Computational design of *de novo* protein nanoparticles for RNA delivery

Christian Richardson

A dissertation

submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:

David Baker, Chair

Daniel Ratner

Neil King

Program Authorized to Offer Degree:

Bioengineering

University of Washington

Abstract

Computational design of *de novo* protein nanoparticles for RNA delivery

Christian Richardson

Chair of the Supervisory Committee

David Baker

Department of Biochemistry

Several non-porous RNA encapsulating protein nanoparticles have been reported, but none have been shown to deliver nucleic acids to mammalian cells. Here, I will discuss my efforts to optimize the topology of a *de novo* pH-responsive trimer for docking into larger protein nanoparticles, and the design and screening of a library of such nanoparticles using a peptide barcoding method. I will also discuss my work engineering a previously reported RNA-encapsulating nanoparticle for RNA delivery in immortalized cell lines. Finally I will discuss my work combining knowledge from all of these efforts into the design of a novel non-porous protein nanoparticle that may provide an optimal scaffold for future efforts in nucleic acid delivery using protein nanoparticles.

Table of Contents

List of Figures

Chapter 1. *De novo* self-assembling protein nanoparticles as nucleic acid delivery vehicles

1.1 Abstract

As viral vector and synthetic nanoparticle technologies have matured, they have brought the first waves of nucleic acid therapies to the clinic in the form of both gene therapies and vaccines. However, there is still significant opportunity for growth. Self-assembling protein nanoparticles are an emerging technology that blends the advantages of both viral vectors and synthetic nanoparticle formulations, creating a promising new area of development in nanotherapeutics.

1.2 Introduction

Millenia of evolution and decades of rational development have made viral vectors one of the most mature nucleic acid delivery technologies. Viral structural proteins have evolved complex and often poorly understood structural rearrangements and progressive uncoating strategies that deliver nucleic acid cargo to target cells with unparallelled efficiency and minimal toxicity. Because these proteins are genetically encodable, they can also be used in powerful directed evolution experiments to evolve new desirable features[1].

Advances in polymer chemistry and liposomal formulations have also led to the development of entirely synthetic nucleic acid delivery vehicles. These often offer a higher degree of modularity, allowing the blending of various polymers or lipids with different physicochemical properties. Many of these synthetic delivery vehicles are also highly tolerant to the fusion of protein targeting or other functional domains to the particle. However, synthetic polymer and lipid

derived nanoparticles are highly intrinsically disordered which can make it challenging to design the types of complex structural rearrangements found in viral vectors. This lack of fine tuned designability may contribute to the reduced delivery efficiencies seen with synthetic nanoparticle platforms[2,3].

Advances in protein engineering have enabled the design of entirely de novo self-assembling protein nanoparticles. These self-assembling protein nanoparticles resemble viral capsid structures and offer a high degree of tunability that may enable the design of highly complex functions similar to that of viral vectors. Additionally, these protein nanoparticles are genetically encodable allowing for the evolution of new and highly complex features. Self-assembling protein nanoparticles can be engineered to be highly stable and tolerant to the fusion of arbitrary functional protein domains. The subunits comprising these nanoparticles can be expressed separately and mixed at different ratios allowing for fine control over nanoparticle assembly and the blending of multiple functions into one particle. Taken together, self-assembling protein nanoparticles offer a unique combination of the fine designability and evolvability of viral vectors with the control and modularity of synthetic nanoparticle platforms[4–6].

Despite these advantages, the design of self-assembling protein nanoparticles is a relatively nascent technology with many opportunities for fundamental development that can improve both the efficacy of this technology and its accessibility to other researchers. Sections 1.3-1.5 will outline the current state of protein nanoparticle design for reference.

1.3 Designable features of self-assembling protein nanoparticles

Adapted from review. Olshefsky, Richardson, Pun, King. Engineering Self-Assembling Protein Nanoparticles for Therapeutic Delivery. Bioconjug Chem. 2022 Nov 16;33(11):2018-2034.

Protein nanoparticle scaffold design

Improved methods in computational protein engineering have allowed for bottom-up engineering of protein nanomaterials greatly improving the protein nanoparticle design space. Most of these methods take a molecular docking approach in which a symmetry operation is performed on an existing natural or de novo protein oligomer in order to arrange them into a closed nanomaterial. The molecules are then designed to bind each other in this conformation through either rigid helical fusion or hydrophobic interface design. The complexity of these design strategies necessitates the use of computational tools for interface design and many of these methods require highly specific technical skills.

Directed evolution is a common method for refining protein nanoparticles in sequence space. Directed evolution methods have been reported to improve the biodistribution of a nucleic acid packaging protein nanoparticle in one case and drastically increase the number of subunits and interior volume of a nucleic acid packaging protein nanoparticle in another[7,8]. An important aspect of designing any directed evolution experiment is the selection assay used and the starting point of the molecule being evolved. Both choices require careful consideration as directed evolution methods tend to be extremely labor intensive. However, there are many features in protein nanoparticles that cannot be easily computationally designed. For example, negative design for a lack of nanoparticle binding against all human cell surface receptors is not a reasonable computational design task given the high complexity. However, directed evolution

approaches have made progress in this area[8]. AAV delivery studies have also demonstrated that the coupling of an effective selection strategy with a relatively fit starting point can effectively evaluate millions of designs in a highly relevant in vivo context[9].

Protein nanoparticle architecture and geometry

Protein nanoparticles are generally atomically defined and highly symmetric nanoparticles reminiscent of non-enveloped viral capsids. Nanoparticle geometry refers to the symmetric arrangement of protein subunits to form a closed three dimensional nanoparticle. Tetrahedral, octahedral and icosahedral symmetries are the most common closed protein nanoparticle geometries. If it is desired to display a cyclic homo-oligomer on the surface of a protein nanoparticle, it is generally desirable to match the geometry of the homo-oligomer to that of the subunit it is being fused to as is commonly seen in antigen displaying protein nanoparticles[10]. The symmetry group also determines the valency of asymmetric units in each closed assembly and thus the valency of any externally fused ligands and internally fused packaging domains. Icosahedral symmetry is generally required for most macromolecular delivery applications as this symmetry group provides the largest interior volume per subunit size thus allowing for the packing of larger cargos or more copies of packaged macromolecules.

The asymmetric unit of a protein nanomaterial may be composed of one or more amino acid chains and along with the symmetry group and relevant axes of symmetry, comprises the minimum set of information needed to define all atomic coordinates in a protein nanomaterial[11]. If a nanoparticle is designed from two or more amino acid chains, the nanoparticle can in principle be assembled in vitro[4]. This allows for a high degree of flexibility in cargo packaging and nanoparticle assembly quality. Additionally the location of the N and C termini determine

fusion points for any desired external or internal functional domains such as receptor binding domains or cargo packaging tags.

Protein nanoparticle interface design and modifications

Due to their highly symmetric nature, protein nanomaterials are highly cooperative. As a result the design of protein nanomaterial interfaces requires special consideration as overly strong nanoparticle interfaces can result in kinetic trapping of partially assembled intermediates while overly weak interfaces will prevent assembly of the nanomaterial in the first place. The large hydrophobic interfaces often found in computationally designed protein nanoparticles can also sometimes interfere with expression of individual subunits which is a prerequisite for in vitro assembly and thus the packaging of arbitrary cargos.

Protein nanoparticle interfaces can also be used to design in environmental responsiveness or otherwise confer nanoparticle functional features. For example the scavenging of a protein A antibody binding interface and fusion to de novo protein oligomers resulted in the scaffolding of an antibody nanoparticle. A natural protein nanoparticle, ferritin, has also been engineered to have pH responsive disassembly in order to enable the packaging of arbitrary cargos. Ionic, pH and metal dependent protein nanomaterials have also been reported for a wide ranging set of applications[12].

Chapter 2: Redesign of a pH-responsive trimeric protein for optimal nanoparticle docking topology

2.1 Abstract

The pH responsive disassembly of nanoparticles has a high potential as an actuation mechanism for endosome escape. We have previously reported a de novo pH responsive trimeric protein capable of disassembly at endosomally relevant pH, but it does not dock well into icosahedral protein nanoparticles. Here we report the fusion of this pH responsive trimer to de novo helical repeat proteins in order to create stellate trimeric protein assemblies that make ideal scaffolds for docking into large icosahedral nanoparticles.

2.2 Introduction

Endosomal degradation remains a significant barrier to nucleic acid delivery for many non-viral and non liposomal delivery vehicles[2]. Given the need for a negatively charged hydrophilic particle for effective biodistribution and the need for a positively charged hydrophobic material for endosomal disruption, a state switching mechanism may be desirable for improved nucleic acid delivery efficiency[13]. Given the acidity of the endosomal compartment and the extensive prior work in protein based pH switches[14].

Our group previously reported a de novo designed, pH-responsive trimeric protein capable of disassembly at endosomally relevant pH named pH192[15]. However, the tightly packed cylindrical topology of pH192 is not amenable to assembly into larger protein nanoparticles with large hydrophobic interfaces and a large interior lumen. We have also recently reported a set of de

novo helical repeat proteins(DHRs) as well as a rigid helical fusion method[16,17]. In order to increase the utility of the pH responsive trimer for docking into higher order symmetric protein nanomaterials, we sought to rigidly fuse a variety of DHRs to variants of the pH responsive trimer.

2.3 Results

2.3.1 Computational design of stellate, pH-responsive trimers

We first designed stellate, pH-responsive trimers compatible with nanoparticle design by fusing previously reported *de novo* heterodimers to the N or C terminus of pH192. However, the C terminus of trimer pH192 is located on the inner helices and has a very short region in which rigid helical fusion can occur. To expand the repertoire of C terminal rigid helical fusions available, the loop between the two helices of the asymmetric unit of pH192 was eliminated to create a C terminus on the distal, more fusible, helix, and the original N and C termini were linked using an unstructured GGSGG linker to produce pH192rl. An additional version of pH192 was designed by removing the last heptad from the distal helix producing a truncated variant with a more accessible C terminus called pH192t.

In order to improve our success rate we only used DHRs with solved crystal structures that matched our design models. To fuse these DHRs to the pH192 variants, we aligned all possible sets of 5 helical residues in the N terminal helix of every DHR to all possible sets of 5 helical residues in each pH192 variant [Figure 2.1B]. After alignment the third residue of the DHR was fused to the 4th residue of the pH192 variant and the remaining residues of each protein were deleted [Figure 2.1 A-C]. This fusion is hereafter referred to as the fusion junction. To ensure the

fusion junction would be rigid, we then filtered these backbones for designs that had at least two helices on each side of the fusion junction that had residues within 10 Å of residues on the other side of the fusion junction ensuring. This was intended to ensure that at least two helices on each side of the fusion junction were contributing to the fusion junction interface. Backbones passing this filter were then designed using Rosetta allowing for mutation and rotamer repacking within 10 Å of the fusion junction and no mutation but rotamer repacking within 25 Å of the fusion junction [Figure 2.1C]. Designs with fewer than 2 clashing residues and no buried unsatisfied polar residues were selected for visual filtering. Using visual filtering for interdigitation of hydrophobic residues across the fusion junction, 24 designs were selected for experimental characterization.

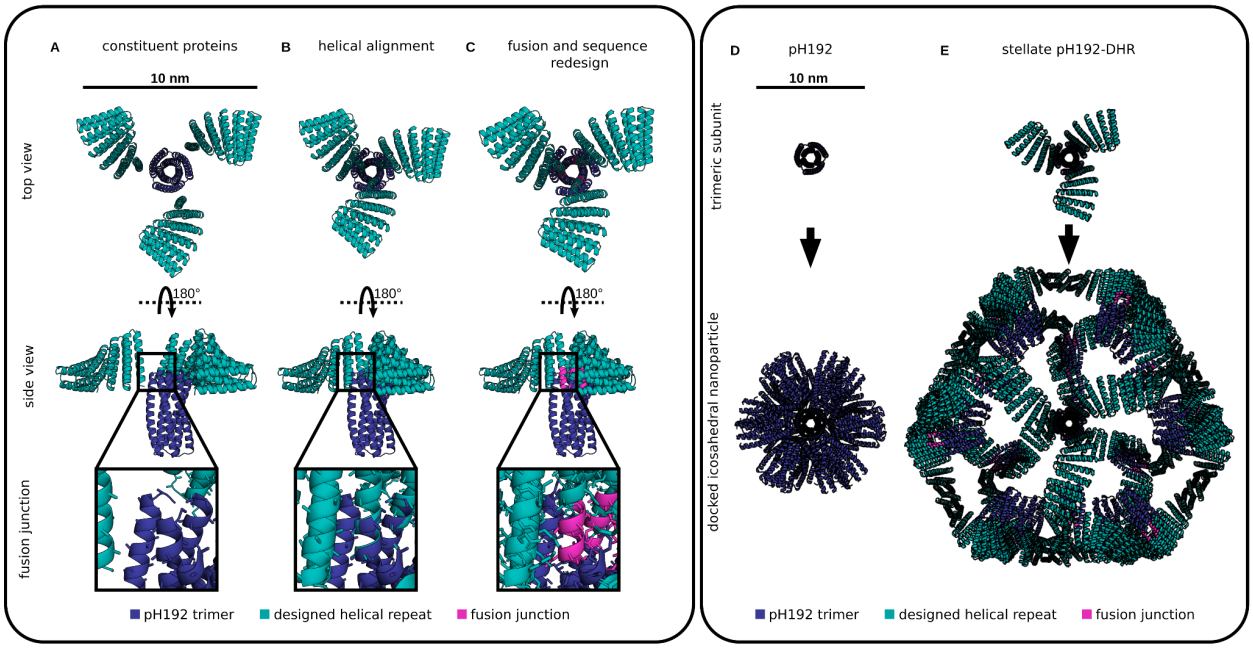


Figure 2.1. Rigid helical fusion of DHRs to cyclic trimers creates a stellate topology that is optimal for docking into larger protein nanoparticles. (A) Three DHRs arranged cyclically in close proximity to pH192. (B) alignment of the backbones of 5 residues in the C terminal, inner helix of pH192 to the backbones of 5 residues in the N terminal helix of a DHR. The beta carbons of each sidechain are shown. (C) Truncated, fused, and sequence redesigned model of pH192-DHR. The fusion junction is shown in magenta. (D) pH192 trimer (top) and pH192 trimer docked into an icosahedral nanoparticle (bottom). (E) Stellate pH192-DHR trimer (top) and pH192-DHR trimer docked into an icosahedral nanoparticle (bottom).

