Intracellular Delivery of Functional RNA by Designed Protein Assemblies

Issa George Benna

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2022

Reading Committee:

David Baker, Chair

Neil King

Daniel Ratner

Suzie Pun

Jesse Zalatan

Program Authorized to Offer Degree:

Department of Bioengineering
University of Washington

Abstract

Intracellular Delivery of Functional RNA by Designed Protein Assemblies

Issa George Benna

Chair of the Supervisory Committee:

David Baker

Department of Biochemistry

Synthetic nucleocapsids have great potential to be developed into next generation medicines, however like all macromolecules they become trapped in the endosome-lysosome pathway and are degraded by lysosomal proteases [1]. For this reason, endosomal escape has been a major longstanding bottleneck for drug delivery [1-3]. In order to deliver therapeutics intracellularly, this limiting step must be overcome [1,4]. Viruses have evolved mechanisms over long time periods to escape the endosome, therefore it is not surprising that many efforts at intracellular delivery and gene editing have taken a 'top-down' approach of re-engineering naturally occurring viruses [5-6]. However, this approach comes with its own challenges and safety risks [8-9]. Here we build on recent advances in protein design to create synthetic targeted nucleocapsids capable of delivering functional RNA cargo intracellularly. To this end we computationally designed proteins capable of disrupting membranes in a pH-driven manner and displayed these components on nucleocapsids encapsulating an RNA payload. To assess and quantify the ability of nucleocapsids to deliver RNA, we describe a novel prime-editing CRISPR/Cas9-based reporter system for high-throughput screening and sensitive detection of intracellular RNA delivery. This assay can detect functional RNA delivery as low as 5 fmols, and is capable of screening large RNA-barcoded libraries. We generated several nucleocapsid assemblies composed of purely computationally designed components, with both targeting and membrane-disrupting domains, and screened them for cytosolic RNA delivery. We report that EGFR and transferrin targeted assemblies containing membrane-disrupting domains delivered 31-fold and 12-fold more RNA, respectively, than their untargeted counterparts. These delivery vehicles offer a modular 'bottom-up' approach to creating tailor-made therapeutics and make significant headway towards developing next generation medicines.

Table of Contents

List of Figures

List of Tables

ACKNOWLEDGEMENTS

I am grateful for having completed my PhD in the Baker Lab at the Institute of Protein Design. The lab is a unique place that has allowed me to pursue research at the cutting edge of computational protein design and drug delivery. There is no other place that provides such a stimulating and exciting environment with the potential to not only develop novel therapeutics but the resources, culture, and mission to spin them out for further development into the clinic.

I would like to recognize the outstanding group of scientists, mentors, and colleagues I have had the pleasure of working with. These include Gabe Butterfield, whose work this project was built on, for the many fruitful discussions and knowledge transfer when I first started working on this project. Scott Boyken, whose work this project was also built on, for always being open to answer questions and provide guidance. I would also like to acknowledge others who served as my early mentors in the lab, including Jorgen Nelson, and Stephanie Berger.

I'd also like to thank several people who have intrepidly joined me in tackling different parts of this project, including Chad Miller and James Lazarovits. I'd like to thank my undergraduate, June Kim, who conducted much of the mRNA packaging efficiency experiments. Shunzhi Wang for always being willing to collaborate, and for collecting much of the EM images. Jason Zhang for collecting nucleocapsid and RNA localization images.

And David Feldman, who was critical in helping me develop the RNA delivery reporter assay, and served as an invaluable scientific mentor.

And my peers who worked on different aspects of endosome escape and understood its many challenges, including Audrey Olshefsky, Josh Lubner, and Christian Richardson.

And others who contributed to parts of this work including Basile Wicky for generating a library of parametrically defined helices and Phil Leung for various AlphaFold script templates.

I would also like to thank Nicolas Goldbach who bravely joined me in tackling a high-risk project of developing novel enhanced membrane-permeabilizing proteins for his Master's thesis, as well as collecting and analyzing data, and always being open to fruitful and thoughtful scientific discussions.

Thank you to the collaborators in the Kelly Lee lab for their guidance on liposome assays, particularly Jacob Croft for collecting cryo-electron tomograms with a quick turnaround.

Thank you to the mammalian protein core, especially Maggie Ahlrichs for maintaining the TC room, and always being there to assist with mammalian cell culture. Justin Decarreau for his assistance and persistence with imaging experiments.

And to Carl Walkey, Rubul Mout, Thomas Schlichthaerle, Danny Sahtoe, Wei Yang, Erin Yang, Brian Coventry, Daniel-Adriano Silva and numerous other colleagues in the lab who are too many to name, for various scientific discussions, wet-lab troubleshooting, and Rosetta knowledge among other things.

Thank you to members of the supervisory committee including David Baker, Neil King, Dan Ratner, Suzie Pun, and Jesse Zalatan for their time and guidance.

To Lance Stewart for bringing in interesting speakers from industry, providing invaluable commercialization support, and attracting key funding to the IPD.

A huge thank you to my advisor, David Baker, who has created a lab environment of excellence where novel ideas are entertained, grand challenges are tackled, and perseverance is common. I am thankful for the space he has provided me to explore various ideas, many of which did not pan out but from which I have learned a lot, as well as for his guidance and encouragement.

I am grateful for the various relationships I have formed outside of the lab, which have contributed tremendously to my growth. These include Will Canestaro, Kim Emmons, and Ron Howell. I am grateful for Will Canestaro and Kim Emmons, who served as mentors and from whom I have learned a lot about research commercialization, life science patents, the Seattle biotech scene, and evaluating startups. My interactions with them provided a much-needed respite from my lab work. And to Ron Howell for his kindness, wisdom, and eagerness to help from which I have learned a lot.

I would also like to thank my colleagues who served with me on the executive team of the non-profit organization, the Science and Engineering Business Association (SEBA) at the University of Washington, through which I have become a much better leader. And to Elizabeth Davisson, who adventured with me throughout Seattle and beyond and helped keep me sane.

And to my parents George and Liza whose sacrifices and early guidance I owe much of my success. And to my siblings Mariam and Marta for our inside jokes and their friendship.

CHAPTER 1. Nature's mechanisms for intracellular entry and applications for gene therapy

1.1 Viral mechanisms for endosomal escape

Viruses have evolved various mechanisms over large periods of time to enter cells and overcome the limiting barrier of endosome escape [10-11]. Their strategies fall into two categories: direct membrane fusion and endosomal membrane penetration [11-12]. Enveloped viruses, such as influenza viruses, use sophisticated machinery to drive fusion with the plasma membrane and release their genomic content into the cytosol [11-12]. In contrast, non-enveloped viruses like adenoviruses have developed methods to pierce the endosomal membrane to enable release of their genomic material into the cytoplasm [10-11]. The life cycle for these viruses begins with attachment to surface receptors on host cells, and in the case of nonenveloped viruses this results in internalization into the endosome pathway. Various viral endocytic routes of entry have been described, which include clathrin-mediated and caveolin-mediated endocytosis [10-11]. Nonenveloped viruses then employ several methods to escape the endosome: adenoviruses use amphipathic proteins that expose hydrophobic sequences to destabilize host cell membranes, in contrast parvoviruses employ phospholipases to do the same, while polioviruses induce pores through which their viral RNA is extruded into the cytoplasm [11]. While much of the mechanism of viral membrane fusion of enveloped viruses has been elucidated, we still do not fully understand the viral mechanisms of membrane penetration utilized by nonenveloped viruses [9-10]. This lack of understanding makes it challenging to design therapeutics for membrane penetration and endosome escape.

1.2 Virus-like particles for drug delivery

With the immense complexity involved in intracellular cargo delivery, virus-like particles (VLPs) have emerged as attractive candidates to overcome many of the challenges faced in the development of next generation medicines [13-15]. VLPs are largely derived from naturally occurring viruses but do not contain the viral genetic material, and typically range in size between 20 - 200 nm in diameter [13]. The main advantage of VLPs is that they come pre-programmed with the cellular entry and transport machinery of the wild-type virus [13,16]. However the downsides are that VLPs also retain the tropism of the wild-type virus, which presents a disadvantage when trying to repurpose them to target other desirable sites [14, 16]. Furthermore, they are inherently immunogenic [14,16]. Pre-existing immunity or successive doses often results in their rapid clearance and reduced circulation time which limits their efficiency as drugs [16]. They also suffer from other drawbacks, including difficult production as well as unpredictability when it comes to modification [16].

1.3 Development of protein assemblies packaging their mRNA genome

Computationally designed protein assemblies offer a complementary alternative to VLPs with several advantages that include modularity, a blank slate from which to engineer desired features, ease of production, and a lack of pre-existing immunogenicity. One such designed two-component assembly, I53-50, was engineered to package its own mRNA genome, resist RNase treatment, and increase its *in vivo* circulation time to 4.5 hours among other features [18]. This nucleocapsid provides a starting point from which to develop a carrier capable of RNA delivery.

CHAPTER 2. Functionalizing synthetic nucleocapsids for pH-mediated membrane lysis and disassembly

2.1 Motivation and results of functionalizing nucleocapsids

Advances in computational protein design have led to the development of synthetic icosahedral protein assemblies that resemble viral capsids [17], and these assemblies have subsequently been evolved to package their own RNA genome [18]. Similarly, pH-responsive proteins have been designed that lyse synthetic liposomes under acidic conditions [19]. Building on these two advances, we hypothesized that we could utilize the powerful link between genotype and phenotype afforded by synthetic nucleocapsids to design and evolve a carrier capable of permeabilizing endosomes to intracellularly deliver an RNA payload.

To investigate this possibility, we first tried to display the pH-responsive pRO-2.3 homotrimer [19] through a direct fusion with the trimeric component of the I53-50-V4 nucleocapsid. However, this approach failed to form nucleocapsids likely due to the strained geometries imposed by the fusion. To overcome this limitation, the pRO-2.3 homotrimer was redesigned into a single-chain (**fig. 2.1a**) that can be displayed on a nucleocapsid via a flexible linker. The loops were also made to be flexible via 2xGGS linkers to counteract the rigidity of a single-chain. To reduce the potential for aggregation between single-chains, the surface was redesigned to be asymmetric with reduced hydrophobicity and increased charge. The final single-chain bundle, pRO-2.3-sc-flex, contained three buried hydrogen bond networks (**fig. 2.1b**) with two networks containing histidines. Versions were also made with three histidine networks, including with mutations that reduce the hydrophobic packing like the most pH-responsive and most active published design, pRO-2 I56V [19].

We hypothesized that to successfully form nucleocapsids that display pRO-2.3-sc-flex, we needed to display these domains at reduced valency to reduce the potential for steric hindrance that may limit assembly. To achieve this, we fused pRO-2.3-sc-flex to the C-terminus of the I53-50-V4 pentamer (**fig. 2.1c**) and used a programmed ribosomal frameshift (PRF) motif to cause ribosomal slip at a ~40% rate (**fig. 2.1d**). Under these expression conditions, the resulting nucleocapsids will contain the membrane-permeabilizing domain on 40% of the pentamers which translates to an average of 24 per nucleocapsid.

The construct was then expressed and intracellularly assembled in *E. coli* (**fig. 2.2a**). Nucleocapsids were purified from cell lysates using immobilized metal affinity chromatography (IMAC), and intact nucleocapsids were observed by negative-stain transmission electron microscopy (**fig. 2.2b**). Assemblies were further purified by size exclusion chromatography (SEC) and eluted as a single peak. Due to their larger size, these nucleocapsids eluted at an earlier retention volume than the bare I53-50-V4 design (**fig. 2.2c**). The purified assemblies were treated with RNase A for 20 minutes at 25 °C to determine whether they still package and protect RNA. The RNA and protein co-migrated on native agarose gels (**fig. 2.2e**), confirming that RNA was packaged and protected. Sanger sequencing and reverse transcription quantitative PCR (RT-qPCR) confirmed packaging of the full-length genome as described below and figure 2.4.

Disruption of synthetic membranes

We next sought to determine whether I53-50-V4-sc-pRO2.3-flex has been successfully functionalized to disrupt membranes in a pH-mediated fashion. The nucleocapsid was incubated with synthetic liposomes encapsulating self-quenching sulforhodamine B (SRB) dye, and the fluorescence of released dye was measured after acidification which correlates with the level of liposome disruption [19-20]. Surprisingly, the modified nucleocapsid led to the same level of liposome rupture as the pRO2.3 homotrimer but with slower kinetics. While pRO2.3 reached its maximal lysis in ~10 seconds, the single-chain version displayed on V4 took about 1.5 minutes. This is likely due to the added constraints of a single-chain slowing the disassembly.

Effect of mRNA length on packaging efficiency

To better understand the effect of genome length on packaging efficiency, we constructed six versions of the bare nucleocapsid with varying lengths of expressed mRNA. These constructs were the original nucleocapsid with a ~1.5 kb genome, as well as versions with 2.5 kb, 3.5 kb, 4.5 kb and 5.5 kb genomes that have padded nonfunctional mRNA between the stop codon and the T7 terminator. A sixth version, V4-dualT7, was made where the genome was split into two ~800bp mRNA genomes each driven by a separate T7 promoter that expressed the pentamer and the trimer. All versions were then expressed in *E. coli* and purified by IMAC and SEC. The constructs eluted as a single peak at the same retention volume as the original V4 nucleocapsid (**fig. 2.4a**). RNase treatment suggested that all versions package and protect RNA (**fig. 2.4b**). However, there was an inverse relationship between packaging efficiency and mRNA length. To investigate this further, RT-qPCR was performed on the purified nucleocapsids. A 16-fold reduction in packaging efficiency is observed as the genome expands from 2.5 kb to 3.5 kb (**fig. 2.4c**). Similarly, another 4-fold reduction in packaging efficiency is observed as the genome expands from 4.5 kb to 5.5 kb. As expected, packaging efficiency takes a hit as the length of mRNA genome increases. The first major reduction in packaging efficiency occurs at genome lengths above 2.5 kb. How does this affect the functionalized nucleocapsid? The fusion of pRO-2.3-flex to the pentamer expands the genome to about 2kb in total. This results in a 1.6-fold reduction in packaging efficiency as compared to the bare nucleocapsid (**fig. 2.4c**). While packaging efficiency and RNA protection integrity is fortunately maintained, the reduced packaging efficiency means that the sensitivity for a library-based evolution is reduced by at least 1.6-fold. Interestingly, going the other direction and reducing the length of mRNA to ~0.8 kb by way of the V4-dualT7 construct led to a 10-fold increase in packaging efficiency (**fig 2.5**).

