Matthew P. Thompson, Julia Oktawiec, Steven Weigand, Omar M. Ebrahim, Monica Olvera de la Cruz, and Nathan C. Gianneschi. Origin of proteolytic stability of

- peptide-brush polymers as globular proteomimetics. *ACS Cent. Sci.*, 7(12):2063–2072, 2021.
- [40] Justin M. Horn, Rachel A. Kapelner, and Allie C. Obermeyer. Macro- and microphase separated protein-polyelectrolyte complexes: Design parameters and current progress. *Polymers*, 11(4), 2019.
- [41] Trung Dac Nguyen, Baofu Qiao, and Monica Olvera de la Cruz. Efficient encapsulation of proteins with random copolymers. *Proc. Natl. Acad. Sci. U.S.A.*, 115(26):6578–6583, 2018.
- [42] Tyler K. Lytle, Li-Wei Chang, Natalia Markiewicz, Sarah L. Perry, and Charles E. Sing. Designing electrostatic interactions via polyelectrolyte monomer sequence. *ACS Cent. Sci.*, 5(4):709–718, apr 2019.
- [43] Clifford P. Brangwynne, Christian R. Eckmann, David S. Courson, Agata Rybarska, Carsten Hoege, Jöbin Gharakhani, Frank Jülicher, and Anthony A. Hyman. Germline p granules are liquid droplets that localize by controlled dissolution/condensation. *Science*, 324(5935):1729–1732, 2009.
- [44] Denis L. J. Lafontaine, Joshua A. Riback, Rümeyza Bascetin, and Clifford P. Brangwynne. The nucleolus as a multiphase liquid condensate. *Nature Reviews Molecular Cell Biology*, 22(3):165–182, Mar 2021.
- [45] Salman F. Banani, Hyun O. Lee, Anthony A. Hyman, and Michael K. Rosen. Biomolecular condensates: Organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.*, 18(5):285–298, 2017.
- [46] Paul Anderson and Nancy Kedersha. Stress granules: the tao of rna triage. *Trends in Biochemical Sciences*, 33(3):141–150, Mar 2008.
- [47] Christina M Agapakis, Patrick M Boyle, and Pamela A Silver. Natural strategies for the spatial optimization of metabolism in synthetic biology. *Nature Chemical Biology*, 8(6):527–535, Jun 2012.
- [48] Jared Strong, Iowa Capital Dispatch January 11, and 2022. Dnr: Ames drinking water has ‘forever chemicals’, Jan 2022.
- [49] Maocai Shen, Yaxin Zhang, Yuan Zhu, Biao Song, Guangming Zeng, Duofei Hu, Xiaofeng Wen, and Xiaoya Ren. Recent advances in toxicological research of nanoplastics in the environment: A review. *Environmental Pollution*, 252:511–521, 2019.

- [50] Harry P. Austin, Mark D. Allen, Bryon S. Donohoe, Nicholas A. Rorrer, Fiona L. Kearns, Rodrigo L. Silveira, Benjamin C. Pollard, Graham Dominick, Ramona Duman, Kamel El Omari, Vitaliy Mykhaylyk, Armin Wagner, William E. Michener, Antonella Amore, Munir S. Skaf, Michael F. Crowley, Alan W. Thorne, Christopher W. Johnson, H. Lee Woodcock, John E. McGeehan, and Gregg T. Beckham. Characterization and engineering of a plastic-degrading aromatic polyesterase. *Proc. Nat. Acad. Sci.*, 115(19):E4350–E4357, 2018.
- [51] Tobias Fecker, Pablo Galaz-Davison, Felipe Engelberger, Yoshie Narui, Marcos Sotomayor, Loreto P. Parra, and César A. Ramírez-Sarmiento. Active site flexibility as a hallmark for efficient pet degradation by *i. sakaiensis* petase. *Biophys. J.*, 114(6):1302–1312, 2018.
- [52] Hyeoncheol Francis Son, In Jin Cho, Seongjoon Joo, Hogyun Seo, Hye-Young Sagong, So Young Choi, Sang Yup Lee, and Kyung-Jin Kim. Rational Protein Engineering of Thermo-Stable PETase from *Ideonella sakaiensis* for Highly Efficient PET Degradation. *ACS Catal.*, 9(4):3519–3526, April 2019.
- [53] Lars Mathias Blank, Tanja Narancic, Jörg Mampel, Till Tiso, and Kevin O’Connor. Biotechnological upcycling of plastic waste and other non-conventional feedstocks in a circular economy. *Curr. Opin. Biotechnol.*, 62:212 – 219, 2020. Energy Biotechnology Environmental Biotechnology.
- [54] Ian A. Kane, Michael A. Clare, Elda Miramontes, Roy Wogelius, James J. Rothwell, Pierre Garreau, and Florian Pohl. Seafloor microplastic hotspots controlled by deep-sea circulation. *Science*, 368(6495):1140–1145, 2020.
- [55] Nolan W Kennedy, Carolyn E Mills, Taylor M Nichols, Charlotte H Abrahamson, and Danielle Tullman-Ercek. Bacterial microcompartments: Tiny organelles with big potential. *Curr. Opin. Microbiol.*, 63:36–42, 2021.
- [56] Jefferson S Plegaria and Cheryl A Kerfeld. Engineering nanoreactors using bacterial microcompartment architectures. *Curr. Opin. Biotechnol.*, 51:1–7, 2018. Systems biology • Nanobiotechnology.
- [57] 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. *Proc. Natl. Acad. Sci. U. S. A.*, 112(10):2990–2995, 2015.
- [58] Roya Zandi, Bogdan Dragnea, Alex Travesset, and Rudolf Podgornik. On virus growth and form. *Physics Reports*, 847:1–102, 2020. On virus growth and form.

- [59] Jason D. Perlmutter and Michael F. Hagan. Mechanisms of virus assembly. *Annu. Rev. Phys. Chem.*, 66(1):217–239, 2015. PMID: 25532951.
- [60] Stanislav Kler, Roi Asor, Chenglei Li, Avi Ginsburg, Daniel Harries, Ariella Oppenheim, Adam Zlotnick, and Uri Raviv. Rna encapsidation by sv40-derived nanoparticles follows a rapid two-state mechanism. *J. Am. Chem. Soc.*, 134(21):8823–8830, 2012. PMID: 22329660.
- [61] Jason D Perlmutter, Cong Qiao, and Michael F Hagan. Viral genome structures are optimal for capsid assembly. *eLife*, 2:e00632, jun 2013.
- [62] Roi Asor, Lisa Selzer, Christopher John Schlicksup, Zhongchao Zhao, Adam Zlotnick, and Uri Raviv. Assembly reactions of hepatitis b capsid protein into capsid nanoparticles follow a narrow path through a complex reaction landscape. *ACS Nano*, 13(7):7610–7626, 2019. PMID: 31173689.
- [63] Roi Asor, Surendra W. Singaram, Yael Levi-Kalisman, Michael F. Hagan, and Uri Raviv. Effect of ionic strength on the assembly of simian vacuolating virus capsid protein around poly(styrene sulfonate). *bioRxiv*, 2022.
- [64] D.L.D. Caspar and A. Klug. Physical principles in the construction of regular viruses. *Cold Spring Harbor Symposia of Quantitative Biology*, 1(27):1–24, 1962.
- [65] Sanaz Panahandeh, Siyu Li, Laurent Marichal, Rafael Leite Rubim, Guillaume Tresset, and Roya Zandi. How a virus circumvents energy barriers to form symmetric shells. *ACS Nano*, 14(3):3170–3180, 2020. PMID: 32115940.
- [66] Sanaz Panahandeh, Siyu Li, Bogdan Dragnea, and Roya Zandi. Virus assembly pathways inside a host cell. *ACS Nano*, 16(1):317–327, 2022. PMID: 35019271.
- [67] Siyu Li, Polly Roy, Alex Travesset, and Roya Zandi. Why large icosahedral viruses need scaffolding proteins. *Proc. Natl. Acad. Sci. U.S.A.* ., 2018.
- [68] Farzaneh Mohajerani and Michael F. Hagan. The role of the encapsulated cargo in microcompartment assembly. *PLoS Comput. Biol.*, 14(7):e1006351, July 2018.
- [69] Farzaneh Mohajerani, Evan Sayer, Christopher Neil, Koe Inlow, and Michael F. Hagan. Mechanisms of scaffold-mediated microcompartment assembly and size control. *ACS Nano*, 15(3):4197–4212, 2021. PMID: 33683101.
- [70] Feng Li, Ke Li, Zong-Qiang Cui, Zhi-Ping Zhang, Hong-Ping Wei, Ding Gao, Jiao-Yu Deng, and Xian-En Zhang. Viral coat proteins as flexible nano-building-blocks for nanoparticle encapsulation. *Small*, 6(20):2301–2308, 2010.

- [71] Yuba Raj Dahal and Monica Olvera de la Cruz. Crystallizing protein assemblies *via* free and grafted linkers. *Soft Matter*, 15:4311–4319, 2019.
- [72] Nolan W. Kennedy, Carolyn E. Mills, Charlotte H. Abrahamson, Andre G. Archer, Sasha Shirman, Michael C. Jewett, Niall M. Mangan, and Danielle Tullman-Ercek. Linking the salmonella enterica 1,2-propanediol utilization bacterial microcompartment shell to the enzymatic core via the shell protein pduB. *J. Bacteriol.*, pages e00576–21, May 2022.
- [73] Nolan W. Kennedy, Svetlana P. Ikononova, 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. *J. Mol. Biol.*, 433(2):166721, 2021.
- [74] Mengru Yang, Nicolas Wenner, Gregory F. Dykes, Yan Li, Xiaojun Zhu, Yaqi Sun, Fang Huang, Jay C. D. Hinton, and Lu-Ning Liu. Biogenesis of a bacterial metabolosome for propanediol utilization. *Nat. Commun.*, 13(1):2920, May 2022.
- [75] William M. Jacobs and Daan Frenkel. Phase transitions in biological systems with many components. *Biophys. J.*, 112(4):683–691, 2017.
- [76] Martin Girard. On kinetics and extreme values in systems with random interactions. *Phys. Biol.*, 20(1):016006, dec 2022.
- [77] Jeremy D. Schmit, Marina Feric, and Miroslav Dunder. How hierarchical interactions make membraneless organelles tick like clockwork. *Trends Biochem. Sci.*, 46(7):525–534, 2021.
- [78] Grant M. Rotskoff and Phillip L. Geissler. Robust nonequilibrium pathways to microcompartment assembly. *Proc. Natl. Acad. Sci. U.S.A.* ., 115(25):6341–6346, 2018.
- [79] Maelenn Chevreuil, Didier Law-Hine, Jingzhi Chen, Stéphane Bressanelli, Sophie Combet, Doru Constantin, Jéril Degrouard, Johannes Möller, Mehdi Zeghal, and Guillaume Tresset. Nonequilibrium self-assembly dynamics of icosahedral viral capsids packaging genome. *Biophys. J.*, 114:60a, 02 2018.
- [80] Jeremy Wang, Curt Waltmann, Han Umama-Kossio, Monica Olvera de la Cruz, and John M. Torkelson. Heterogeneous charged complexes of random copolymers for the segregation of organic molecules. *ACS Cent. Sci.*, 7(5):882–891, 2021.
- [81] Simcha Srebnik, Arup K. Chakraborty, and Eugene I. Shakhnovich. Adsorption-freezing transition for random heteropolymers near disordered 2d manifolds due to “pattern matching”. *Phys. Rev. Lett.*, 77:3157–3160, Oct 1996.