2.3.2 Biophysical characterization of stellate, pH-responsive trimer designs

Synthetic genes encoding each of the 24 designs were obtained, expressed in *E. coli* and purified by nickel nitrilotriacetic acid(NiNTA) chromatography. All designs were soluble, and size exclusion chromatography (SEC) yielded a monodisperse peak for 6 of the 24 designs (Figure 2.2A-B). Small angle x-ray scattering (SAXS) closely matched the design model in 4 of the designs, with pH192t-DHR03 likely mismatching due to radiation damage and pH192rl-DHR79 likely failing due to a highly populated off-target structure. SEC purified samples were run on mass photometry which showed a dominant trimeric state for all designs except pH192rl-DHR79 which had a high prevalence of a monomeric species.

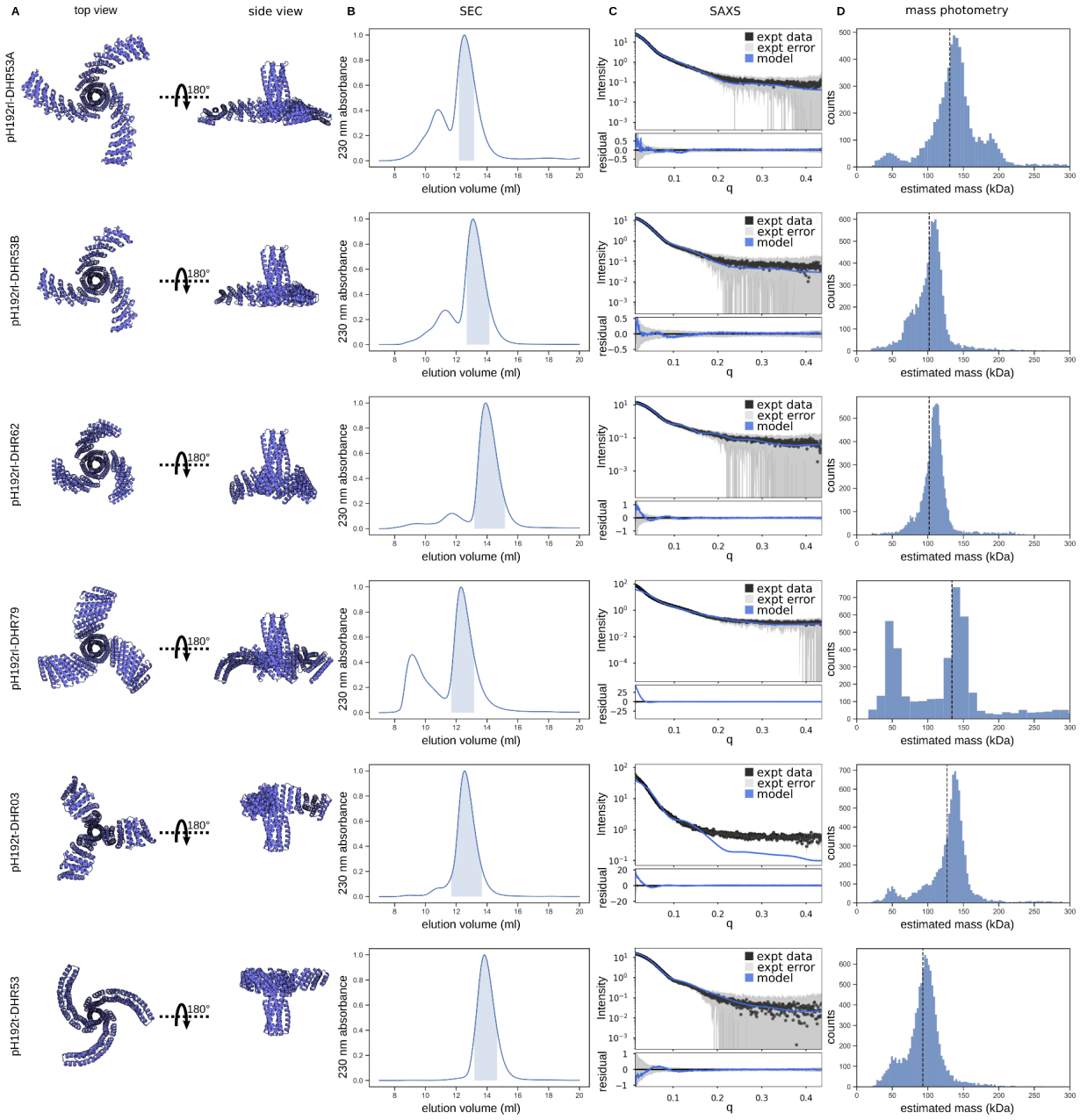


Figure 2.2. Biophysical characterization validates structure and oligomeric state of stellate cyclic oligomers. (A) Structure from the top (left) and side (right) of 6 stellate trimer designs that eluted as a monodisperse peak on SEC. (B) SEC trace of 230 nm absorbance. Fractions collected for later analysis are shaded in blue. (C) Small angle X-ray scattering traces averaged across 33 frames except in the case of pH192t-DHR03 which includes only one frame due to radiation damage. The mean experimental value is in black, standard deviation of the 33 frames collected is in gray, and expected scattering based on the design model is in blue. The residual is shown below along with standard deviation. (D) Histogram of mass photometer measurements of the mass of SEC purified trimeric assemblies. The expected mass of each trimer is indicated with a dotted line in each plot.

2.4 Discussion

Here we demonstrate the redesign of a pH-responsive trimeric helical bundle to have a more stellate topology. Five out of 24 designs ran as a single species on SEC and closely matched the

expected conformation and oligomeric state, demonstrating the high success rate of rigid helical fusion with Rosetta redesign relative to the design of *de novo* oligomers and docking of higher order protein assemblies[18–20]. Nearly every design with a monodisperse SEC trace adopted the expected conformation suggesting SEC is a reliable screening method for protein folding and protein complex assembly.

The fusion of various versions of pH192 to various DHRs further validates the modularity of de-novo protein nanomaterials, and indeed such methods are now being reported for a wide variety of applications[17,21,22]. Interestingly, DHR53 was highly represented in the successful designs suggesting that it may be more amenable to rigid fusion and mutation than other DHRs. It is possible that in the future other high performing de novo protein nanomaterial building blocks will be identified, and alongside an increasingly diverse set of pre-existing rigid protein fusions may greatly increase the success rates of protein nanomaterial design in the near future.

Notably the pH sensitivity of the stellate trimers was not evaluated. These designs are intended for use as scaffolds for large and highly symmetric nanomaterials which are known to be highly cooperative[23–25]. As such, the pH at which disassembly occurs and the nature of that disassembly are likely to be significantly altered by the context of the nanomaterial making such measurements superfluous at this stage. The base pH192 scaffold was designed with three hydrophilic networks in its core that can optionally incorporate histidine. The more networks that incorporate histidine, the higher the pH at which trimer disassembly occurs[15]. This will allow for minimally perturbative tuning of the pH response of symmetric protein nanomaterials after design and validation of the overall architecture.

Chapter 3. Design and screening of a self-assembling protein nanoparticle library designed from pH-responsive trimers

3.1 Abstract

Until very recently, self-assembling protein nanoparticle design has been limited to low throughput testing of tens of designs at a time. Development of library approaches to testing a wide variety of self-assembling protein nanoparticles is desirable to elucidate design principles and tackle the design of complex nanoparticles with low success rates. Here we demonstrate the multiplexed expression, purification, and analytical size exclusion chromatography of peptide barcoded self-assembling protein nanoparticles. We tested a small set of 11 previously characterized self-assembling protein nanoparticles and a large set of 1173 newly designed self-assembling protein nanoparticles to demonstrate the utility of this method. Low throughput follow up revealed internal start sites in most of the 1148 designs in the large set. However, there were still self-assembling protein nanoparticles that could be rescued by the removal of the alternate start sites.

3.2 Introduction

We have previously reported several protein nanoparticles, but these efforts typically suffer from low success rates on the order of 1-10% of ordered designs[20,26,27]. Recent methods using rigid helical fusion of previously validated heterodimers to previously validated homo-oligomers has shown a higher success rate for lower order symmetry groups, but these results have not translated into high success rates for nanomaterials with icosahedral symmetry, likely due to the high binding affinity of de novo heterodimers combined with the high cooperativity of the

icosahedral symmetry group[18,25,28]. Given the low number of successful protein nanoparticle assemblies, more data is needed to determine what intersubunit interface metrics are important for design success.

Egloff et al reported a novel peptide barcoding scheme for the analysis of binding protein libraries[29]. This method utilized SEC separation of the protein binder-target complexes from protein binder alone before liquid chromatography with tandem mass spectrometry (LC-MS-MS). We hypothesized that a similar method could be used to distinguish assembly-competent nanomaterials from assembly-incompetent nanomaterials if we ran multiple SEC fractions on LC-MS-MS [Fig. 3.1 A]. Furthermore, we hypothesized that such a method could be used to design libraries of icosahedral protein nanoparticles designed from stellate trimeric components.

3.3 Results

3.3.1 Small scale validation of the barcoding technique with existing nanoparticles

To validate our hypothesis, we chose an initial set of 10 nanoparticles to validate the mass spectrometry barcoding assay. They comprised 2 nanoparticles that were known to assemble [Fig. 3.1 B-C], 4 nanoparticles that were known to not assemble [Fig. 3.1 D-G], and 4 nanoparticles that were known to aggregate [Fig 3.1 H-K]. To ensure that the peptide barcodes did not modulate protein nanoparticle expression or assembly, 8 unique barcodes were cloned onto each protein nanoparticle and separately purified from 1mL cultures. SDS and Native PAGE confirmed that there were no expression or assembly differences across the 8 barcodes fused to each design [Supplementary Figure 2.1]. All samples were pooled and expressed in a

500mL culture, purified, and separated by SEC. The barcodes in each fraction were isolated and run on LC-MS to quantify relative barcode representation in each fraction. Barcode intensity in each SEC fraction as measured by LC-MS was then compared to the 280 nm and 230 nm absorbance of each nanoparticle when individually run on SEC [Fig. 3.1 B-K]. All nanoparticles except for 3.1 C were two-component nanoparticles and all two-component nanoparticles had varying numbers of tryptophan in each subunit resulting in distinct 280 nm and 230 nm traces. The two nanoparticles that were expected to assemble had barcode intensities that closely matched the peak corresponding to the nanoparticle assembly [Fig. 3.1 B-C]. Non-assembling but otherwise well behaved nanoparticle designs had barcode intensities that closely matched only one of the two component traces [Fig. 3.1 D-G]. This is consistent with our expectations as only one of the two components was barcoded. Nanoparticle designs that were known to aggregate showed barcode agreement with nanoparticle component peaks but not with the void fraction associated with protein aggregation [Fig 3.1 H-K]. This is likely due in part to the much longer handling time for the low throughput designs, due to simultaneous purification and use of an autosampler, which allowed for more aggregation to occur.