Redesigning nucleocapsids for pH-mediated disassembly

Early on in this investigation, we hypothesized that nucleocapsids must be designed to disassemble for them to deliver their encapsulated cargo. To this end, we redesigned the pentamer of the I53-50-V4 nucleocapsid to have buried hydrogen-bond networks with histidines participating in the networks (**fig. 2.6a**). Furthermore, designs were made where histidines that do not participate in hydrogen bond networks were rationally introduced into the trimeric component of nucleocapsids **(fig. 2.6a**). These designs were expressed in *E. coli* bicistronicly and purified from lysate by IMAC (**fig. 2.6b**) and SEC. Six designs eluted as single peaks at the expected retention volume (**fig. 2.6c-d**), and were verified to package and protect RNA by RNase treatment and native gel co-migration of RNA and protein (not shown).

To assess monodispersity and pH-mediated disassembly, designs were analyzed at varying pH conditions by dynamic light scattering (DLS). Bare nucleocapsids I53-50-V4, and a version with a circularly permuted pentamer (V4-cp) were monodisperse and had peaks at the expected size of ~28 nm. Both V4 and V4-cp maintained their monodispersity and average diameter at pH of 6.0, 5.5, and 5.0 (**fig. 2.7 a-b**). At pH 5.0 there was a slight broadening of monodispersity in V4 but not V4-cp. These data confirm that V4 is not inherently pH responsive. In contrast, designs p7, p18, p21, p23, and p34 were all pH-responsive to varying degrees (**fig. 2.7 c-h**). One clear example is p28 (**fig. 2.7g**), which begins to aggregate at pH 6.0 and reaches an average diameter >1000 at pH 5.0. Because only interface residues were mutated during the redesign, and no surface charge was changed we can rule out pH-mediated surface aggregation of intact nucleocapsids. Further, it is not surprising that we see aggregation because while the pentameric interface is disrupted by charged histidines it is still largely hydrophobic. While we were able to redesign I53-50-V4 to disassemble and aggregate in a pH-dependent manner, later results reveal that this feature is not necessary for RNA delivery.

2.2 Figures

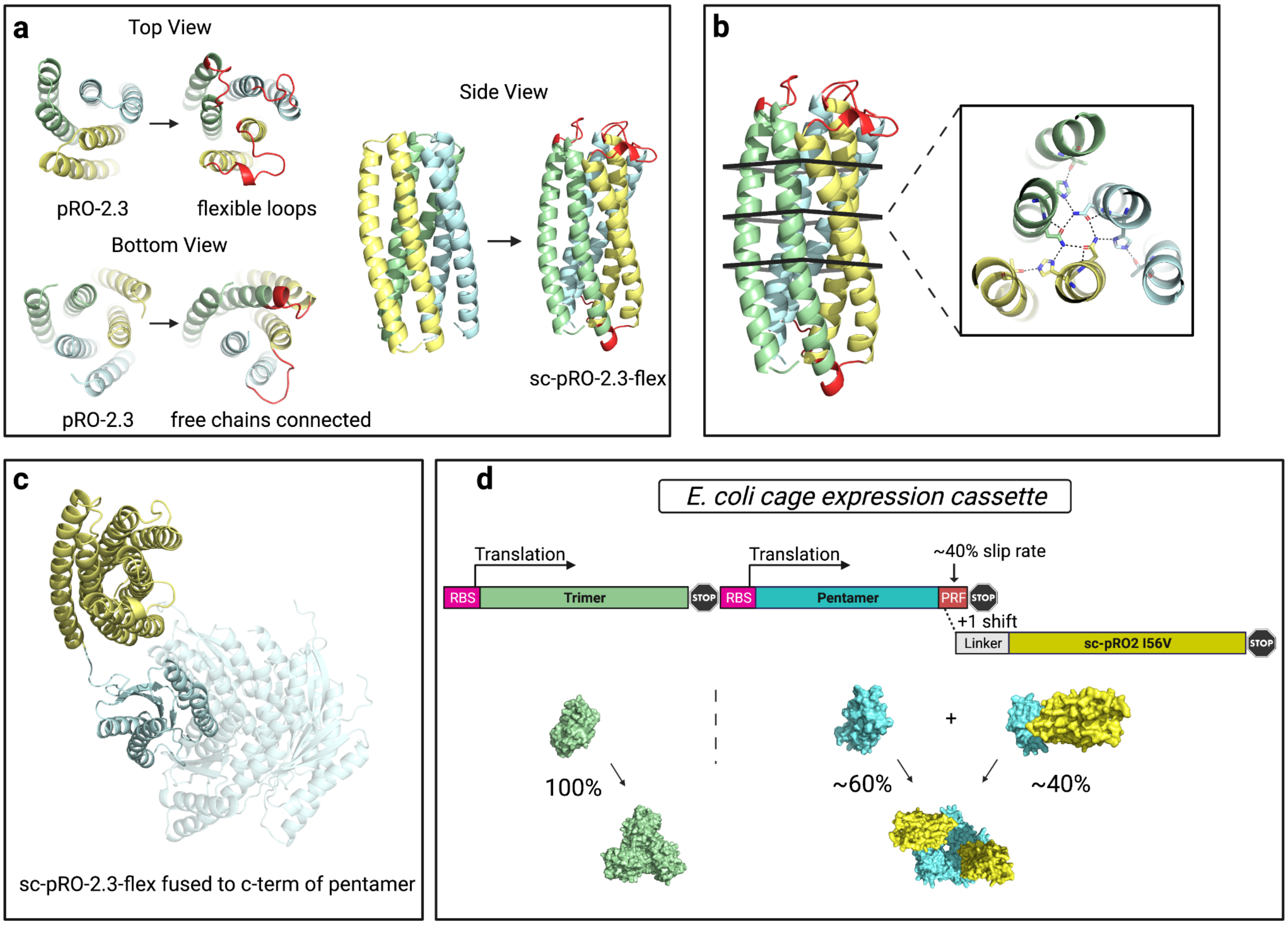


Figure 2.1. Functionalizing the synthetic I53-50-V4 nucleocapsid with a single-chain pH-responsive domain

a. pRO-2.3 is looped into a single chain with flexible loops, and the exterior is redesigned to be asymmetric, less hydrophobic and more charged. **b.** sc-pRO-2.3 contains three hydrogen bond networks, two of which contain histidines which result in the pRO bundle's pH-responsiveness. **c.** sc-pRO-2.3 is fused to the C-terminus of the pentameric component of I53-50-V4 for display on the nucleocapsid. **d.** sc-pRO-2.3 is expressed with a programmed ribosomal frameshift slip sequence to reduce the number of displayed bundles on the surface of nucleocapsids and thereby preventing any potential for steric hindrance or aggregation. A slip sequence is used such that 40% of the time the ribosome skips the stop codon, and continues to translate the sc-pRO-2 bundle. This results in 40% of the expressed pentamers having a fusion to the sc-pRO-2.

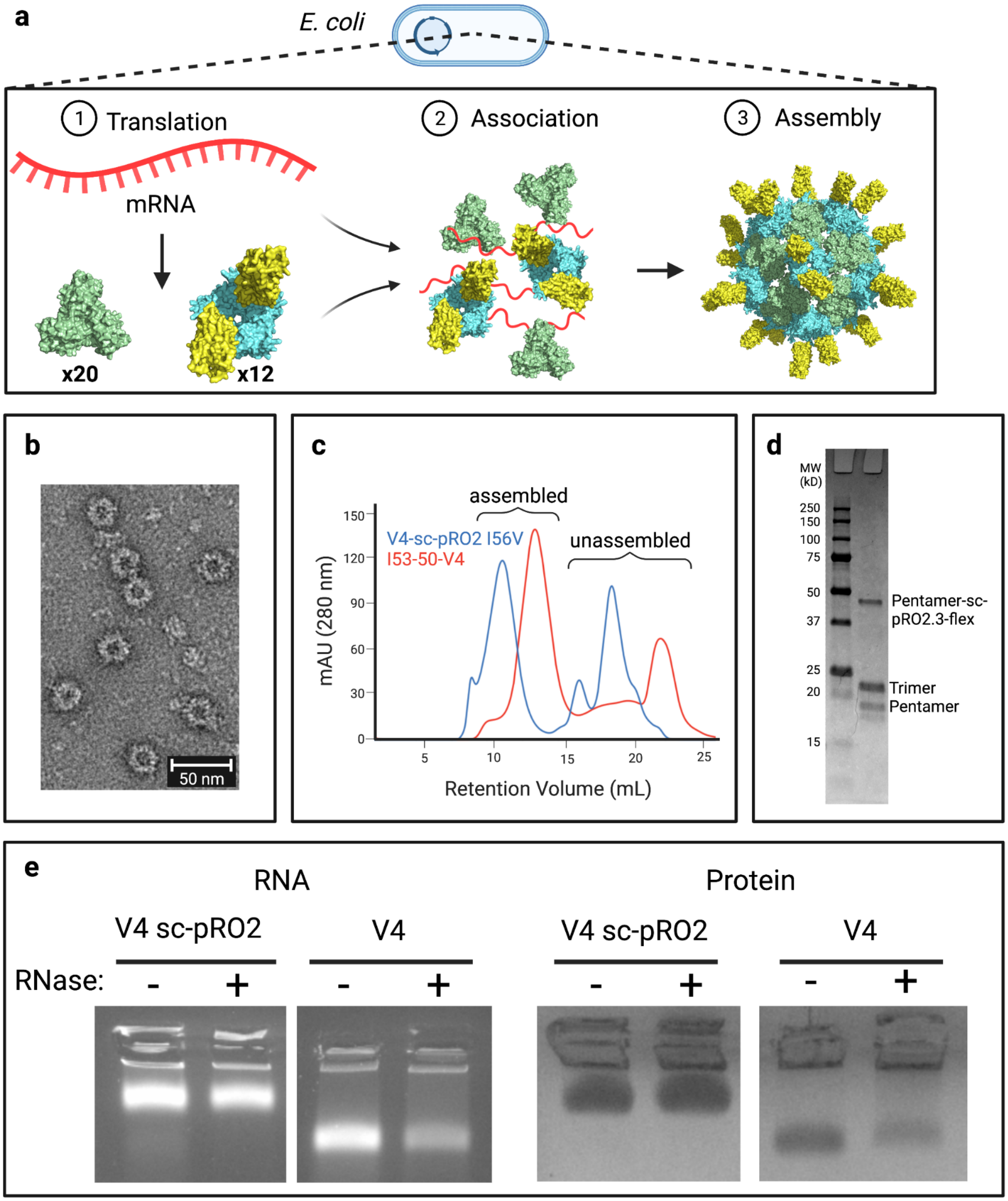


Figure 2.2. Biochemical characterization of functionalized synthetic nucleocapsids

a. nucleocapsids encapsulate their own mRNA genomes during bicistronic expression in *E. coli*, and assemble into icosahedral capsids. **b.** negative-stain electron microscopy of nucleocapsids displaying sc-pRO-2.3-flex. **c.** size exclusion chromatography on an S6 column of nucleocapsids displaying sc-pRO-2.3-flex as compared to the bare I53-50-V4 nucleocapsid. The larger V4-sc-pRO-2.3-flex elutes earlier than V4 as expected. **d.** SDS-PAGE gel confirming the presence of the three expected species co-expressed and assembled in *E. coli*. **e.** native agarose gel stained with SYBR Gold for RNA and Coomassie for protein showing co-migration of RNA and protein, suggesting encapsulation. Further, the functionalized V4 robustly protects RNA against RNase A.

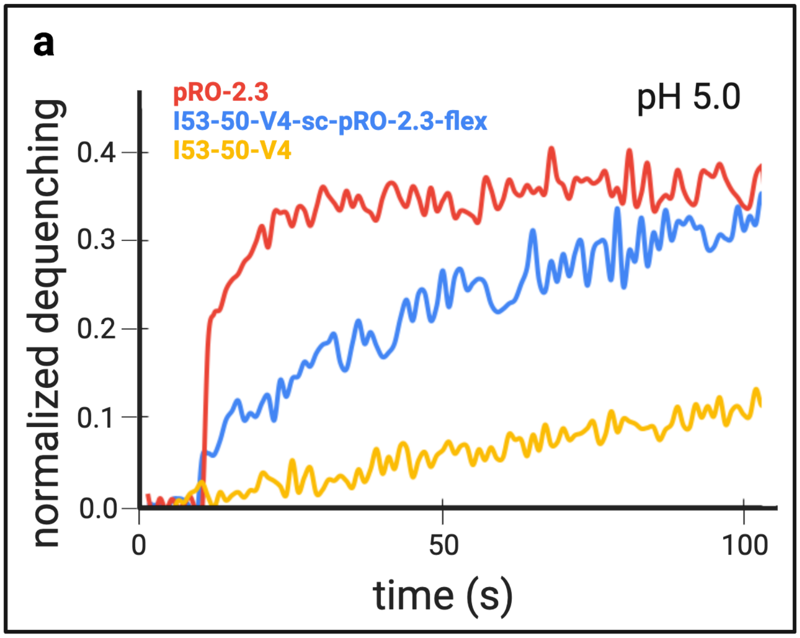


Figure 2.3. Comparing pH-driven membrane disruption

a. Functionalized V4 is compared to the original V4 as well as pRO-2.3 alone. Proteins were incubated at 4 uM with respect to the pH-responsive domain. In the case of V4, the concentration was normalized to V4-sc-pRO-2.3-flex. The functionalized nucleocapsid reaches the same level of liposome disruption after acidification at pH 5.0 as pRO-2.3 but with slower kinetics. In contrast, V4 alone minimally disrupts liposomes.