- [82] Arup K. Chakraborty. Disordered heteropolymers: models for biomimetic polymers and polymers with frustrating quenched disorder. *Physics Reports*, 342(1):1–61, 2001.
- [83] Alexey Polotsky, Friederike Schmid, and Andreas Degenhard. Polymer adsorption onto random planar surfaces: Interplay of polymer and surface correlations. *The Journal of Chemical Physics*, 121(10):4853–4864, 2004.
- [84] D Bratko, Arup K Chakraborty, and Eugene I Shakhnovich. Recognition between random heteropolymers and multifunctional disordered surfaces. *Chemical Physics Letters*, 280(1):46–52, 1997.
- [85] Brian Panganiban, Baofu Qiao, Tao Jiang, Christopher DelRe, Mona M. Obadia, Trung Dac Nguyen, Anton A. A. Smith, Aaron Hall, Izaac Sit, Marquise G. Crosby, Patrick B. Dennis, Eric Drockenmuller, Monica Olvera de la Cruz, and Ting Xu. Random heteropolymers preserve protein function in foreign environments. *Science*, 359(6381):1239–1243, mar 2018.
- [86] Charles E. Sing and Sarah L. Perry. Recent progress in the science of complex coacervation. *Soft Matter*, 16(12):2885–2914, 2020.
- [87] Andreas F. Thünemann, Martin Müller, Herbert Dautzenberg, Jean-François Joanny, and Hartmut Löwen. *Polyelectrolyte Complexes*, pages 113–171. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.
- [88] Dimitrios Priftis, Xiaoxing Xia, Khatcher O. Margossian, Sarah L. Perry, Lorraine Leon, Jian Qin, Juan J. de Pablo, and Matthew Tirrell. Ternary, tunable polyelectrolyte complex fluids driven by complex coacervation. *Macromolecules*, 47(9):3076–3085, 2014.
- [89] Sean Friedowitz, Ali Salehi, Ronald G. Larson, and Jian Qin. Role of electrostatic correlations in polyelectrolyte charge association. *J. Chem. Phys.*, 149(16):163335, 2018.
- [90] Venkata S. Meka, Manprit K.G. Sing, Mallikarjuna R. Pichika, Srinivasa R. Nali, Venkata R.M. Kolapalli, and Prashant Kesharwani. A comprehensive review on polyelectrolyte complexes. *Drug Discovery Today*, 22(11):1697–1706, nov 2017.
- [91] Abhijeet D. Kulkarni, Yogesh H. Vanjari, Karan H. Sancheti, Harun M. Patel, Veena S. Belgamwar, Sanjay J. Surana, and Chandrakantsing V. Pardeshi. Polyelectrolyte complexes: mechanisms, critical experimental aspects, and applications. *Artificial Cells, Nanomedicine, and Biotechnology*, 44(7):1615–1625, 2016. PMID: 26757773.

- [92] Jasper van der Gucht, Evan Spruijt, Marc Lemmers, and Martien A. Cohen Stuart. Polyelectrolyte complexes: Bulk phases and colloidal systems. *Journal of Colloid and Interface Science*, 361(2):407–422, 2011.
- [93] Artem M. Rumyantsev, Nicholas E. Jackson, and Juan J. de Pablo. Polyelectrolyte complex coacervates: Recent developments and new frontiers. *Annual Review of Condensed Matter Physics*, 12(1):155–176, 2021.
- [94] Thierry Delair. Colloidal polyelectrolyte complexes of chitosan and dextran sulfate towards versatile nanocarriers of bioactive molecules. *European Journal of Pharmaceutics and Biopharmaceutics*, 78(1):10–18, may 2011.
- [95] Sergei V. Raik, Ekaterina R. Gasilova, Natallia V. Dubashynskaya, Anatoliy V. Dobrodumov, and Yury A. Skorik. Diethylaminoethyl chitosan–hyaluronic acid polyelectrolyte complexes. *International Journal of Biological Macromolecules*, 146:1161–1168, 2020.
- [96] Federica Vecchies, Pasquale Sacco, Eva Decleva, Renzo Menegazzi, Davide Porrelli, Ivan Donati, Gianluca Turco, Sergio Paoletti, and Eleonora Marsich. Complex coacervates between a lactose-modified chitosan and hyaluronic acid as radical-scavenging drug carriers. *Biomacromolecules*, 19(10):3936–3944, 2018. PMID: 30204431.
- [97] Mengmeng Zhao and Nicole S. Zacharia. Protein encapsulation via polyelectrolyte complex coacervation: Protection against protein denaturation. *The Journal of Chemical Physics*, 149(16):163326, oct 2018.
- [98] Saskia Lindhoud and Mireille M. A. E. Claessens. Accumulation of small protein molecules in a macroscopic complex coacervate. *Soft Matter*, 12:408–413, 2016.
- [99] Karina K. Nakashima, Jochem F. Baaij, and Evan Spruijt. Reversible generation of coacervate droplets in an enzymatic network. *Soft Matter*, 14:361–367, 2018.
- [100] Gudrun Petzold and Simona Schwarz. Polyelectrolyte complexes in flocculation applications. In *Advances in Polymer Science*, pages 25–65. Springer Berlin Heidelberg, 2013.
- [101] Ana Balea, M. Concepcion Monte, Elena Fuente, Jose Luis Sanchez-Salvador, Angeles Blanco, and Carlos Negro. Cellulose nanofibers and chitosan to remove flexographic inks from wastewaters. *Environmental Science: Water Research & Technology*, 5(9):1558–1567, 2019.

- [102] Mriganka M. Ghosh, Chris D. Cox, and Temkar M. Prakash. Polyelectrolyte selection for water treatment. *Journal - American Water Works Association*, 77(3):67–73, mar 1985.
- [103] Vahid Vajihinejad, Sarang P. Gumfekar, Behnaz Bazoubandi, Zahra Rostami Najafabadi, and João B. P. Soares. Water soluble polymer flocculants: Synthesis, characterization, and performance assessment. *Macromolecular Materials and Engineering*, 304(2):1800526, 2019.
- [104] Li Yu, Xiaokong Liu, Weichang Yuan, Lauren Joan Brown, and Dayang Wang. Confined flocculation of ionic pollutants by poly(l-dopa)-based polyelectrolyte complexes in hydrogel beads for three-dimensional, quantitative, efficient water decontamination. *Langmuir*, 31(23):6351–6366, jun 2015.
- [105] Mengmeng Zhao and Nicole S. Zacharia. Sequestration of methylene blue into polyelectrolyte complex coacervates. *Macromolecular Rapid Communications*, 37(15):1249–1255, 2016.
- [106] Shuyue Huang, Mengmeng Zhao, Mahesh B. Dawadi, Yuhang Cai, Yakov Lapitsky, David A. Modarelli, and Nicole S. Zacharia. Effect of small molecules on the phase behavior and coacervation of aqueous solutions of poly(diallyldimethylammonium chloride) and poly(sodium 4-styrene sulfonate). *Journal of Colloid and Interface Science*, 518:216–224, 2018.
- [107] Mengmeng Zhao, Seyed Ali Eghtesadi, Mahesh B. Dawadi, Chao Wang, Shuyue Huang, Amy E. Seymore, Bryan D. Vogt, David A. Modarelli, Tianbo Liu, and Nicole S. Zacharia. Partitioning of small molecules in hydrogen-bonding complex coacervates of poly(acrylic acid) and poly(ethylene glycol) or pluronic block copolymer. *Macromolecules*, 50(10):3818–3830, 2017.
- [108] Takaya Terashima, Takanori Sugita, Kaoru Fukae, and Mitsuo Sawamoto. Synthesis and single-chain folding of amphiphilic random copolymers in water. *Macromolecules*, 47(2):589–600, jan 2014.
- [109] 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, mar 2018.
- [110] Boran Ma, Trung Dac Nguyen, and Monica Olvera de la Cruz. Control of ionic mobility via charge size asymmetry in random ionomers. *Nano Lett.*, 20(1):43–49, nov 2019.
- [111] Didier Cochin, Frans C. De Schryver, André Laschewsky, and Jan van Stam. Polysoaps in aqueous solutions: intermolecular versus intramolecular hydrophobic

- aggregation studied by fluorescence spectroscopy. *Langmuir*, 17(9):2579–2584, may 2001.
- [112] Rati Ranjan Nayak, Sumita Roy, and Joykrishna Dey. Characterization and self-assembly of poly[sodium n-(11-acrylamidoundecanoyl)-l-alaninate] in water. *Polymer*, 46(26):12401–12409, 2005.
- [113] Cansu Aydogan, Mustafa Ciftci, and Yusuf Yagci. Hyperbranched polymers by type II photoinitiated self-condensing vinyl polymerization. *Macromolecular Rapid Communications*, 37(7):650–654, jan 2016.
- [114] Alison R. Maniego, Adam T. Sutton, Marianne Gaborieau, and Patrice Castignolles. Assessment of the branching quantification in poly(acrylic acid): Is it as easy as it seems? *Macromolecules*, 50(22):9032–9041, 2017.
- [115] Anu Matilainen, Mikko Vepsäläinen, and Mika Sillanpää. Natural organic matter removal by coagulation during drinking water treatment: A review. *Advances in Colloid and Interface Science*, 159(2):189–197, 2010.
- [116] Herbert Dautzenberg and Natalia Karibyants. Polyelectrolyte complex formation in highly aggregating systems. effect of salt: response to subsequent addition of NaCl. *Macromolecular Chemistry and Physics*, 200(1):118–125, jan 1999.
- [117] Adina Roxana Petcu, Elena Adina Rogozea, Cosmina Andreea Lazar, Nicoleta Liliana Olteanu, Aurelia Meghea, and Maria Mihaly. Specific interactions within micelle microenvironment in different charged dye/surfactant systems. *Arabian Journal of Chemistry*, 9(1):9–17, 2016.
- [118] Kateryna V. Roshchyna, Sergey V. Eltsov, Anna N. Laguta, and Nikolay O. Mchedlov-Petrosyan. Micellar rate effects in the alkaline fading of crystal violet in the presence of various surfactants. *Journal of Molecular Liquids*, 201:77–82, jan 2015.
- [119] Alexander Kudlay and Monica Olvera de la Cruz. Precipitation of oppositely charged polyelectrolytes in salt solutions. *The Journal of Chemical Physics*, 120(1):404–412, jan 2004.
- [120] Artem M. Rumyantsev, Nicholas E. Jackson, Boyuan Yu, Jeffrey M. Ting, Wei Chen, Matthew V. Tirrell, and Juan J. de Pablo. Controlling complex coacervation via random polyelectrolyte sequences. *ACS Macro Lett.*, 8(10):1296–1302, 2019.
- [121] Paul C. Painter and Michael M. Coleman. *Fundamentals of polymer science: an introductory text*. Technomic Pub. Co, Lancaster, Pa, 2nd ed edition, 1997.