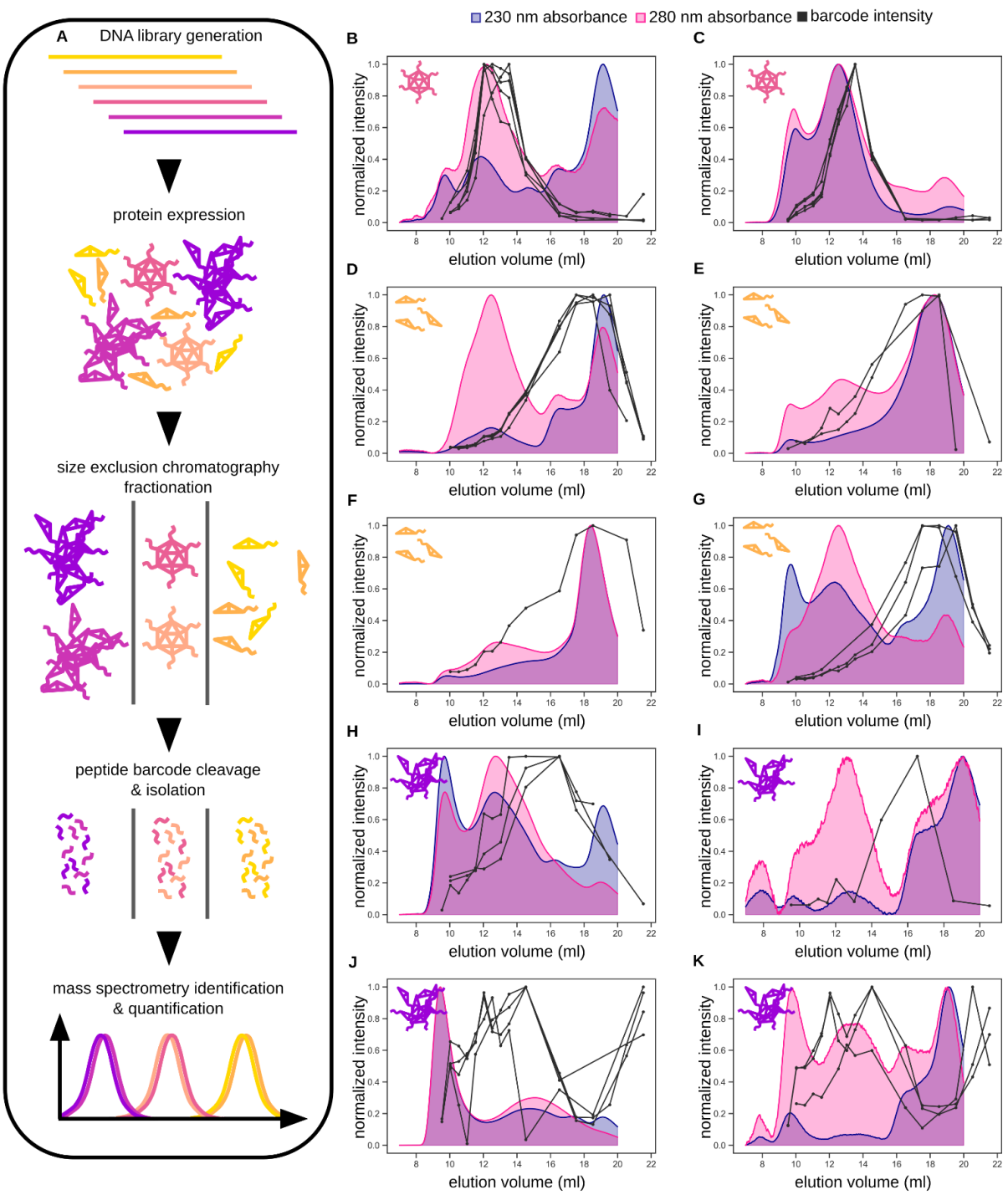


Figure 3.1. Validation of of a peptide barcoding strategy for nanoparticle screening using a small nanoparticle library. (A) schematic showing the peptide barcoded nanoparticle screening strategy. Genes encoding each nanoparticle design are expressed and assembly competent designs are separated from component and aggregate by SEC. NiNTA pulldown of His tagged peptide barcodes and subsequent His tag cleavage results in variable enrichment of each barcode that is representative of nanoparticle migration on SEC and can be matched to the low throughput SEC trace. (B-K) Low throughput 230 nm absorbance (blue) and 280 nm absorbance (pink) traces on SEC compared to one or more peptide barcodes measured by LC-MS (black).

3.3.2 Design of a de novo, self-assembling protein nanoparticle library from stellate, pH-responsive trimer subunits

After validating the peptide barcoding method for a small pool of protein nanoparticles, we set out to design a nanoparticle library based on a previously validated stellate pH trimer, pH192_0431 [Fig. 3.1 A]. pH192_0431 was docked against itself using residue pair transform docking (RPX dock) on grid search mode to increase viable docks [Fig. 3.2 B][30]. DNA oligonucleotide array limitations required that we use a variable region smaller than 400 base pairs. To accommodate this constraint, docks with residues outside the 78 C-terminal amino acids that were within 8 angstroms of the intratrimer interface were dropped. The remaining docks were then sequence designed using a layer design script in Rosetta. Briefly, residues in contact with adjacent subunits were mutated to hydrophobic residues and packed tightly using a monte carlo method and pairwise decomposable energy function for residue mutation and rotamer repacking. Hydrogen bonded polar resides were added along the boundary of the interface, and the neighborhood within 10 angstroms of any mutated residue was repacked without sequence changes to allow for small adjustments to side chain orientation. Designed sequences were filtered for a Rosetta computed ddG < 0 and shape complementarity > 0.4. These designs were deduplicated, and any designs with mutations outside the 78 C-terminal amino acids were removed resulting in 1173 unique protein sequences.

3.3.2 Nanoparticle encoding DNA library construction

DNA encoding for the approximately 1 kb region that is constant among all protein nanoparticle designs was obtained and cloned into a backbone with golden gate cloning adapter sites for insertion of the variable C-terminal region. The 1173 unique protein sequences were each associated with 5 unique barcodes and the resulting 5865 sequences were split into two

overlapping sequences with minimally homology in the overlap region across designs. These sequences were ordered as oligomers and assembled into the variable regions in 3 subpools by polymerase chain reaction (PCR). The PCR assembled variable region library was Golden Gate cloned into the constant region to produce a library of genes encoding the nanoparticle designs [Fig. 3.2 C][31]. The library was then transformed into BL21 DE3 *E. coli* by electroporation and a sample of the resulting culture was sequenced using an Illumina MiSeq. Nearly all sequences were represented, and relative representation was within two orders of magnitude as determined by next gen sequencing (NGS) read count [Fig. 3.2 D]. Pairwise reads were computationally reconstructed using PEAR[32]. Of the successfully reconstructed reads, 50.4% had early stop codons and were not expected to produce a barcode by our purification scheme, 33.1% had an incorrect sequence that was expected to produce a barcode matching one of the designed barcodes, 9.2% had a correct sequence and were expected to produce a matching correct barcode, and 7.3% were incorrect sequences that were expected to produce a detectable barcode that did not match any of the barcodes in our design library [Fig. 3.2 E]. Notably, 0% contained a correct sequence with an incorrect or mismatched barcode.

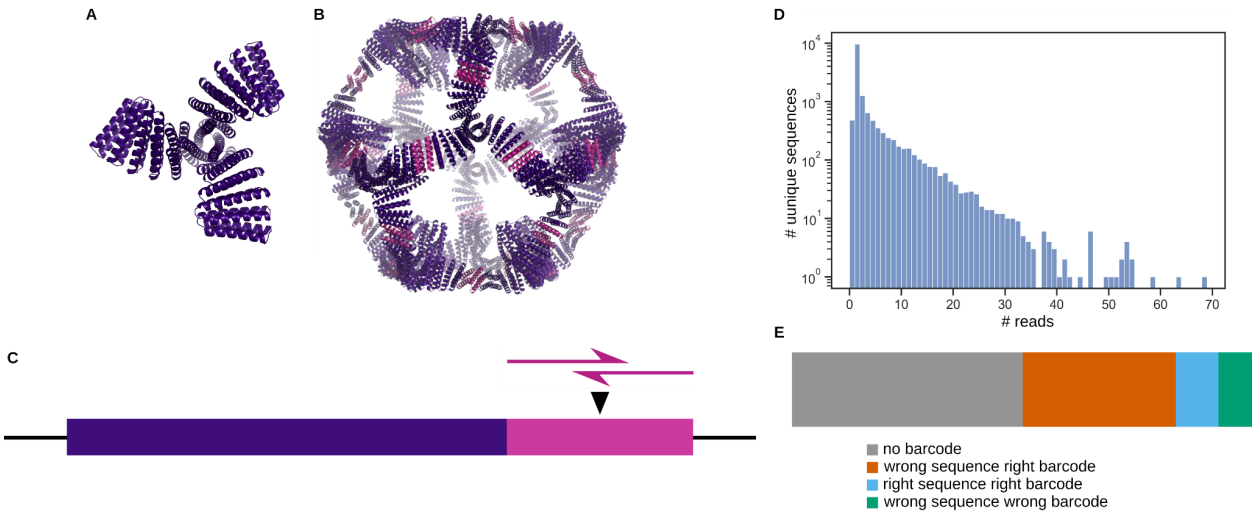


Figure 3.2. The design, assembly, and quality control of a nanoparticle encoding gene library for later screening. (A) pH192_0431 stellate pH responsive trimer. (B) I3 nanoparticle designed from stellate pH responsive trimers with the mutated residues highlighted in pink and the constant region shown in purple. (C) Schematic of gene encoding the protein nanoparticle library with the variable region in pink and the constant region in purple. (D) Representation plot of the number of unique reads matching the design library and counts for each of those reads. (E) Plot of the state of expected barcodes or all genes identified by NGS.

3.3.3 Screening of the protein nanoparticle library for self-assembly

Genes encoding the protein nanoparticle library were expressed and purified from BL21 DE3 *E. coli*, purified by Ni-NTA chromatography and then run on SEC. Peptide barcodes were then isolated from the SEC fractions and barcodes isolated from each fraction. The added complexity of the larger library necessitated secondary mass spectrometry on the peptide barcodes which were thus run on LC-MS-MS using an orbitrap detector. Barcodes that were detected by LC-MS-MS in 10 or more SEC fractions were plotted using LC-MS-MS peak intensity on the y-axis and elution volume on the x axis. Samples were given void, assembly, and component scores equal to the fraction of the area under the curve present in the 8-11 mL region, 11-14.5 mL region, and 14.5-19.5 mL regions respectively. Genes encoding designs with the 12 highest assembly scores and 6 highest component scores were ordered for low throughput analysis. Because the time from bacterial lysis to freezing of the SEC fractions was less than two hours, minimal aggregation was found to have occurred so no samples representing the aggregation failure mode were synthesized for low-throughput validation.

3.3.4 Low throughput follow-up of hits and misses in protein nanoparticle library

The 18 genes synthesized for low-throughput validation were expressed in 500 mL cultures, purified by Ni-NTA chromatography and then run on SEC. Few designs contained tryptophans so only the 230 nm absorbance as measured on SEC was used to compare against peptide barcode intensities for each design. The 6 designs with high component scores had barcode intensities that closely matched the 230 nm absorbance trace acquired by SEC [Fig 3.3 M-R]. The 12 designs with high assembly scores had lower agreement between barcode intensities and the 230 nm SEC absorbance trace. However, The 230 nm absorbance traces were consistent with some form of higher order intramolecular assembly as determined by SEC migration. Dynamic light scattering (DLS) followup on the SEC purified samples verified a higher assembly size but also higher polydispersity relative to trimers alone [Supplementary Fig. 3.1 A]. SDS-PAGE of SEC purified protein nanoparticles showed lower molecular weight contaminants that may have been poisoning nanoparticle assembly in the low throughput experiments. SDS-PAGE of cells lysed by boiling in SDS showed that the low molecular weight contaminants were present in the *E. coli* before lysis by microfluidizer [supplementary Fig. 3.2 A-B].

3.3.5 Correction of non-canonical start codons in protein nanoparticle library

There have been previous reports of degenerate ribosomal binding sites (RBS) followed by TTG and GTG codons acting as alternate start codons in de novo proteins[33]. Retrospective analysis of the constant region of the protein nanoparticle library revealed several RBSs followed by non canonical start codons that were consistent with the detected degradation products. Four nanoparticles with high assembly scores were ordered with all TTG codons silently mutated to TTT and all GTG codons silently mutated to GTT. Upon purification all four showed a reduction in degradation products and one of the samples showed intact nanoparticles by negative stain

electron microscopy (nsEM) [Fig. 3.3 S]. Two-dimensional and three-dimensional averaging of the micrographs yielded structures that closely matched the design model [Fig. 3.3 T-U].

