Extending coarse-grained modeling methods to improve reliability and yield of designed protein nanocages

Susan Curry Kleinfelter

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2026

Committee:

Neil King

Charles Asbury

Frank DiMaio

Kelly K. Lee

Abhinav Nath

Program Authorized to Offer Degree:

Department of Biochemistry

University of Washington

Abstract

Extending coarse-grained modeling methods to improve

reliability and yield of designed protein nanocages

Susan Curry Kleinfelter

Chair of the Supervisory Committee:

Neil King

Biochemistry

The King lab designs protein nanocages as vaccine platforms and for other downstream applications. In order to characterize and compare vaccine candidates, we need the nanocage assembly process to be robust and reliable. Our academic and industry partners require uniform production and scalability. However, protein cage assembly is a complex process and often fails for unknown reasons. Unpredictable yield and polydispersity, two of the most common failure modes, may be caused by kinetic factors such as trapping in metastable states or unexpected energy barriers. In the past we have focused on improving interface affinity to make cages more stable, without addressing kinetics or assembly dynamics. Expanding our protein design toolkit to include coarse-grained molecular dynamics (CGMD) and Markov state modeling (MSM) will improve our understanding of the geometric and kinetic factors leading to poor assembly. These insights will guide adjustments to individual design campaigns for the purpose of improving nanocage yield and assembly-product uniformity.

Motivation

Infectious diseases continue to take a heavy toll on human health worldwide. Vaccines are needed to reduce the spread and severity of infectious diseases. In multiple trials, vaccines displaying multiple copies of antigen arrayed on nanoparticles have been shown to elicit a stronger and/or broader immune response than vaccines without multivalent display. In 2022, a protein nanoparticle vaccine against SARS-CoV2 was the first de novo designed protein therapeutic approved for use in humans. However, the assembly process of protein cages is incompletely understood and often fails for unknown reasons. Unpredictable yield and polydispersed product frequently hinder characterization efforts.

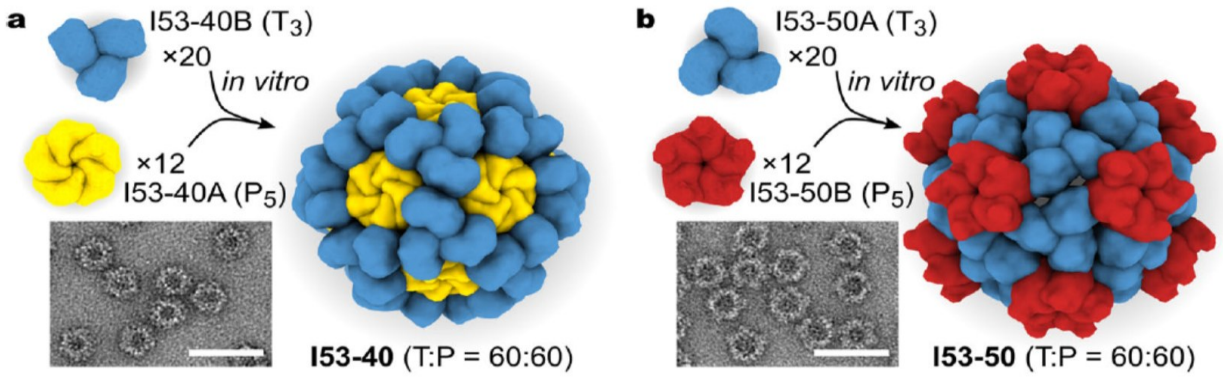
Our lab is uniquely experienced in designing protein nanocages as a vaccine platform. We have several active vaccine design campaigns and efficacy studies underway at any given time. In addition, we have a unique collection of experimental data on a variety of de novo designed protein nanocages with and without antigen or adjuvant attached. We want to leverage this data to expand our understanding of what goes wrong during protein assembly, and how to proactively design for robust assembly pathways.

Description of the model protein nanocages

Icosahedral protein nanocages I53-50 and I53-40 (Fig. 1) have been extensively characterized and have been used by several academic groups and our industry partners. (Lacasta 2023, Sun 2023, Brouwer 2022, McLeod 2022, Schoeder 2022, Sliepen 2022, Kang 2022, Lai 2021, Brouwer 2019). Each consists of 12 homo-pentamers with a C5 interface (P:P:P:P:P) and 20 homo-trimers with a C3 interface (T:T:T); the two components interact with each other via a heterodimeric T:P interface. The components are expressed and purified separately and rapidly form either the pentamer or the trimer; these subunits are then combined in controlled stoichiometry and form icosahedral cages. They were designed for hierarchical assembly, in that the C5 and C3 interfaces are stable oligomers for which monomers are not observed. They have high thermal stability: I53-40P_5 and I53-50T_3 remained folded up to at least 90°C;

I53-40T_3 and I53-50P_5 were stable to 55° and 70°C, respectively. The T:P interfaces, in contrast, are relatively weak. Their effective affinities have been estimated at -4.1 kcal/mol for I53-50 and -3.6 kcal/mol for I53-40. (Wargacki 2021)