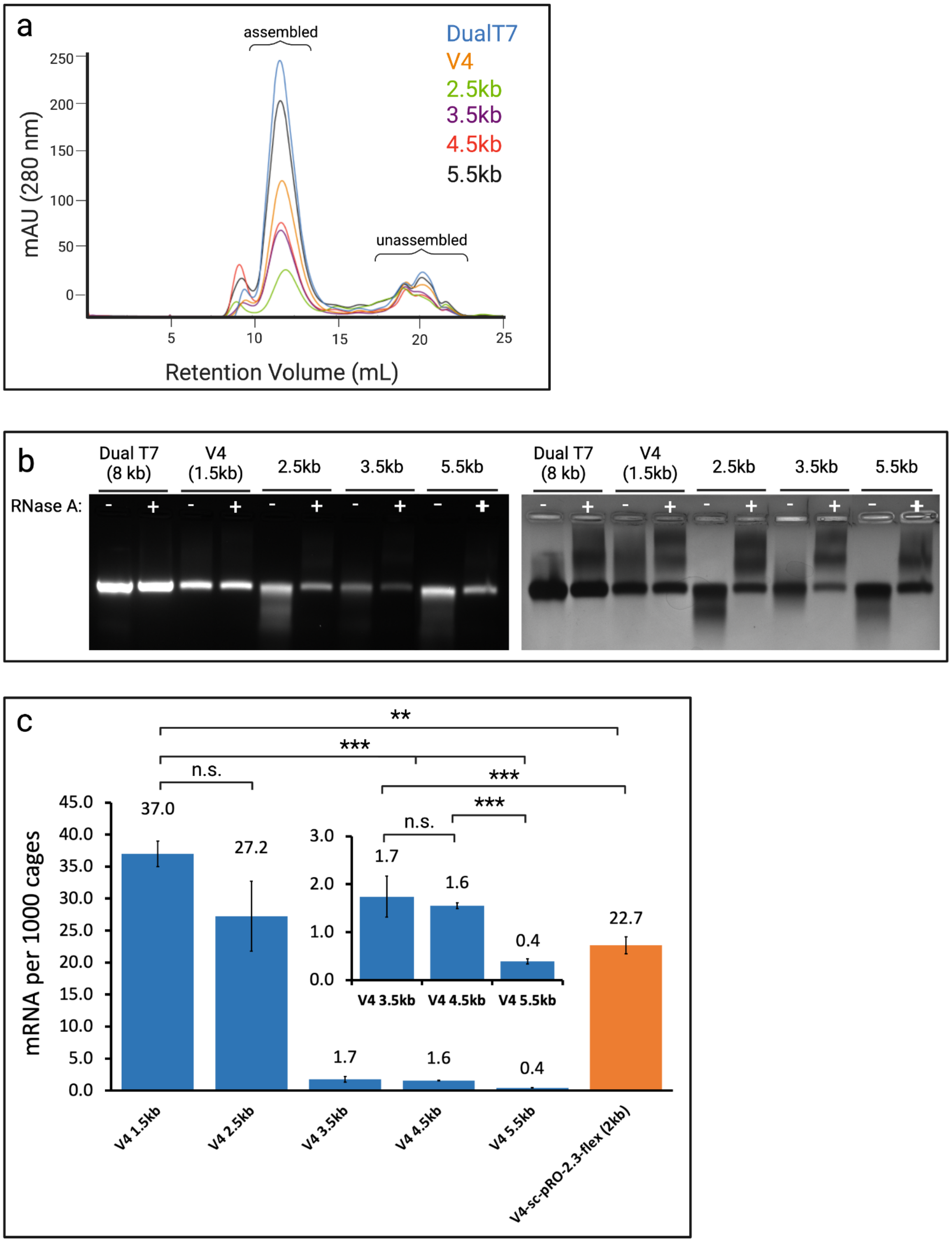


Figure 2.4. Effect of mRNA length on packaging efficiency

a. V4 constructs with varying mRNA genomes all eluted at the expected retention volume. **b.** all variants packaged RNA, and protected against RNase A as seen by native agarose gel. **c.** RT-qPCR of each cage shows a significant inverse correlation between mRNA genome length and packaging efficiency. A large drop in packaging is seen as the genome lengthens from 2.5kb to 3.5 kb. Another significant drop in packaging efficiency is observed as the genome lengthens from 4.5kb to 5.5 kb. V4 displaying sc-pRO-2.3-flex has a 2kb genome that resulted in a 1.6-fold reduction in packaging efficiency as compared to V4.

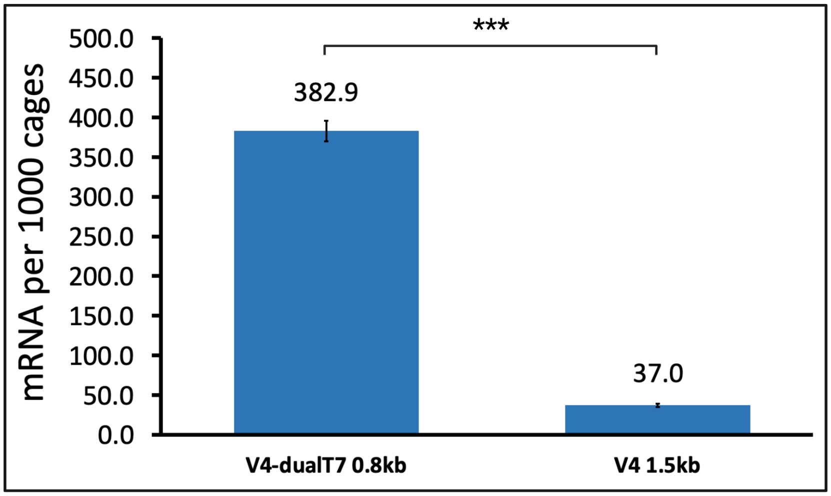


Figure 2.5. Substantial improvement in RNA packaging efficiency with reduced mRNA length

a. Reducing the genome size to 0.8 kb by splitting the mRNA into two strands that express the pentameric or trimeric component results in a significant increase in mRNA packaging.

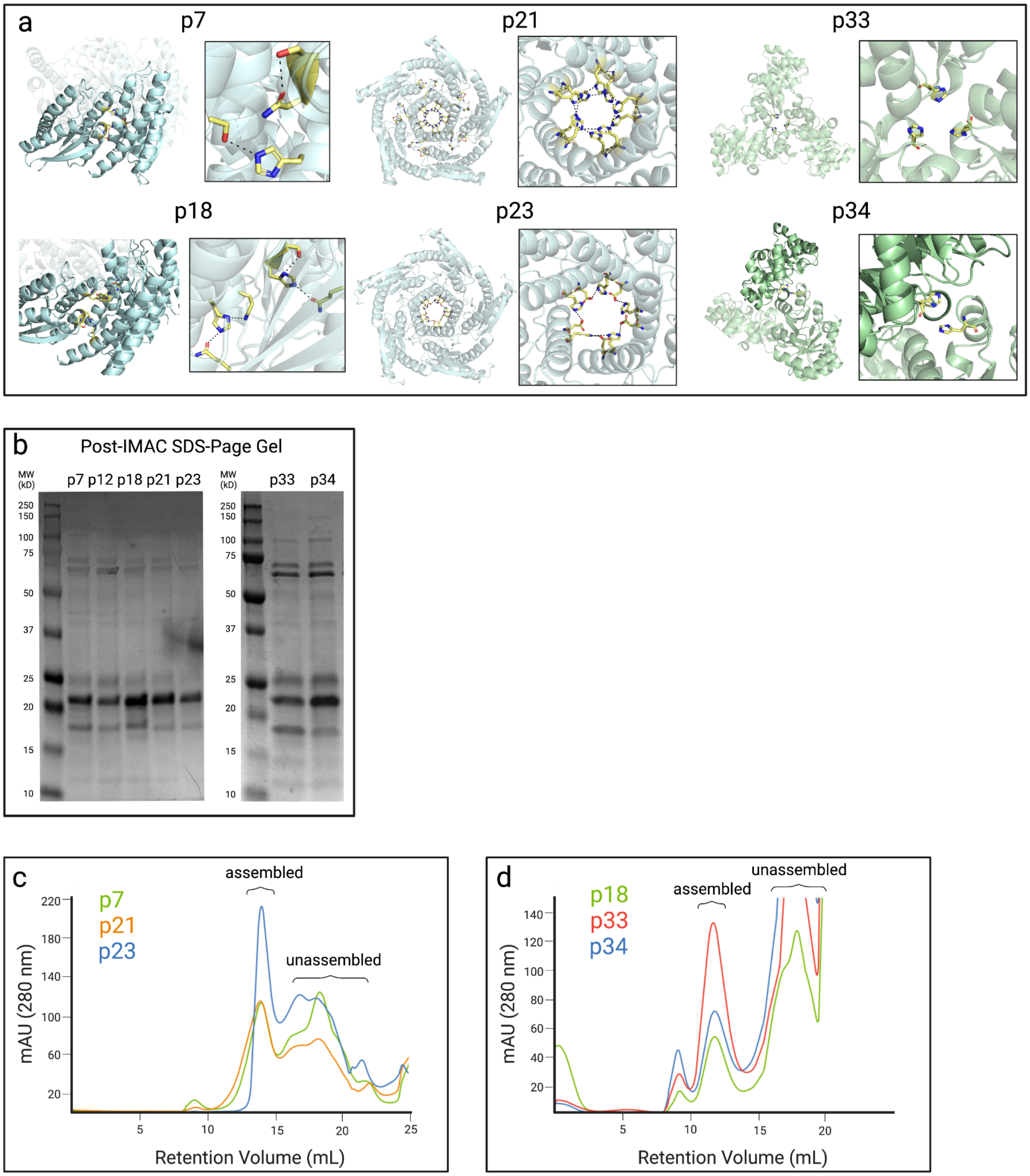
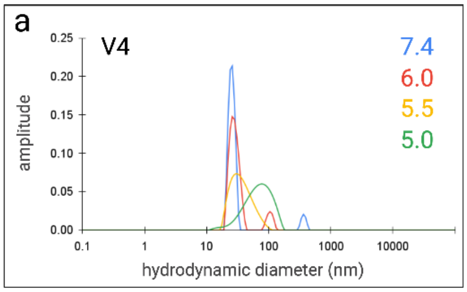
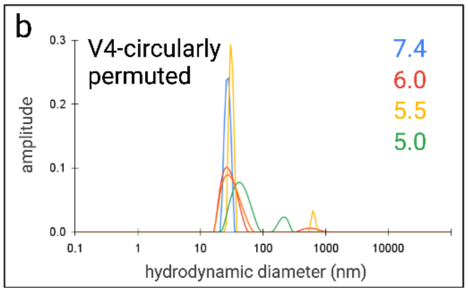


Figure 2.6. Redesign of the I53-50-V4 pentamer and trimer for pH-driven disassembly

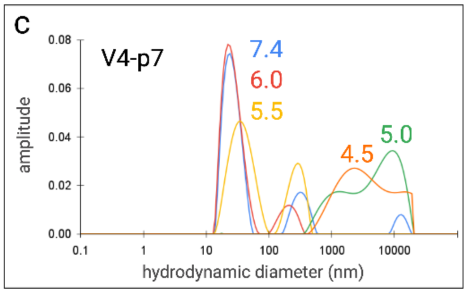
a. Design models of redesigned V4 pentamers with buried hydrogen bond networks at the interface. Histidines were rationally introduced into trimers to also make them pH-responsive. **b.** SDS-PAGE gel of pH-sensitive nucleocapsids after IMAC purification. **c, d.** SEC purification confirms elution of assembled nucleocapsids. P7, p21, and p23 eluted later than expected, however they were of the right size when measured by DLS as seen below. This is likely due to column-specific drifting that could occur over time through repeated use.

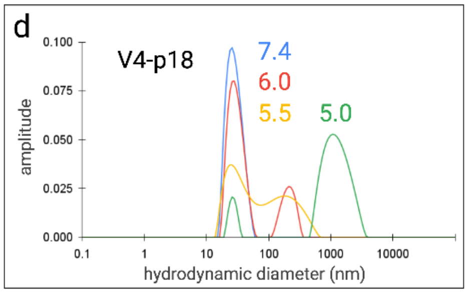




Cage	pH	Z-Ave. Dia. (nm)	PDI	Pk 1 Mode Dia. (nm)
V4	7.4	27.92	0.03	26.93
	6.0	30.51	0.06	26.94
	5.5	35.72	0.22	31.45
	5.0	58.55	0.58	79.74

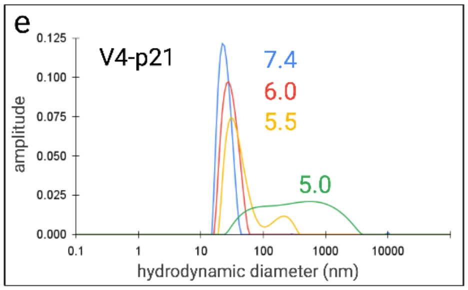
Cage	pH	Z-Ave. Dia. (nm)	PDI	Pk 1 Mode Dia. (nm)
V4-cp	7.4	29.68	0.28	29.10
	6.0	30.33	0.07	26.93
	5.5	31.43	0.24	29.10
	5.0	37.98	0.04	31.44

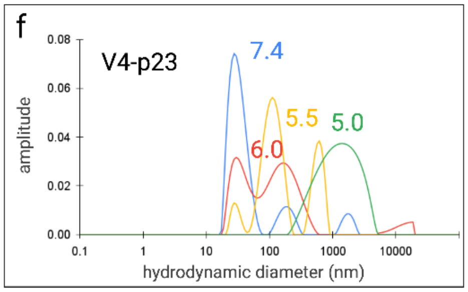




Cage	pH	Z-Ave. Dia. (nm)	PDI	Pk 1 Mode Dia. (nm)
V4-p7	7.4	33.27	0.10	24.92
	6.0	28.31	0.22	23.05
	5.5	51.69	0.01	36.73
	5.0	>1000	1.62	>1000
	4.5	>1000	1.91	>1000

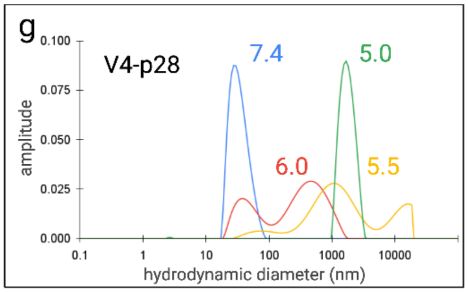
Cage	pH	Z-Ave. Dia. (nm)	PDI	Pk 1 Mode Dia. (nm)
V4-p18	7.4	28.76	0.07	26.82
	6.0	36.98	0.11	29.07
	5.5	56.58	0.18	27.11
	5.0	394.00	3.51	26.93

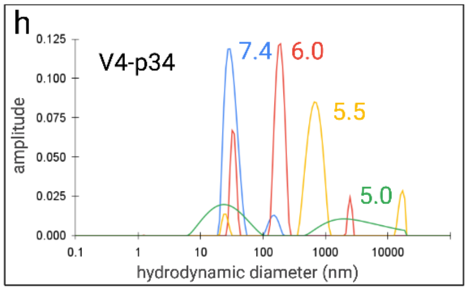




Cage	pH	Z-Ave. Dia. (nm)	PDI	Pk 1 Mode Dia. (nm)
V4-p21	7.4	28.31	0.10	25.07
	6.0	31.17	0.17	29.10
	5.5	43.47	0.13	31.45
	5.0	193.47	0.42	599.19

Cage	pH	Z-Ave. Dia. (nm)	PDI	Pk 1 Mode Dia. (nm)
V4-p23	7.4	39.33	1.10	28.68
	6.0	71.07	1.75	30.98
	5.5	121.77	0.60	28.66
	5.0	944.52	1.20	>1000





Cage	pH	Z-Ave. Dia. (nm)	PDI	Pk 1 Mode Dia. (nm)
V4-p28	7.4	34.12	0.23	28.66
	6.0	139.92	0.24	39.07
	5.5	881.05	2.96	78.51
	5.0	>1000	1.26	>1000

Cage	pH	Z-Ave. Dia. (nm)	PDI	Pk 1 Mode Dia. (nm)
V4-p34	7.4	34.78	0.12	30.96
	6.0	95.73	0.96	33.46
	5.5	406.71	2.44	26.52
	5.0	100.85	2.76	24.54

Figure 2.7. pH-driven disassembly and aggregation I53-50-V4 with redesigned pentamer a-h. Dynamic light scattering of assemblies after a 3-hour incubation at 25 C varying pHs of 7.4, 6.0, 5.5, and 5.0. V4 and V4-cp remained monodisperse and did not aggregate at low pH, while nucleocapsids with pH-responsive pentamers and trimers aggregated at different pHs.