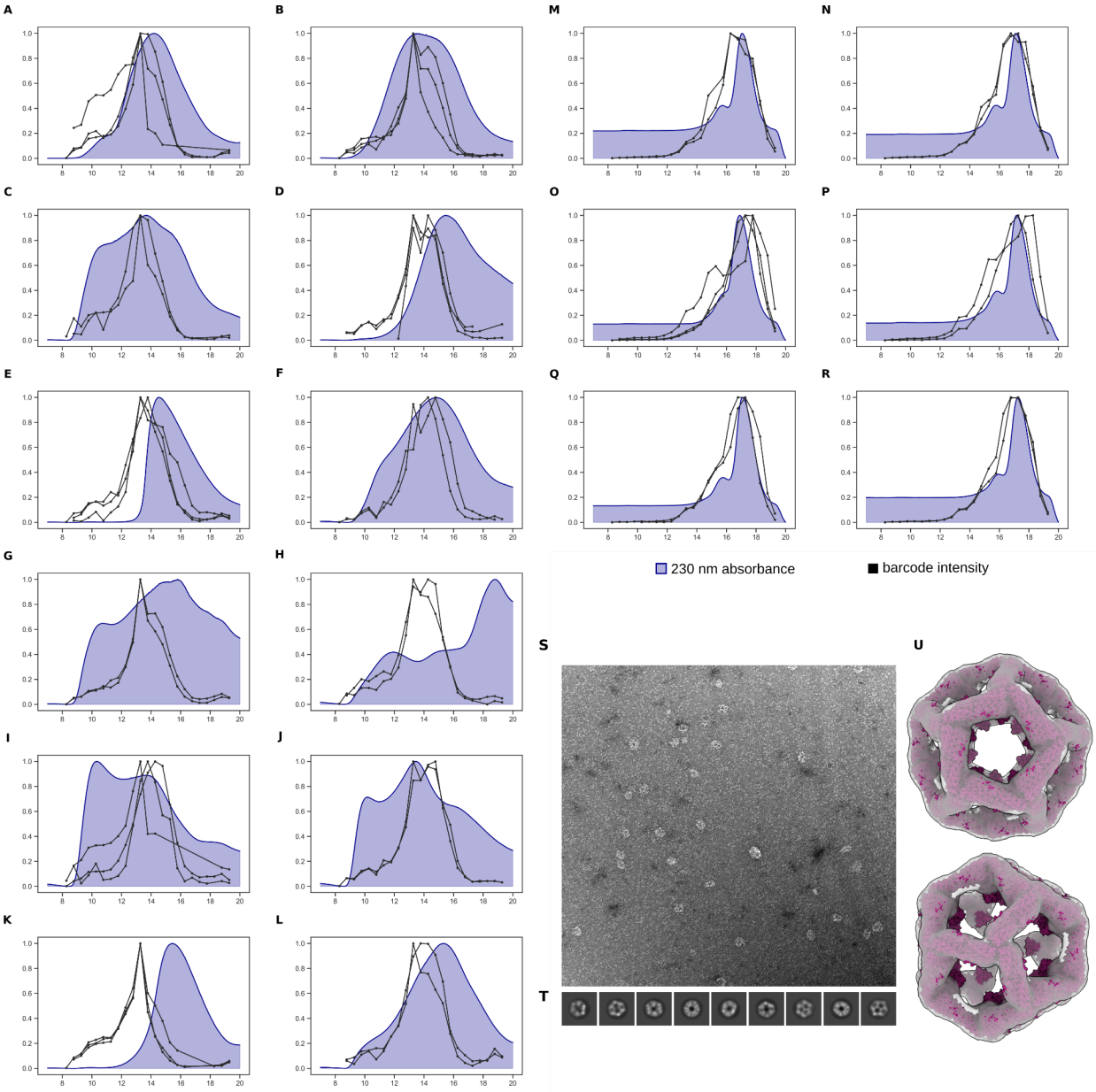


Figure 3.3. Low throughput characterization of nanoparticle library hits agrees with high throughput data. (A-L) Expected nanoparticle assemblies with low throughput 230 nm absorbance trace by SEC (blue) with two or more peptide barcode intensity traces overlaid. (M-R) Expected non-assembling component with low throughput 230 nm absorbance trace by SEC (blue) with two or more peptide barcode intensity traces overlaid. (S) Negative stain electron micrographs with 2D (T) and 3D (U) averages overlaid on the design model.

3.4 Discussion

The small dataset of 11 nanoparticles with 8 barcodes each demonstrates that fusion of a peptide barcode will typically not modulate the underlying assembly and SEC migration patterns of protein nanoparticles. Furthermore, it shows that SEC traces can be accurately recapitulated, at least in small sample sets of approximately 100 barcodes. Our ability to differentiate the partially assembling pHW2.2 from the efficiently assembling I3-01 and I53-50_v4 indicates that this method can distinguish between partial and complete assemblies. An important caveat of barcoding two-component nanoparticles is that the barcode will migrate with only one of the two subunits, giving incomplete information about assembly characteristics.

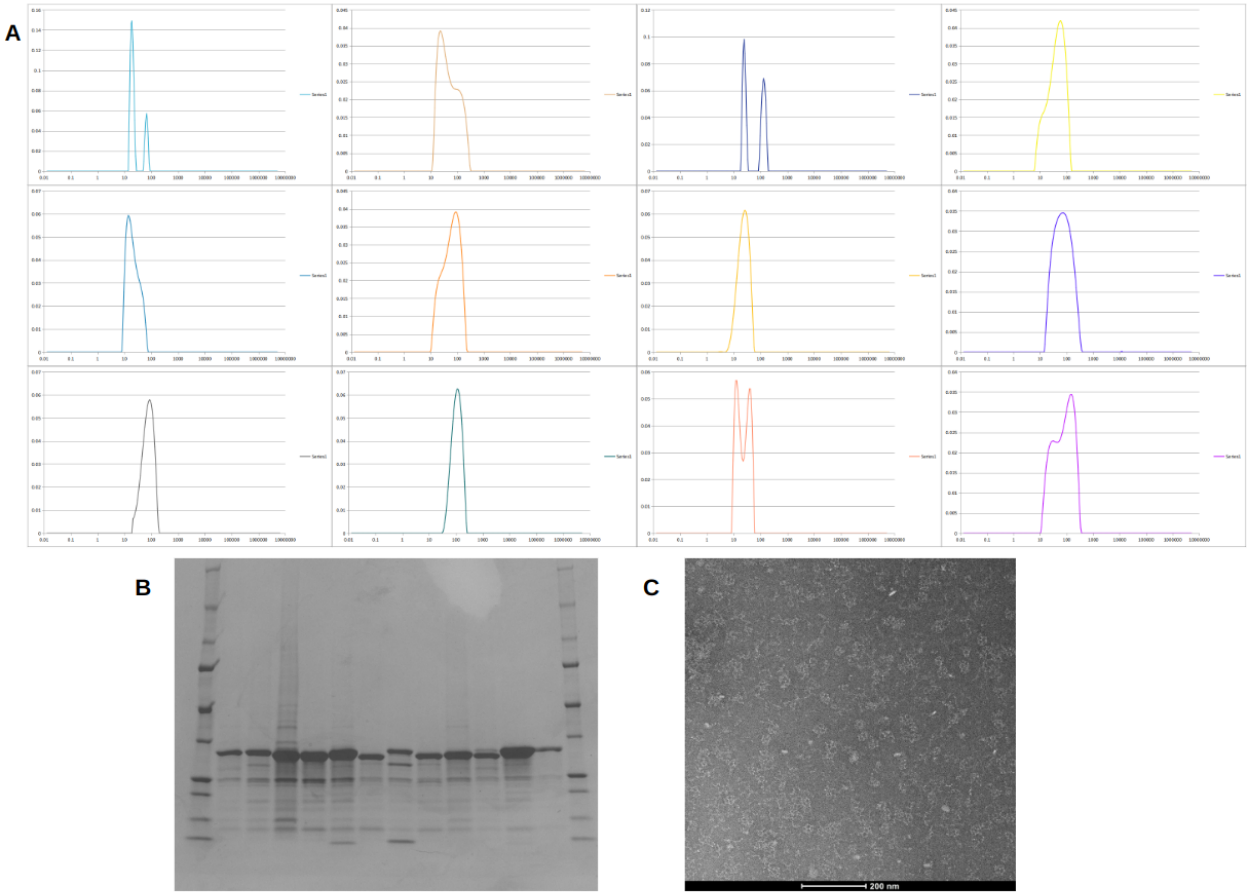
Because the internal start sites occurred at multiple loci in the constant region of the protein nanoparticle library,it is unlikely that any protein nanoparticles in the library formed complete assemblies. This makes it difficult to identify designs that would otherwise be complete assemblies. However, it is likely that the partially assembled nanoparticle fraction is enriched for designs that are capable of complete assembly when alternate start sites are removed. Moreover, unassembled components were tightly correlated with their parent trimer pH192_0431, suggesting that this method is highly effective at identifying lower order assemblies even in highly complex protein mixtures. High assembly score designs were also highly enriched in designs with a contact molecular surface area above 200 suggesting that this is a good minimum interface size for protein nanoparticle designs in the future[30].

The multiplex PCR assembly of oligonucleotide pools has been previously successful when using NGS readouts. NGS methods can generally tolerate high rates of off-target gene synthesis as the readout is generally capable of detecting any deviations from the designed sequence and

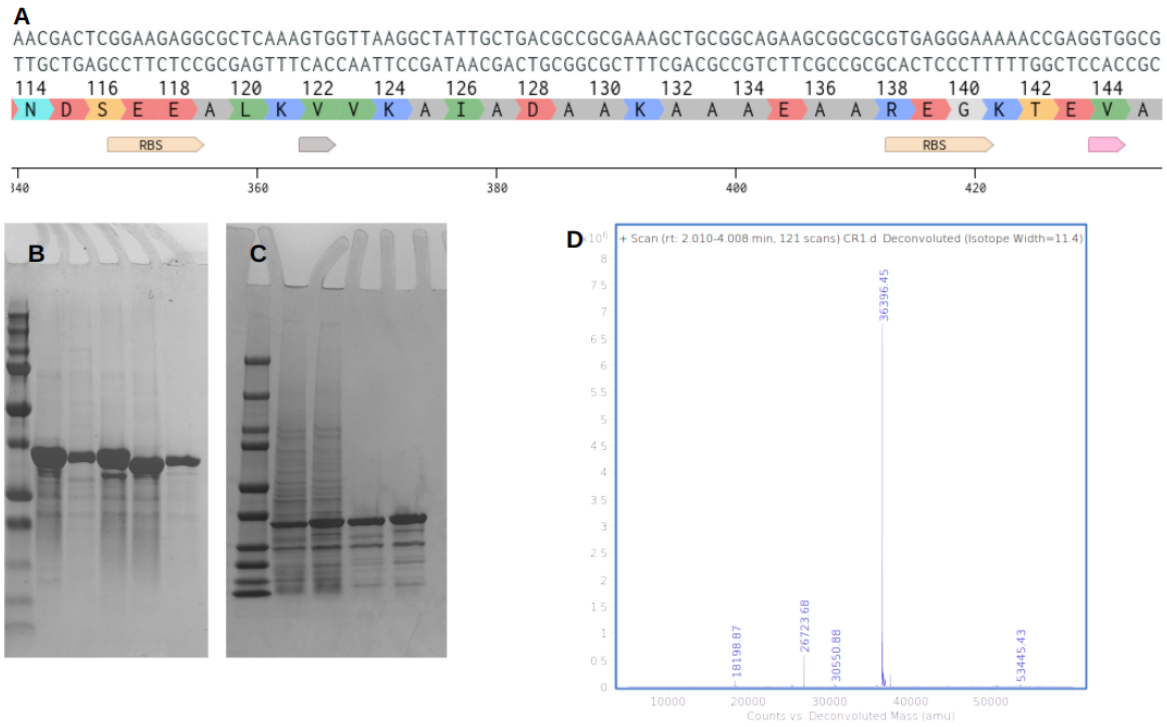
correlating those deviations with outputs. Because the peptide barcodes are orthogonal to any sequence deviations in other parts of the assembled library they are likely to be less tolerant of low library quality. However, we recovered a working nanoparticle design despite less than 10% of the library being correct design models associated with their respective barcodes. This suggests that this method is still relatively tolerant of point mutations. A likely cause of this is a high transformation efficiency relative to the number of designs in the library. This will result in any degenerate point mutants or chimeras being far less represented than the intended design model. However, it remains to be seen what the maximum complexity can be for a peptide barcoded de novo protein library such as this.

Protein nanoparticle design methods have been rapidly evolving and success rates for de novo protein nanoparticle design have been rapidly increasing. This trend is driven both by library efforts such as this one, by efforts in low throughput nanoparticle design and characterization, and by the refinement of de novo protein design methods in general. With increasing success rates in general protein nanoparticle design library methods are no longer strictly necessary, but library methods will likely remain useful in more complex protein nanoparticle design such as flexible and state switching protein nanoparticles.

3.5 Supplemental information



Supplementary Figure 3.1. Additional characterization of expected nanoparticle assemblies from the peptide barcoding library. (A) DLS of the 12 expected assemblies. (B) SDS PAGE of all 12 expected protein nanoparticle assemblies. (C) Negative stain electron micrograph of example nanoparticle assembly before correction of the alternate start sites.



Supplementary Figure 3.2. Identification of alternate start sites in the constant region of the I3 nanoparticle library. (A) Sequence view showing two internal RBSs and alternate start codons in the consta region. (B) SDS-PAGE after mutation to remove alternate start codons and (C) before removal of the alternate start codons. (D) LC-MS validating the molecular weight of a corrected assembling nanoparticle.

Chapter 4. Engineering an RNA encapsulating protein nanoparticle, I53-50_v3, for cellular RNA delivery

4.1 Abstract

We have previously reported an RNA packaging protein nanoparticle, and a previous graduate student has collected promising data on functionalized variants of this nanoparticle for RNA delivery. Here we report the evaluation of several RNA packaging protein nanoparticles for efficient RNA encapsulation and aggregation free assembly. We selected I53-50_v3IY for further stabilization and assembly optimization. Stabilized variants of this construct were then shown to be efficiently endocytosed by HeLa cells and deliver a small but significant amount of pegRNA for prime editing in a HeLa reporter cell line.