- [122] A. Halperin. On the collapse of multiblock copolymers. *Macromolecules*, 24(6):1418–1419, nov 1991.
- [123] A. Nesarikar, M. Olvera de la Cruz, and B. Crist. Phase transitions in random copolymers. *The Journal of Chemical Physics*, 98(9):7385–7397, may 1993.
- [124] Siewert J. Marrink, H. Jelger Risselada, Serge Yefimov, D. Peter Tieleman, and Alex H. de Vries. The martini force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B*, 111(27):7812–7824, 2007. PMID: 17569554.
- [125] Semen O. Yesylevskyy, Lars V. Schäfer, Durba Sengupta, and Siewert J. Marrink. Polarizable water model for the coarse-grained martini force field. *PLoS Computat. Biol.*, 6(6):e1000810, Jun 2010.
- [126] Gerardo Campos-Villalobos, Flor R. Siperstein, and Alessandro Patti. Transferable coarse-grained MARTINI model for methacrylate-based copolymers. *Mol. Syst. Des. Eng.*, 4(1):186–198, 2019.
- [127] Joshua A. Anderson, Chris D. Lorenz, and A. Travesset. General purpose molecular dynamics simulations fully implemented on graphics processing units. *J. Comput. Phys.*, 227(10):5342 – 5359, 2008.
- [128] David N. LeBard, Benjamin G. Levine, Philipp Mertmann, Stephen A. Barr, Arben Jusufi, Samantha Sanders, Michael L. Klein, and Athanassios Z. Panagiotopoulos. Self-assembly of coarse-grained ionic surfactants accelerated by graphics processing units. *Soft Matter*, 8(8):2385–2397, 2012.
- [129] John D. Weeks, David Chandler, and Hans C. Andersen. Role of repulsive forces in determining the equilibrium structure of simple liquids. *The Journal of Chemical Physics*, 54(12):5237–5247, jun 1971.
- [130] Fabian Grunewald, Giulia Rossi, Alex H. de Vries, Siewert J. Marrink, and Luca Monticelli. Transferable MARTINI Model of Poly(ethylene Oxide). *J. Phys. Chem. B*, 122(29):7436–7449, July 2018.
- [131] Kai Zhang, Amir Hossein Hamidian, Aleksandra Tubić, Yu Zhang, James K. H. Fang, Chenxi Wu, and Paul K. S. Lam. Understanding plastic degradation and microplastic formation in the environment: A review. *Environ. Pollut.*, 274:116554, April 2021.
- [132] P. Tziourrou, S. Kordella, Y. Ardali, G. Papatheodorou, and H. K. Karapanagioti. Microplastics formation based on degradation characteristics of beached plastic bags. *Mar. Pollut. Bull.*, 169:112470, August 2021.

- [133] Rahul A. Naik, Lewis Stetson Rowles, Areeb I. Hossain, Megan Yen, Ruba M. Aldossary, Onur G. Apul, Jeremy Conkle, and Navid B. Saleh. Microplastic particle versus fiber generation during photo-transformation in simulated seawater. *Sci. Total Environ.*, 736:139690, September 2020.
- [134] Prabhat Kumar Rai, Jechan Lee, Richard J. C. Brown, and Ki-Hyun Kim. Micro- and nano-plastic pollution: Behavior, microbial ecology, and remediation technologies. *J. Cleaner Prod.*, 291:125240, April 2021.
- [135] Marie Enfrin, Ludovic F. Dumée, and Judy Lee. Nano/microplastics in water and wastewater treatment processes – Origin, impact and potential solutions. *Water Res.*, 161:621–638, September 2019.
- [136] Albert A. Koelmans, Adil Bakir, G. Allen Burton, and Colin R. Janssen. Microplastic as a Vector for Chemicals in the Aquatic Environment: Critical Review and Model-Supported Reinterpretation of Empirical Studies. *Environ. Sci. Technol.*, 50(7):3315–3326, April 2016. Publisher: American Chemical Society.
- [137] Baorong Jiang, Alexandra E Kauffman, Lei Li, Wayne McFee, Bo Cai, John Weinstein, Jamie R Lead, Saurabh Chatterjee, Geoffrey I Scott, and Shuo Xiao. Health impacts of environmental contamination of micro- and nanoplastics: a review. *Environ. Health Prev. Med.*, 25(1):29, July 2020.
- [138] Xing Gao, Iram Hassan, Yutao Peng, Shouliang Huo, and Lan Ling. Behaviors and influencing factors of the heavy metals adsorption onto microplastics: A review. *J. Cleaner Prod.*, 319:128777, October 2021.
- [139] Lina Ramirez Arenas, Stéphan Ramseier Gentile, Stéphane Zimmermann, and Serge Stoll. Nanoplastics adsorption and removal efficiency by granular activated carbon used in drinking water treatment process. *Sci. Total Environ.*, 791:148175, October 2021.
- [140] Fei Yu, Changfu Yang, Zhilin Zhu, Xueting Bai, and Jie Ma. Adsorption behavior of organic pollutants and metals on micro/nanoplastics in the aquatic environment. *Sci. Total Environ.*, 694:133643, December 2019.
- [141] Maxine Swee-Li Yee, Ling-Wei Hii, Chin King Looi, Wei-Meng Lim, Shew-Fung Wong, Yih-Yih Kok, Boon-Keat Tan, Chiew-Yen Wong, and Chee-Onn Leong. Impact of Microplastics and Nanoplastics on Human Health. *Nanomaterials*, 11(2):496, February 2021. Number: 2 Publisher: Multidisciplinary Digital Publishing Institute.

- [142] Denise M. Mitrano, Peter Wick, and Bernd Nowack. Placing nanoplastics in the context of global plastic pollution. *Nature Nanotechnology*, 16(5):491–500, May 2021. Number: 5 Publisher: Nature Publishing Group.
- [143] Jian Kang, Li Zhou, Xiaoguang Duan, Hongqi Sun, Zhimin Ao, and Shaobin Wang. Degradation of Cosmetic Microplastics via Functionalized Carbon Nanosprings. *Matter*, 1(3):745–758, September 2019.
- [144] Tajkia Syeed Tofa, Karthik Laxman Kunjali, Swaraj Paul, and Joydeep Dutta. Visible light photocatalytic degradation of microplastic residues with zinc oxide nanorods. *Environ. Chem. Lett.*, 17(3):1341–1346, September 2019.
- [145] Ceyhun Akarsu and Fatma Deniz. Electrocoagulation/Electroflotation Process for Removal of Organics and Microplastics in Laundry Wastewater. *CLEAN – Soil, Air, Water*, 49(1):2000146, 2021. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/clen.202000146>.
- [146] William Perren, Arkadiusz Wojtasik, and Qiong Cai. Removal of Microbeads from Wastewater Using Electrocoagulation. *ACS Omega*, 3(3):3357–3364, March 2018. Publisher: American Chemical Society.
- [147] Yongli Zhang, Allison Diehl, Ashton Lewandowski, Kishore Gopalakrishnan, and Tracie Baker. Removal efficiency of micro- and nanoplastics (180 nm–125 m) during drinking water treatment. *Sci. Total Environ.*, 720:137383, June 2020.
- [148] Mriganka M. Ghosh, Chris D. Cox, and Temkar M. Prakash. Polyelectrolyte Selection for Water Treatment. *Journal AWWA*, 77(3):67–73, 1985. _eprint: <https://awwa.onlinelibrary.wiley.com/doi/pdf/10.1002/j.1%551-8833.1985.tb05510.x>.
- [149] Gudrun Petzold and Simona Schwarz. Polyelectrolyte Complexes in Flocculation Applications. In Martin Müller, editor, *Polyelectrolyte Complexes in the Dispersed and Solid State II: Application Aspects*, Advances in Polymer Science, pages 25–65. Springer, Berlin, Heidelberg, 2014.
- [150] John Bratby. *Coagulation and flocculation in water and wastewater treatment*. IWA publishing, 2016.
- [151] Yunhai Zhang, Xinjie Wang, Ying Li, Hao Wang, Yuexiao Shi, Yang Li, and Yongjun Zhang. Improving nanoplastic removal by coagulation: Impact mechanism of particle size and water chemical conditions. *Journal of Hazardous Materials*, 425:127962, March 2022.

- [152] Ziying Chen, Junhong Liu, Chengyu Chen, and Zhujian Huang. Sedimentation of nanoplastics from water with Ca/Al dual flocculants: Characterization, interface reaction, effects of pH and ion ratios. *Chemosphere*, 252:126450, August 2020.
- [153] C. H. M. Hofman-Caris, P. S. Bäuerlein, W. G. Siegers, S. M. Mintenig, R. Messina, S. C. Dekker, Ch Bertelkamp, E. R. Cornelissen, and A. P. van Wezel. Removal of nanoparticles (both inorganic nanoparticles and nanoplastics) in drinking water treatment – coagulation/flocculation/sedimentation, and sand/granular activated carbon filtration. *Environmental Science: Water Research & Technology*, 8(8):1675–1686, 2022. Publisher: Royal Society of Chemistry.
- [154] Bolormaa Bayarkhuu and Jeehye Byun. Optimization of coagulation and sedimentation conditions by turbidity measurement for nano- and microplastic removal. *Chemosphere*, 306:135572, November 2022.
- [155] Audrey Murray and Banu Örmeci. Removal Effectiveness of Nanoplastics (<400 nm) with Separation Processes Used for Water and Wastewater Treatment. *Water*, 12(3):635, March 2020. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.
- [156] Vikramjit S. Rathee, Aristotle J. Zervoudakis, Hythem Sidky, Benjamin J. Sikora, and Jonathan K. Whitmer. Weak polyelectrolyte complexation driven by associative charging. *The Journal of Chemical Physics*, 148(11):114901, mar 2018.
- [157] Pengfei Zhang, Nayef M. Alsaifi, Jianzhong Wu, and Zhen-Gang Wang. Salting-out and salting-in of polyelectrolyte solutions: A liquid-state theory study. *Macromolecules*, 49(24):9720–9730, 2016.
- [158] Mengmeng Zhao, Chao Wang, Haowei Jiang, Mahesh B. Dawadi, Bryan D. Vogt, David A. Modarelli, and Nicole S. Zacharia. Polyelectrolyte–micelle coacervates: intrapolymer-dominant vs. interpolymer-dominant association, solute uptake and rheological properties. *Soft Matter*, 15(14):3043–3054, 2019.
- [159] Annalisa Cardellini, Felipe Jiménez-Ángeles, Pietro Asinari, and Monica Olvera de la Cruz. A modeling-based design to engineering protein hydrogels with random copolymers. *ACS Nano*, 15(10):16139–16148, 2021. PMID: 34644059.
- [160] Yuri A Kriksin, Pavel G Khalatur, and Alexei R Khokhlov. Recognition of complex patterned substrates by heteropolymer chains consisting of multiple monomer types. *J. Chem. Phys.*, 124(17):174904, 2006.