Table 2.1 Sequences of single-chain pRO bundles redesigned I53-50-V4 components

Design	AA sequence
pRO-2.3-sc-flex	DGSGSGSGSEEEIKRLLEELRKSSEELRRITKELDDLSKELRVGGSGSGSEALVEHNKLISEHNRIIVENNRIIVEILEAIARVGGSGSGSVEVERILDELRKSSEELDRVTKELKKLTEELDVGGSGSGSENLYFQGGSGSGSVEALVRHNVLITRHNDIIVKNNDIINKILKLIAEAVGGSGSGSELERILRELEESTKELRKATEELRRLSEELKVGGSGSGSVEALVRHNEAIVEHNKIIVKNNDIIVKILELITERI
pRO-2-sc-flex	EEEIKRLLEELRKASEELRRITKELDDLSKELRVGGSGSGSEALVEHNKLISEHNRIVVEHNRIIVEILEAIARVGGSGSGSVEVERILDELRKASEELDRVTKELKKLTEELDVGGSGSGSVEALVRHNVLITRHNDIVVKHNDIIAKILKLIAEAVGGSGSGSELERILRELEEATKELRKATEELRRLSEELKVGGSGSGSVEALVRHNEAIVEHNKIVVKHNDIIVKILELITERI
circularly permuted V4 pentamer (V4-cp)	MDKSNAKTLLFLALFAVKGMEAARACVEILAAREKIAAGSGGSGNQHSQKDQETVRIAVVRARWHAFIVDACVSAFEAAMRKIGGERFAVDVFDVPGAYEIPLHARTLAKTGRYGAVLGTAFVVNGGIYRHEFVASAVIDGMMNVQLDTGVPVLSAVLTPHNY
p7	GSSHHHHHHSSGENLYFQGNQHSQKDQETVRIAVVRARIHAFIVDACVSAFEAAMRKIGGERFAVDVFDVPSANSIPLHARTLAKTGRYGAVLGTAFVVNGGIYRHEFHASAVIDGMMNVALDTGVPVLSAVLIPHNYDKSNAKTLLFLALFAVKGMEAARACVELLAAREKIAAGS
p18	GSSHHHHHHSSGENLYFQGNQHSQKDQETVRIAVVRARIHAFIVDACVSAFEAAMRKIGGERFAVDVFDVPTAHSIPLHARTLAKTGRYGAVLGTAFVVNGGIYRHETQASAVHDGMMNVALDTGVPVLSAVLQPHNYDKSNAKTLLFLALFAVKGMEAARQCVELLAAREKIAAGS
p21	GSSHHHHHHSSGENLYFQGNQHSQKDQETVRIAVVRARWHAFIVDACVSAFEAAMRKIGGERFAVDVFDVPGAYEIPLHARTLAKTGRYGAVLGTAFVVNGGIYRHEFVAHAVIHGMMNVALDTGVPVLSAVLTPHNYDKSNAKTLLFLALFAVKGMEAARACVEILAAREKIAAGS

p23	GSSHHHHHHSSGENLYFQGNQHSQKDQETVRIAVVRARWHAFIVDACVSAFEAAMRKIGGERFAVDVFDVPGAYEIPLHARTLAKTGRYGAVLGTAFVVNGGIYRHEFVADSVIHGMMNVQLDTGVPVLSAVLTPHNYDKSNAKTLLFLALFAVKGMEAARACVEILAAREKIAAGS
p28	GSSHHHHHHSSGENLYFQGNQHSQKDQETVRIAVVRARWHAFIVDACVSAFEAAMRKIGGERFAVDVFDVPGAYEIPLHARTLAKTGRYGAVLGTAFVVNGGIYRHEFHASAVIHGMMNVQLDTGVPVLSAVLTPHNYDKSNAKTLLFLALFAVKGMEAARACVEILAAREKIAAGS

2.3 METHODS

Protein expression and preparation

Plasmids coding for synthetic genes that express the designed proteins were ordered through Genscript Inc. (Piscataway, N.J., USA) or IDT (Newark, NJ) into a pET29+ vector. These plasmids were transformed into chemically competent E. coli expression strains BL21(DE3)Star (Invitrogen). After transformation, single colonies were picked from agar plates and grown overnight in 10 mL starter cultures of Luria-Bertani (LB) medium containing 50 µg/mL kanamycin with shaking at 225 rpm for 12-18 hours at 37°C. 5 mL starter cultures were added to 500 mL of TB supplemented with 50 µg/mL kanamycin, and grown to an optical density of 0.6 at 37 °C with shaking at 225 rpm. Expression was induced by the addition of IPTG (500 µM final) when the optical density was between 0.6 and 0.8. Incubation continued for another 4 hours at 37 °C with shaking at 225 rpm. Cultures were harvested by centrifugation at 4,000 rcf for 10 minutes and stored at -80 °C. Cell pellets were resuspended in Tris buffer (25 mM Tris, 150 mM NaCl), and lysed by sonication or microfluidizer. The resulting lysate was clarified by centrifugation at 24,000 rcf for 30 minutes and passed through 2 mL of Nickel-Nitrilotriacetic acid agarose (Ni-NTA) (Qiagen), followed by 3 washes (25 mM Tris, 150 mM NaCl, 40 mM Imidazole), and eluted into 2x3 mL of Tris elution buffer (25 mM Tris, 150 mM NaCl, 300 mM

Imidazole). EDTA was immediately added to 5mM final concentration to prevent Ni-mediated aggregation. Constructs were then filtered by passing through 0.2 um filter, and purified by SEC on a Superose 6 Increase 10/300 column (GE) equilibrated with Tris buffer (25 mM Tris, 150 mM NaCl).

mRNA PRF slip sequence

A PRF slip sequence was utilized to generate proteins with pre-determined ratios of fused to unfused species according to empirically determined PRF slip sequences with different ribosomal slip rates (40-41).

Protein SDS-PAGE

SDS-PAGE was performed using 4-20% polyacrylamide gels (Bio-Rad) in tris-glycine buffer. Gels were electrophoresed at 160V for 45 minutes before staining with Genscript's eStain Protein Staining System.

Measuring protein concentration

Qubit (ThermoFisher) was used according to the manufacturer's instructions.

Gel electrophoresis

Native agarose gels were prepared using 1% ultrapure agarose (Invitrogen) in TAE buffer. 20 uL of purified nucleocapsids were treated with 10 µg/mL RNase A at 25 °C for 20 minutes, mixed with 3 µL 6x loading dye containing no SDS (NEB), and electrophoresed at 120 volts for 30 minutes. Gels were then stained with SYBR gold (Thermo-fisher) for RNA staining followed by incubation with GelCode Blue (Pierce) for a minimum of 3 hours for protein staining.

RNA purification and reverse transcription

RNA was purified using TRIzol (Thermo fisher Scientific) and the Qiagen RNeasy kit (Qiagen), as described previously (18). 100 µL synthetic nucleocapsid samples were mixed with 500 µL TRIzol and vortexed, followed by the addition of 100µL chloroform and another round of vortexing. The solution was centrifuged for 10 min at 24,000 rcf. 150 uL of the resulting aqueous phase was mixed with 150 µL of 100% RNase-free ethanol, transferred to a RNeasy spin column for purification according to manufacturer's protocol, and eluted in 50 µL nuclease-free dH2O.

RNA was reverse transcribed using Thermoscript's reverse transcriptase for one hour at 53 °C as described previously (18). A 10 µL reaction contained: 1 µL dNTPs (10 mM each), 1 µL DTT (100 µM), 1 µL Thermoscript reverse transcriptase, 2 µL cDNA synthesis buffer, 1 µL RNase-Out, 1 µL reverse primer (10 µM), 2 µL purified RNA template, and 1 µL nuclease-free dH2O. Controls without the reverse transcriptase were also set up identically.

Quantitative PCR

Quantitative PCR was conducted as described previously (18). In summary, a 10 µL reaction was mixed using a High Fidelity PCR kit (Kapa Biosystems) according to the manufacturer's instructions with the addition of SYBR green at 1x concentration and 0.5 µM forward and reverse primers. Thermocycling and Cq calculations were performed on a Bio-Rad CFX96 with the following protocol: 5 min at 95 °C, followed by 40 cycles of 98 °C for 20 seconds, 64 °C for 15 seconds, 72 °C for 90 seconds. The thermocycling protocol was modified to increase extension time as needed for longer mRNA quantifications.

Dynamic Light Scattering

Sample concentrations at 0.5 mg/mL were analyzed by DLS data using an UNcle instrument (UNchained Labs), with eight 5s acquisitions. Average hydrodynamic diameter, mode, and PDI for each sample was calculated using the UNcle software. For pH disassembly measurements, nucleocapsids in Tris buffer (25 mM Tris, 150 mM NaCl) were acidified to the desired pH by mixing with citrate buffer (0.1M, pH 2.2) at empirically determined ratios and incubated for 3 hours before measuring on DLS.

Liposome preparation and characterization

Liposomes composed of DOPC (1,2-dioleoyl-sn-glycero-3-phosphocholine) lipids were prepared to a final concentration of 7.5 mM as previously described (19-20). Lipids purchased from Avanti Polar Lipids were solubilized in chloroform and dried under vacuum for at least 12 hours to remove residual solvent. The dried lipid film was resuspended in Tris buffer (25 mM Tris, 150 mM NaCl, and 0.02% NaN3 pH 8.2) containing 25 mM Sulforhodamine B (SRB) fluorophore (Sigma) and subjected to a minimum of 10 sequential freeze thaw cycles in liquid nitrogen and a 42 °C water bath. Liposomes were extruded a minimum of 15 times through 100 nm pore size polycarbonate filters (Avanti Polar Lipids) and purified from free fluorophore using a PD-10 gel filtration column (GE Healthcare) into Tris storage buffer (25 mM Tris, 150 mM NaCl, and 0.02% NaN3 pH 8.2). Liposome size and homogeneity was analyzed by dynamic light scattering (DLS) using an UNcle instrument (Unchained Labs). On average, liposome diameter ranged from 100-110 nm with low polydispersity ≤0.05 PDI. Liposomes were stored at 4°C and used within 5 days of preparation.

Liposome disruption assay

Liposome disruption was assessed by measuring the fluorescence of released SRB dye as described previously (19-20). Liposomes containing SRB fluorophore at self-quenching concentrations were incubated with 4 µM protein in a quartz cuvette, with respect to monomer, at 25°C and pH 8.2 in Tris buffer (25 mM Tris, 150 mM NaCl, 0.02% NaN3, pH 8.2) for 5 minutes. The solution was rapidly acidified to the target pH by addition of a fixed volume of acidification buffer and incubated for 15 minutes. Acidification buffers are mixtures of the Tris pH 8.2 buffer and citrate buffer pH 3.0 (10 mM HEPES, 150 mM NaCl, 50 mM citrate and 0.02% NaN3 pH 3.0) in experimentally determined ratios to achieve the target pH. SRB fluorescence is independent of the lowest pH used in this study. Finally, Triton X-100 (Sigma) was added to a final concentration of 1% to fully disrupt liposomes. Liposome disruption as indicated by content leakage and SRB dequenching was normalized using the formula $[F_{(t)}-F_{(0)}]/[F_{(Max)}-F_{(0)}]$ where $F_{(0)}$ is the average fluorescence intensity before acidification and $F_{(Max)}$ is the average fluorescence intensity after addition of Triton X-100. All measurements were carried out on a Fluoromax-4 series spectrophotometer (HORIBA) using an excitation/emission measurement of 565/586 and 1 nm slit widths.

Visualization and figures

Structural protein models were generated using PyMOL and figures were assembled with BioRender.

CHAPTER 3. De novo design of proteins for enhanced membrane lysis

3.1 Motivation and results for designing proteins for enhanced membrane lysis

While we now have nucleocapsids that can disrupt synthetic membranes while maintaining their genotype to phenotype link, we wanted to determine whether their potential for membrane disruption could be enhanced as this would inform the design of a library by which to evolve for potent intracellular delivery. To this end, we computationally grafted dozens of motifs derived from naturally occurring viruses that are used by viruses and microbes to penetrate membranes. One such motif is from the adeno-associated virus (AAV) protein VI. These sequences were grafted onto the N-terminus of pRO-2-sc-flex and surrounding residues were redesigned (**fig. 3.1a).** However, it became apparent that the core was limited in the size of residues that it would accept. This was an issue because one striking feature shared by these motifs is the presence of bulky hydrophobic residues, specifically phenylalanine and tryptophan. Because of the limited size of the core, the N-terminal helix was pushed out in the grafted design model as compared to the crystal (**fig 3.1a**). In total, seven motifs were successfully grafted (**fig 3.1b)** and were predicted to fold to their design model with a predicted per-residue confidence > 88 on a scale of 0 - 100, with RMSD to design model <2.0 Å as predicted by AlphaFold[36]**.** These designs were expressed in *E. coli*, and purified by IMAC and SEC. All grafts contained a peak at the expected retention volume by SEC (**fig. 3.1d).** However, these designs also formed higher order species and tended to be unstable, as evident by the multiple peaks observed by SEC. To address this, graft 484/499 was redesigned to reduce its Rosetta surface aggregation propensity (SAP) score. This resulted in design 533, which eluted as a single monomeric peak by SEC (**fig. 3.1e-f).** Despite this improvement, these designs tended to precipitate over time at 4 °C.