4.2 Introduction

Nucleic acid packaging and protection is another essential feature of any protein nanoparticle intended for the delivery of nucleic acids. We have previously reported an RNA encapsidating protein nanoparticle, I53-50_v3[8]. Additionally other members of our lab have engineered two other protein nanoparticles for encapsulation of non-self RNA, I53-565 and O432-59 (unpublished). Prior work has been done in the Baker Lab focusing on fusing pH sensitive proteins to the exterior of I53-50_v3, I53-50_v3IY with promising initial data for delivery of prime editor guide RNA in a reporter cell line[34,35].

While endosomally responsive state switching is likely necessary for both favorable biodistribution and efficient nucleic acid delivery, it is desirable to validate endosomolytic

domains independently of state switching materials before combining the two. While the pathways for nucleic acid release from non-pH-responsive protein nanoparticles are not obvious, it is possible that the protein nanoparticle will be degraded in a way that leaves the nucleic acid cargo intact. Here we report the evaluation of several protein nanoparticles for assembly efficiency and nucleic acid cargo packaging and protection from nucleases, as well as the characterization of an optimal candidate for cellular entry and nucleic acid delivery.

4.3 Results

4.3.1 Comparison of previously reported nucleic acid packaging protein nanoparticles

We selected 3 previously characterized self-assembling protein nanoparticles based on the availability of termini for the fusion of cellular targeting domains and endosomolytic domains for cytosolic delivery as well as their non-porosity and prior success at packaging and protecting nucleic acid cargo. I53-50_v3 was chosen as it had initially been reported to efficiently package and protect its own mRNA genome and was supported by further efforts and promising initial data suggesting that variants of this nanoparticle were capable of delivering prime editor guide RNA in a cell based assay. O432-59 was also selected as antibodies could be used to modularly target many different cellular receptors and initial efforts had been made to develop it as a nucleic acid packaging platform. I53-565 was selected as a non-porous nanoparticle incorporating pH responsive and moderately endosomolytic trimers as one of its subunits though it had not currently been engineered to package RNA.

I53-565 was composed of a pH responsive trimer and a pentamer that was known to express poorly and aggregate. The surface of the pentamer was redesigned using a message passing

neural network for protein design in the hopes of improving solubility. In addition, multiple loci on the interior of the protein nanoparticle were mutated to arginine and lysine to allow for electrostatic packaging of nucleic acids. Plasmids encoding 8 surface redesigned pentamers with positively charged interior surfaces were ordered as well as a plasmid encoding a trimer carrying positively charged interior surface mutations. Purification by hexahistidine tag chromatography and subsequent SDS-PAGE revealed moderate improvements in solubility and an unexpected lower molecular weight species in both the original I53-565 pentamer and its redesigned variants.

To package nucleic acids, the subunits for I53-565 and O432-59 were mixed at 1 μM nanoparticle concentrations with either no RNA, 5 μM prime editor guide RNA (111 nucleotides) or 5-methoxyuridine modified firefly luciferase RNA (1929 nucleotides). As per previously developed assembly protocols, these materials were mixed in a high salt and high CHAPS (zwitterionic detergent) buffer at pH 8.0 and dialyzed into Tris-buffered saline (TBS) (50mM TrisHCl pH 8.0, 150mM NaCl) at room temperature overnight. I53-50_v3 was assembled at 1μM nanoparticle concentration with the same RNA conditions as above for 30 minutes at room temperature in TBS. Visible precipitation occurred in the I53-565 and O432-59 assemblies. Assemblies were clarified by centrifugation and the pellet and supernatant were run on SDS-PAGE to analyze the extent of the precipitation [Fig. 4.1 A-D]. I53-565 was found to have precipitated entirely and was not detected in subsequent analysis. More than 50% of O432-59 appeared to have precipitated but some material remained in the soluble fraction and was sufficient for subsequent analysis. Little to no precipitation was detected in I53-50_v3. The soluble fraction of each assembly was treated with 20 units of benzonase for 30 minutes at room

temperature before a native agarose gel and DLS [Fig. 4.2 A-D]. The native gel showed prime editor guide RNA packaging and protection in both I53-50_v3 and O432-59. In the presence of benzonase, both nanoparticles appeared to package and protect mRNA. However, in the absence of benzonase, nanoparticles appeared to assemble inefficiently but despite inefficient assembly, mRNA appeared to be protected and migrated far more rapidly. DLS on non-benzonase-treated protein nanoparticles confirmed increased protein nanoparticle heterogeneity when assembled with mRNA.

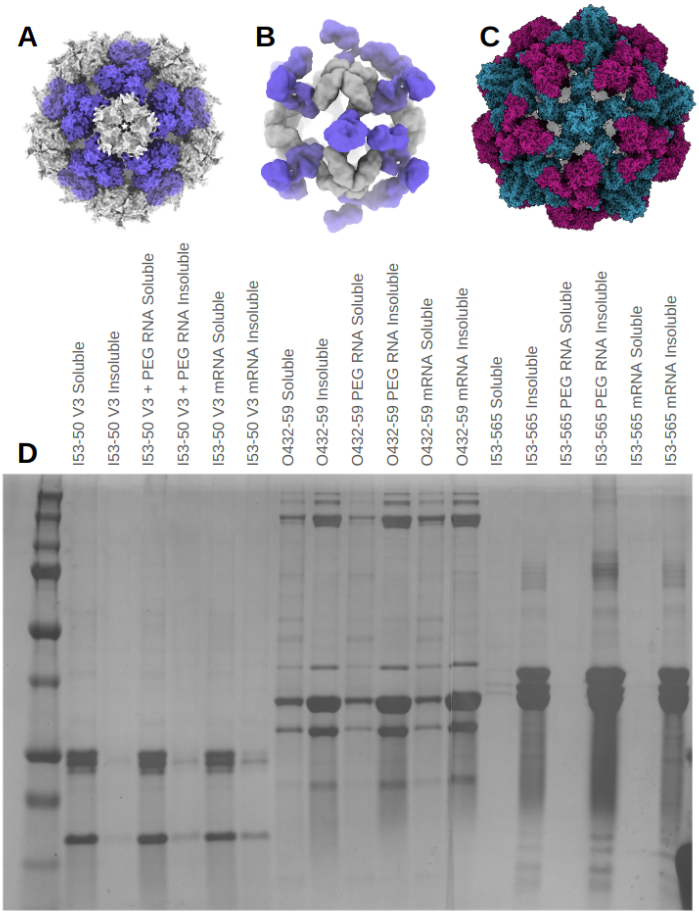


Figure 4.1. Existing self-assembling protein nanoparticles showed varying degrees of aggregation when undergoing in-vitro assembly with and without RNA. (A) Model of I53-50 V3 nanoparticle structure. (B) Model of O432-59 nanoparticle structure(C) Model of I53-565 nanoparticle structure (D) SDS-PAGE of soluble and insoluble fractions of all three candidate protein nanoparticles mixed with either no RNA, ~100bp prime editor guide RNA (PEG RNA), or ~1900bp firefly luciferase mRNA (mRNA).

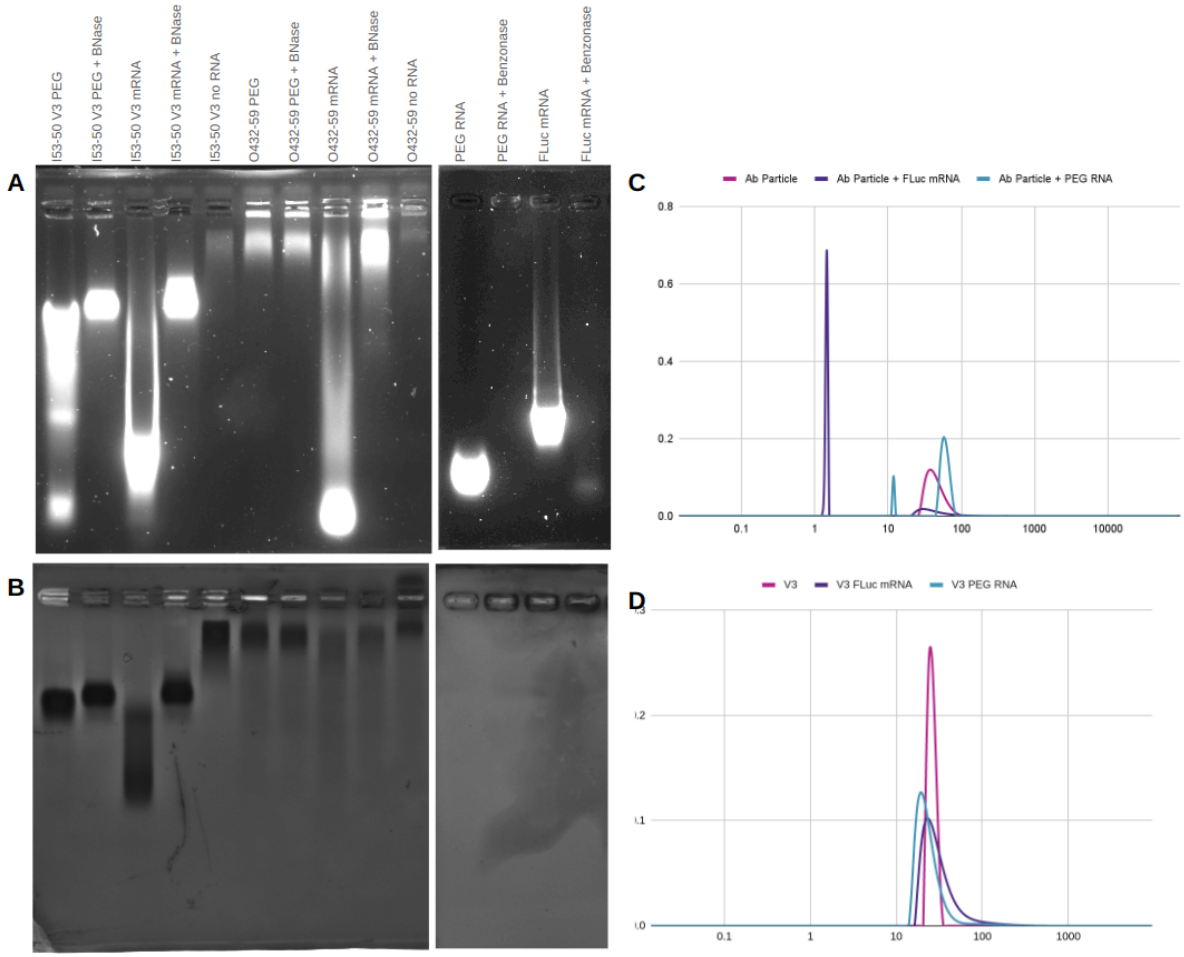


Figure 4.2. I53-50 V3 and O432-59 appear to efficiently package and protect short but not long RNA. (A) 0.8% agarose gel loaded with various native protein nanoparticle assemblies and stained with SYBR Gold nucleic acid dye. (B) The same 0.8% agarose gel as in A stained with coomassie blue protein stain (C) DLS of O432-59 (Ab Particle) assembled with no RNA, PEG RNA, and FLuc mRNA but not treated with benzonase nuclease (D) DLS of I53-50 V3 nanoparticle assembled with no RNA, PEG RNA, and FLuc mRNA.

4.3.2 Optimization of expression and assembly of I53-50_v3

Plasmids encoding I53-50_v3IY were obtained from a previous graduate student. These constructs included I53-50_v3IY trimer, I53-50_v3IY_T; pentamer, I53-50_v3IY_P; trimer fused to a single chain variant of pH192, I53-50_v3IY_T_scpH192; trimer fused to a transferrin receptor minibinder, I53-50_v3IY_T_TFRmb; and trimer fused to an epidermal growth factor receptor minibinder, I53-50_v3IY_T_EGFRmb. Notably the v3 pentamer subunit does not tolerate minibinder fusions to either of its externally facing termini (data not shown).

All components were expressed in 500mL cultures and purified by Ni-NTA chromatography. SEC of the eluate and subsequent non reducing SDS-PAGE showed that the individual components were polydisperse and many of the trimeric components had formed aberrant intersubunit disulfide bonds, inconsistent with the design model [Fig. 4.3 A-H, N]. SEC purified components were assembled overnight at room temperature and the resulting assemblies were run on DLS and nsEM. DLS showed significant polydispersity and nsEM micrographs revealed significant flocculation and aggregation [Fig. 4.3 O]. Sanger sequencing of the received constructs revealed non-cleavable decahistidine tags and 15-20 amino acid glycine-serine linkers for the minibinder fusions. Additionally, a methionine appeared to be unintentionally retained in a helical extension of the trimer subunit which may have further destabilized the component.

We designed and ordered new variants of each trimer with tobacco etch virus (TEV) protease cleavable hexahistidne tags, shortened 5 residue glycine serine linkers, replacement of all cysteines in the trimer with alanines, and removal of the potentially destabilizing methionine insertion. The stabilized versions of I53-50_v3IY_P, I53-50_v3IY_T, I53-50_v3IY_T_scpH192, I53-50_v3IY_T_TFRmb, I53-50_v3IY_T_EGFRmb were named v3s_P, v3s_T, v3s_T_scpH192, v3s_T_TFRmb, and v3s_T_EGFRmb respectively. An additional trimeric subunit fused to an EGFR DARPin, v3s_T_EGFRda, was ordered. Stabilized components were expressed in 500mL cultures, purified by Ni-NTA chromatography and hexahistidine tags were cleaved with TEV protease. Subsequent SEC purification showed monodisperse peaks for all stabilized components except for v3s_T_scpH192 [Fig. 4.3 I-M]. Negative stain electron micrographs of overnight assemblies of stoichiometrically mixed v3s_T and v3s_P showed

reduced aggregation but a large population of off-target assembly states [Fig. 4.3 P]. Other variants of the I53-50 nanoparticle benefitted from an increased concentration of CHAPS detergent, 0.375% instead of 0.1875%, and assembly overnight at 37°C rather than room temperature. Assembly of v3s_T and v3s_P under such conditions generated highly monodisperse nanoparticles by nsEM [Fig 4.3 Q].