Figure 1

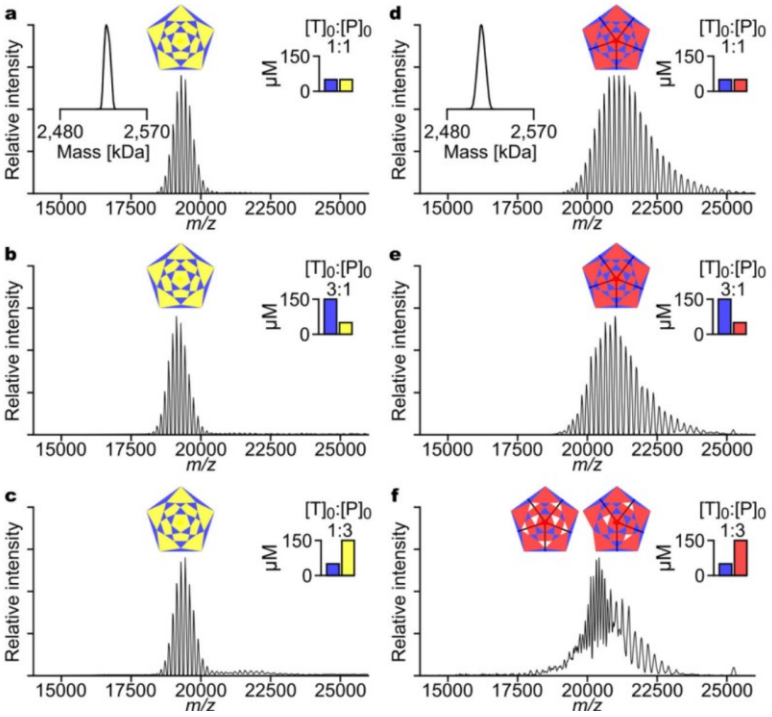


Wargacki *et al.* 2021

Evidence of pathway dependence in nanocage assembly

I53-50, but not I53-40, shows non-cooperativity in some stoichiometric regimes, when pentamer is present in excess (Fig. 2). In that case more near-cage-sized assemblies form than are expected based on limited trimer availability, and some of those assemblies lack between one and eight trimeric subunits. This suggests the rate of formation of incomplete capsids outstrips the rate of completion (Endres 2002), resulting in the trimer concentration falling below that needed to complete a cage. The incomplete cages represent an off-target metastable state in which assemblies can be trapped at experimental timescales. Assuming the complete cage is the lowest energy arrangement, such kinetic trapping in off-target states indicates the system as a whole is unable to equilibrate under this regime (Wargacki 2021).

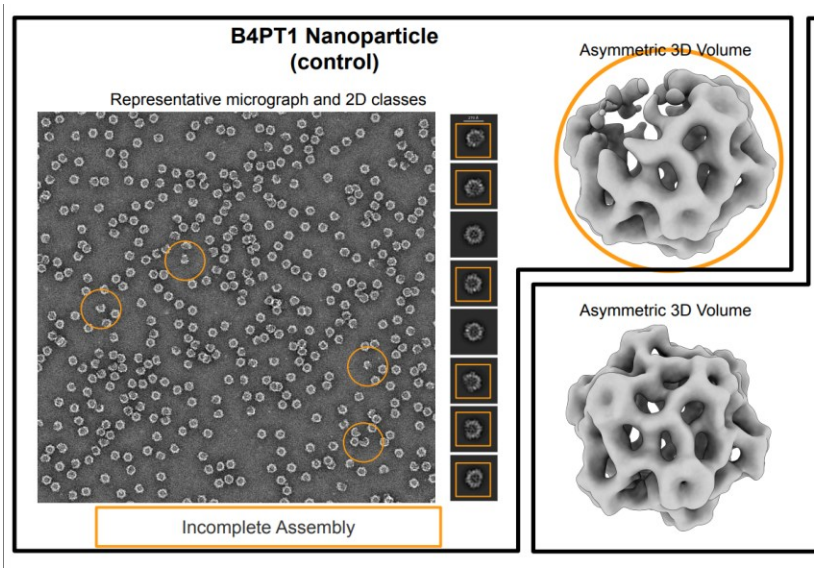
Figure 2



Wargacki *et al.* 2021

Stalled assembly of I53-50 in the excess pentamer regime can be cured by adding more trimer. Curing assembly, which consists mostly of cage completion events, is substantially slower than the average assembly reaction, which includes subunit addition events at all intermediate assembly sizes. One possible explanation is that steric hindrance emerges from the positioning of the I53-50 trimer (underneath the pentamer, closer to the cage center), slowing addition of trimer when all pentameric subunits are already in place. More recent data from cryo-EM reveal that I53-50 can in some instances form nearly complete cages with missing pentamers as a minor product, suggesting that which metastable state a particular cage stalls in depends on the sequence in which subunits are added (Fig. 3).

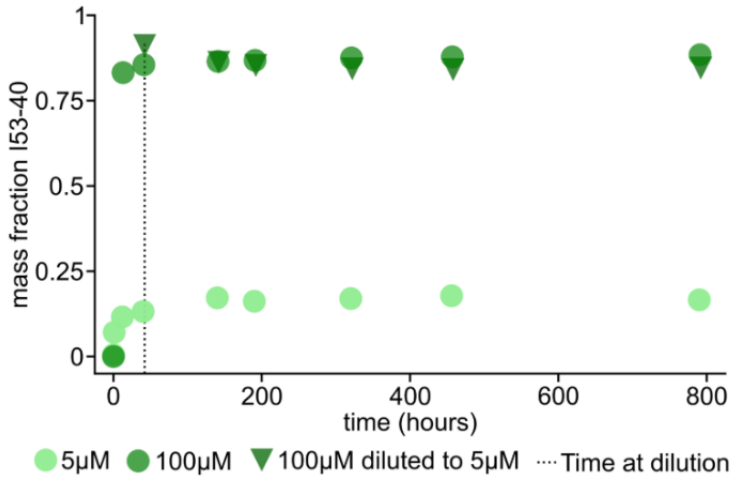
Figure 3



Data courtesy of Kenneth D. Carr

I53-40 and I53–50 both show significant hysteresis of dissociation. In the case of I53-40, assembly occurs at low subunit concentration ($[T]_0=[P]_0=5\mu M$) and within ~48 hours stalls with ~20% mass fraction cage (mfc) (Fig. 4). Higher starting concentrations of ($[T]_0=[P]_0=100\mu M$) result in yield of ~80% mfc, but subsequent dilution to 5μM at ~48 hours does not trigger disassembly, as it would if the forward and backward processes were in equilibrium. This is evidence for an energy barrier in the backward direction (disassembly) that is not observed in the forward direction (assembly). Non-reversibility indicates the latter part of the assembly process is governed primarily by kinetic factors rather than interface strength.

Figure 4.



Wargacki 2021

Our typical design pipeline relies on proxies for stability, such as AlphaFold (AF) confidence or Rosetta energy units, to select computational designs for experimental characterization. Both AF and Rosetta base their scoring on structures found in the Protein Data Bank (PDB), in which completed assemblies predominate. We do not currently have any design metrics that reflect typical yield, nor metrics based on ensembles of metastable conformations. We cannot screen either for or against designs likely to experience kinetic trapping. Our design response to low yield has historically been to try to increase stability (i.e. lower free energy) of the protein-protein interface, the single component, or both. However, this more-is-better approach presents other problems (Khmelinskaia 2021).

Goldilocks hypothesis vs. Hierarchical assembly hypothesis

There is evidence from viral capsid assembly that too-strong interfaces can lead to non-reversibility and off-target products. One hypothesis we have wanted to explore with our nanocages is that there is a fairly narrow range of interface affinity that optimizes assembly yield. Outside that range, lower affinity is expected to result in low cage yield relative to free component concentration, while higher affinity would tend toward non-reversibility, noncooperative assembly and kinetic trapping. We call this the Goldilocks

hypothesis. If this turns out to be true, we want to know what the approximate upper and lower bounds of the Goldilocks range are.