- [161] Christopher DelRe, Yufeng Jiang, Philjun Kang, Junpyo Kwon, Aaron Hall, Ivan Jayapurna, Zhiyuan Ruan, Le Ma, Kyle Zolkin, Tim Li, and et al. Near-complete depolymerization of polyesters with nano-dispersed enzymes. *Nature*, 592(7855):558–563, Apr 2021.
- [162] Curt Waltmann, Carolyn E. Mills, Jeremy Wang, Baofu Qiao, John M. Torkelson, Danielle Tullman-Ercek, and Monica Olvera de la Cruz. Functional enzyme polymer complexes. *Proceedings of the National Academy of Sciences*, 119(13):e2119509119, 2022.
- [163] Christopher Balzer, Pengfei Zhang, and Zhen-Gang Wang. Wetting behavior of polyelectrolyte complex coacervates on solid surfaces. *Soft Matter*, 18:6326–6339, 2022.
- [164] Siewert J. Marrink, H. Jelger Risselada, Serge Yefimov, D. Peter Tieleman, and Alex H. de Vries. The martini force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B*, 111(27):7812–7824, 2007. PMID: 17569554.
- [165] Kenneth R. Shull, Meredith Taghon, and Qifeng Wang. Investigations of the high-frequency dynamic properties of polymeric systems with quartz crystal resonators. *Biointerphases*, QCM2020(1):021012, March 2020.
- [166] P.G. De Gennes. Scaling theory of polymer adsorption. *Journal de Physique*, 37(12):1445–1452, 1976.
- [167] A. Johner and J. F. Joanny. Polymer adsorption in a poor solvent. *Journal de Physique II*, 1(2):181–194, Feb 1991.
- [168] Ting Ge and Michael Rubinstein. Strong selective adsorption of polymers. *Macromolecules*, 48(11):3788–3801, 2015. PMID: 27065491.
- [169] Jiayu Cao, Qing Yang, Jie Jiang, Tatenda Dalu, Aliaksei Kadushkin, Joginder Singh, Rawil Fakhrullin, Fangjun Wang, Xiaoming Cai, and Ruibin Li. Coronas of micro/nano plastics: a key determinant in their risk assessments. *Particle and Fibre Toxicology*, 19(1):55, August 2022.
- [170] Djurre H. de Jong, Gurpreet Singh, W. F. Drew Bennett, Clement Arnarez, Tsjerk A. Wassenaar, Lars V. Schäfer, Xavier Periole, D. Peter Tieleman, and Siewert J. Marrink. Improved parameters for the martini coarse-grained protein force field. *J. Chem. Theory Comput.*, 9(1):687–697, 2013. PMID: 26589065.

- [171] Semen O. Yesylevskyy, Lars V. Schäfer, Durba Sengupta, and Siewert J. Marrink. Polarizable water model for the coarse-grained MARTINI force field. *PLoS Comput. Biol.*, 6(6):e1000810, jun 2010.
- [172] J. P. Jog. Crystallization of polyethyleneterephthalate. *J. Macromol. Sci., Part C*, 35(3):531–553, 1995.
- [173] Martin Vögele, Christian Holm, and Jens Smiatek. Coarse-grained simulations of polyelectrolyte complexes: Martini models for poly(styrene sulfonate) and poly(diallyldimethylammonium). *The Journal of Chemical Physics*, 143(24):243151, 2015.
- [174] Felipe Jiménez-Ángeles, Ha-Kyung Kwon, Kazi Sadman, Thomas Wu, Kenneth R. Shull, and Monica Olvera de la Cruz. Self-assembly of charge-containing copolymers at the liquid–liquid interface. *ACS Cent. Sci.*, 5(4):688–699, 2019.
- [175] In-Chul Yeh and Max L. Berkowitz. Ewald summation for systems with slab geometry. *The Journal of Chemical Physics*, 111(7):3155–3162, 1999.
- [176] Joshua A. Anderson, Chris D. Lorenz, and A. Travesset. General purpose molecular dynamics simulations fully implemented on graphics processing units. *J. Comput. Phys.*, 227(10):5342–5359, may 2008.
- [177] Minor J. Coon. Cytochrome p450: Nature’s most versatile biological catalyst. *Annu. Rev. Pharmacol. Toxicol.*, 45(1):1–25, 2005. PMID: 15832443.
- [178] Carlos A. Martinez and Sanjeewa G. Rupasinghe. Cytochrome P450 bioreactors in the pharmaceutical industry: Challenges and opportunities. *Curr. Trends Med. Chem.*, 13(12):1470–1490, 2013.
- [179] Andrés Cózar, Fidel Echevarría, J. Ignacio González-Gordillo, Xabier Irigoien, Bárbara Úbeda, Santiago Hernández-León, Álvaro T. Palma, Sandra Navarro, Juan García-de Lomas, Andrea Ruiz, María L. Fernández-de Puellas, and Carlos M. Duarte. Plastic debris in the open ocean. *Proc. Nat. Acad. Sci.*, 111(28):10239–10244, 2014.
- [180] Kara Lavender Law, Skye Morét-Ferguson, Nikolai A. Maximenko, Giora Proskurowski, Emily E. Peacock, Jan Hafner, and Christopher M. Reddy. Plastic accumulation in the north atlantic subtropical gyre. *Science*, 329(5996):1185–1188, 2010.

- [181] Jenna R. Jambeck, Roland Geyer, Chris Wilcox, Theodore R. Siegler, Miriam Perryman, Anthony Andrady, Ramani Narayan, and Kara Lavender Law. Plastic waste inputs from land into the ocean. *Science*, 347(6223):768–771, 2015.
- [182] Kieran D. Cox, Garth A. Covernton, Hailey L. Davies, John F. Dower, Francis Juanes, and Sarah E. Dudas. Human consumption of microplastics. *Environ. Sci. Technol.*, 53(12):7068–7074, 2019. PMID: 31184127.
- [183] Joleah B. Lamb, Bette L. Willis, Evan A. Fiorenza, Courtney S. Couch, Robert Howard, Douglas N. Rader, James D. True, Lisa A. Kelly, Awaludinnoer Ahmad, Jamaluddin Jompa, and C. Drew Harvell. Plastic waste associated with disease on coral reefs. *Science*, 359(6374):460–462, 2018.
- [184] An Su, Serpil Kiokekli, Mariam Naviwala, Abhijit N. Shirke, Ioannis V. Pavlidis, and Richard A. Gross. Cutinases as stereoselective catalysts: Specific activity and enantioselectivity of cutinases and lipases for menthol and its analogs. *Enzyme Microb. Technol.*, 133:109467, 2020.
- [185] Abhijit N. Shirke, Christine White, Jacob A. Englaender, Allison Zwarycz, Glenn L. Butterfoss, Robert J. Linhardt, and Richard A. Gross. Stabilizing leaf and branch compost cutinase (lcc) with glycosylation: Mechanism and effect on pet hydrolysis. *Biochemistry*, 57(7):1190–1200, 2018. PMID: 29328676.
- [186] Åsa M. Ronkvist, Wenchun Xie, Wenhua Lu, and Richard A. Gross. Cutinase-catalyzed hydrolysis of poly(ethylene terephthalate). *Macromolecules*, 42(14):5128–5138, 2009.
- [187] Stephen B. Lamb, David C. Lamb, Steven L. Kelly, and David C. Stuckey. Cytochrome P450 immobilisation as a route to bioremediation/biocatalysis. *FEBS Lett.*, 431(3):343–346, July 1998.
- [188] Kai Chen and Frances H. Arnold. Engineering cytochrome p450s for enantioselective cyclopropanation of internal alkynes. *J. Am. Chem. Soc.*, 142(15):6891–6895, 2020. PMID: 32223130.
- [189] Frances H. Arnold. Directed evolution: Bringing new chemistry to life. *Angew. Chem., Int. Ed.*, 57(16):4143–4148, 2018.
- [190] Nidhi C. Dubey and Bijay P. Tripathi. Nature inspired multienzyme immobilization: Strategies and concepts. *ACS Appl. Bio Mater.*, 4(2):1077–1114, 2021.
- [191] Brandon C. Knott, Erika Erickson, Mark D. Allen, Japheth E. Gado, Rosie Graham, Fiona L. Kearns, Isabel Pardo, Ece Topuzlu, Jared J. Anderson, Harry P.

- Austin, Graham Dominick, Christopher W. Johnson, Nicholas A. Rorrer, Carolyn J. Szostkiewicz, Valérie Copié, Christina M. Payne, H. Lee Woodcock, Bryon S. Donohoe, Gregg T. Beckham, and John E. McGeehan. Characterization and engineering of a two-enzyme system for plastics depolymerization. *Proc. Natl. Acad. Sci. U. S. A.*, 117(41):25476–25485, October 2020.
- [192] Lance A. Wollenberg, Jarod L. Kabulski, Matthew J. Powell, Jifeng Chen, Darcy R. Flora, Timothy S. Tracy, and Peter M. Gannett. The Use of Immobilized Cytochrome P450C9 in PMMA-Based Plug Flow Bioreactors for the Production of Drug Metabolites. *Appl. Biochem. Biotechnol.*, 172(3):1293–1306, February 2014.
- [193] Julie Ducharme and Karine Auclair. Use of bioconjugation with cytochrome P450 enzymes. *Biochim. Biophys. Acta, Proteins Proteomics*, 1866(1):32–51, January 2018.
- [194] An Su, Abhijit Shirke, Joshua Baik, Yi Zou, and Richard Gross. Immobilized cutinases: Preparation, solvent tolerance and thermal stability. *Enzyme Microb. Technol.*, 116:33–40, 2018.
- [195] Yinglu Cui, Yanchun Chen, Xinyue Liu, Saijun Dong, Yu’e Tian, Yuxin Qiao, Ruchira Mitra, Jing Han, Chunli Li, Xu Han, Weidong Liu, Quan Chen, Wangqing Wei, Xin Wang, Wenbin Du, Shuangyan Tang, Hua Xiang, Haiyan Liu, Yong Liang, Kendall N. Houk, and Bian Wu. Computational redesign of a petase for plastic biodegradation under ambient condition by the grape strategy. *ACS Catal.*, 0(0):1340–1350, 0.
- [196] Chi W. Pak, Martyna Kosno, Alex S. Holehouse, Shae B. Padrick, Anuradha Mittal, Rustam Ali, Ali A. Yunus, David R. Liu, Rohit V. Pappu, and Michael K. Rosen. Sequence determinants of intracellular phase separation by complex coacervation of a disordered protein. *Molecular Cell*, 63(1):72–85, jul 2016.
- [197] Shifan Mao, Quinn J. MacPherson, Steve S. He, Elyse Coletta, and Andrew J. Spakowitz. Impact of conformational and chemical correlations on microphase segregation in random copolymers. *Macromolecules*, 49(11):4358–4368, 2016.
- [198] Baofu Qiao, Felipe Jiménez-Ángeles, Trung Dac Nguyen, and Monica Olvera de la Cruz. Water follows polar and nonpolar protein surface domains. *Proc. Nat. Acad. Sci. U.S.A.*, 116(39):19274–19281, Sep 2019.
- [199] Sieun Kim, Hursh V. Sureka, A. Basak Kayitmazer, Gang Wang, James W. Swan, and Bradley D. Olsen. Effect of protein surface charge distribution on