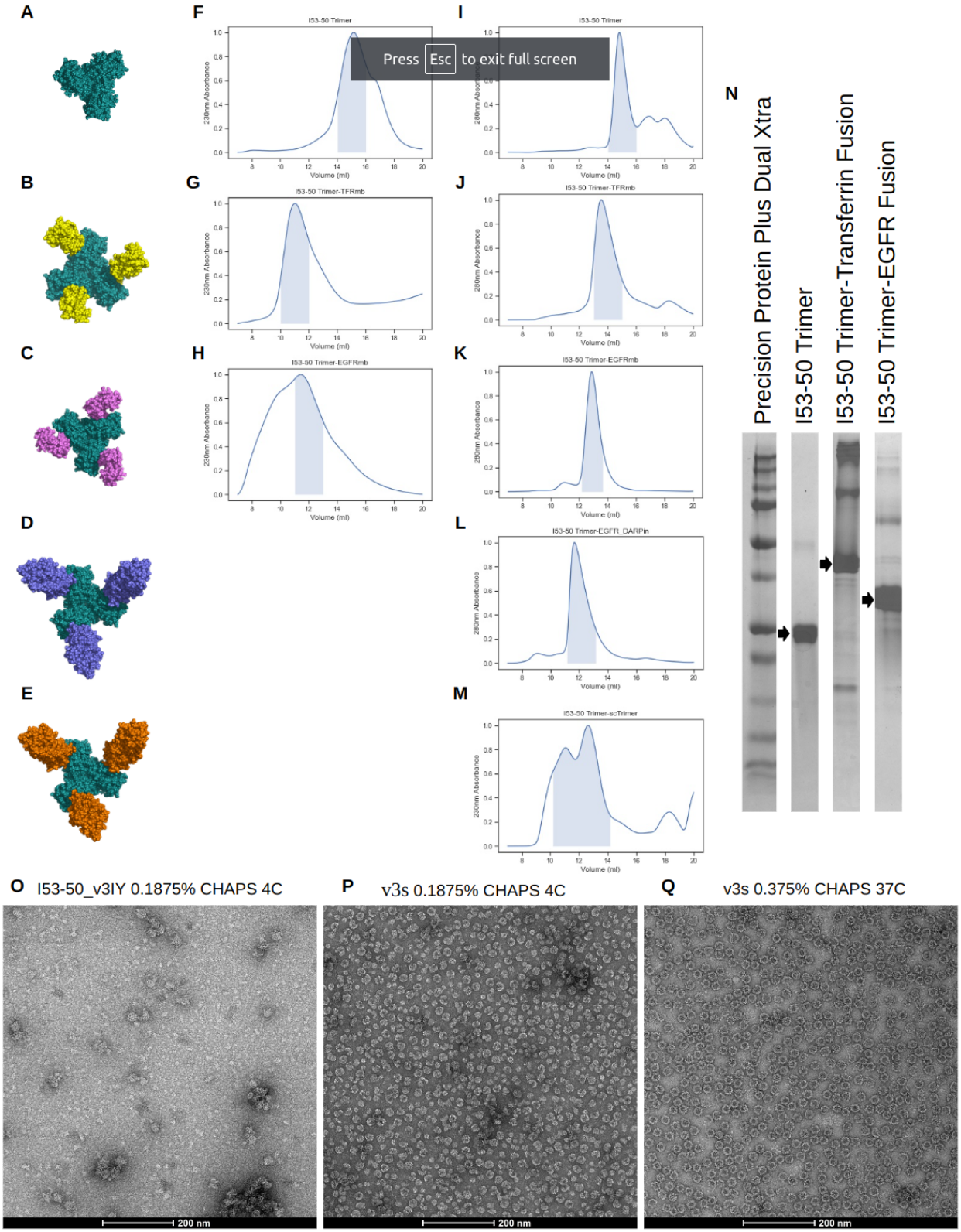


Figure 4.3. Stabilization of I53-50 variant subunits and optimization of I53-50 variant assembly conditions. (A-E) Models of I53-50 trimer alone and fused to TFR minibinder, EGFR minibinder, EGFR DARPin, and scpH192 respectively. (F-H) SEC traces of non-stabilized trimer designs with fractions collected for later assembly highlighted in blue. (I-M) SEC traces of stabilized I53-50 trimer variants with fractions collected for later assembly highlighted in blue. (N) Non-reducing SDS-PAGE of various I53-50 trimer variants with precision protein plus dual xtra ladder. (O-Q) Negative stain electron micrographs of assemblies of various I53-50 nanoparticle variants under increasingly optimized conditions.

4.3.4 Cellular entry of I53-50_v3 nanoparticles displaying protein minibinders

We next sought to evaluate the ability of minibinder displaying protein nanoparticles to be efficiently endocytosed by mammalian cells. To this end, we designed and ordered an I53-50 trimer variant with an internally facing green fluorescent protein(GFP), v3s_T_GFP. We assembled 60μM v3s_P, 15 μM v3s_T, and 30 μM v3s_T_GFP with 15μM of either v3s_T, v3s_T_TFRmb, v3s_T_EGFRmb, or v3s_T_EGFRda. SEC purification of assemblies with 480 nm absorbance measurements validated that GFP was associated with the nanoparticle assemblies although a significant component peak was found as well suggesting that fewer than half of the trimeric components of each nanoparticle included fused GFP [Fig. 4.4 A]. Notably, the v3s_EGFRmb containing assembly had significantly reduced nanoparticle assembly efficiency by SEC potentially due to aggregation related to the hydrophobic surface of the fused minibinder. Nevertheless, DLS and nsEM of the SEC purified nanoparticle fraction showed highly monodisperse nanoparticle assemblies in all cases [Fig. 4.4 B-C].

HeLa cells were transfected with 100 nM final concentration of assemblies containing either no minibinder fusions, v3s_T_TFRmb, or v3s_T_EGFRda for 45 minutes then washed with PBS, fixed, and stained with DAPI and anti Rab7 antibodies. Subcellular distribution of GFP containing nanoparticles was determined with fluorescence microscopy [Fig 4.4 D-G]. Both v3s_T_TFRmb and v3s_T_EGFRda containing nanoparticles appeared to be efficiently endocytosed into intracellular vesicles. The GFP signal did not appear to co-localize with anti Rab7 antibodies suggesting that the endocytic vesicles containing the GFP nanoparticles were not late endosomes. The nanoparticle assembly containing no displayed minibinders showed

significant fluorescence in the vicinity of the plasma membrane and extracellular matrix but reduced presence in endocytic vesicles.

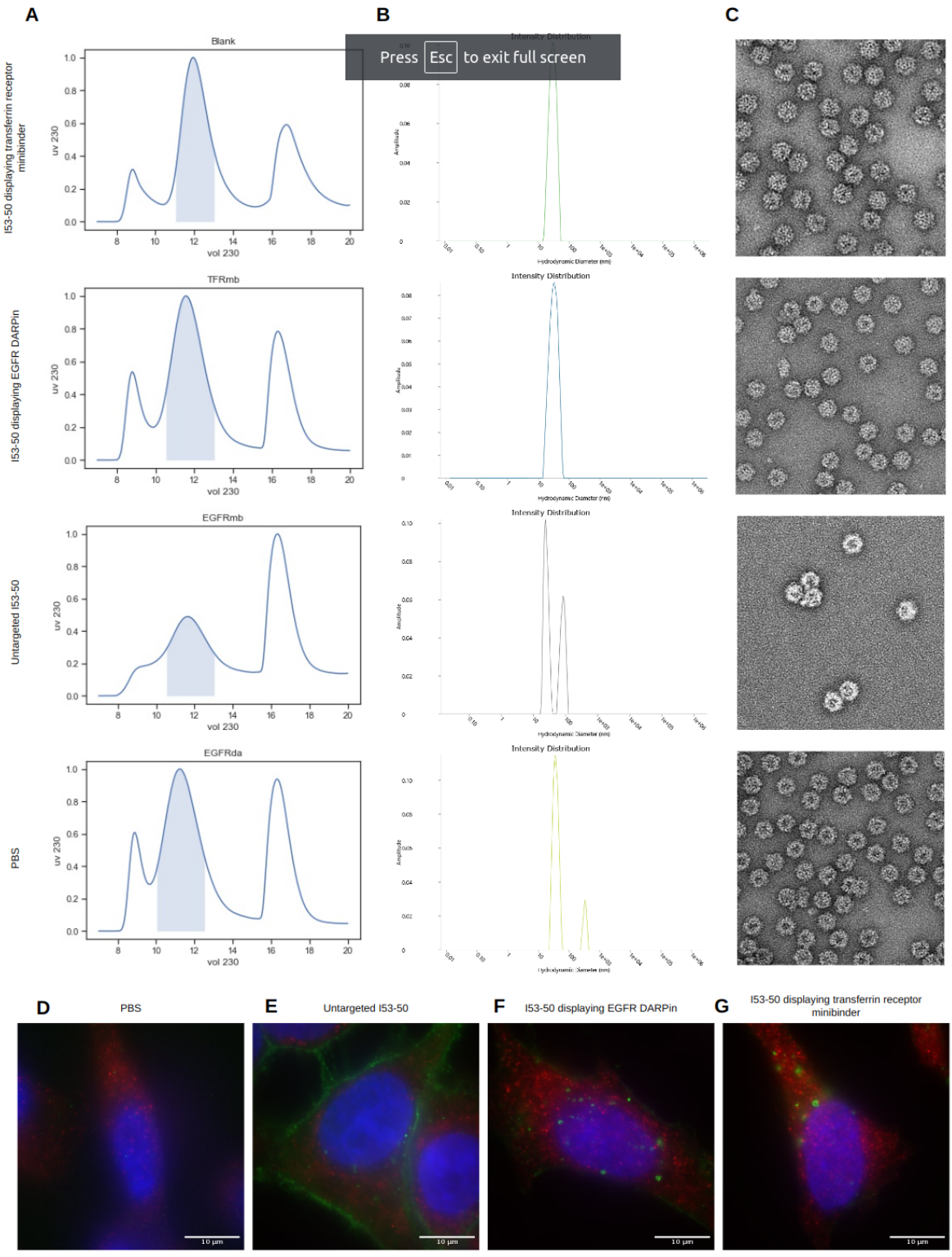


Figure 4.4. GFP labeled minibinder displaying v3s nanoparticle variants are efficiently taken up by HeLa cells. (A) SEC traces of each nanoparticle assembly with the fractions collected for characterization and cellular delivery highlighted in blue. (B) DLS of the selected SEC fractions. (C) Negative stain electron micrographs of v3s nanoparticle variants. (D-G) Fluorescence microscopy of v3s nanoparticle designs (green) overlaid on anti Rab7 antibody (red) and DAPI nuclear stain (blue).

4.3.5 Modest RNA delivery with functionalized chimeric I53-50_v3 nanoparticles

To evaluate the efficacy of v3s nanoparticle variants to deliver nucleic acids we sought to deliver prime editor guide RNA(pegRNA) to HeLa cells transduced with a previously reported frameshift prime editing assay[34]. Assemblies were made of 60 μM v3s_P, 15 μM v3s_T, either 30 μM v3s_T_scpH192 or 30 μM v3s_T, either 15 μM v3s_T_TFRmb, or 15 μM v3s_T, and 60μM pegRNA. An additional nanoparticle assembly was made of 60 μM v3s_T, 60 μM v3s_P, and 60 μM pegRNA without cleaved hexahistidine tags to evaluate the role of such tags in endosome escape. SEC showed monodisperse peaks for all constructs except for v3s_T_scpH192 containing constructs which had a significant void peak [Fig. 4.5 C-F]. In the case of these constructs the expected void peak and cage peak were taken for downstream analysis. HeLa reporter cells were then transfected with plasmids encoding the prime editor protein and incubated for 24 hours to allow for prime editor protein expression. Transfected cells were then treated with a 100nM final concentration of each nanoparticle assembly and cultured for 72 hours without passaging. A pegRNA-only negative control and a lipofectamine pegRNA positive control were included as well. Flow cytometry was then used to assess the loss of red fluorescence and gain of green fluorescence associated with pegRNA delivery and subsequent prime editing. After gating for gain of green fluorescence, the pegRNA transfection positive control showed 19.2% editing of reporter cells [Fig. 4.5 A]. The pegRNA only negative control showed 0.00472% editing of reporter cells [Fig. 4.5 B]. Nanoparticle containing only v3s_T and v3s_P showed 0.29% editing efficiency while the version of this nanoparticle without cleaved hexahistidine tags showed 0.37% editing efficiency, suggesting a moderate effect caused by the hexahistidine tags [Fig. 4.5 G-H]. The expected nanoparticle peak of v3s_T_TFRmb, v3s_T_scpH192 containing nanoparticles showed 0.5% editing efficiency while the expected

aggregate showed 0.37% editing efficiency [Fig. 4.5 I-J]. Hexahistidine tag cleavage was validated by mass spectrometry [Supplementary Fig. 4.1 A-C].