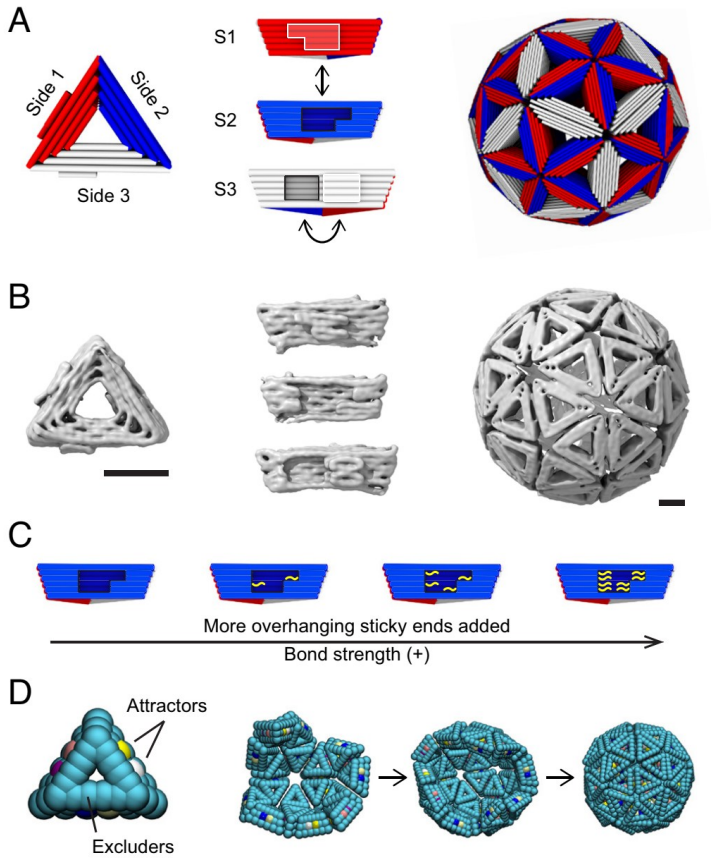
Another possibility is that the relative strengths of the various interfaces in a cage, rather than absolute strengths, determine whether assembly progresses primarily to on-target cages or not. I53-50 and I53-40 were designed hierarchically, with high affinity at the 5-fold and 3-fold interfaces and lower at the T:P interface, which is expected to form more slowly and be more readily reversible. We do not currently have enough variants of any one cage with different relative interface strengths with which to test this possibility. It is possible that both hypotheses are true, with hierarchical assembly being more favorable and also a limited range of affinities.

Two recent papers were able to provide evidence about both questions as they relate not to designed protein nanocages but to designed DNA nanocages.

Hierarchical assembly more robust in DNA origami cages

Wei *et al.* (2024) and Trubiano & Hagan (2024a) used a combination of dynamic modeling and experimental verification to investigate the effects of different per-interface affinities in self assembly of DNA origami nanocages. Their experimental model was an icosahedral cage consisting of 20 triangular faces, each of which was a pseudotrimer constructed from DNA. Each triangle had three similar but not identical interfaces (A, B and C), one on each edge. Connectivity was controlled by recessed or protruding areas in the interface, such that A would stick to B, and C to C, but no other combinations were energetically favorable. Strength of affinity was controlled by having 0, 2, 4 or 8 sticky DNA overhangs exposed in the interface (Fig. 5). The triangles could form pentamers via the AB interface or dimers via the CC interface, and these intermediates would go on to form the icosahedral cages.

Figure 5.

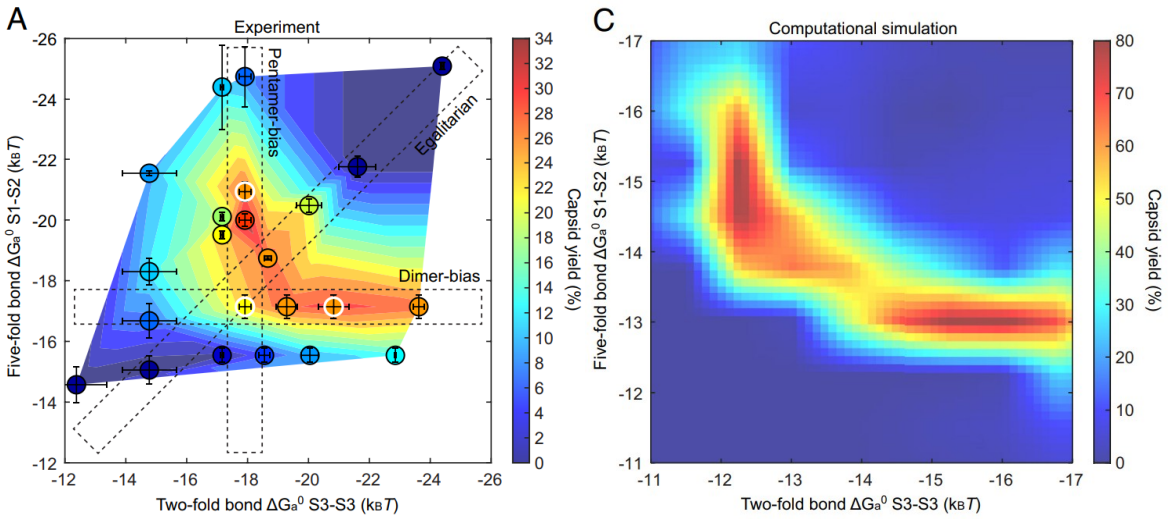


Wei *et al.* 2024

By combining different strengths of interface the authors were able to compare assembly yield of completed cages across several affinity regimes: “egalitarian,” in which the AB and CC interfaces had the same affinity; “pentamer-first,” in which the AB interface was stronger than the CC; and “dimer-first,” in which the CC interface was stronger than the AB. They found that hierarchical assembly was faster and had better yield than egalitarian, regardless which of the two interfaces was the stronger. Assembly was also more tolerant of changes in the affinity of the non-weakest interface (Fig. 6). Their data provide strong support for both the Goldilocks hypothesis and the hierarchical assembly hypothesis, in the case of DNA origami cages. When interface affinities were equal, high yield was found only at a narrow range of affinities and fell off at higher or lower affinities, which is consistent with the Goldilocks model. In the hierarchical regimes, high yield was found at a narrow range of affinity for the weaker interface, but a

wide range of affinities for the stronger interface regardless which geometric position had the stronger affinity. This robustness to variation, and overall faster assembly, provides support for the hierarchical assembly model when applied to DNA cages.