De novo design of tailor-made pH-driven membrane disrupting proteins

Because pRO bundles are limited in their capacity to be redesigned to include endosome permeabilizing motifs, evolving a diverse library of nucleocapsids displaying these bundles will likely also be limited in yielding potent delivery vehicles. To address this, we sought to design **p**H **r**esponsive membrane **l**ytic **b**undles (pRLBs) tailor made for endosome-permeabilization. We required that these proteins have a large core that can accommodate bulky aromatic residues, and have a long stretch of hydrophobic residues to increase their amphipathicity and thereby increasing their potential to embed deeply into membranes and disrupt them (**fig. 3.2a)**.

In contrast to pRO bundles which are homotrimers made up of two-helix monomers, pRLBs were designed to be a single-chain four-helix bundle with a lytic C-terminal helix that is ejected in a pH-dependent manner to disrupt membranes (**fig. 3.3a)**. This C-terminal lytic helix can be further redesigned to accommodate a variety of endosome-permeabilizing motifs or novel lytic sequences due to the larger core, including large hydrophobic residues that can serve as an 'anchor' to embed deeply into the hydrophobic lipid bilayer of membranes.

To generate pRLBs, we started with a library of one million four-helix bundles with varied superhelical parameters and searched for buried hydrogen bond networks using Rosetta HBNet as described previously [19, 21]. We required the networks to contain histidines, and contact at least three of the helices (**fig. 3.3b)**, one of which will become the C-terminal helix. We hypothesized that at least two histidine-containing hydrogen bond networks are required for pH-driven helical ejection at endosome relevant pHs between 5 and 6. We found 36,000 scaffolds that met this criteria, of which 150 passed our filtering criteria. These 150 scaffolds were then connected into a single chain through a designed helix-loop-helix. The sequence of the rest of the protein was then optimized using RosettaDesign as described previously, while keeping the histidine networks intact [19, 22-24]. During design, we required the presence of at least one bulky hydrophobic residue, specifically phenylalanine, on the lytic helix. Designs were then screened by AlphaFold with the following cutoffs: average pLDDT > 90, pTM > 0.8, and RMSD to design model <1.5 (**Table 3.1**). Synthetic genes encoding 26 total designs were constructed and expressed *E. coli* and purified by IMAC and SEC. Of these designs 13 were monomeric by SEC.

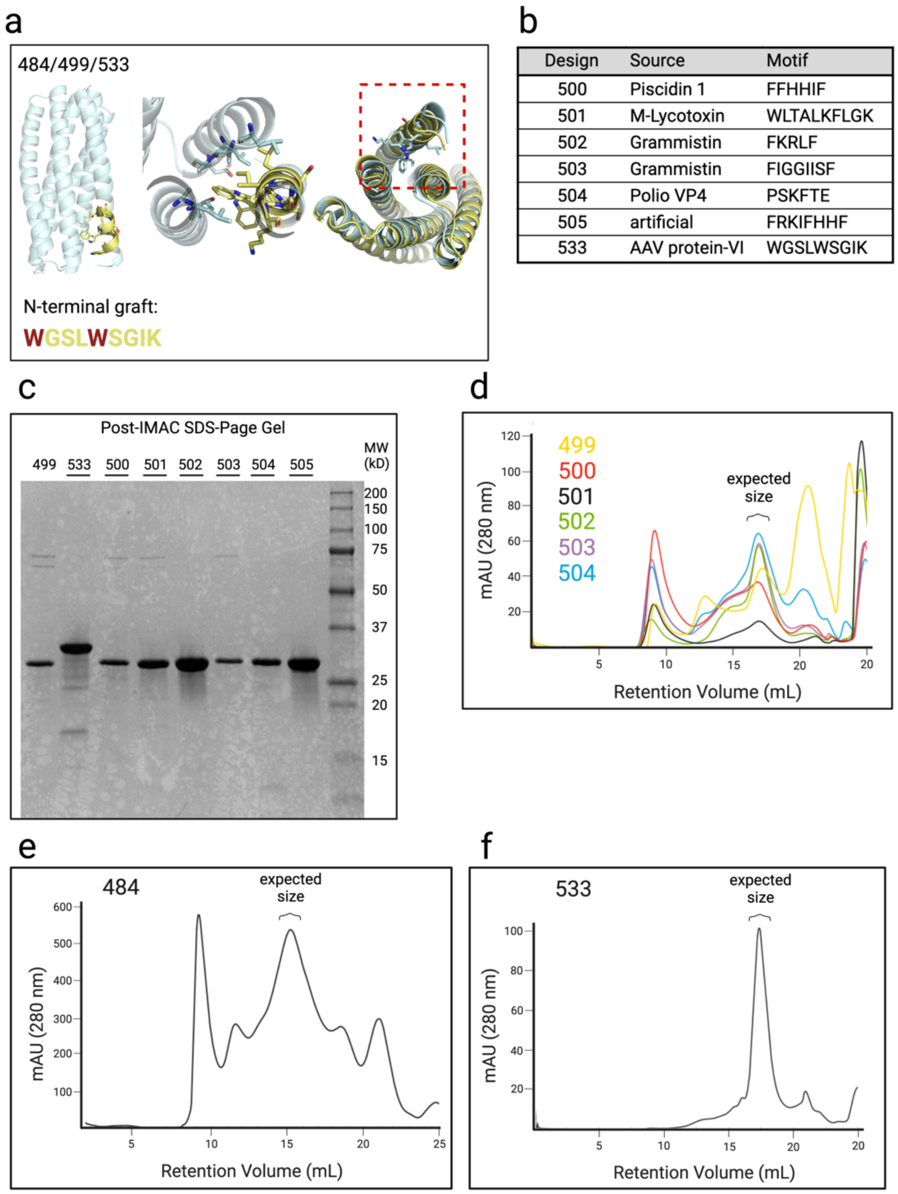
The designs varied by the types of histidine-containing hydrogen-bond networks, number and position of networks, and the number and position of phenylalanine residues on the lytic helix (**fig. 3.3c-f).** All pRLBs were highly stable and maintained their helical propensity at pH 5 (**fig. 3.3c-f)** as measured by circular dichroism (CD). This suggests that pRLBs do not become disordered at low pH but remain helical, like naturally occurring viral proteins involved in endosome permeabilization such as protein VI of AAV [25-27]. pRLBs also maintained their helical content at temperatures up to 95C, suggesting 'rock' like stability (**fig. 3.3g-h)**.

Disruption of synthetic membranes by pRLBs

We then investigated whether pRLBs function as desired, and disrupt synthetic membranes in a pH-driven manner by ejecting the C-terminal helix which includes a stretch of hydrophobic residues. We tested four unique designs that vary in the type, number of pH-responsive networks, and number and location of phenylalanine residues. Designs pRLB-513, pRLB-515, pRLB-518, and pRLB-540 all caused pH-driven liposome disruption at pH 5.0 (**fig. 3.4a).** Their levels of activity correlated with the number of pH-networks, and number of phenylalanines. With three networks and three phenylalanines on the lytic helix, pRLB-540 was the most active. This was followed by pRLB-518, which has two networks and three phenylalanines on the lytic helix. pRLB-515 and pRLB-513 had similar levels of activity, and also contained two networks but with only two phenylalanines on the lytic helix. pRLB-540 also has the longest stretch of contiguous hydrophobic residues due to its networks being placed higher up in the protein, which may have also contributed to its high level of activity.

To confirm disruption was due to the histidine networks, we constructed a pH-KO version of pRLB-540 with the histidines mutated to asparagines. We also compared pRLB-540 to the most active pRO design, pRO-2 I56V. Encouragingly, pRLB-540 was just as active as pRO-2 I56V at pH 5.0 while the pH-KO version was not (**fig. 3.4b**). While pRLB-540 was just as active as pRO-2 I56V, we hypothesized that it may eclipse the activity of pRO-2 I56V at higher pHs due to its phenylalanine content, longer stretch of contiguous hydrophobic residues, and three networks. This turned out to be the case, whereby at pH 5.5 pRLB-540 was more active than pRO I56V (**fig 3.4c)**. To visualize this disruption, pRLB-540 was incubated with liposomes under the same conditions and imaged via cryo-electron tomography. At pH 8.0 there are clear liposomes, while at pH 5.0 the liposomes are completely disrupted **(fig 3.5**).

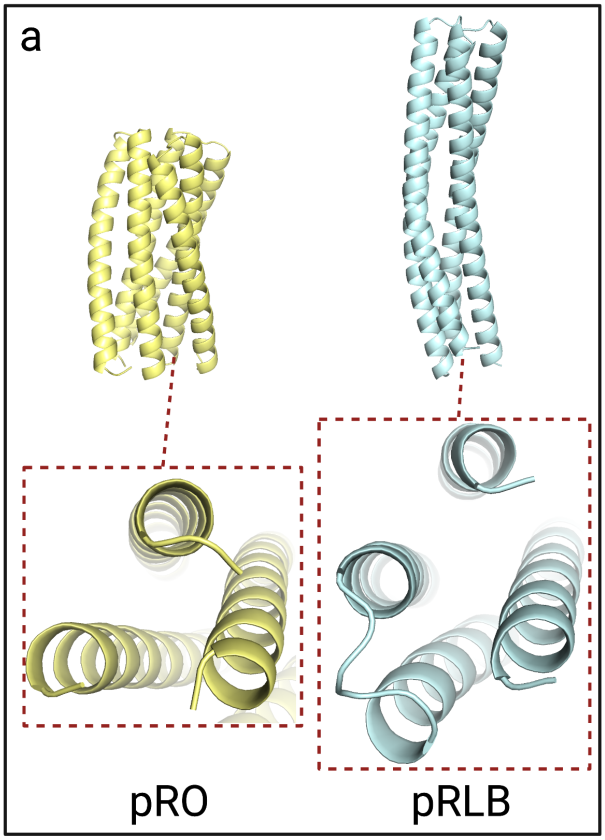
3.2 FIGURES



Design	Source	Motif
500	Piscidin 1	FFHHIF
501	M-Lycotoxin	WLTALKFLGK
502	Grammistin	FKRLF
503	Grammistin	FIGGIISF
504	Polio VP4	PSKFTE
505	artificial	FRKIFHHF
533	AAV protein-VI	WGSLWSGIK

Figure 3.1. Grafting naturally occurring endosome-lytic motifs onto pRO-2-sc-flex

a. Model sc-pRO-2 grafted on the N-terminus with endosome-disrupting motif from adenovirus protein VI. **b.** Endosome disrupting motifs from naturally occurring proteins that were grafted onto sc-pRO bundles. **c.** SDS-Page gel post-IMAC shows robust expression. **d.** Grafted pRO bundles eluted at the expected sizes. **e,f.** Exterior surface charge redesign resulted in the stabilization of protein VI grafted sc-pRO



a. Core size comparison between pRO bundles and pRLBs, showing that pRLBs can accommodate larger hydrophobic residues that are common in naturally occurring endosome disrupting proteins.

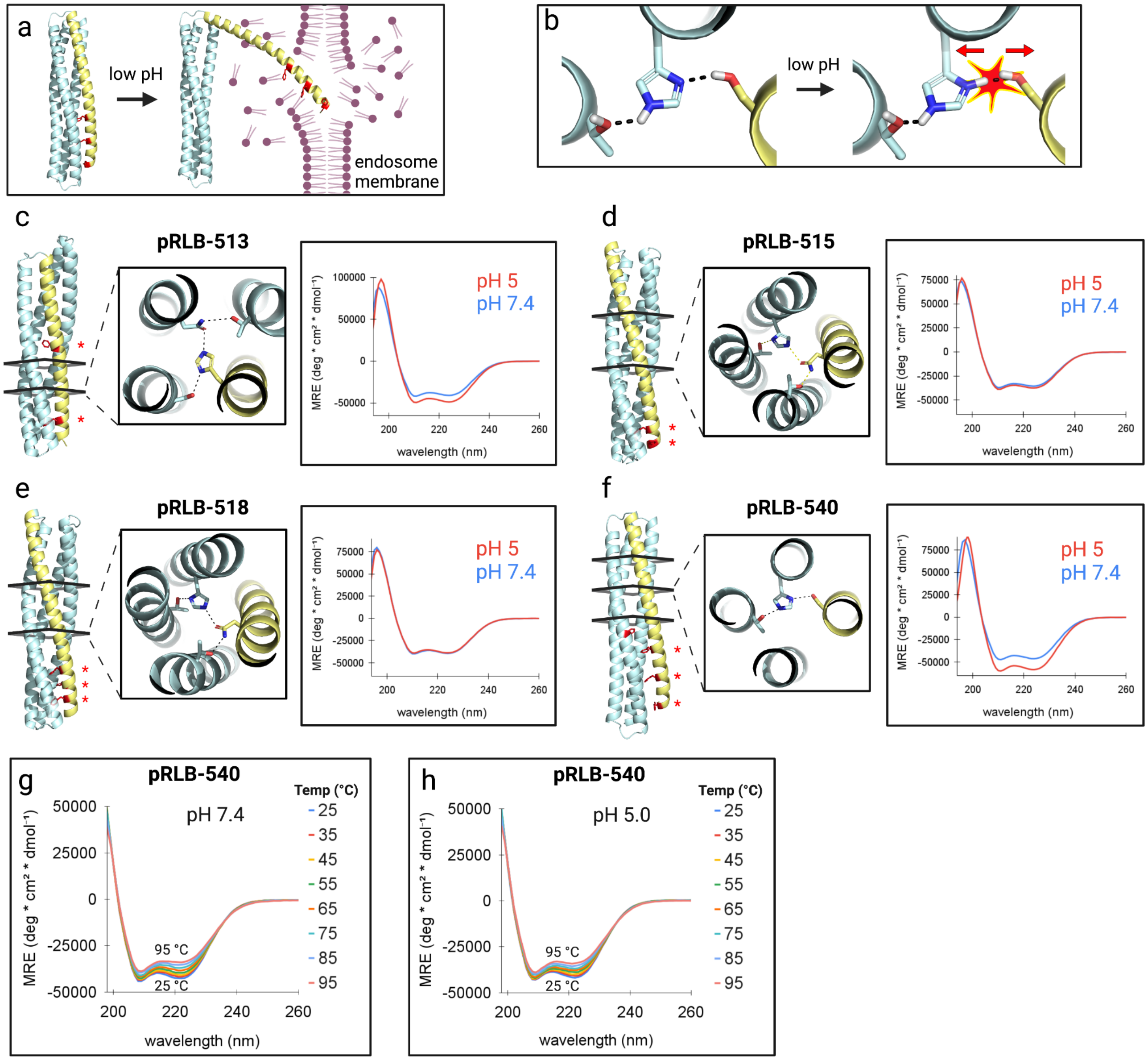


Figure 3.3. De novo design of pH-driven helical bundle for enhanced membrane lysis

a. Design principle behind pRLBs, whereby a lytic amphipathic helix is ejected at low pH to disrupt membranes. **b.** At low pH, histidines participating in buried hydrogen bond networks are protonated leading to electrostatic repulsion and steric clashing that result in a conformational change, ejecting the lytic helix. **c-f.** circular dichroism (CD) measurements at pH 5.0 and pH 7.4

demonstrating the integrity of the helices at low pH. **g-h.** CD measurements demonstrating the rock-like stability of pRLBs at high temperatures at both pH 7.4 and pH 5.0.