- protein–polyelectrolyte complexation. *Biomacromolecules*, 21(8):3026–3037, 2020. PMID: 32672952.
- [200] Christopher DelRe, Brian Panganiban, Tim Li, Charley Huang, Monica Olvera de la Cruz, Patrick Dennis, and Ting Xu. Rational design of a synthetic peg-like polymer for protein stabilization. *Biophys. J.*, 112(3, Supplement 1):59a, 2017.
- [201] Adolfo B. Poma, Marek Cieplak, and Panagiotis E. Theodorakis. Combining the MARTINI and Structure-Based Coarse-Grained Approaches for the Molecular Dynamics Studies of Conformational Transitions in Proteins. *J. Chem. Theory Comput.*, 13(3):1366–1374, March 2017.
- [202] Minkyu Kim, Manos Gkikas, Aaron Huang, Jeon Woong Kang, Nisaraporn Suthiwangcharoen, Ramanathan Nagarajan, and Bradley D. Olsen. Enhanced activity and stability of organophosphorus hydrolase via interaction with an amphiphilic polymer. *Chem. Commun.*, 50:5345–5348, 2014.
- [203] David Blaha-Nelson, Dennis M. Krüger, Klaudia Szeler, Moshe Ben-David, and Shina Caroline Lynn Kamerlin. Active site hydrophobicity and the convergent evolution of paraoxonase activity in structurally divergent enzymes: The case of serum paraoxonase 1. *J. Am. Chem. Soc.*, 139(3):1155–1167, 2017. PMID: 28026940.
- [204] Wynand B. L. Alkema, Anne-Jan Dijkhuis, Erik de Vries, and Dick B. Janssen. The role of hydrophobic active-site residues in substrate specificity and acyl transfer activity of penicillin acylase. *Eur. J. Biochem.*, 269(8):2093–2100, 2002.
- [205] Phillip W. Snyder, Jasmin Mecinović, Demetri T. Moustakas, Samuel W. Thomas, Michael Harder, Eric T. Mack, Matthew R. Lockett, Annie Héroux, Woody Sherman, and George M. Whitesides. Mechanism of the hydrophobic effect in the biomolecular recognition of arylsulfonamides by carbonic anhydrase. *Proc. Natl. Acad. Sci. U. S. A.*, 108(44):17889–17894, 2011.
- [206] Shubhasish Goswami, Shibashis Das, and Supratim Datta. Understanding the role of residues around the active site tunnel towards generating a glucose-tolerant -glucosidase from *Agrobacterium tumefaciens* 5A. *Protein Eng., Des. Sel.*, 30(7):523–530, 07 2017.
- [207] Curt Waltmann, Roi Asor, Uri Raviv, and Monica Olvera de la Cruz. Assembly and stability of simian virus 40 polymorphs. *ACS Nano*, 14(4):4430–4443, 2020. PMID: 32208635.

- [208] Berk Hess, Carsten Kutzner, David van der Spoel, and Erik Lindahl. Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.*, 4(3):435–447, Mar 2008.
- [209] Jing Huang, Sarah Rauscher, Grzegorz Nawrocki, Ting Ran, Michael Feig, Bert L. de Groot, Helmut Grubmüller, and Alexander D. MacKerell. Charmm36m: An improved force field for folded and intrinsically disordered proteins. *Nat. Methods*, 14(1):71–73, Jan 2017.
- [210] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B*, 102(18):3586–3616, 1998. PMID: 24889800.
- [211] Shuichi Miyamoto and Peter A. Kollman. Settle: An analytical version of the shake and rattle algorithm for rigid water models. *Journal of Computational Chemistry*, 13(8):952–962, 1992.
- [212] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An $n\log(n)$ method for ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089–10092, 1993.
- [213] Ulrich Essmann, Lalith Perera, Max L. Berkowitz, Tom Darden, Hsing Lee, and Lee G. Pedersen. A smooth particle mesh ewald method. *The Journal of Chemical Physics*, 103(19):8577–8593, 1995.
- [214] Jumin Lee, Xi Cheng, Jason M. Swails, Min Sun Yeom, Peter K. Eastman, Justin A. Lemkul, Shuai Wei, Joshua Buckner, Jong Cheol Jeong, Yifei Qi, Sunhwan Jo, Vijay S. Pande, David A. Case, Charles L. Brooks, Alexander D. MacKerell, Jeffery B. Klauda, and Wonpil Im. Charmm-gui input generator for namd, gromacs, amber, openmm, and charmm/openmm simulations using the charmm36 additive force field. *Journal of Chemical Theory and Computation*, 12(1):405–413, 2016. PMID: 26631602.
- [215] Baofu Qiao, Luis Lopez, and Monica Olvera de la Cruz. “mirror”-like protein dimers stabilized by local heterogeneity at protein surfaces. *The Journal of Physical Chemistry B*, 123(18):3907–3915, 2019. PMID: 30998380.

- [216] Baofu Qiao and Monica Olvera de la Cruz. Enhanced binding of sars-cov-2 spike protein to receptor by distal polybasic cleavage sites. *ACS Nano*, 14(8):10616–10623, 2020. PMID: 32806067.
- [217] Tao Jiang, Aaron Hall, Marco Eres, Zahra Hemmatian, Baofu Qiao, Yun Zhou, Zhiyuan Ruan, Andrew D. Couse, William T. Heller, Haiyan Huang, Monica Olvera de la Cruz, Marco Rolandi, and Ting Xu. Single-chain heteropolymers transport protons selectively and rapidly. *Nature*, 577(7789):216–220, Jan 2020.
- [218] Luca Monticelli, Senthil K. Kandasamy, Xavier Periole, Ronald G. Larson, D. Peter Tieleman, and Siewert-Jan Marrink. The martini coarse-grained force field: Extension to proteins. *J. Chem. Theory Comput.*, 4(5):819–834, 2008. PMID: 26621095.
- [219] Paulo C. T. Souza, Sebastian Thallmair, Siewert J. Marrink, and Raúl Mera-Adasme. An allosteric pathway in copper, zinc superoxide dismutase unravels the molecular mechanism of the g93a amyotrophic lateral sclerosis-linked mutation. *The Journal of Physical Chemistry Letters*, 10(24):7740–7744, Dec 2019.
- [220] Sebastian Thallmair, Petteri A. Vainikka, and Siewert J. Marrink. Lipid fingerprints and cofactor dynamics of light-harvesting complex ii in different membranes. *Biophys. J.*, 116(8):1446–1455, 2019.
- [221] Djurre H. de Jong, Svetlana Baoukina, Helgi I. Ingólfsson, and Siewert J. Marrink. Martini straight: Boosting performance using a shorter cutoff and gpu. *Computer Physics Communications*, 199:1–7, 2016.
- [222] Cheryl A. Kerfeld, Clement Aussignargues, Jan Zarzycki, Fei Cai, and Markus Sutter. Bacterial microcompartments. *Nat. Rev. Microbiol.*, 16(5):277–290, May 2018.
- [223] Seth D Axen, Onur Erbilgin, and Cheryl A Kerfeld. A taxonomy of bacterial microcompartment loci constructed by a novel scoring method. *PLoS computational biology*, 10(10):e1003898, 2014.
- [224] Markus Sutter, Matthew R Melnicki, Frederik Schulz, Tanja Woyke, and Cheryl A Kerfeld. A catalog of the diversity and ubiquity of bacterial microcompartments. *Nature communications*, 12(1):3809, 2021.
- [225] George S Espie and Matthew S Kimber. Carboxysomes: cyanobacterial rubisco comes in small packages. *Photosynthesis Research*, 109:7–20, 2011.
- [226] Aaron Kaplan and Leonora Reinhold. Co₂ concentrating mechanisms in photosynthetic microorganisms. *Annual review of plant biology*, 50(1):539–570, 1999.

- [227] Katie L. Stewart, Andrew M. Stewart, and Thomas A. Bobik. Prokaryotic organelles: Bacterial microcompartments in *ij.e. coli**ij* and *ij.salmonella**ij*. *EcoSal Plus*, 9(1), 2020.
- [228] Chiranjit Chowdhury, Sharmistha Sinha, Sunny Chun, Todd O Yeates, and Thomas A Bobik. Diverse bacterial microcompartment organelles. *Microbiology and Molecular Biology Reviews*, 78(3):438–468, 2014.
- [229] Michael B Prentice. Bacterial microcompartments and their role in pathogenicity. *Current Opinion in Microbiology*, 63:19–28, 2021.
- [230] Zhe Zeng, Eddy J Smid, Sjeff Boeren, Richard A Notebaart, and Tjakko Abee. Bacterial microcompartment-dependent 1, 2-propanediol utilization stimulates anaerobic growth of *listeria monocytogenes* egde. *Frontiers in microbiology*, 10:2660, 2019.
- [231] Christopher M Jakobson, Danielle Tullman-Ercek, and Niall M Mangan. Spatially organizing biochemistry: choosing a strategy to translate synthetic biology to the factory. *Scientific reports*, 8(1):8196, 2018.
- [232] Alex P. Lundin, Katie L. Stewart, Andrew M. Stewart, Taylor I. Herring, Chiranjit Chowdhury, and Thomas A. Bobik. Genetic characterization of a glyceryl radical microcompartment used for 1,2-propanediol fermentation by uropathogenic *escherichia coli* cft073. *Journal of Bacteriology*, 202(9):e00017–20, 2020.
- [233] Taylor I. Herring, Tiffany N. Harris, Chiranjit Chowdhury, Sujit Kumar Mohanty, and Thomas A. Bobik. A bacterial microcompartment is used for choline fermentation by *escherichia coli* 536. *Journal of Bacteriology*, 200(10):e00764–17, 2018.
- [234] Jan Zarzycki, Markus Sutter, Niña Socorro Cortina, Tobias J Erb, and Cheryl A Kerfeld. In vitro characterization and concerted function of three core enzymes of a glyceryl radical enzyme-associated bacterial microcompartment. *Scientific reports*, 7(1):42757, 2017.
- [235] 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.
- [236] Francisco Zaera. Shape-controlled nanostructures in heterogeneous catalysis. *ChemSusChem*, 6(10):1797–1820, 2013.
- [237] Anirban Sen Gupta. Role of particle size, shape, and stiffness in design of intravascular drug delivery systems: insights from computations, experiments, and nature.

- Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology*, 8(2):255–270, 2016.
- [238] Thomas A Bobik, Gregory D Havemann, Robert J Busch, Donna S Williams, and Henry C Aldrich. The propanediol utilization (pdu) operon of salmonella 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.
- [239] Edith M Sampson and Thomas A Bobik. Microcompartments for b12-dependent 1, 2-propanediol degradation provide protection from dna and cellular damage by a reactive metabolic intermediate. *Journal of bacteriology*, 190(8):2966–2971, 2008.
- [240] Nolan W. Kennedy, Jasmine M. Hershewe, Taylor M. Nichols, Eric W. Roth, Charlene D. Wilke, Carolyn E. Mills, Michael C. Jewett, and Danielle Tullman-Ereck. Apparent size and morphology of bacterial microcompartments varies with technique. *PLoS One*, 15(3):1–17, 03 2020.
- [241] Joshua B Parsons, Stefanie Frank, David Bhella, Mingzhi Liang, Michael B Prentice, Daniel P Mulvihill, and Martin J Warren. Synthesis of empty bacterial microcompartments, directed organelle protein incorporation, and evidence of filament-associated organelle movement. *Molecular cell*, 38(2):305–315, 2010.
- [242] Shouqiang Cheng, Sharmistha Sinha, Chenguang Fan, Yu Liu, and Thomas A Bobik. Genetic analysis of the protein shell of the microcompartments involved in coenzyme b12-dependent 1, 2-propanediol degradation by salmonella. *Journal of bacteriology*, 193(6):1385–1392, 2011.
- [243] Shiho Tanaka, Michael R Sawaya, and Todd O Yeates. Structure and mechanisms of a protein-based organelle in escherichia coli. *Science*, 327(5961):81–84, 2010.
- [244] 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.
- [245] Shiho Tanaka, Cheryl A. Kerfeld, Michael R. Sawaya, Fei Cai, Sabine Heinhorst, Gordon C. Cannon, and Todd O. Yeates. Atomic-level models of the bacterial carboxysome shell. *Science*, 319(5866):1083–1086, 2008.
- [246] Allan Pang, Martin J Warren, and Richard W Pickersgill. Structure of pdut, a trimeric bacterial microcompartment protein with a 4fe–4s cluster-binding site. *Acta Crystallographica Section D: Biological Crystallography*, 67(2):91–96, 2011.

- [247] Christopher S Crowley, Michael R Sawaya, Thomas A Bobik, and Todd O Yeates. Structure of the pdu shell protein from the pdu microcompartment of salmonella. *Structure*, 16(9):1324–1332, 2008.
- [248] 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.
- [249] Cheryl A Kerfeld, Michael R Sawaya, Shiho Tanaka, Chau V Nguyen, Martin Phillips, Morgan Beeby, and Todd O Yeates. Protein structures forming the shell of primitive bacterial organelles. *Science*, 309(5736):936–938, 2005.
- [250] Markus Sutter, Steven C Wilson, Samuel Deutsch, and Cheryl A Kerfeld. Two new high-resolution crystal structures of carboxysome pentamer proteins reveal high structural conservation of ccml orthologs among distantly related cyanobacterial species. *Photosynthesis research*, 118:9–16, 2013.
- [251] Evan Mallette and Matthew S. Kimber. A complete structural inventory of the mycobacterial microcompartment shell proteins constrains models of global architecture and transport. *Journal of Biological Chemistry*, 292(4):1197–1210, 2017.
- [252] Thomas J Keeling, Bożena Samborska, Ryan W Demers, and Matthew S Kimber. Interactions and structural variability of β -carboxysomal shell protein ccml. *Photosynthesis research*, 121:125–133, 2014.
- [253] Alexander G McFarland, Nolan W Kennedy, Carolyn E Mills, Danielle Tullman-Ercek, Curtis Huttenhower, and Erica M Hartmann. Density-based binning of gene clusters to infer function or evolutionary history using genegrouper. *Bioinformatics*, 38(3):612–620, Jan 2022.
- [254] Jeffrey C. Cameron, Steven C. Wilson, Susan L. Bernstein, and Cheryl A. Kerfeld. Biogenesis of a bacterial organelle: The carboxysome assembly pathway. *Cell*, 155(5):1131–1140, 2013.
- [255] Fei Cai, Balaraj B. Menon, Gordon C. Cannon, Kenneth J. Curry, Jessup M. Shively, and Sabine Heinhorst. The pentameric vertex proteins are necessary for the icosahedral carboxysome shell to function as a co₂ leakage barrier. *PLoS One*, 4(10):1–9, 10 2009.
- [256] Christopher M. Jakobson, Danielle Tullman-Ercek, Marilyn F. Slininger, and Niall M. Mangan. A systems-level model reveals that 1,2-propanediol utilization microcompartments enhance pathway flux through intermediate sequestration. *PLoS Comput. Biol.*, 13(5):e1005525, May 2017.

- [257] Chenguang Fan and Thomas A Bobik. The n-terminal region of the medium subunit (pdud) packages adenosylcobalamin-dependent diol dehydratase (pdudce) into the pdu microcompartment. *Journal of bacteriology*, 193(20):5623–5628, 2011.
- [258] Edward Y Kim and Danielle Tullman-Ercek. A rapid flow cytometry assay for the relative quantification of protein encapsulation into bacterial microcompartments. *Biotechnology Journal*, 9(3):348–354, 2014.
- [259] Taylor M. Nichols, Nolan W. Kennedy, and Danielle Tullman-Ercek. Chapter seven - cargo encapsulation in bacterial microcompartments: Methods and analysis. In Claudia Schmidt-Dannert and Maureen B. Quin, editors, *Metabolons and Supramolecular Enzyme Assemblies*, volume 617 of *Methods in Enzymology*, pages 155 – 186. Academic Press, 2019.
- [260] Taylor M Nichols, Nolan W Kennedy, and Danielle Tullman-Ercek. A genomic integration platform for heterologous cargo encapsulation in 1, 2-propanediol utilization bacterial microcompartments. *Biochemical Engineering Journal*, 156:107496, 2020.
- [261] Ismail Uddin, Stefanie Frank, Martin J. Warren, and Richard W. Pickersgill. A generic self-assembly process in microcompartments and synthetic protein nanotubes. *Small*, 14(19):1704020, 2018.
- [262] Basil J. Greber, Markus Sutter, and Cheryl A. Kerfeld. The plasticity of molecular interactions governs bacterial microcompartment shell assembly. *Structure*, 27(5):749–763.e4, 2019.
- [263] Edward Y Kim, Christopher M Jakobson, and Danielle Tullman-Ercek. Engineering transcriptional regulation to control pdu microcompartment formation. *PloS one*, 9(11):e113814, 2014.
- [264] Ping Chen, Dan I Andersson, and John R Roth. The control region of the pdu/cob regulon in salmonella typhimurium. *Journal of bacteriology*, 176(17):5474–5482, 1994.
- [265] MICHELLE R Rondon and JORGE C Escalante-Semerena. The poc locus is required for 1, 2-propanediol-dependent transcription of the cobalamin biosynthetic (cob) and propanediol utilization (pdu) genes of salmonella typhimurium. *Journal of bacteriology*, 174(7):2267–2272, 1992.
- [266] 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 Synth. Biol.*, 6(10):1880–1891, October 2017.

- [267] Allen W Tsang, Alexander R Horswill, and Jorge C Escalante-Semerena. Studies of regulation of expression of the propionate (prpbcd) operon provide insights into how salmonella typhimurium lt2 integrates its 1, 2-propanediol and propionate catabolic pathways. *Journal of bacteriology*, 180(24):6511–6518, 1998.
- [268] Mengru Yang, Deborah M. Simpson, Nicolas Wenner, Philip Brownridge, Victoria M. Harman, Jay C. D. Hinton, Robert J. Beynon, and Lu-Ning Liu. Decoding the stoichiometric composition and organisation of bacterial metabolosomes. *Nat. Commun.*, 11(1):1976, December 2020.
- [269] Markus Sutter, Sean McGuire, Bryan Ferlez, and Cheryl A. Kerfeld. Structural characterization of a synthetic tandem-domain bacterial microcompartment shell protein capable of forming icosahedral shell assemblies. *ACS Synth. Biol.*, 8(4):668–674, 2019. PMID: 30901520.
- [270] Andrew Hagen, Markus Sutter, Nancy Sloan, and Cheryl A. Kerfeld. Programmed loading and rapid purification of engineered bacterial microcompartment shells. *Nat. Commun.*, 9(1):2881, Jul 2018.
- [271] Jonathan K. Lassila, Susan L. Bernstein, James N. Kinney, Seth D. Axen, and Cheryl A. Kerfeld. Assembly of robust bacterial microcompartment shells using building blocks from an organelle of unknown function. *J. Mol. Biol.*, 426(11):2217–2228, 2014.
- [272] Dongseon Lee, Juyong Lee, and Chaok Seok. What stabilizes close arginine pairing in proteins? *Phys. Chem. Chem. Phys.*, 15:5844–5853, 2013.
- [273] Isabel Huber, David J Palmer, Kira N Ludwig, Ian R Brown, Martin J Warren, and Julia Frunzke. Construction of recombinant pdu metabolosome shells for small molecule production in corynebacterium glutamicum. *ACS synthetic biology*, 6(11):2145–2156, 2017.
- [274] 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 Lett.*, 18(11):7030–7037, 2018. PMID: 30346795.
- [275] Bryan Ferlez, Markus Sutter, and Cheryl A Kerfeld. A designed bacterial microcompartment shell with tunable composition and precision cargo loading. *Metabolic engineering*, 54:286–291, 2019.
- [276] Matthew J Lee, Judith Mantell, Ian R Brown, Jordan M Fletcher, Paul Verkade, Richard W Pickersgill, Derek N Woolfson, Stefanie Frank, and Martin J Warren. De