Figure 6



Wei *et al.* 2024

In addition to these experimental results, the authors used a CGMD approach called “patchy particles” (Zhang 2004) combined with Markov State Modeling (MSM) (Bowman 2013, Plattner 2017) to simulate the particle interactions and assembly processes for several affinity regimes. Their simulation pipeline allowed them to predict relative yields for the various affinity regimes and to probe what is likely to occur when affinity is too high or too low. By combining CGMD and MSM, they were able to model complete assembly of structures that would be of intractable size in regular MD simulations. They were able to use the trained MSMs to simulate various starting concentrations rapidly, without a need to retrain.

Coarse graining with patchy particles

Collective motions in polymer assembly span a wide range of timescales. Modeling all pertinent timescales at once is computationally infeasible. This has motivated the development of a variety of

coarse-graining approaches in both sampling of configuration space and dynamic modeling of individual particle trajectories through that space. Typical coarse graining in simulations of proteins involves combining some atoms into groups while tracking others individually. Grouping can be applied to the backbone or the sidechains. This accelerates simulations by reducing degrees of freedom, but simulations still require a timescale suited to individual atomic motions. A review of CGMD for virus capsid assembly can be found in Guo (2024).

Patchy particle modeling (Zhang 2004, Wilber 2009) instead treats the entire polymer subunit as a rigid body, typically formed of overlapping spherical beads, and adds sticky “patches” where subunits contact each other. The sticky behavior is instantiated as an attractive potential between select pairs of patch beads embedded between the structural beads. In addition, the structural beads have a short-range repulsive potential (typically Lennard-Jones or Weeks-Chandler-Anderson) against structural beads of other rigid bodies, for volume exclusion. Atomic-scale motions are assumed to equilibrate at short timescales, allowing slower correlated motion of the larger subunits to be discovered.

Anisotropic placement of interface patches on simplified subunits has been applied several times to model polyhedron assembly from protein or DNA components. The subunits are typically modeled as spheres with anisotropic patches (Glotzer 2004, Wilber 2009, Depta 2022), disks (Klein 2025, Zandi 2004), or idealized triangular or pentagonal tiles (Wei 2024, Trubiano 2024a, Frechette 2025, Tyukodi 2025) that are beveled at the edges to produce a dihedral angle that favors the intended polyhedron. To my knowledge, no one has published simulations that preserve the anisotropic shape of the specific proteins being modeled as well as their patch placement. The closest would be Mohajerani *et al.* (2022), in which dimers of Hepatitis B virus-like particles (HBVLP) were modeled as pairs of triangular tiles with varied dihedral angles within and between dimers, based on experimental values from T4 and T3 HBVLP icosahedra. These tiles also had tunable stretch moduli for the edges and moduli of bending at the dihedrals to allow them to be less rigid.

Because real viral capsid protomers tend to be interdigitated or layered together and to form large, convoluted interfaces (Janin 2008), their native shapes do not lend themselves well to rigid body representation. However, our nanocage components have mostly convex surfaces and show little or no conformational change between the pentamer or trimer state and the completed cage. They form smaller, non-overlapping interfaces that can be more accurately represented by sticky patches.

Markov State Modeling

Wei *et al.* (2024), and Trubiano & Hagan (2024a), combined their patchy particle MD data with Markov State modeling (MSM). MSM relies on the assumption that at certain timescales, faster motions will equilibrate and can therefore be represented as averages when modeling larger timesteps. A state model has the Markov property if its states are defined in such a way that transitions within each state happen rapidly, while transitions between states happen more slowly. This allows the model to treat all particles within a state as having equal probabilities of converting to the next state. That is, the probability of converting to a state at time t+1 depends only on what state a particle is in at time t, without any memory of how it got there. This makes it possible to further reduce degrees of freedom while preserving realistic probabilities of events (Bowman 2013). MSMs were originally applied to MD data to allow researchers to generate trajectories at longer timescales from many short, atomic-scale trajectories.

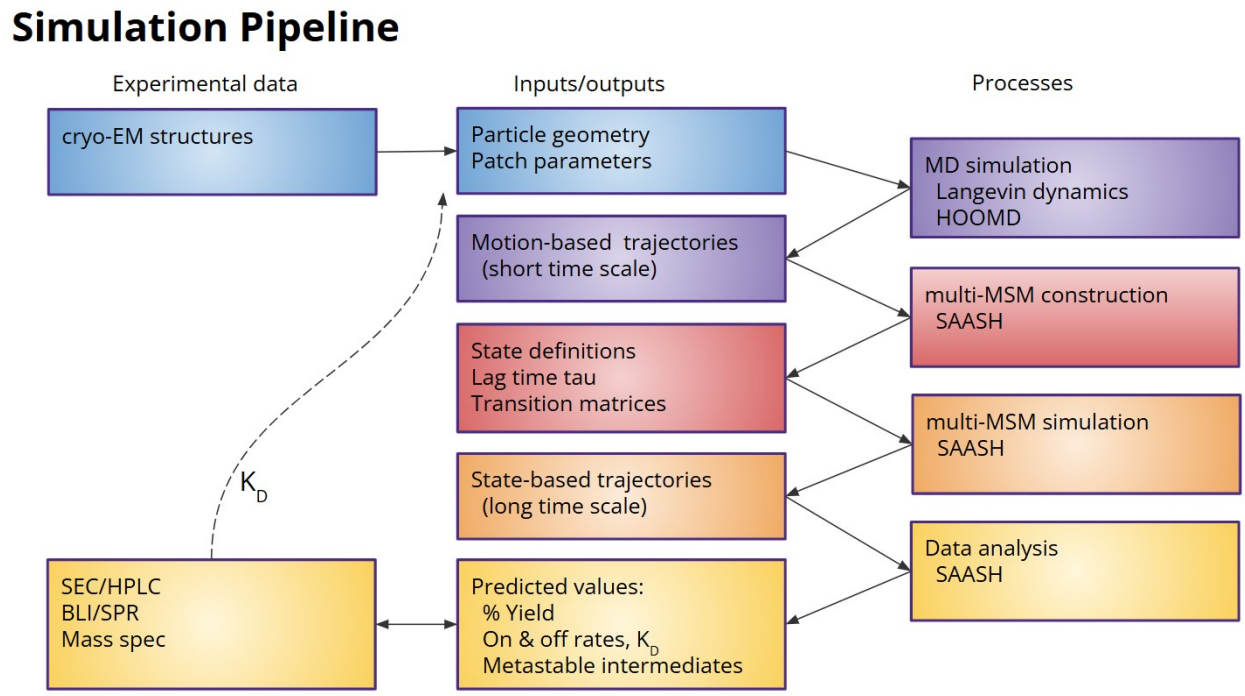
Rather than a single MSM for the entire assembly process, Wei *et al.* built a multi-MSM, which has different transition rate tables for different degrees of reaction progress. Multi-MSM allows transition rates to change as observable properties of the ensemble, such as assembly size or average number of exposed edges per fillable position, change. This approach will allow for the possibility that completion events in I53-50 have a slower rate than other addition events.