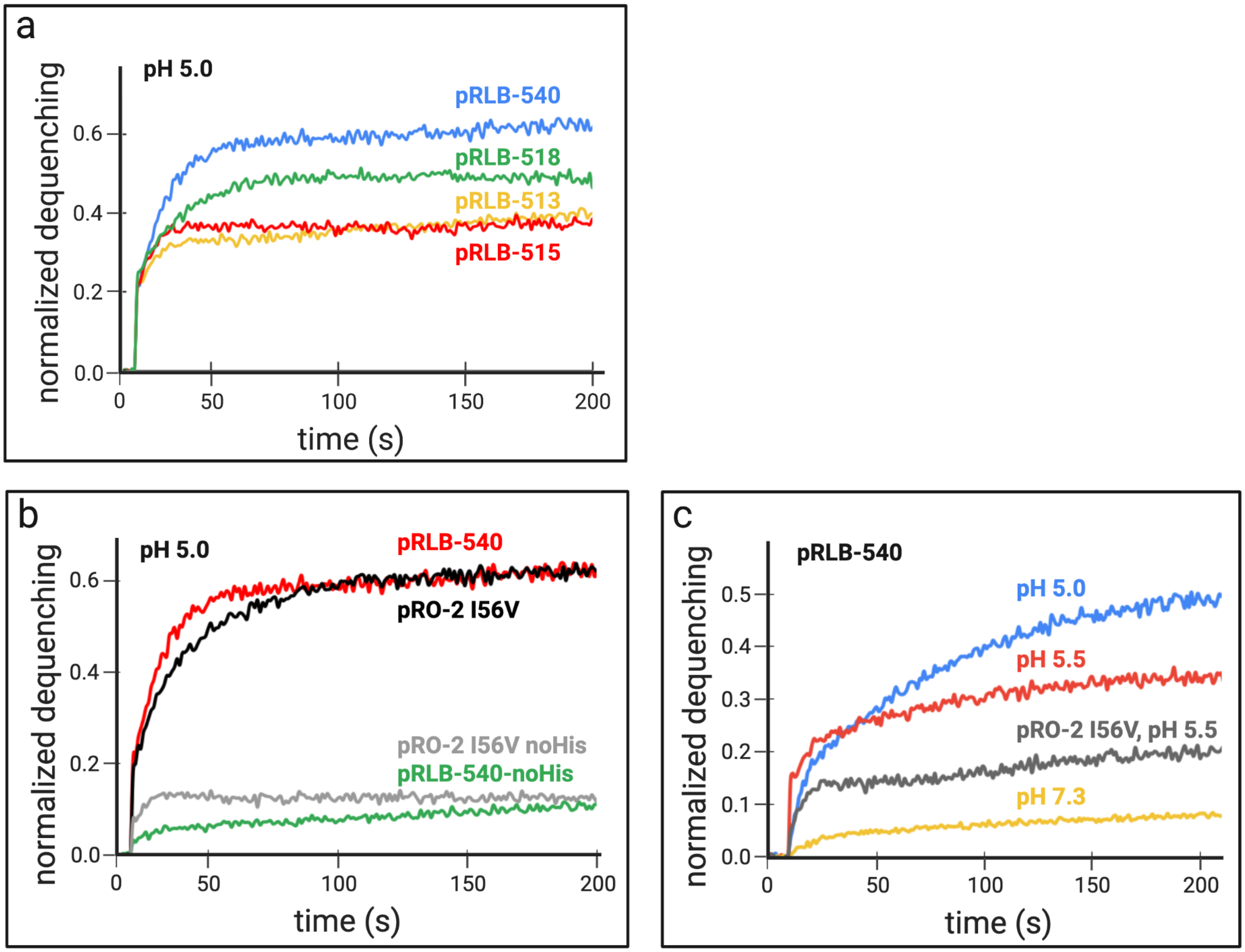


Figure 3.4. Enhanced pH-induced membrane disruption at endosomal pH ranges

a. pRLB designs disrupt synthetic liposomes at pH 5.0 to various degrees that correlate with the number and location of hydrogen bond networks, and phenylalanine residues on the lytic helix. **b.** pRLB-540 disrupts liposomes to the same level as the most active pRO design, and is not active without histidines. **c.** at pH 5.5 pRLB-540 is more active than pRO-2 I56V.

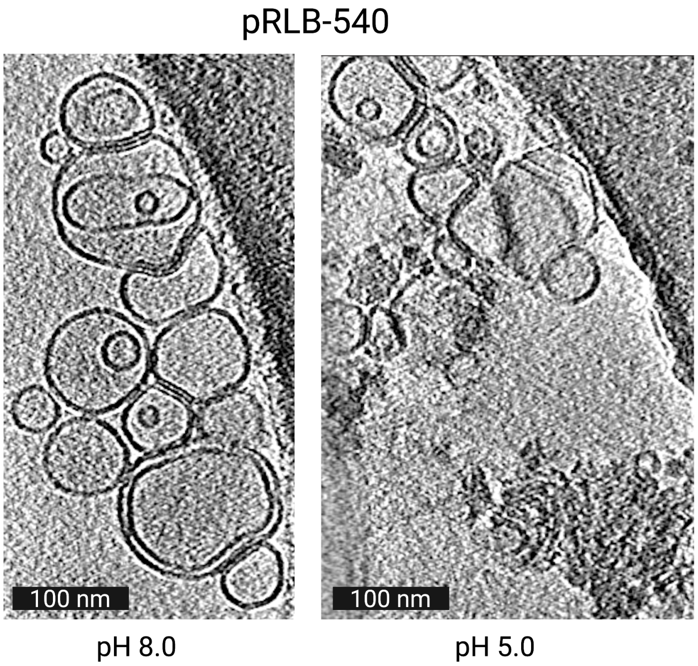


Figure 3.5. Cryo-electron tomography of liposomes incubated with pRLB-540

a. 4 uM of pRLB-540 was incubated with liposomes under the same conditions as the dequenching dye release assay at pH 8.0 and pH 5.0. At pH 8.0 there are clear liposomes, while at pH 5.0 much of the liposomes are disrupted.

Table 3.1. Alpha-Fold scores of pRLB designs

Design	pLDDT	pTM	RMSD to design
pRLB-513	95.39	0.89	0.80
pRLB-515	95.07	0.88	0.98
pRLB-518	95.19	0.89	0.83
pRLB-540	91.62	0.84	2.32

Table 3.2 Sequences of pRO-grafts and pRLB designs

Design	AA sequence
pRO-484	GSEERWGSLWSGIKAASAELKRATASLRAITEELKKGGSGRGDNPGGSEDALVEHNRAIVEHNAIIVEHNRLLLEALELATKDGGSGSGSEEIKKALEELKAASAELKRATASLRAITEELKKGGSGSGSEDALVEHNRAIVEHNAIIVEHNRIIAAVLELIVRALGGSGSGSDEEVRKALEELKAASAELKRATASLRAITEELKKGGSGSGSEDALVEHNRAIVEHNAIIVEHNRILLKVAELISRKG
pRO-499	GSHHHHHHSGENLYFQGGSEERWGSLWSGIKAASAELKRATASLRAITEELKKGGSGRGDNPGGSEDALVEHNRAIVEHNAIIVEHNRLLLEALELATKDGGSGSGSEEIKKALEELKAASAELKRATASLRAITEELKKGGSGSGSEDALVEHNRAIVEHNAIIVEHNRIIAAVLELIVRALGGSGSGSDEEVRKALEELKAASAELKRATASLRAITEELKKGGSGSGSEDALVEHNRAIVEHNAIIVEHNRILLKVAELISRKG
pRO-500	GSHHHHHHSGENLYFQGFFHHIFRKAAEEAKASTAELKRATASLRAITEELKKGGSGRGDNPGGSEDALVEHNRAIVEHNAIVVEHNEALVRALEALVKAGGSGSGSEEIKKALEELKASTAELKRATASLRAITEELKKGGSGSGSEDALVEHNRAIVEHNAIVVEHNRIIAAVLELIVRALGGSGSGSDEEVRKALEELKASTAELKRATASLRAITEELKKGGSGSGSEDALVEHNRAIVEHNAIVVEHNRILLKIIELIAKAA
pRO-501	GSHHHHHHSGENLYFQGGSEEEWLTALKFLGKSTAELKRATASLRAITEELKKGGSGRGDNPGGSEDALVEHNRAIVEHNAIVVEHNRAIVEILRAAVEAGGSGSGSEEIKKALEELKASTAELKRATASLRAITEELKKGGSGSGSEDALVEHNRAIVEHNAIVVEHNRIIAAVLELIVRALGGSGSGSDEEVRKALEELKASTAELKRATASLRAITEELKKGGSGSGSEDALVEHNRAIVEHNAIVVEHNRILLDILRLIARAA
pRO-502	GSHHHHHHSGENLYFQGGSEEEFKRLFEEAKASTAELKRATASLRAITEELKKGGSGRGDNPGGSEDALVEHNRAIVEHNAIVVEHNDVLVKILEALVEKGGSGSGSEEIKKALEELKASTAELKRATASLRAITEELKKGGSGSGSEDALVEHNRAIVEHNAIVVEHNRIIAAVLELIVRALGGSGSGSDEEVRKALEELKASTAELKRATASLRAITEELKKGGSGSGSEDALVEHNRAIVEHNAIVVEHNRILLTILELIAKSA
pRO-503	GSHHHHHHSGENLYFQGGSEEEFIGGIISFKASTAELKRATASLRAITEELKKGGSGRGDNPGGSEDALVEHNRAIVEHNAIVVEHNRALIEALRLLVERGGSGSGSEEIKKALEELKASTAELKRATASLRAITEELKKGGSGSGSEDALVEHNRAIVEHNAIVVEHNRIIAAVLELIVRALGGSGSGSDEEVRKALEELKASTAELKRATASLRAITEELKKGGSGSGSEDALVEHNRAIVEHNAIVVEHNRILEKIIELISRRA
pRO-504	GSHHHHHHSGENLYFQGGSPSKFTEAAEEAKASTAELKRATASLRAITEELKKGGSGRGDNPGGSEDALVEHNRAIVEHNAIVVEHNRALIKALDQIVRAGGSGSGSEEIKKALEELKASTAELKRATASLRAITEELKKGGSGSGSEDALVEHNRAIVEHNAIVVEHNRIIAAVLELIVRALGGSGSGSDEEVRKALEELKASTAELKRATASLRAITEELKKGGSGSGSEDALVEHNRAIVEHNAIVVEHNRILLTIIELIARKA
pRO-505	GSHHHHHHSGENLYFQGGSEEEFRKIFHHFKEASEELKRATEDLREITEELKKGGSGRGDNPGGSEDALVEHNRAIVEHNAIVVEHNRAAVEALKILVRAGGSGSGSEEIKEALEELKEASEELKRATEDLREITEELKKGGSGSGSEDALVEHNRAIVEHNAIVVEHNRIIAAVLELIVRALGGSGSGSDEEVREALEELKEASEELKRATEDLREITEELKKGGSGSGSEDALVEHNRAIVEHNAIVVEHNRILTKIIDLLAKRA

pRO-533	GSHHHHHHSGENLYFQGGSEERWGSLWSGIKKASEELKRATKKLKKKTEELKRGGSGRGDNPGGSEEALVRHNEQIVKHNEIVVEHNRLLLEALELAVKEGGSGSGSEEIKEALKKLREASKRLEEATKKLQKKTEELKKGGSGSGSEEALVRHNEQIVRHNEIVVEHNEIIAEVLELIVKALGGSGSGSEEVKKALEELRKASEELRRATEKLKKKTEELRKGGSGSGSEEALVRHNQQIVRHNEIVVKHNEILLKVAELITRQ
pRLB-513	GSHHHHHHSGENLYFQGSEEEEEQLEKIRKILEETEELAEETERVAREAEEVVKEAQEAAELVEEIAKNPDEEERKRVREEVERILREAERLLRELERLARENKKAVERNQEAVRLAKELAKRKGDEELRRAAKRLEESVRELERSVRELERAIKRAQELLREAQEEVRKGGSGSGENLYFQGGSGSLEELAKLLQRLVELLRQLVELFKRHVELLRRHRELAQEFEEIAKK
pRLB-515	GSHHHHHHSGENLYFQGSEELREEAKRLAEEAERLLREHERILREVREVLEKHEKALEEARKALERAKRGEIPEEELKRVLEKLQRVAKETREALRRLQESLEKAKEALRLLKRALEIAKKLGLSEEEKKLAERAERVAREAEKVLRETERLAREAERILKETEELLREAEELLRGGSGSGENLYFQGGSGSEEEREEVLRKLEEALRRNQEAVKRLQEILRANEEVIKELQEVLKEFKRFMG
pRLB-518	GSHHHHHHSGENLYFQGSEELREEAKRLAEEAQKIAEEHREVLERVEEVLKEHEEALKKAEEMLERLKRGEADDEEARRALEELLRVAEKTERALQELRESLQRLKRALELLRRALKLLKKLGLSEEERKVAEEIEKLAKEAQELAERTEELAKEAQEVLRETQRLLREARELLRNGGSGSGENLYFQGGSGSEEKERAARELEEALRRNKELVRRLQEALRQNKEALKRFEQAFREFQRLLG
pRLB-540	GSHHHHHHSGENLYFQGSEEERARELLKRARELLEEAERLLREHEELLEEHRRALEEHQELLERAKKLGEEELKEAARRSEQVLRITEEVARRSEEIARRVEKAAREIREMAEELRKNVPDSEEARKIAQELREIAERLQEVAKRFREVVQRAREILERIARILREIIELLRGGSGSGENLYFQGGSGSEEALRRARELSEEARRRSEELERQSRKLIREFEQLAREFRRAARKFG

Table 3.3. Mass spectrometry of purified pRLBs

Design	Expected Mass (Da)	Observed Mass (Da)	Difference (-1M)
pRLB-506	27192	27061	-131
pRLB-507	27400	27269	-131
pRLB-511	27348	27217	-131
pRLB-512	27264	27133	-131
pRLB-513	27531	27400	-131
pRLB-514	27139	27008	-131

pRLB-515	28784	28653	-131
pRLB-516	27046	26915	-131
pRLB-517	27402	27271	-131
pRLB-518	28694	28564	-130
pRLB-519	28784	28653	-131
pRLB-539	27290	27160	-130
pRLB-540	28261	28130	-131

3.3 METHODS

Computational Design Methods

Backbone sampling: Oligomeric protein backbones of four α-helices were produced by systematically varying helical parameters using the Crick generating equations, and using ideal values for the supercoil twist (ω0) and helical twist (ω1)(19, 38-39). The offset along the z-axis (Z-offset) for the first helix was fixed to 0 as a reference point, with the rest of the helices independently sampled from -1.51 Å to 1.51 Å, with a step size of 1.51 Å.