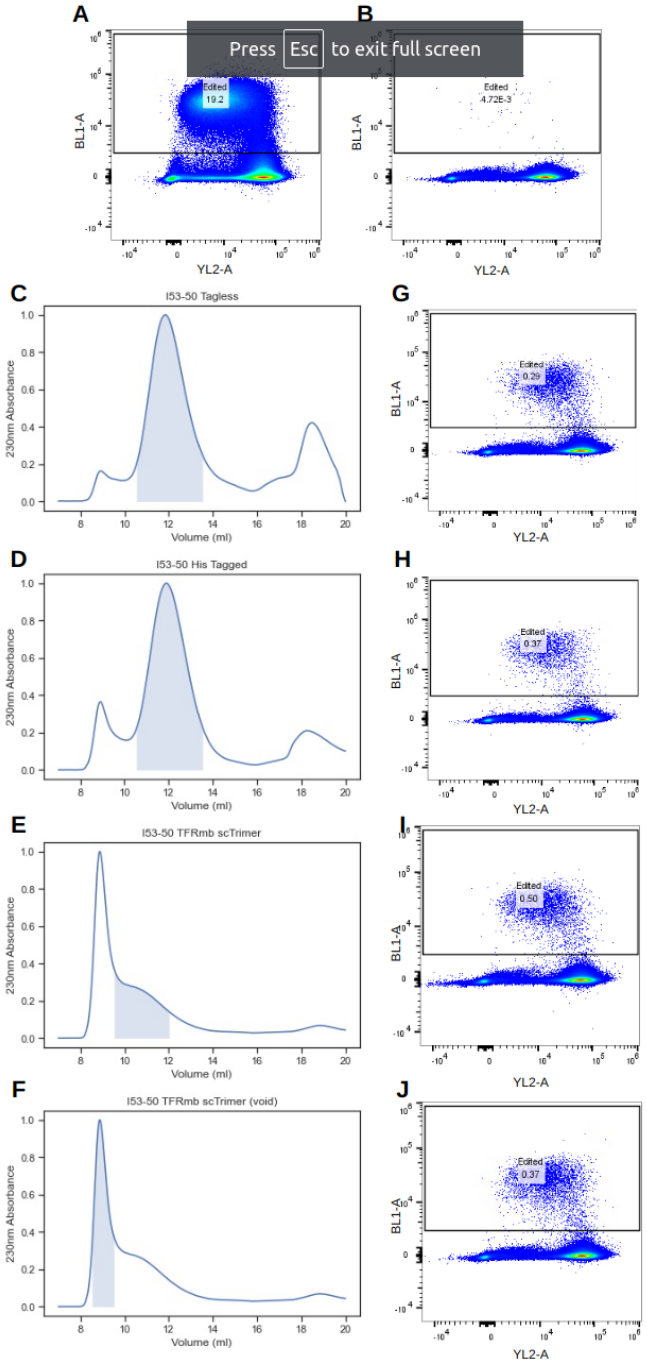


Figure 4.5. Delivery of pegRNA to a frameshfit HeLa reporter cell line using v3s nanoparticle variants. (A) Flow cytometry of positive control of cells transfected with pegRNA using XFect. (B) Flow cytometry of negative control of cells treated with pegRNA suspended in PBS. (C-F) SEC traces of v3s nanoparticle assemblies with the fractions collected for cellular delivery highlighted in blue. (G-J) Flow cytometry of v3s nanoparticle variants with packaged pegRNA.

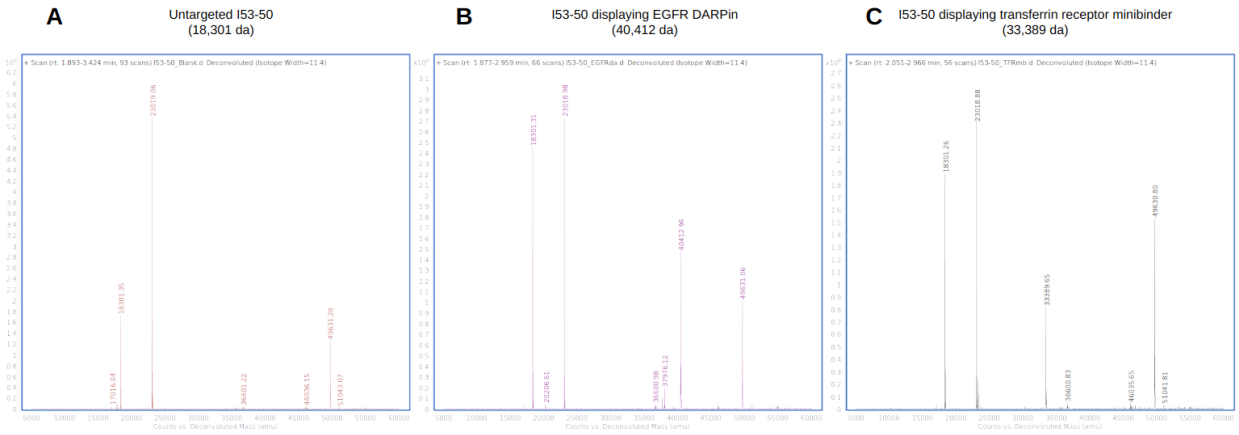
4.4 Discussion

Here we demonstrated that RNA delivery with protein nanoparticles is several orders of magnitude more efficient than unencapsulated RNA but remains several orders of magnitude less efficient than commercially available transfection reagents. It is nevertheless exciting that we have a signal upon which we can optimize. Low but detectable levels of RNA delivery could provide an excellent starting point for a directed evolution assay for RNA delivery. This is particularly true for the I53-50 nanoparticle evaluated in these experiments as it has already been shown to have a genotype phenotype linkage through encapsidation of its own RNA genome[8].

We also demonstrated that the flexible fusion of protein minibinders to the surface of protein nanoparticles can result in efficient endocytosis of the nanoparticles. Even in the absence of endosome escape this could be useful in the delivery of small molecule therapeutics with endosomally cleavable linkers. Notably, the small molecule cargo could be kept in the interior lumen of the nanoparticle. This is likely to allow for large amounts of drug loading without affecting the biodistribution of the nanoparticle. The non-specific association of non-targeted v3s nanoparticles is explicable in that this variant of I53-50 does not include mutations found to improve biodistribution in mice. Electrostatic modeling of the exterior of v3s and the extended half-life I53-50_v4 variant shows an increased exterior negative charge suggesting that the extended half-life and non-specific cell association may be connected (data not shown). Nanoparticle variants incorporating the I53-50_v4 mutations were not developed as these mutations were also found to form off-target nanoparticle assemblies making in vitro nanoparticle assembly extremely difficult(data not shown).

The significant amounts of aggregation observed when comparing nanoparticle assemblies, difficulties packaging longer RNA, and extensive optimization needed to obtain well behaved, monodisperse nanoparticles highlights the difficulties in working on many existing protein nanoparticles. Many of these systems incorporate natural proteins that did not evolve to be structural proteins. Further development of existing nanoparticles or the design of novel, more well behaved protein nanoparticles is needed.

4.5 Supplemental information



Supplementary Figure 4.1. LC-MS validation of hexahistidine tag cleavage for v3s variants. LC-MS was collected on (A) v3s non-targeted nanoparticle, (B) EGFR DARPin displaying v3s nanoparticle, and (C) TFR minibinder displaying v3s nanoparticle.

Chapter 5. A novel non-porous protein nanoparticle

5.1 Abstract

Many self-assembling protein nanoparticles have been previously reported. However many of these protein nanoparticles cannot be assembled in vitro due to either a single component architecture or due to prohibitively unstable protein subunits. Here we report the design and characterization of I53-0, a novel protein nanoparticle with a minimal hydrophobic intersubunit interface for optimal expression and purification of individual components for subsequent in vitro assembly.

5.2 Introduction

Many self-assembling protein nanoparticles have been reported[36]. However, many of these nanoparticles are either one-component nanoparticles that can not be assembled *in vitro* or are two-component nanoparticles in which one or both of the subunits is poorly behaved on its own making nanoparticle preparation very difficult. This poor behavior is likely a result of instability of the natural protein scaffolds used to design the nanoparticle, alternative assembly states exhibited by the natural scaffold that were not successfully ablated, or the large hydrophobic interfaces that are introduced during nanoparticle interface design. Despite these issues, *in vitro* protein nanoparticle assembly promises a highly controllable and versatile product capable of packaging small molecules, proteins, and nucleic acids[37]. De novo protein oligomers are known to be highly stable and may present an opportunity for the development of protein nanoparticles designed from highly stable de novo protein scaffolds[38]. In particular, several *de novo* protein pentamers and trimers were developed using a peptide barcoding strategy that are disclike in

topology and likely to dock into non-porous protein nanoparticles[39]. As previously discussed, much work has also been done in evaluating how strong nanoparticle binding interfaces need to be for efficient nanoparticle assembly. We hypothesized that a protein nanoparticle designed from highly stable *de novo* protein subunits with hydrophobic interfaces tuned to be the minimum required size for nanoparticle assembly would result in a two-component, *in vitro*-assemblable protein nanoparticle with highly stable and well behaved components.

5.3 Results

5.3.1 Low-throughput design of non-porous, self-assembling protein nanoparticles

A set of previously reported disc-like *de novo* protein oligomers was selected as a set of starting component[16,39]. These components were visually selected for topologies that would result in non-porous nanoparticles upon docking. Designs were docked using RPX dock on hierarchical search mode to increase computational efficiency and then filtered for designs with residues from at least three helices within 10 angstroms of the intersubunit interface. Filtered docks were then sequence designed using a layer design script in Rosetta. Briefly, residues in contact with adjacent subunits were mutated to hydrophobic residues and packed tightly using a Monte-Carlo method and pairwise decomposable energy function for residue mutation and rotamer repacking. Hydrogen bonded polar residues were added along the boundary of the interface, and the neighborhood within 10 angstroms of any mutated residue was repacked without sequence changes to allow for small adjustments to side chain orientation. Designed sequences were filtered for a Rosetta computed contact molecular surface area between 200 and 300 to optimize for assembly with minimal exposed hydrophobic surface area in components before assembly.

Designs were also filtered for shape complementarity > 0.4 and visually inspected for interdigitation of hydrophobic residues across the intersubunit interface.

5.3.2 Biophysical characterization of I53-0

Genes encoding each subunit of the 24 designs were ordered, expressed in 500mL cultures, and purified by Ni-NTA chromatography. SEC of the resulting components showed one design, I53-0, in which both components were highly expressed and monodisperse(data not shown). Initial attempts to assemble these components were unsuccessful, but assembly of >60μM of each component in the presence of 5% glycerol and absence of any detergents was found to form a nanoparticle by native gel(data not shown). DLS of the components and nanoparticle assembly showed a high diameter monodisperse material and nsEM micrographs showed a highly monodisperse nanoparticle [Fig. 5.1 A-B]. Two dimensional and three dimensional averaging of nsEM micrographs yielded a structure that closely matched the design model [Fig. 5.1 C-F].