Wei *et al.* showed that rigid particles with identical geometry but differing patch strength could be used to produce semi-quantitative predictions of assembly yield similar to actual yields they found in the lab with

differing AB or CC interface strengths (Fig. 6). The original patchy particles paper (Zhang 2004) demonstrated that different patch placement on identical spherical or triangular particles could lead to different assembly geometries, such as filaments, lattices, or different platonic solids. However, it has yet to be shown how changes to geometry, to reflect actual protein shapes, interact with patch definitions to affect assembly rates or yield.

I propose to adapt Wei and Trubiano's method to model assembly of our de-novo designed protein nanocages, a use that has not been previously demonstrated.

Figure 7



Adapting the pipeline to model our protein nanocages

1. Define patchy particles based on actual protein structures
2. Sanity check definitions by running HPMC and checking that the expected cages are the predominant species produced.

3. Run Langevin MD with various patch parameter values and select high/medium/low attraction strengths to move forward with.
4. Build MSM by clustering assembly states and selecting lag time $\boldsymbol{\tau}$; validate with Chapman-Kolmogorov equation to verify Markovianity at the chosen granularity.
5. Use defined states, $\boldsymbol{\tau}$, and transition matrices to run longer state-based trajectories.
6. Calculate yield, on rates, off rates, and K_D for the most common assembly states.
7. Compare yield and K_D to experimental values; update patch parameters based on K_D.

Setting initial parameters

Trubiano *et al.* (2024a) tried a variety of values for the attractive potentials and chose three with qualitatively different outcomes in terms of yield. I will use the same empirical method for the first round of simulations. After training these first-round MSMs and running longer simulations with them, I will be able to extract on and off rates for events where two monomers associate or a dimer dissociates and use those to calculate the single-interface K_D for each interface type. This will allow me to calibrate the attractive potentials in the model against experimental quantities for $\Delta G_{contact}$.

Validation of model

The validation of the particle designs is qualitative: a fast run in HPMC will indicate if the expected assembly shape predominates, indicating it is geometrically feasible. When the particles are run in the Langevin regime they should produce primarily cages of the expected size and geometry, and it should be possible to find patch values that produce the three different expected outcomes in terms of yield (low, medium, high).

Validation of the discretization of states and the selection of lag time τ consists of checking whether the Chapman-Kolmogorov equation holds true:

$$P(t+s)=P(t)P(s)$$

That is, the matrix of transition probabilities P over a large timestep denoted t+s is the same as P after the two smaller timesteps t and s. If the model's state definitions and time lag satisfy this condition, it indicates that Markovianity has been preserved and no leakage of correlated probability across states and timescale has occurred.

Robustness of the state definitions can be tested by building them from two or more different sets of MD trajectories (training set and test set); the states identified should be consistent across data sets.

Generating new experimental data

We have ordered bicistronic versions of the I53-50 trimers and pentamers in which one T:P interface is passivated by reverting it to wild type. Dual tagging with HIS at one terminal and poly-GLU at the other will allow us to separate particles by how many passivated interfaces they have. We are in the process of characterizing these partial cage constructs and demonstrating that we can purify the different possible valencies. We will also monitor whether component exchange and resulting valency changes are detectable at experimental timescales.

The expected constructs are pentamers or trimers with one or more inactive T:P interfaces; when these pentamers and trimers are combined, completed cages cannot form. Because some pentamers or trimers will be able to attach only by one interface, we expect to see association and dissociation events from which single-interface off rates and K_D can be calculated. Once these constructs have been validated, we will run biolayer interferometry (BLI) to determine both on and off rates for the T:P interface when cage

growth is prevented. The partial cage assemblies will be imaged with negative stain EM and cryo-EM to get detailed atomic structures.

Calibration of model to experiment

Single-interface on rates are determined mostly by concentration, while differences in intrinsic interface affinity show up as single-interface off rates. The simulations produce dissociation events, and therefore dissociation rates, for single interfaces. When experimental off rates become available for our limited-interface variants, the patch strengths can be tuned until the model produces similar off rates. This can be done using simplified particle definitions with interface valencies matching the experimental constructs. Further simulation with cage-competent patchy particles will then reveal whether off rates of single particles having only one interface to be broken remain constant across trajectories, as would be expected if avidity from multiple interfaces is the only factor driving reduction of off rates as a particle grows.

Cryo-EM will confirm whether the partial assemblies maintain the subunit conformations seen in completed cages. The geometry of our modeled particles can be adjusted at this time if needed to better represent either pre-closure or post-closure conformational state if these are found to differ.

A complete cycle through the simulation pipeline with updated patch and geometry values can be run to generate more accurate yield predictions, which can in turn be compared to experimental values.

PROJECT PROPOSAL

Central hypothesis: using patchy-particle dynamic modeling with multi-MSM will improve our understanding of protein nanocage assembly trajectories, allowing us to better predict assembly outcomes and to know which features to tune to improve uniformity and yield.

Aim 1: Model assembly trajectories of our two best-characterized designed icosahedral nanocages using the Hagan lab's patchy particles + multi-MSM pipeline.

-1a. Create rigid body + multi-MSM models of I53-40 and I53-50 based on geometries of experimental structures. Try a range of sticky patch parameters to find ones that reliably reproduce designed cages in silico. Map input patch parameters to predicted yield and look for upper and lower bounds of T:P patch strength that lead to predicted mass fraction cage above 50%.

-1b. Choose 1-3 trained models for each cage that have the most accurate yield predictions. Run these models with various starting stoichiometries of the components as described in prior work (Wargacki 2021). Compare models' predicted yield in each case to experimental results.

-1c. Describe any long-lived metastable assemblies that arise during simulation. Compare these predicted off-target conformations to those detected in experimental samples.

-1d. Determine whether steric hindrance alone is enough to explain noncooperative/incomplete assemblies seen in I53-50 and not in I53-40. Look for reduced on-rates for trimeric subunits as the availability of not-fully-surrounded triangular voids decreases. See if the likelihood of voids remaining open beyond experimental timescale depends on the sequence in which subunits are added or in which voids are filled (i.e. depends on assembly path).

Key Methods

Design rigid bodies with shape and relative size based on our actual cage subunits, including interactive patches matched on size and location to experimental structures. Each protein chain will be represented by overlapping spherical “beads” at the inner surface of the chain volume; these “structural” beads will maintain the rigid volume through a WCA repulsive potential towards structural beads of any other chain. Faces of the subunits that will contain sticky patches will be sized and angled to emulate interactive surfaces found in experimental structures. Stickiness of the patches will consist of attractive Morse potentials centered in non-structural beads embedded in the interface surfaces.