Hydrogen bond networks: the HBNet (19,21) method in Rosetta was used to select hydrogen bond networks that contain one histidine connecting at least three of the four helices. HBNet was used to select backbones that could accommodate a minimum of 2 such networks in the homotetrameric backbones.

Connecting loops: After scaffolds with desired hydrogen bond networks were found, the four helices were connected into a single chain by short structured loops as described previously (19, 21).

Rosetta design: To design the sequence and sidechain rotamer conformations for the rest of the protein surrounding the hydrogen bond networks, the network residues were constrained using AtomPair constraints on the donors and acceptors of the hydrogen bonds and RosettaDesign was carried out using layer design with an added design feature of including phenylalanine on the C-terminal helix. Designs were selected based on their AlphaFold prediction scores (36), with cutoffs of pLDDT > 0.92, pTM > 0.8, and RMSD to design model < 1.5 Å.

Protein expression and purification

Plasmids encoding pRO and pRLBs were ordered from IDT or Genscript as described in Methods of Chapter 2. These plasmids were transformed into chemically competent E. coli expression strains BL21(DE3)(non Star) (Invitrogen) or Lemo21(DE3) (New England Biolabs). After transformation, single colonies were picked from agar plates and grown overnight in 50 mL starter cultures of Luria-Bertani (LB) medium containing 50 µg/mL kanamycin with shaking at 225 rpm for 6-12 hours at 37°C. 500 mL of TBM-5052 supplemented with 50 µg/mL kanamycin was inoculated with 10 mL starter culture for autoinduction. Cultures were incubated at 18 °C with shaking at 225 rpm for 18-24 hours. Cultures were harvested by centrifugation at 4,000 rcf for 10 minutes and stored at -80 °C. Cell pellets were resuspended in Tris buffer (25 mM Tris, 150 mM NaCl, 2% glycerol), and lysed by sonication or microfluidizer. The resulting lysate was clarified by centrifugation at 24,000 rcf for 30 minutes and passed through 2 mL of Nickel-Nitrilotriacetic acid agarose (Ni-NTA) (Qiagen), followed by 3 washes (25 mM Tris, 150 mM NaCl, 40 mM Imidazole, pH 8), and eluted into 2x3 mL of Tris elution buffer (25 mM Tris, 150 mM NaCl, 300 mM Imidazole, pH 8). EDTA was immediately added to 5mM final concentration. Constructs were then filtered by passing through 0.2 um filter, and purified by SEC on a Superose 75 Increase 10/300 or Superose 200 Increase 10/300 column (GE) equilibrated with Tris buffer (25 mM Tris, 150 mM NaCl, 2% glycerol, pH 8). His tags were not removed, as they caused aggregation for the pRO-grafts.

Circular dichroism (CD)

Proteins were buffered-exchanged into 10 mM phosphate, 100 mM NaCl pH 7.4. Samples were diluted to about 0.2 mg/mL in the appropriate phosphate buffer, and the CD spectrum from 260 nm to 190 nm was recorded. CD analysis was carried out on a JASCO J-1500, using a 1 mm quartz cuvette. For temperature melt analysis, samples were heated at a ramp rate of 1 °C/min while monitoring ellipticity at 222 nm. Data were collected at 1 °C intervals without pausing temperature ramping. To vary the pH, pRLBs were acidified for 3 hours before measurement, as described in chapter 2. Units were converted to mean residue ellipticity (MRE) by dividing the raw spectra by the formula $N \times C \times L \times 10$, where N is the number of residues, C is the concentration of protein, and L is the path length (0.1 cm).

Liposome preparation and disruption assay

See Methods section of Chapter 2.

CHAPTER 4. Development of a prime editing CRISPR/Cas9-based reporter system for high-throughput screening and sensitive detection of functional RNA delivery

4.1 Motivation and results for RNA delivery reporter system

While we now have a nucleocapsid that can disrupt synthetic membranes and new enhanced pH-driven membrane permeabilizing domains with greater potential for evolution, we need a method by which to benchmark our designs in a cell based context and ultimately screen a library for potent cargo delivery. However, after trying several cell-based approaches including split-GFP complementation [28], biotin ligase assay [29], and Gal8 endosome localization [30] among others, none were sensitive enough and high throughput enough to screen designs for intracellular delivery. The ability to test designs is made difficult by the lack of assays to detect and quantify endosome escape and RNA delivery [31]. There have been no reports of an assay that is both highly sensitive and high-throughput enough to screen and evolve a library of vehicles for RNA delivery. To address this, we developed a prime editing[32] CRISPR/Cas9-based reporter system for high throughput screening and sensitive detection of intracellular RNA delivery. Stable monoclonal cell lines were generated using lentivirus to insert the reporter system cassette into both HEK293T and HeLa cells (**fig. 4.1a**).

The reporter system works in the following way (**fig. 4.1b**): in the off state the cells express mCherry and have blasticidin resistance. After successful delivery of prime-editing gRNA (pegRNA), the genome is edited to introduce a missing cysteine residue that is part of the catalytic triad necessary for puromycin N-acetyltransferase (PAC) to function. Further, the genome is frame-shifted by one residue, which turns off mCherry and Blasticidin S deaminase

(BSD) and turns on expression of mNeon. Further, prime editing enables editing up to 80 bp of the genome [32] which can be utilized to introduce an RNA barcode into each pegRNA that is then written into the genome upon successful RNA delivery. Edited cells can then be selected and quantified through several methods to easily detect functional RNA delivery, including next generation sequencing (NGS), fluorescence activated cell sorting (FACS), high-throughput imaging, and antibiotic selection. This reporter system enables us to screen a large library of membrane-permeabilizing nucleocapsids that encapsulate barcoded pegRNA. Because of the genotype to phenotype link of nucleocapsids, we can assess enrichment of potent designs that led to RNA delivery through an NGS analysis of enriched barcodes.

To validate this reporter system, successive 10-fold dilutions of RNA starting at 5 pmol down to 5 fmol were delivered via the commercial RNA transfection reagent XFect in a 96-well format in triplicates. Cells were then incubated for 3 days before imaging for mNeon+ cells. 21 images were collected per replicate, or 63 per condition. Untargeted pegRNA led to no visible edits (**fig 4.2)**, while 5 pmol of commercially transfected pegRNA led to about 50% frameshifts in both HEK293 (**fig 4.2a)** and HeLa (**fig 4.2b**) cell lines. On-target edits were confirmed by NGS and sanger sequencing. Successive dilutions of 500, 50 and 5 fmol led to frameshifts in about 7 - 12%, 0.5 - 1.8%, and 0.07% - 0.16% of cells in HEK293s and HeLas, respectively. A statistical image analysis showed that RNA delivery can detect frameshifts with as little as 5 fmol of RNA delivered in both HEK293s and HeLas (**fig 4.3a-b**).

4.2 FIGURES

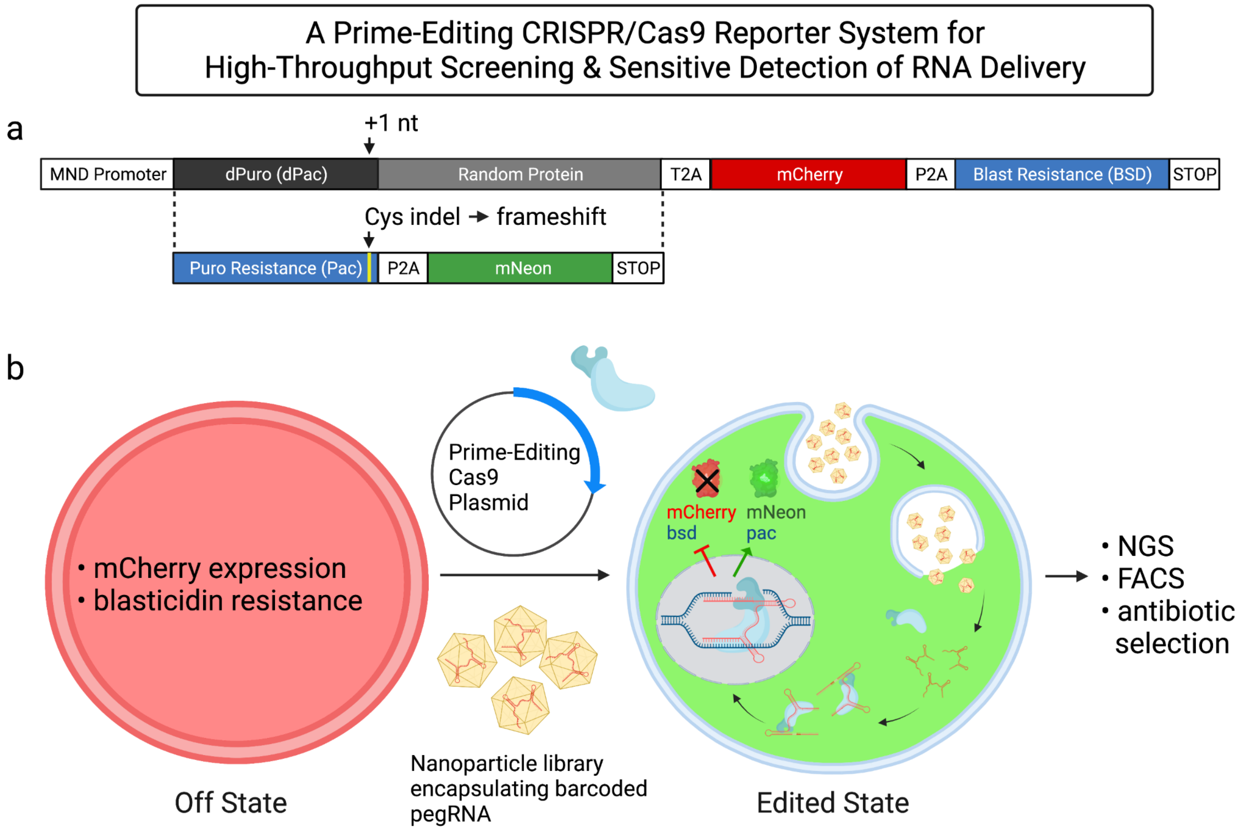
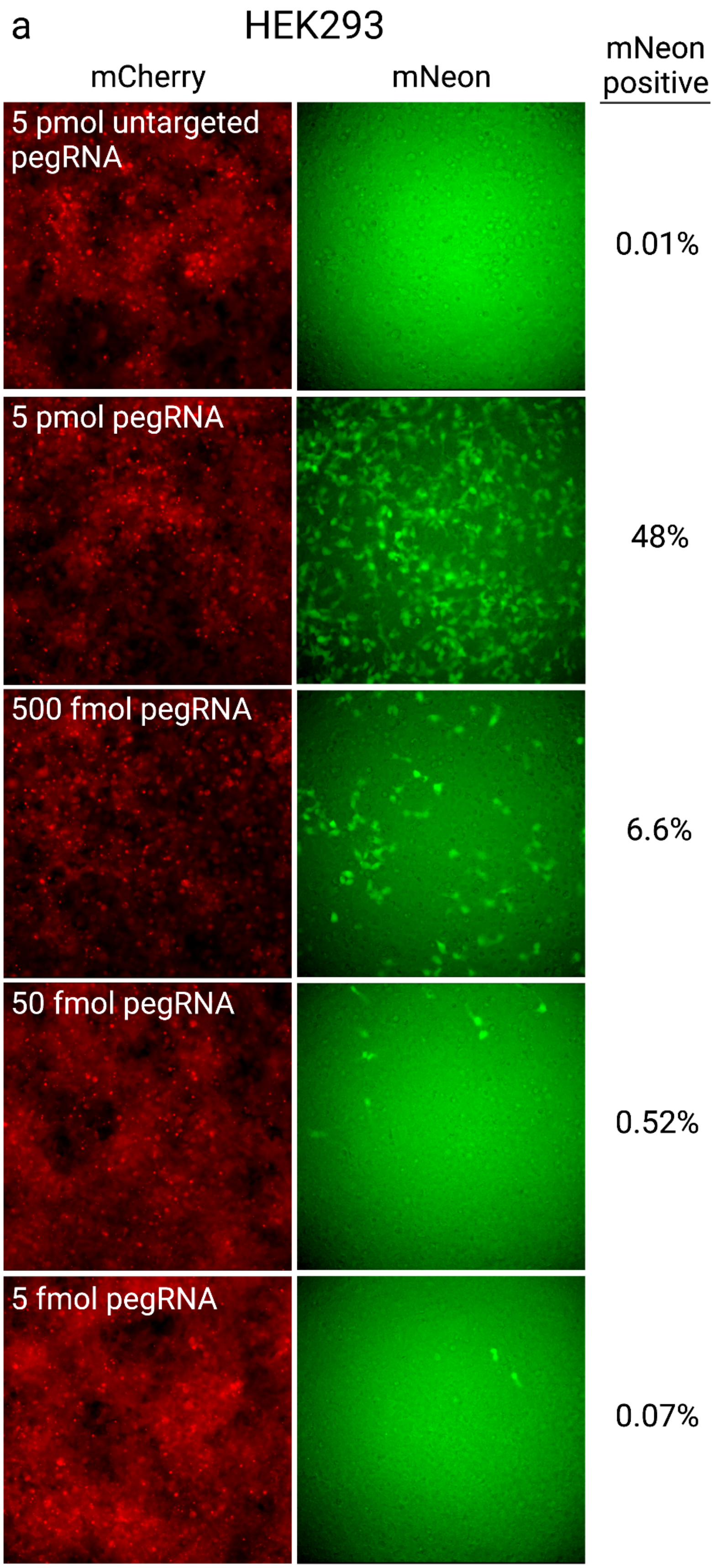