- novo targeting to the cytoplasmic and luminal side of bacterial microcompartments. *Nature communications*, 9(1):3413, 2018.
- [277] Lawrence A Kelley, Stefans Mezulis, Christopher M Yates, Mark N Wass, and Michael J E Sternberg. The phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, 10(6):845–858, Jun 2015.
- [278] Eric F. Pettersen, Thomas D. Goddard, Conrad C. Huang, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng, and Thomas E. Ferrin. Ucsf chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612, 2004.
- [279] Elaine C Meng, Eric F Pettersen, Gregory S Couch, Conrad C Huang, and Thomas E Ferrin. Tools for integrated sequence-structure analysis with ucsf chimera. *BMC Bioinformatics*, 7(1):339, Dec 2006.
- [280] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C Smith, Berk Hess, and Erik Lindahl. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1:19–25, 2015.
- [281] Baofu Qiao, Felipe Jiménez-Ángeles, Trung Dac Nguyen, and Monica Olvera de la Cruz. Water follows polar and nonpolar protein surface domains. *Proc. Nat. Acad. Sci. U.S.A.*, 116(39):19274–19281, September 2019.
- [282] Siyu Li, Polly Roy, Alex Travesset, and Roya Zandi. Why large icosahedral viruses need scaffolding proteins. *Proc. Natl. Acad. Sci.*, 115(43):10971–10976, 2018.
- [283] Benjamin Schwarz and Trevor Douglas. Development of virus-like particles for diagnostic and prophylactic biomedical applications. *Wiley Interdiscip. Rev.: Nanomed. Nanobiotechnol.*, 7(5):722–735, 2015.
- [284] Amy M. Wen, Rudolf Podgornik, Giuseppe Strangi, and Nicole F. Steinmetz. Photonics and plasmonics go viral: Self-assembly of hierarchical metamaterials. *Atti Accad. Naz. Lincei, Cl. Sci. Fis., Mat. Nat., Rend.*, 26(2):129–141, Aug 2015.
- [285] Cheng Zeng, Mercedes Hernando-Pérez, Bogdan Dragnea, Xiang Ma, Paul van der Schoot, and Roya Zandi. Contact mechanics of a small icosahedral virus. *Phys. Rev. Lett.*, 119:038102, Jul 2017.
- [286] Xinlei Huang, Barry D. Stein, Hu Cheng, Andrey Malyutin, Irina B. Tsvetkova, David V. Baxter, Nicholas B. Remmes, Jeanmarie Verchot, Cheng Kao, Lyudmila M. Bronstein, and Bogdan Dragnea. Magnetic virus-like nanoparticles in n.

- benthamiana plants: A new paradigm for environmental and agronomic biotechnological research. *ACS Nano*, 5(5):4037–4045, 2011. PMID: 21452886.
- [287] Dahyun Oh, Jifa Qi, Binghong Han, Geran Zhang, Thomas J. Carney, Jacqueline Ohmura, Yong Zhang, Yang Shao-Horn, and Angela M. Belcher. M13 virus-directed synthesis of nanostructured metal oxides for lithium–oxygen batteries. *Nano Lett.*, 14(8):4837–4845, 2014. PMID: 25058851.
- [288] Yun Jung Lee, Hyunjung Yi, Woo-Jae Kim, Kisuk Kang, Dong Soo Yun, Michael S. Strano, Gerbrand Ceder, and Angela M. Belcher. Fabricating genetically engineered high-power lithium-ion batteries using multiple virus genes. *Science*, 324(5930):1051–1055, 2009.
- [289] Joshua W. Wilkerson, Seung-Ook Yang, Parker J. Funk, Steven K. Stanley, and Bradley C. Bundy. Nanoreactors: Strategies to encapsulate enzyme biocatalysts in virus-like particles. *New Biotechnol.*, 44:59 – 63, 2018.
- [290] Luminita Eid, Zohar Bromberg, Mahmoud Abd EL-Latif, Evelyn Zeira, Ariella Oppenheim, and Yoram G. Weiss. Simian virus 40 vectors for pulmonary gene therapy. *Respir. Res.*, 8(1):74, Oct 2007.
- [291] Masaaki Kawano, Masanori Matsui, and Hiroshi Handa. Sv40 virus-like particles as an effective delivery system and its application to a vaccine carrier. *Expert Rev. Vaccines*, 12(2):199–210, 2013.
- [292] B. H. Sweet and M. R. Hilleman. The vacuolating virus, s.v.40. *Proc. Soc. Exp. Biol. Med.*, 105(2):420–427, 1960.
- [293] Eric Raspaud, M Olvera De La Cruz, J-L Sikorav, and Françoise Livolant. Precipitation of dna by polyamines: a polyelectrolyte behavior. *Biophys. J.*, 74(1):381–393, 1998.
- [294] Roi Asor, Daniel Khaykelson, Orly Ben-nun Shaul, Ariella Oppenheim, and Uri Raviv. Effect of calcium ions and disulfide bonds on swelling of virus particles. *ACS Omega*, 4(1):58–64, 2019.
- [295] Ruben D. Cadena-Nava, Mauricio Comas-Garcia, Rees F. Garmann, A. L. N. Rao, Charles M. Knobler, and William M. Gelbart. Self-assembly of viral capsid protein and rna molecules of different sizes: Requirement for a specific high protein/rna mass ratio. *J. Virol.*, 86(6):3318–3326, 2012.
- [296] Feng Li, Zhi-Ping Zhang, Jun Peng, Zong-Qiang Cui, Dai-Wen Pang, Ke Li, Hong-Ping Wei, Ya-Feng Zhou, Ji-Kai Wen, and Xian-En Zhang. Imaging viral behavior

- in mammalian cells with self-assembled capsid–quantum-dot hybrid particles. *Small*, 5(6):718–726.
- [297] Tingjuan Wang, Zhiping Zhang, Ding Gao, Feng Li, Hongping Wei, Xiaosheng Liang, Zongqiang Cui, and Xian-En Zhang. Encapsulation of gold nanoparticles by simian virus 40 capsids. *Nanoscale*, 3:4275–4282, 2011.
- [298] Teruya Enomoto, Maskai Kawano, Hajime Fukuda, Wataru Sawada, Takasam Inoue, Kok Chee Haw, Yoshinori Kita, Satoshi Sakamoto, Yuki Yamaguchi, Takeshi Imai, Mamoru Hatakeyama, Shigeyoshi Saito, Ardasha Sandhu, Mansanori Matsui, and Ichoi Aoki. Viral protein-coating of magnetic nanoparticles using simian virus 40 vp1. *J. Biotechnol.*, 167(1):8 – 15, 2013.
- [299] Masaaki Kawano, Koji Doi, Hajime Fukuda, Yoshinori Kita, Kensuke Imai, Takamasa Inoue, Teruya Enomoto, Masanori Matsui, Mamoru Hatakeyama, Yuki Yamaguchi, and Hiroshi Handa. Sv40 vp1 major capsid protein in its self-assembled form allows vp1 pentamers to coat various types of artificial beads *In Vitro* regardless of their sizes and shapes. *Biotechnology Reports*, 5:105 – 111, 2015.
- [300] T. Tarnai, Z. Gáspár, and L. Szalai. Pentagon packing models for all-pentamer virus structures. *Biophys. J.*, 69(2):612 – 618, 1995.
- [301] Orly Ben nun Shaul, Hagit Bronfeld, Dan Reshef, Ora Schueler-Furman, and Ariella Oppenheim. The sv40 capsid is stabilized by a conserved pentapeptide hinge of the major capsid protein vp1. *J. Mol. Biol.*, 386(5):1382 – 1391, 2009.
- [302] Vladimir A. Belyi and M. Muthukumar. Electrostatic origin of the genome packing in viruses. *Proc. Natl. Acad. Sci. U.S.A.* ., 103(46):17174–17178, 2006.
- [303] Peggy P. Li, Akira Nakanishi, Dorothy Shum, Peter C.-K. Sun, Adler M. Salazar, Cesar F. Fernandez, Sze-Wai Chan, and Harumi Kasamatsu. Simian virus 40 vp1 dna-binding domain is functionally separable from the overlapping nuclear localization signal and is required for effective virion formation and full viability. *J. Virol.*, 75(16):7321–7329, 2001.
- [304] Vered Roitman-Shemer, Jitka Stokrova, Jitka Forstova, and Ariella Oppenheim. Assemblages of simian virus 40 capsid proteins and viral dna visualized by electron microscopy. *Biochem. Biophys. Res. Commun.*, 353(2):424 – 430, 2007.
- [305] Michael F. Hagan. Modeling viral capsid assembly. pages 1–68, 2014.
- [306] R. F. Bruinsma, M. Comas-Garcia, R. F. Garmann, and A. Y. Grosberg. Equilibrium self-assembly of small rna viruses. *Phys. Rev. E*, 93:032405, Mar 2016.

- [307] Roya Zandi, David Reguera, Robijn F. Bruinsma, William M. Gelbart, and Joseph Rudnick. Origin of icosahedral symmetry in viruses. *Proc. Natl. Acad. Sci. U.S.A.*, 101(44):15556–15560, 2004.
- [308] Hung D. Nguyen, Vijay S. Reddy, and Charles L. Brooks III. Invariant polymorphism in virus capsid assembly. *J. Am. Chem. Soc.*, 131(7):2606–2614, 2009. PMID: 19199626.
- [309] Paul F. McMillan, David C. Clary, and David J. Wales. The energy landscape as a unifying theme in molecular science. *Philos. Trans. R. Soc., A*, 363(1827):357–377, 2005.
- [310] Szilard N. Fejer, Tim R. James, Javier Hernández-Rojas, and David J. Wales. Energy landscapes for shells assembled from pentagonal and hexagonal pyramids. *Phys. Chem. Chem. Phys.*, 11:2098–2104, 2009.
- [311] Iain G Johnston, Ard A Louis, and Jonathan P K Doye. Modelling the self-assembly of virus capsids. *J. Phys.: Condens. Matter*, 22(10):104101, feb 2010.
- [312] Donald L.D Caspar. Virus structure puzzle solved. *Curr. Biol.*, 2(4):169 – 171, 1992.
- [313] Roi Asor, Daniel Khaykelson, Orly Ben-nun Shaul, Yael Levi-Kalishman, Ariella Oppenheim, and Uri Raviv. ph stability and disassembly mechanism of wild-type simian virus 40. *Soft Matter*, 16:2803–2814, 2020.
- [314] Roi Asor, Lisa Selzer, Christopher John Schlicksup, Zhongchao Zhao, Adam Zlotnick, and Uri Raviv. Assembly reactions of hepatitis b capsid protein into capsid nanoparticles follow a narrow path through a complex reaction landscape. *ACS Nano*, 13(7):7610–7626, 2019. PMID: 31173689.
- [315] Jason D. Perlmutter, Matthew R. Perkett, and Michael F. Hagan. Pathways for virus assembly around nucleic acids. *J. Mol. Biol.*, 426(18):3148 – 3165, 2014.
- [316] Mariska G.M. van Rosmalen, Chenglei Li, Adam Zlotnick, Gijs J.L. Wuite, and Wouter H. Roos. Effect of dsdna on the assembly pathway and mechanical strength of sv40 vp1 virus-like particles. *Biophys. J.*, 115(9):1656 – 1665, 2018.
- [317] Gadiel Saper, Stanislav Kler, Roi Asor, Ariella Oppenheim, Uri Raviv, and Daniel Harries. Effect of Capsid Confinement on the Chromatin Organization of the SV40 Minichromosome. *Nucleic Acids Res.*, 41(3):1569–1580, 12 2012.
- [318] Serban Ilca, Xiaoyu Sun, Kamel El Omari, Abhay Kotecha, Felix de Haas, Frank DiMaio, Jonathan M Grimes, David I. Stuart, Minna Marjetta Poranen, and Juha