Interface specificity will be set to recap that of the modeled cages. That is, trimer interfaces T1 and T2 will attract each other; pentamer interfaces P1 and P2 will attract each other; and interface T3 on the trimer and P3 on the pentamer will attract each other but not other interfaces.

Coarse-grained MD simulations using these patchy particles will be run in HOOMD-blue (Anderson 2020) using Langevin dynamics with settings based initially on those in Trubiano (2024a). Trubiano’s published SAASH software (Trubiano 2024b, 2024c) will be used to train the multi-MSMs from the MD data. Once the multi-MSM is constructed for a given parameter set, it can be used to rapidly run trajectories over longer timescales, and to run trajectories with different starting concentrations of each component type. From the MSM-generated trajectories, I will calculate yield over reaction progress, single-interface on and off rates, and single-interface K_D in HOOMD’s arbitrary energy units. These can be compared semi-quantitatively to calculated affinity from experiments.

Possible pitfalls and alternative approaches

1. Recent cryo-EM structures of I53-50 show a very small interface between the trimeric subunits at the symmetric 2-fold axis of the icosahedron, which was not previously known. The existence of a fourth (presumably weakest) interface may mean a model based on only three interfaces is

unable to accurately predict assembly kinetics. If so, we will add a fourth interface to the rigid body model, and see if predicted yields and metastable products are more aligned with experiment. We will vary patch strength for this interface to predict the effect on yield of strengthening this interface.

2. Recent sequence variants of the I53-50B pentamer have a much more negative net charge than the wild-type lumazine synthase from which I53-50B was derived or previously designed variants. There is some evidence suggesting charge repulsion decreases nanoparticle yield in these more charged variants, a phenomenon that Morse potentials on sticky patches may not accurately reflect. If we find that this approach is insufficiently predictive, we will add a coarse electrostatic repulsion potential to the pentameric subunit of I53-50 to emulate the significantly more negative charge of recently designed variants. We will determine whether such a potential enables more accurate prediction of the loss of assembly we have seen for some variants in recent experiments.

Aim 2: Model assembly trajectories of a more recently designed set of single-component icosahedral nanocages from our lab. These cages lack the hierarchical structure and therefore forced sequential interface formation of our two-component cages and therefore present different design challenges and opportunities.

-2a. I will partner with a labmate (likely Sam Tipps or Cyrus Haas) on their current cage design campaigns to apply both their usual computational filters (AF confidence, RMSD from design model, etc.) and the CGMD + multi-MSM pipeline to the same set of designed backbones. I expect the CGMD pipeline to flag some backbone shapes as having low yield or poor monodispersity. I will select candidate designs that score well a) in AlphaFold pipeline but not CGMD, b) in CGMD but not AF, and c) in both pipelines. My labmate will express, purify, and characterize the designs, enabling comparison of experimental yield, adoption of the target assembly state, and monodispersity to modeling predictions. These experiments

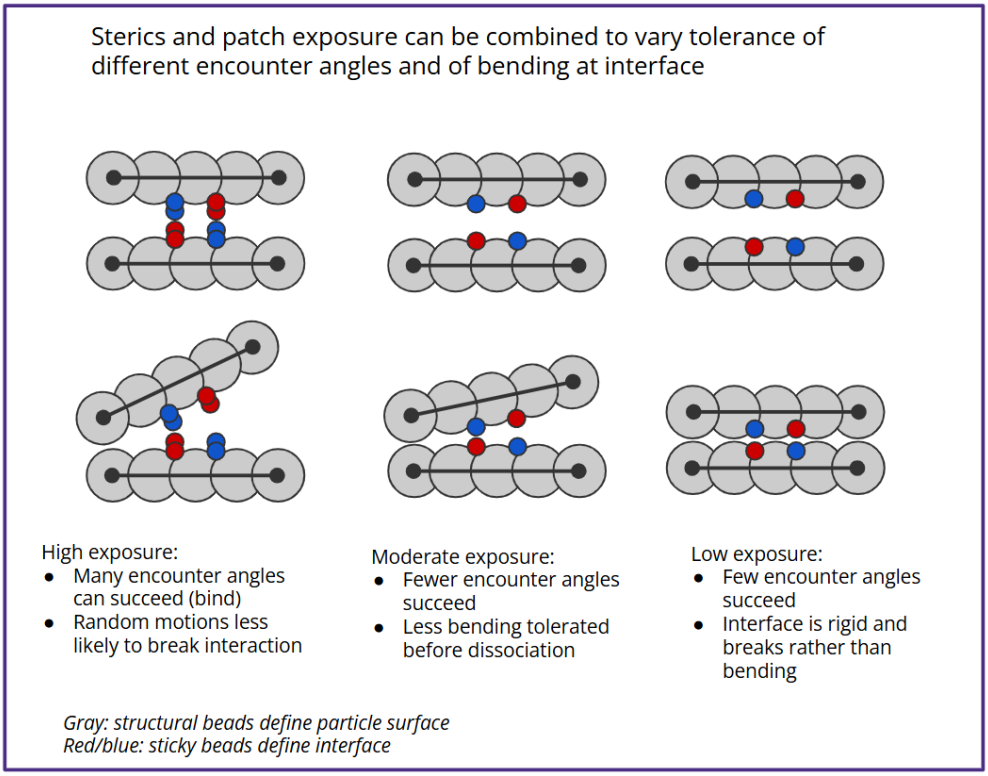
will show how dynamic modeling complements our current design pipeline by providing an orthogonal filter for computational designs. I hypothesize that the combination of AF and CGMD will be better than either alone for improving our wet lab success rate, because we can filter out candidates that may be stable when assembled but may not assemble reliably.

-2c. The CGMD + multi-MSM pipeline will allow me to describe long-lived metastable intermediates that are predicted to occur for these design candidates. My labmate and I will check for the presence of these intermediates in experimental products. Ideally, these studies will yield an awareness that these off-target structures are likely, which can be used to improve purification protocols for such designs in future efforts. For example, can buffer conditions be altered to slow nucleation, or can additional steps in purification remove the expected intermediates from samples?

Aim 3: Vary the model's tolerance for dihedral bending at the interface and different angles of encounter, and assess effects of this flexibility on assembly trajectories.