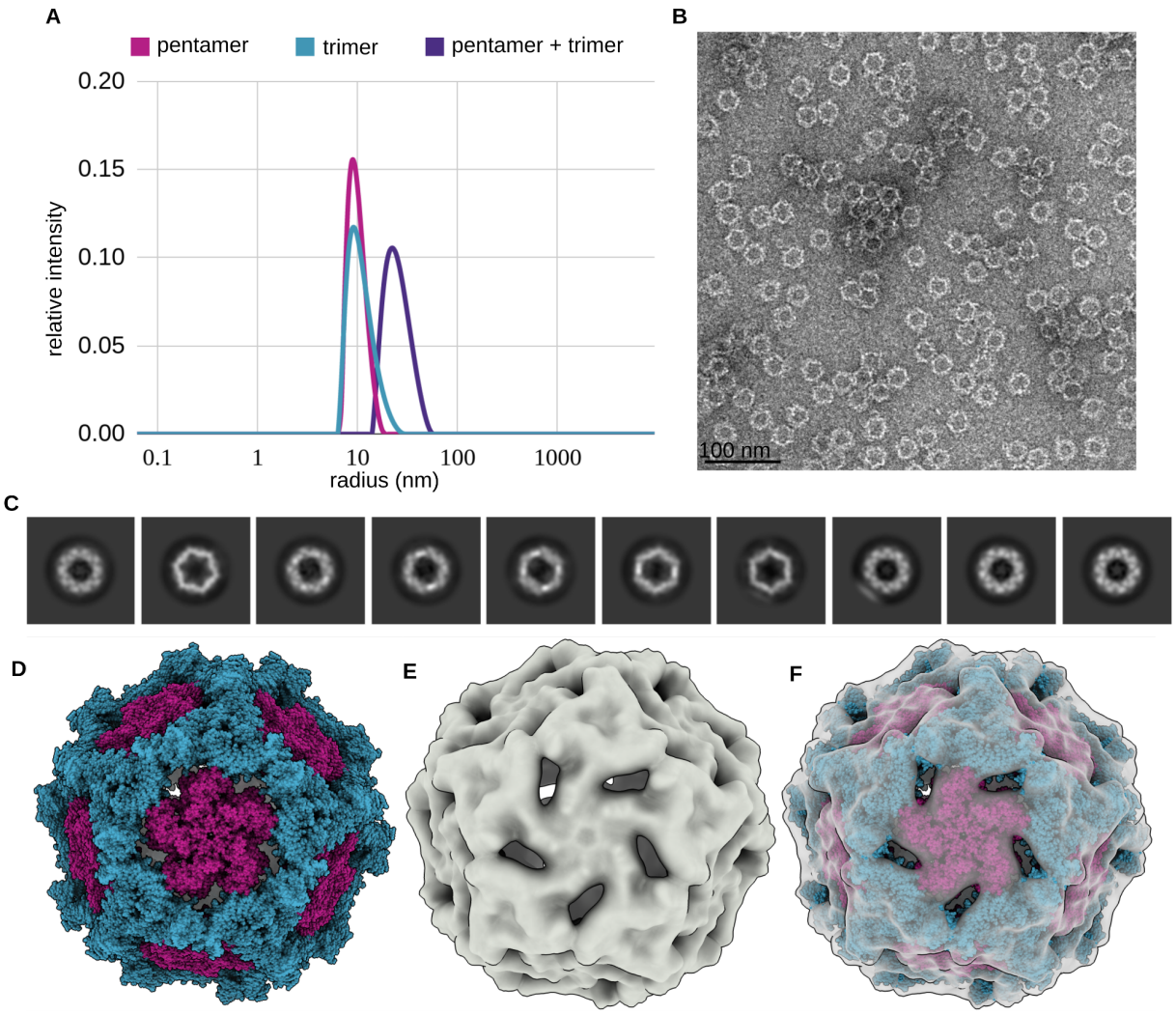


Figure 5.1. Characterization of I53-0, a novel, non-porous self-assembling de novo designed protein nanoparticle. (A) DLS of pentamer (pink) trimer (cyan) and trimer-pentamer mixture (purple). (B) Negative stain electron micrograph of I53-0 nanoparticles. (C) Two dimensional micrograph averages. (D - F) I53-0 design model, 3-D averaging of negative stain micrographs, and the 3-D average overlaid on the design model.

5.4 Discussion

Here we demonstrate that I53-0 is a novel, non-porous and fully de novo self-assembling protein nanoparticle. This nanoparticle has components that can be highly expressed and remain stable when purified individually but still assemble efficiently into a protein nanoparticle. In vitro assembly of this nanoparticle does not require the use of detergents which can be expensive and

difficult to remove. The relative simplicity of handling for the base variant of this nanoparticle rivals that of other nanoparticles that have been extensively developed for production at scale[40].

I53-0 also includes three externally facing termini for fusion of functional domains for modulation of targeting, biodistribution, and effector functions as well as one internally facing terminus that can be used for the fusion of packaging domains, polycysteine for small molecule conjugation, or direct fusion of protein cargo. The pores are smaller when compared with the pores of I53-50 suggesting that this nanoparticle could be engineered for nucleic acid packaging. The intersubunit interface of I53-0 is also quite small. Many protein nanomaterials that have been engineered to be pH responsive aggregate upon disassembly[15]. The large hydrophobic interfaces that are exposed upon disassembly are likely at least partially responsible for this behavior. The relatively small and weak interfaces that make I53-0 subunits well behaved may also reduce aggregation if the interface were designed to be pH responsive making this an interesting material in which to investigate pH sensitivity. These combined features give I53-0 a high potential for future engineering as a nucleic acid delivery vehicle as well as a useful scaffold for many other protein nanomaterials.

Acknowledgements

My PhD would not have been possible without the help and camaraderie of many people along the way.

I am extremely grateful to my postdoc mentors. Joshua Lubner provided many hours of training on nsEM and SEC instruments as well as continuous feedback on my progress in protein design. David Feldman made the mass spec barcoding chapter possible and his help and countless hours of effort working on the mass spec barcoding system with me were essential in the development of that work. Carl Walkey introduced me to Rosetta during my rotation and got me started working on the pH responsive trimer developed by Scott Boyken. Sam Hendel provided invaluable feedback and support when outlining my dissertation and planning my final experiments.

Many other labmates taught me an enormous amount about protein design. In particular I'd like to thank Robby Divine and Yang Hsia for help with protocols for computational rigid helical fusion of the pH trimers. Erin Yang, Quinton Dowling, Alena Khmelinskaia, and Will Sheffler were indispensable in adapting the RPX dock protocol to one component nanoparticle library design. And thanks to TJ Brunette and Stacey Gerben for providing protein oligomers for docking into I53-0. Isaac Sappington and Issa Yousif laid the groundwork for the I53-50 work that I did and provided invaluable insights when optimizing nanoparticle stability and assembly conditions. Thanks to David Chmielewski for helping me learn to use the mass photometer for some last minute measurements. Ryan Kibler and SIBYLS in general made it possible for me to collect high quality SAXS data on the stellate pH responsive trimers as well as providing

essential support analyzing the subsequent data. Jeremiah Simms has also provided invaluable and continuing support for my continued work on the peptide barcoding systems and subsequent analyses.

The IPD is a special place where students can leverage an enormous amount of support from many of the staff scientists on hire. Another indispensable member of the peptide barcoding team is Xinting Li who has spent many many hours running samples on the orbitrap at SLU for me, in addition to learning and doing some of the peptide barcode isolations. I would like to specifically thank Aza Allen for support teaching me to do NGS preparation and analysis. Mengyu Wu, Andrew Borst, and Rebecca Skotheim taught me everything I know about electron microscopy data collection. Stacey Gerben and the general protein production core has provided invaluable support purifying some of the nanoparticle components at scale. Finally, I would like to thank microscopy core Justin Decarreau and postdoc Hojun Choi for their help with immunostaining and fluorescence microscopy.

I would like to thank Will White for always being willing to throw around ideas and troubleshoot random protein design problems despite not having any stake in the outcomes of my experiments. He also provided critical mutual aid during the pandemic when we adapted to severely limited hours by doing simple lab tasks for each other during our incubations. I am not sure if I would have been able to make any significant progress during the pandemic without that support.

I would also like to specifically thank Audrey Olshefsky for help in more ways than I can count. She provided indispensable support in the characterization and design of the nanoparticle library as well as critical advice when optimizing both I53-0 and I53-50 assembly protocols. It would have been much more difficult to complete my PhD without her continued support and patience.

Thank you to my advisor David Baker and shadow advisor Neil King for their support and insight throughout my PhD. I would also like to thank my committee members Suzie Pun and Daniel Ratner for their continued direction and support. Finally, I would like to thank Jennifer Hyde for being willing to trek all the way over from SLU just to watch my back as the GSR of my committee.

References

1. Li, C. & Samulski, R. J. Engineering adeno-associated virus vectors for gene therapy. *Nat. Rev. Genet.* **21**, 255–272 (2020).
2. Mitchell, M. J. *et al.* Engineering precision nanoparticles for drug delivery. *Nat. Rev. Drug Discov.* **20**, 101–124 (2021).
3. Wilhelm, S. *et al.* Analysis of nanoparticle delivery to tumours. *Nature Reviews Materials* **1**, 1–12 (2016).
4. Olshefsky, A., Richardson, C., Pun, S. H. & King, N. P. Engineering Self-Assembling Protein Nanoparticles for Therapeutic Delivery. *Bioconjug. Chem.* (2022) doi:10.1021/acs.bioconjchem.2c00030.
5. Khmelinskaia, A., Wargacki, A. & King, N. P. Structure-based design of novel polyhedral protein nanomaterials. *Curr. Opin. Microbiol.* **61**, 51–57 (2021).
6. Basu Mallik, B., Stanislaw, J., Alawathurage, T. M. & Khmelinskaia, A. De novo design of polyhedral protein assemblies: before and after the AI revolution. *Chembiochem* e202300117 (2023).
7. Tetter, S. *et al.* Evolution of a virus-like architecture and packaging mechanism in a repurposed bacterial protein. *Science* **372**, 1220–1224 (2021).
8. Butterfield, G. L. *et al.* Evolution of a designed protein assembly encapsulating its own RNA genome. *Nature* **552**, 415–420 (2017).
9. Gonzalez, T. J. *et al.* Cross-species evolution of a highly potent AAV variant for therapeutic gene transfer and genome editing. *Nat. Commun.* **13**, 5947 (2022).
10. Walls, A. C. *et al.* Elicitation of Potent Neutralizing Antibody Responses by Designed Protein Nanoparticle Vaccines for SARS-CoV-2. *Cell* **183**, 1367–1382.e17 (2020).

11. Laniado, J. & Yeates, T. O. A complete rule set for designing symmetry combination materials from protein molecules. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 31817–31823 (2020).
12. Van de Steen, A. *et al.* Bioengineering bacterial encapsulin nanocompartments as targeted drug delivery system. *Synth Syst Biotechnol* **6**, 231–241 (2021).
13. Peeler, D. J., Sellers, D. L. & Pun, S. H. pH-Sensitive Polymers as Dynamic Mediators of Barriers to Nucleic Acid Delivery. *Bioconjug. Chem.* **30**, 350–365 (2019).
14. Podinovskaia, M., Prescianotto-Baschong, C., Buser, D. P. & Spang, A. A novel live-cell imaging assay reveals regulation of endosome maturation. *Elife* **10**, (2021).
15. Boyken, S. E. *et al.* De novo design of tunable, pH-driven conformational changes. *Science* **364**, 658–664 (2019).
16. Brunette, T. J. *et al.* Exploring the repeat protein universe through computational protein design. *Nature* **528**, 580–584 (2015).
17. Vulovic, I. *et al.* Generation of ordered protein assemblies using rigid three-body fusion. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).
18. Chen, Z. *et al.* Programmable design of orthogonal protein heterodimers. *Nature* **565**, 106–111 (2019).
19. Boyken, S. E. *et al.* De novo design of protein homo-oligomers with modular hydrogen-bond network–mediated specificity. *Science* **352**, 680–687 (2016).
20. King, N. P. *et al.* Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* **336**, 1171–1174 (2012).
21. Sahtoe, D. D. *et al.* Reconfigurable asymmetric protein assemblies through implicit negative design. *Science* **375**, eabj7662 (2022).
22. Courbet, A. *et al.* Computational design of mechanically coupled axle-rotor protein

assemblies. *Science* **376**, 383–390 (2022).

23. Zlotnick, A. Are weak protein-protein interactions the general rule in capsid assembly? *Virology* **315**, 269–274 (2003).
24. Perlmutter, J. D. & Hagan, M. F. Mechanisms of virus assembly. *Annu. Rev. Phys. Chem.* **66**, 217–239 (2015).
25. Hagan, M. F. & Chandler, D. Dynamic pathways for viral capsid assembly. *Biophys. J.* **91**, 42–54 (2006).
26. Hsia, Y. *et al.* Design of a hyperstable 60-subunit protein dodecahedron. [corrected]. *Nature* **535**, 136–139 (2016).
27. Divine, R. *et al.* Designed proteins assemble antibodies into modular nanocages. *Science* (2021) doi:10.1126/science.abd9994.
28. Hsia, Y. *et al.* Design of multi-scale protein complexes by hierarchical building block fusion. *Nat. Commun.* **12**, 2294 (2021).
29. Egloff, P. *et al.* Engineered peptide barcodes for in-depth analyses of binding protein libraries. *Nat. Methods* **16**, 421–428 (2019).
30. Sheffler, W. *et al.* Fast and versatile sequence-independent protein docking for nanomaterials design using RPXDock. *bioRxiv* 2022.10.25.513641 (2022) doi:10.1101/2022.10.25.513641.
31. Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS One* **4**, e5553 (2009).
32. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**, 614–620 (2014).
33. Dantas, G. *et al.* Mis-translation of a computationally designed protein yields an

exceptionally stable homodimer: implications for protein engineering and evolution. *J. Mol. Biol.* **362**, 1004–1024 (2006).

34. Nelson, J. W. *et al.* Engineered pegRNAs improve prime editing efficiency. *Nat. Biotechnol.* **40**, 402–410 (2022).
35. Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
36. Molino, N. M. & Wang, S.-W. Caged protein nanoparticles for drug delivery. *Curr. Opin. Biotechnol.* **28**, 75–82 (2014).
37. Chen, H., Zhang, S., Xu, C. & Zhao, G. Engineering protein interfaces yields ferritin disassembly and reassembly under benign experimental conditions. *Chem. Commun.* **52**, 7402–7405 (2016).
38. Huang, P.-S., Boyken, S. E. & Baker, D. The coming of age of de novo protein design. *Nature* **537**, 320–327 (2016).
39. Gerben, S. R. *et al.* Design of Diverse Asymmetric Pockets in De Novo Homo-oligomeric Proteins. *Biochemistry* **62**, 358–368 (2023).
40. Boyoglu-Barnum, S. *et al.* Quadrivalent influenza nanoparticle vaccines induce broad protection. *Nature* **592**, 623–628 (2021).