- Tapio Huiskonen. Multiple liquid crystalline geometries of highly compacted nucleic acid in a dsrna virus. *Nature*, 2019.
- [319] Liang Tang, Karyn N Johnson, L. Andrew Ball, Tianwei Lin, Mark Yeager, and John E. Johnson. The structure of pariacoto virus reveals a dodecahedral cage of duplex rna. *Nat. Struct. Biol.*, 8:77–83, 2000.
- [320] Stanislav Kler, Roi Asor, Chenglei Li, Avi Ginsburg, Daniel Harries, Ariella Oppenheim, Adam Zlotnick, and Uri Raviv. Rna encapsidation by sv40-derived nanoparticles follows a rapid two-state mechanism. *J. Am. Chem. Soc.*, 134(21):8823–8830, 2012.
- [321] S Alexander, PM Chaikin, P Grant, GJ Morales, P Pincus, and D Hone. Charge Renormalization, Osmotic-Pressure, and Bulk Modulus of Colloidal Crystals - Theory. *Journal of Chemical Physics*, 80(11):5776–5781, 1984.
- [322] P. González-Mozuelos and M. Olvera de la Cruz. Ion condensation in salt-free dilute polyelectrolyte solutions. *J. Chem. Phys.*, 103(8):3145–3157, 1995.
- [323] M. Olvera de la Cruz. Electrostatic control of self-organization: The role of charge gradients in heterogeneous media. *Soft Matter*, 4:1735–1739, 2008.
- [324] Francisco J. Solis and Monica Olvera de la Cruz. Collapse of flexible polyelectrolytes in multivalent salt solutions. *J. Chem. Phys.*, 112(4):2030–2035, 2000.
- [325] Alpha A. Lee, Carla S. Perez-Martinez, Alexander M. Smith, and Susan Perkin. Scaling analysis of the screening length in concentrated electrolytes. *Phys. Rev. Lett.*, 119:026002, Jul 2017.
- [326] Alpha A. Lee, Carla S. Perez-Martinez, Alexander M. Smith, and Susan Perkin. Underscreening in concentrated electrolytes. *Faraday Discuss.*, 199:239–259, 2017.
- [327] Rikkert Nap, Anze Losdorfer Bozic, Igal G Szleifer, and Rudolf Podgornik. The role of solution conditions in the bacteriophage pp7 capsid charge regulation. *Biophys. J.*, 107(8):1970–1979, 10 2014.
- [328] Yaohua Li, Martin Girard, Meng Shen, Jaime Andres Millan, and Monica Olvera de la Cruz. Strong attractions and repulsions mediated by monovalent salts. *Proc. Natl. Acad. Sci. U.S.A.* ., 114(45):11838–11843, 2017.
- [329] Markus Sutter, Thomas G. Laughlin, Nancy B. Sloan, Daniel Serwas, Karen M. Davies, and Cheryl A. Kerfeld. Structure of a Synthetic -Carboxysome Shell. *Plant Physiol.*, 181(3):1050–1058, 09 2019.

- [330] Onur Erbilgin, Kent L. McDonald, and Cheryl A. Kerfeld. Characterization of a planctomycetal organelle: a novel bacterial microcompartment for the aerobic degradation of plant saccharides. *Appl. Environ. Microbiol.*, 80(7):2193–2205, 2014.
- [331] Yaqi Sun, Adam J. M. Wollman, Fang Huang, Mark C. Leake, and Lu-Ning Liu. Single-Organelle Quantification Reveals Stoichiometric and Structural Variability of Carboxysomes Dependent on the Environment. *Plant Cell*, 31(7):1648–1664, 05 2019.
- [332] Lu-Ning Liu, Mengru Yang, Yaqi Sun, and Jing Yang. Protein stoichiometry, structural plasticity and regulation of bacterial microcompartments. *Curr. Opin. Microbiol.*, 63:133–141, 2021.
- [333] Chiranjit Chowdhury, Sunny Chun, Michael R. Sawaya, Todd O. Yeates, and Thomas A Bobik. The function of the pduj microcompartment shell protein is determined by the genomic position of its encoding gene. *Mol. Microbiol.*, 101(5):770–783, 2016.
- [334] Hanna J. Wagner, Charlotte C. Capitain, Karsten Richter, Michelle Nessling, and Jörg Mampel. Engineering bacterial microcompartments with heterologous enzyme cargos. *Eng. Life Sci.*, 17(1):36–46, 2017.
- [335] Siyu Li, Daniel A. Matoz-Fernandez, and Monica Olvera de la Cruz. Effect of mechanical properties on multicomponent shell patterning. *ACS Nano*, 15(9):14804–14812, 2021. PMID: 34402621.
- [336] Rastko Sknepnek, Graziano Vernizzi, and Monica Olvera de la Cruz. Buckling of multicomponent elastic shells with line tension. *Soft Matter*, 8:636–644, 2012.
- [337] Graziano Vernizzi, Rastko Sknepnek, and Monica Olvera de la Cruz. Platonic and archimedean geometries in multicomponent elastic membranes. *Proc. Natl. Acad. Sci. U. S. A.*, 108(11):4292–4296, 2011.
- [338] Daniel S. Trettel, William Resager, Beatrix M. Ueberheide, Conor C. Jenkins, and Wade C. Winkler. Chemical probing provides insight into the native assembly state of a bacterial microcompartment. *Structure*, 2022.
- [339] Matthias J. Mayer, Rokas Juodeikis, Ian R. Brown, Stefanie Frank, David J. Palmer, Evelyne Deery, David M. Beal, Wei-Feng Xue, and Martin J. Warren. Effect of bio-engineering on size, shape, composition and rigidity of bacterial microcompartments. *Sci. Rep.*, 6(1):36899, December 2016.

- [340] C. Raul Gonzalez-Esquer, Sarah E. Newnham, and Cheryl A. Kerfeld. Bacterial microcompartments as metabolic modules for plant synthetic biology. *Plant J.*, 87(1):66–75, 2016.
- [341] Jason D Perlmutter, Farzaneh Mohajerani, and Michael F Hagan. Many-molecule encapsulation by an icosahedral shell. *eLife*, 5:e14078, May 2016.
- [342] Michael F Hagan and Roya Zandi. Recent advances in coarse-grained modeling of virus assembly. *Curr. Opin. Virol.*, 18:36 – 43, 2016.
- [343] Roi Asor, Christopher John Schlicksup, Zhongchao Zhao, Adam Zlotnick, and Uri Raviv. Rapidly forming early intermediate structures dictate the pathway of capsid assembly. *J. Am. Chem. Soc.*, 142(17):7868–7882, 2020. PMID: 32233479.
- [344] Patricia Lam, Neetu M. Gulati, Phoebe L. Stewart, Ruth A. Keri, and Nicole F. Steinmetz. Bioengineering of tobacco mosaic virus to create a non-infectious positive control for ebola diagnostic assays. *Sci. Rep.*, 6(1):23803, Mar 2016.
- [345] Jodi A Hadden, Juan R Perilla, Christopher John Schlicksup, Balasubramanian Venkatakrisnan, Adam Zlotnick, and Klaus Schulten. All-atom molecular dynamics of the hbv capsid reveals insights into biological function and cryo-em resolution limits. *eLife*, 7:e32478, apr 2018.
- [346] Brent P. Lehman, Chiranjit Chowdhury, and Thomas A. Bobik. The n terminus of the pduB protein binds the protein shell of the pdu microcompartment to its enzymatic core. *J. Bacteriol.*, 199(8), Apr 2017.
- [347] Adam Zlotnick, J. Zachary Porterfield, and Joseph Che-Yen Wang. To build a virus on a nucleic acid substrate. *Biophys. J.*, 104(7):1595–1604, 2013.
- [348] Mariska GM van Rosmalen, Chenglei Li, Adam Zlotnick, Gijs JL Wuite, and Wouter H Roos. Effect of dsdna on the assembly pathway and mechanical strength of sv40 vp1 virus-like particles. *Biophys. J.*, 115(9):1656–1665, 2018.
- [349] Wenjing Zhang, Qingyan Jia, Yibo Teng, Mengsi Yang, Hui Zhang, Xian-En Zhang, Pengfei Wang, Jiechao Ge, Sheng Cao, and Feng Li. An ultrastable virus-like particle with a carbon dot core and expanded sequence plasticity. *Small*, n/a(n/a):2101717.
- [350] Lye Siang Lee, Nicholas Brunk, Daniel G. Haywood, David Keifer, Elizabeth Pierson, Panagiotis Kondylis, Joseph Che-Yen Wang, Stephen C. Jacobson, Martin F. Jarrold, and Adam Zlotnick. A molecular breadboard: Removal and replacement of subunits in a hepatitis b virus capsid. *Protein Sci.*, 26(11):2170–2180, 2017.

- [351] Markus Sutter, Matthew Faulkner, Clément Aussignargues, Bradley C. Paasch, Steve Barrett, Cheryl A. Kerfeld, and Lu-Ning Liu. Visualization of bacterial microcompartment facet assembly using high-speed atomic force microscopy. *Nano Lett.*, 16(3):1590–1595, 2016. PMID: 26617073.
- [352] Clément Aussignargues, Bradley C Paasch, Raul Gonzalez-Esquer, Onur Erbilgin, and Cheryl A Kerfeld. Bacterial microcompartment assembly: The key role of encapsulation peptides. *Communicative & Integrative Biology*, 8(3):e1039755, 2015. PMID: 26478774.
- [353] Curt Waltmann, Nathan Horst, and Alex Travesset. Capping ligand vortices as “atomic orbitals” in nanocrystal self-assembly. *ACS Nano*, 11(11):11273–11282, 2017. PMID: 29077382.
- [354] Curt Waltmann, Nathan Horst, and Alex Travesset. Potential of mean force for two nanocrystals: Core geometry and size, hydrocarbon unsaturation, and universality with respect to the force field. *J. Chem. Phys.*, 149(3):034109, 2018.
- [355] Luis F. Garcia-Alles, Miguel Fuentes-Cabrera, Gilles Truan, and David Reguera. Inferring assembly-curving trends of bacterial micro-compartment shell hexamers from crystal structure arrangements. *PLoS Computational Biology*, 19(4):e1011038, Apr 2023.