-3a. Several recent design campaigns in the lab have yielded both on- and off-target assembly products, either through within-subunit flexibility (Khmelinskaia 2025) or unanticipated interfaces between building blocks (Haas 2026; Chelsea Fries, unpublished data). I will attempt to explicitly model the latter by exposing a wider or narrower conic subsection of the spherical Morse potential to increase or decrease dihedral flexibility at the interface (Fig. 8). My hypothesis is that a geometry that tolerates more encounter angles will also tolerate more random dihedral variation at a bound interface without unbinding, and that these flexible geometries will be the ones that yield off-target assemblies in the wet lab.

Figure 8



I will vary the allowed conic subsection of the Morse potential in my pipeline to model designs that have been experimentally confirmed to yield off-target assemblies because of unanticipated interfaces. I will compare the modeling results on these assemblies to controls that have been confirmed to adopt solely the target assembly state. I note that the goal distance between attractive beads may need to be adjusted to maintain constant expected distance between neighboring tiles and resulting overall cage size. I expect to be able to improve both accuracy and recall of on-target and off-target assemblies in at least three cases. If successful, these studies will provide a tool for predicting in advance whether a given design will form multiple assembly states or only one, an important step towards predictive design of environmentally responsive, state-switching protein assemblies.

-3b. I will select a cage in which we have previously observed dihedral bending at the interface in order to demonstrate that I can tune dihedral flexibility in the modeled particle without altering K_D. Tolerance for more encounter angles is expected to increase on rate, while tolerance for more bending in the interface without breaking the bond is expected to decrease off rate. Both rate changes will tend to increase K_D in the model. The patch strengths will need to be adjusted in the opposite direction to offset these K_D changes. I will determine empirically how much patch strength reduction offsets how much increased flexibility and vice versa.

-3c. Too much flexibility at the interface is expected to lead to aggregation. Demonstrate this in the model and determine initial estimates of how much flexibility is too much or not enough in terms of acceptable assembly rates.

REFERENCES

Anderson, J.A., Glaser, J. and Glotzer, S.C., (2020). HOOMD-blue: A Python package for high-performance molecular dynamics and hard particle Monte Carlo simulations. *Computational Materials Science*, *173*, p.109363.

Asor, R., Loewenthal, D., Melnyk, D., Tan, T. K., & Kukura, P. (2026). Molecular-level observation of the self-assembly of a virus-like particle. *bioRxiv*, 2026-02.

Asor, R., Selzer, L., Schlicksup, C. J., Zhao, Z., Zlotnick, A., & Raviv, U. (2019). 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.

Ban, Y. E. A., Edelsbrunner, H., & Rudolph, J. (2006). Interface surfaces for protein-protein complexes. *Journal of the ACM (JACM)*, *53*(3), 361-378.

Bowman, G. R., Pande, V. S., & Noé, F. (Eds.). (2013). *An introduction to Markov state models and their application to long timescale molecular simulation*. Springer Science & Business Media.

Brouwer, P. J., Antanasijevic, A., Ronk, A. J., Müller-Kräuter, H., Watanabe, Y., Claireaux, M., ... & Sanders, R. W. (2022). Lassa virus glycoprotein nanoparticles elicit neutralizing antibody responses and protection. *Cell Host & Microbe*, *30*(12), 1759-1772.

Brouwer, P. J., Antanasijevic, A., Berndsen, Z., Yasmeen, A., Fiala, B., Bijl, T. P., ... & Sanders, R. W. (2019). Enhancing and shaping the immunogenicity of native-like HIV-1 envelope trimers with a two-component protein nanoparticle. *Nature Communications*, *10*(1), 4272.

Chodera, J. D., Singhal, N., Pande, V. S., Dill, K. A., & Swope, W. C. (2007). Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *The Journal of chemical physics*, *126*(15).

Chong, L. T., & Zuckerman, D. M. (2025). Weighted Ensemble Simulation: Advances in methods, software, and applications. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, *15*(6), e70055.

Christians, L. F., Halingstad, E. V., Kram, E., Okolovitch, E. M., & Pak, A. J. (2024). Formalizing coarse-grained representations of anisotropic interactions at multimeric protein interfaces using virtual sites. *The Journal of Physical Chemistry B*, *128*(6), 1394-1406.

Depta, P. N., Dosta, M., Wenzel, W., Kozlowska, M., & Heinrich, S. (2022). Hierarchical coarse-grained strategy for macromolecular self-assembly: application to hepatitis B virus-like particles. *International journal of molecular sciences*, *23*(23), 14699.

Endres, D., & Zlotnick, A. (2002). Model-based analysis of assembly kinetics for virus capsids or other spherical polymers. *Biophysical journal*, *83*(2), 1217-1230.

Frechette, L. B., Sundararajan, N., Caballero, F., Trubiano, A., & Hagan, M. F. (2025). Computer simulations show that liquid–liquid phase separation enhances self-assembly. *ACS nano*, *19*(33), 30275-30291.

Grason, G., Rogers, W. B., & Hagan, M. (2025). Teaching nanostructures to build themselves. *Physics today*, *78*(12).

Guo, W., Alarcon, E., Sanchez, J. E., Xiao, C., & Li, L. (2024). Modeling viral capsid assembly: A review of computational strategies and applications. *Cells*, *13*(24), 2088.

Hagan, M. F., & Chandler, D. (2006). Dynamic pathways for viral capsid assembly. *Biophysical journal*, *91*(1), 42-54.

Janin, J., Bahadur, R. P., & Chakrabarti, P. (2008). Protein–protein interaction and quaternary structure. *Quarterly reviews of biophysics*, *41*(2), 133-180.

Kang, Y. F., Sun, C., Sun, J., Xie, C., Zhuang, Z., Xu, H. Q., ... & Zeng, M. S. (2022). Quadrivalent mosaic HexaPro-bearing nanoparticle vaccine protects against infection of SARS-CoV-2 variants. *Nature Communications*, *13*(1), 2674.

Khmelinskaia, A., Wargacki, A., & King, N. P. (2021). Structure-based design of novel polyhedral protein nanomaterials. *Current opinion in microbiology*, *61*, 51-57.

Klein, S. M., Patterson, A., Young, K., Biever, M. P., Miller, L. M., Zlotnick, A., ... & Jarrold, M. F. (2025). Speed Matters: Directed Assembly of Icosahedral HPV Virus-Like Particles. *Journal of the American Chemical Society*, *147*(28), 24950-24957.

Lacasta, A., Kim, H. C., Kepl, E., Gachogo, R., Chege, N., Ojuok, R., ... & Nene, V. (2023). Design and immunological evaluation of two-component protein nanoparticle vaccines for East Coast fever. *Frontiers in Immunology*, *13*, 1015840.

Lagunes, L., Briggs, K., Martin-Holder, P., Xu, Z., Maurer, D., Ghabra, K., & Deeds, E. J. (2024). Modeling reveals the strength of weak interactions in stacked-ring assembly. *Biophysical Journal*, *123*(13), 1763-1780.

Lai, S. H., Tamara, S., & Heck, A. J. (2021). Single-particle mass analysis of intact ribosomes by mass photometry and Orbitrap-based charge detection mass spectrometry. *IScience*, *24*(11).

McLeod, B., Mabrouk, M. T., Miura, K., Ravichandran, R., Kephart, S., Hailemariam, S., ... & Julien, J. P. (2022). Vaccination with a structure-based stabilized version of malarial antigen Pfs48/45 elicits ultra-potent transmission-blocking antibody responses. *Immunity*, *55*(9), 1680-1692.

Mohajerani, F., Tyukodi, B., Schlicksup, C. J., Hadden-Perilla, J. A., Zlotnick, A., & Hagan, M. F. (2022). Multiscale modeling of hepatitis B virus capsid assembly and its dimorphism. *ACS nano*, *16*(9), 13845-13859.

Nitskansky, N., Clein, K., & Raveh, B. (2026). Building multiscale Markov state models by systematic mapping of temporal communities. *Bioinformatics*, *42*(1), btaf585.

Noé, F., Horenko, I., Schütte, C., & Smith, J. C. (2007). Hierarchical analysis of conformational dynamics in biomolecules: Transition networks of metastable states. *The Journal of chemical physics*, *126*(15).

Patel, J. S., & Ytreberg, F. M. (2018). Fast calculation of protein–protein binding free energies using umbrella sampling with a coarse-grained model. *Journal of chemical theory and computation*, *14*(2), 991-997.

Perlmutter, J. D., & Hagan, M. F. (2015). Mechanisms of virus assembly. *Annual review of physical chemistry*, *66*(1), 217-239.

Perlmutter, J. D., Qiao, C., & Hagan, M. F. (2013). Viral genome structures are optimal for capsid assembly. *elife*, *2*, e00632.

Plattner, N., Doerr, S., De Fabritiis, G., & Noé, F. (2017). Complete protein–protein association kinetics in atomic detail revealed by molecular dynamics simulations and Markov modelling. *Nature chemistry*, *9*(10), 1005-1011.

Price, M., Hayakawa, D., Videbæk, T. E., Saha, R., Tyukodi, B., Fraden, S., ... & Rogers, W. B. (2025). From toroids to helical tubules: Kirigami-inspired programmable assembly of two-periodic curved crystals from DNA origami. *Proceedings of the National Academy of Sciences*, *122*(49), e2516695122.

Saunders, M. G., & Voth, G. A. (2013). Coarse-graining methods for computational biology. *Annual review of biophysics*, *42*, 73-93.

Schluttig, J., Alamanova, D., Helms, V., & Schwarz, U. S. (2008). Dynamics of protein-protein encounter: A Langevin equation approach with reaction patches. *The Journal of chemical physics*, *129*(15).

Schoeder, C. T., Gilchuk, P., Sangha, A. K., Ledwitch, K. V., Malherbe, D. C., Zhang, X., ... & Meiler, J. (2022). Epitope-focused immunogen design based on the ebolavirus glycoprotein HR2-MPER region. *PLoS pathogens*, *18*(5), e1010518.

Sigl, C., Willner, E. M., Engelen, W., Kretzmann, J. A., Sachenbacher, K., Liedl, A., ... & Dietz, H. (2021). Programmable icosahedral shell system for virus trapping. *Nature materials*, *20*(9), 1281-1289.

Sliepen, K., Radić, L., Capella-Pujol, J., Watanabe, Y., Zon, I., Chumbe, A., ... & Sanders, R. W. (2022). Induction of cross-neutralizing antibodies by a permuted hepatitis C virus glycoprotein nanoparticle vaccine candidate. *Nature Communications*, *13*(1), 7271.

Starr, C. A., Barnes, L. F., Jarrold, M. F., & Zlotnick, A. (2022). Hysteresis in hepatitis B virus (HBV) requires assembly of near-perfect capsids. *Biochemistry*, *61*(7), 505-513.

Sun, C., Kang, Y. F., Fang, X. Y., Liu, Y. N., Bu, G. L., Wang, A. J., ... & Zeng, M. S. (2023). A gB nanoparticle vaccine elicits a protective neutralizing antibody response against EBV. *Cell Host & Microbe*, *31*(11), 1882-1897.

Trubiano, A., & Hagan, M. F. (2024a). Markov state model approach to simulate self-assembly. *Physical Review X*, *14*(4), 041063.

Trubiano, A. (2024b). Self-assembly analysis suite for HOOMD, https://github.com/onehalfatsquared/SAASH.

Trubiano, A. (2024c). MultiMSM, https://github.com/onehalfatsquared/MultiMSM.

Tyukodi, B., Caballero, F., Hayakawa, D., Hall, D. M., Rogers, W. B., Grason, G. M., & Hagan, M. F. (2025). Magic sizes enable minimal-complexity high-fidelity assembly of programmable shells. *Physical review letters*, *135*(11), 118203.

Videbæk, T. E., Hayakawa, D., Hagan, M. F., Grason, G. M., Fraden, S., & Rogers, W. B. (2025). Measuring multisubunit mechanics of geometrically programmed colloidal assemblies via cryo-EM multi-body refinement. *Proceedings of the National Academy of Sciences*, *122*(37), e2500716122.

Wargacki, A. J., Wörner, T. P., van de Waterbeemd, M., Ellis, D., Heck, A. J., & King, N. P. (2021). Complete and cooperative in vitro assembly of computationally designed self-assembling protein nanomaterials. *Nature Communications*, *12*(1), 883.

Wei, W. S., Trubiano, A., Sigl, C., Paquay, S., Dietz, H., Hagan, M. F., & Fraden, S. (2024). Hierarchical assembly is more robust than egalitarian assembly in synthetic capsids. *Proceedings of the National Academy of Sciences*, *121*(7), e2312775121.

Wilber, A. W., Doye, J. P., Louis, A. A., & Lewis, A. C. (2009). Monodisperse self-assembly in a model with protein-like interactions. *The Journal of chemical physics*, *131*(17).

Zandi, R., Reguera, D., Bruinsma, R. F., Gelbart, W. M., & Rudnick, J. (2004). Origin of icosahedral symmetry in viruses. *Proceedings of the National Academy of Sciences*, *101*(44), 15556-15560.

Zhang, Z., & Glotzer, S. C. (2004). Self-assembly of patchy particles. *Nano letters*, *4*(8), 1407-1413.