Figure 4.1. Development of novel cell-based reporter system for HTS and sensitive detection of functional RNA delivery

a. Gene cassette encoding the frameshift reporter. **b.** In the off-state, mCherry is expressed along with blasticidin resistance. Upon successful gene editing, a missing cysteine part of a catalytic triad is introduced to save the dead puromycin (dPuro). The edit causes a frameshift that turns off mCherry expression and blasticidin resistance while switching to the correct frame encoding mNeon. Synonymous mutations or linkers can be introduced into the genome between the puromycin gene and mNeon, which can act as RNA barcodes for a powerful library-based screen. Edited cells can be quantified through high-throughput imaging, next generation sequencing (NGS), fluorescence activated cell sorting (FACS), or selected with antibiotics. This system can be used for screening a large library delivering barcoded pegRNA, or to quantify and benchmark delivery efficiency of different designs.


a
HEK293
mCherry
mNeon
mNeon positive
5 pmol untargeted pegRNA
0.01%
5 pmol pegRNA
48%
500 fmol pegRNA
6.6%
50 fmol pegRNA
0.52%
5 fmol pegRNA
0.07%

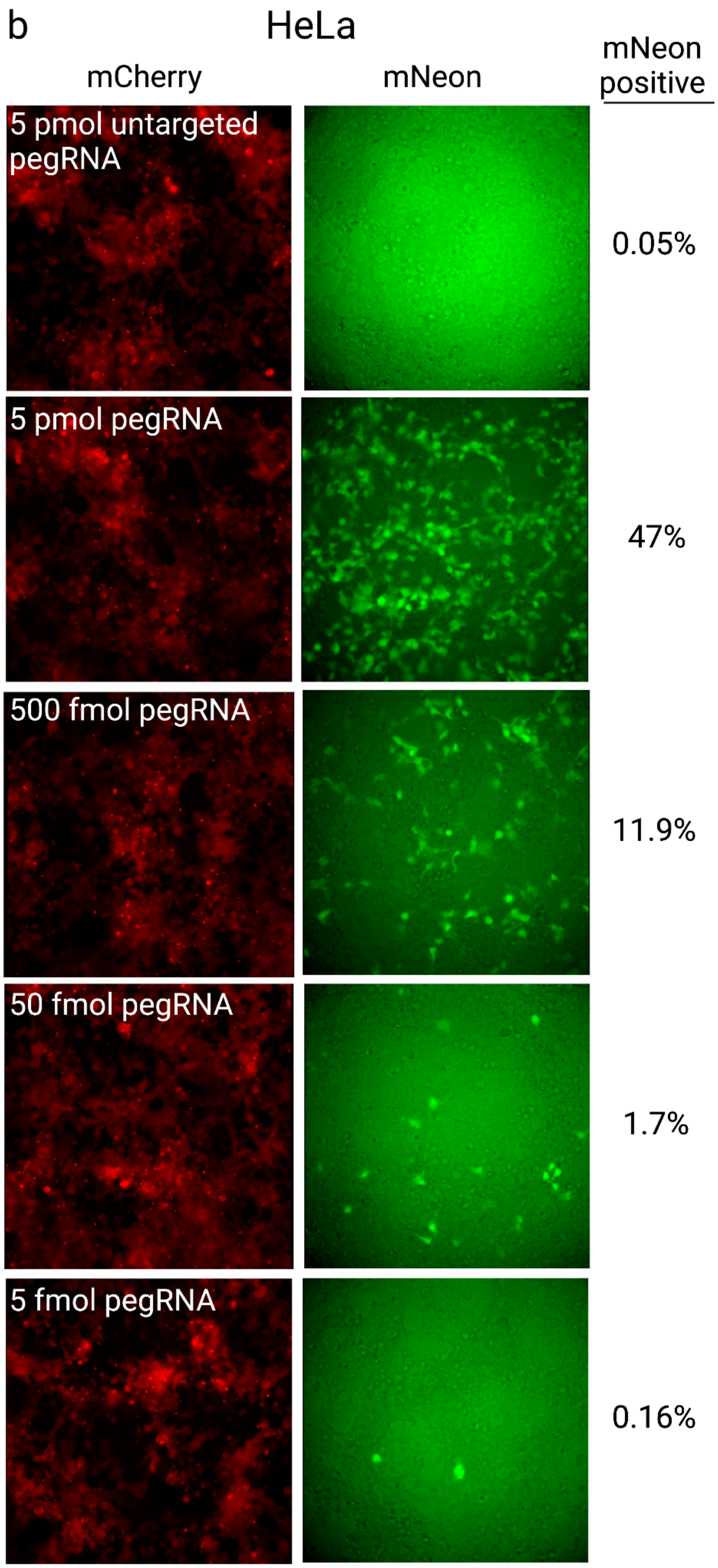

b
HeLa
mCherry
mNeon
mNeon positive
5 pmol untargeted pegRNA
0.05%
5 pmol pegRNA
47%
500 fmol pegRNA
11.9%
50 fmol pegRNA
1.7%
5 fmol pegRNA
0.16%

Figure 4.2. Validation and sensitivity of reporter system

a, b. A titration of commercially transfected pegRNA starting at 5 pmol with four successive 10-fold dilutions down to 5 fmol. Each conditions was done in triplicates in a 96-well format. In both HEK293 and HeLa cells frameshifts were detected with as low as 5 fmol of RNA. High-throughput imaging, followed by analysis of mNeon+ cells revealed the percent of edited cells at each amount of pegRNA delivered. 21 images were collected per triplicate, and analyzed by CellProfiler. To determine the percent of edited cells, we estimated each well contained 30,000 cells.

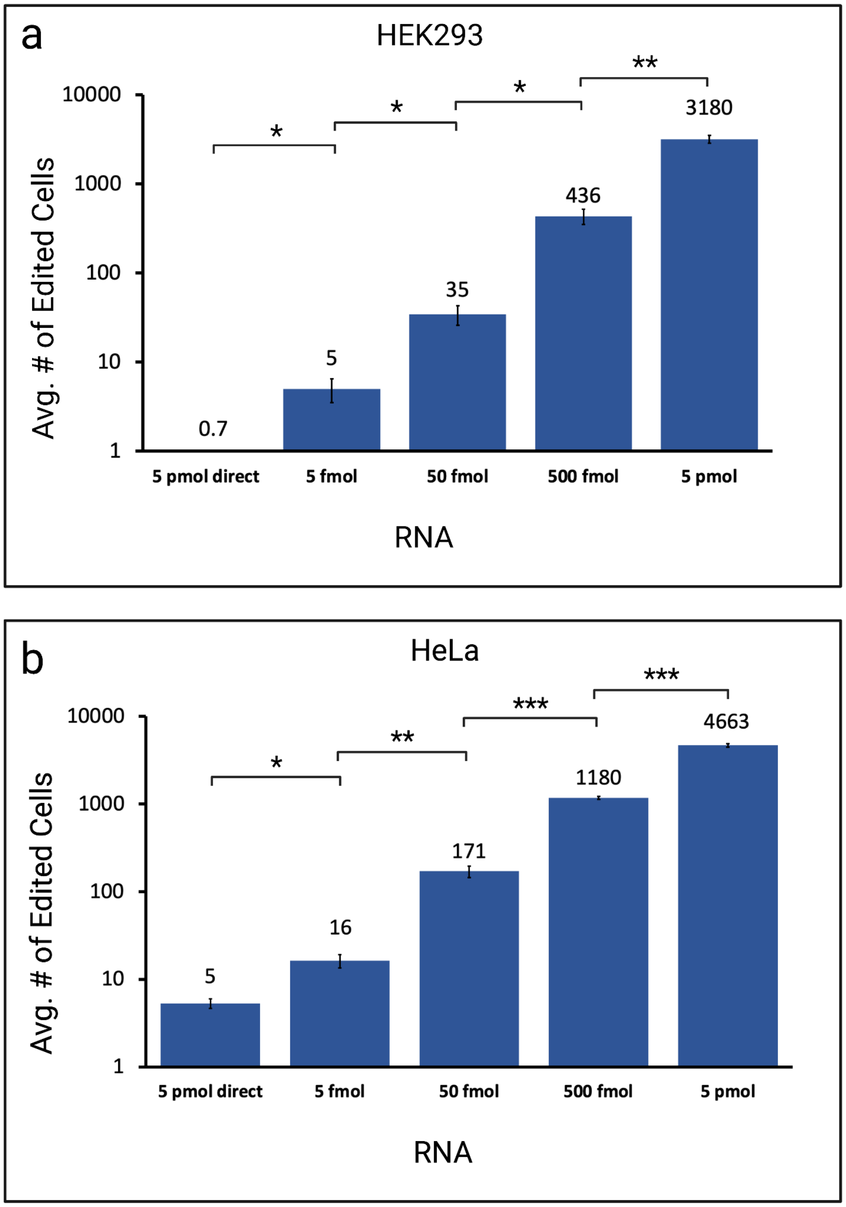


Figure 4.3. Image analysis and quantification of RNA delivery sensitivity

a,b. CellProfiler image analysis to quantify the number of mNeon+ cells per condition. Data was used to estimate the % of edited cells in **fig. 4.2.** An ANOVA test was performed on all conditions and a t-test with equal variance was performed between conditions.

4.3 METHODS

Generating mammalian monoclonal cell lines

Human embryonic kidney cells (HEK293T) and HeLa cells were acquired from ATCC. Cells were transduced with lentiviral particles delivering an MND-driven gene encoding the reporter cassette. Cells were then selected with blasticidin over a period of two weeks. To generate monoclonal cell lines, cells were plated at <1 cell/well in a 96-well format and grown to confluence. Monoclonal cells were selected for robust expression of mCherry as well as mNeon after editing. To verify that the reporter cassette is functional, cells were treated with puromycin before editing to ensure complete cell death.

pegRNA

Stabilized pegRNA was ordered from IDT, constructed such that the first three and last three bases are modified with 2'OMethylations and contain phosphorothioate bonds.

Testing detection limits of frameshift assay

Cells were plated at 15k cells/well onto a Greiner Poly-L-Lysine coated 96-well plate (ref 655936). 24 hours after plating, cells were transfected with the PE2 and PE3b plasmids using Lipofectamine 3000 (ThermoFisher) according to the manufacturer's instructions. 48 hours after plating, the media was replenished with 50% phenol-free DMEM containing 10% FBS and 1% penicillin-streptomycin and 50% OptiMEM. Cells were then transfected with pegRNA at 10-fold dilutions starting at 5 pmol down to 5 fmol using XFect Transfection Reagent (Takara) in triplicates. Cells were imaged periodically using the IN Cell 2000 analyzer, with 21 images collected per replicate. Cells were cultured in phenol-free DMEM containing 10% FBS and 1% penicillin-streptomycin, and incubated at 37 °C and 5% CO2.

CellProfiler image analysis

A CellProfiler protocol was used to analyze images for edited cells expressing mNeon. The following modules were used: 1. EnhanceOrSuppressFeatures (operation: supress, features size: of 10), 2. Opening (structuring element: disk, feature size: 10), 3. IdentifyPrimaryObjects (typical diameter of objects in pixels: 20 - 300, discard objects outside of the diameter range: yes, discard objects touching border: no, threshold strategy: global, thresholding method: minimum cross-entropy, threshold smoothing scale: 1.2, threshold correction factor: 0.5, lower and upper bounds on threshold: 0.14 - 1.0, log transform before thresholding: yes, method to distinguish clumped objects: intensity, method to draw dividing lines between clumped objects: intensity), 4. MeasureObjectIntensity, 5. ExportToSpreadsheet.

CellProfiler settings must be normalized to negative and positive controls between different imaging experiments on different plates, thus these values are starting recommendations but must be slightly adjusted between different imaging experiments.

An ANOVA test was performed on all conditions and a t-test with equal variance was performed between conditions.

CHAPTER 5. Targeted Delivery of RNA via synthetic nucleocapsids

5.1 Motivation and results for RNA delivery by nucleocapsid assemblies

In order to use this reporter system for screening a large library of nucleocapsids, we must be able to package barcoded pegRNA such that each barcoded pegRNA corresponds to a particular nucleocapsid design. To achieve this, we must express the pegRNA at the same time as the nucleocapsid components for co-assembly in microbes. Based on previous results on the relationship between mRNA length and packaging efficiency **(fig. 2.4 and fig. 2.5)** we hypothesized that overexpression of a ~200 bp pegRNA would outcompete the genome's mRNA and lead to enriched packaging in *E. coli*. Further, we hypothesized that utilizing specific packaging mechanisms may also lead to greater pegRNA loading.

Specific and Nonspecific Packaging of pegRNA in E. coli

To investigate these possibilities, we constructed two variants of I53-50-V4 for both specific and nonspecific packaging. A version of V4 was generated with a circularly permuted pentamer (V4-cp), such that the N- and C-termi face inward. This enabled us to fuse the 22 amino acid RNA-binding domain (RBD) λN derived from bacteriophage [33-35] to the inside of V4 with a 20% slip sequence, such that every pentamer had on average 1 λN fusion facing the interior. Further, the complementary ~20 bp BoxB RNA hairpin recognized by λN was added onto the 3' end of a T7-driven pegRNA for specific packaging (V4-cp-RBD). Finally, a third version was constructed with a V4-cp that co-expresses a pegRNA for nonspecific packaging (**fig. 5.1a**).

These constructs were expressed in *E. Coli* and purified by IMAC and SEC. Nucleocapsids were then quantified for pegRNA packaging using RT-qPCR. Nonspecific packaging of co-expressed pegRNA led to loading of pegRNA in 120/1000 capsids which corresponds to 1 out of every 8 nucleocapsids containing a full-length pegRNA. Specific packaging via V4-cp-RBD led to 2.4-fold greater loading at ~287/1000 capsids which corresponds to 1 out of every 3.5 nucleocapsids containing pegRNA. V4-cp without a co-expressed pegRNA had a background level of 3.8/1000. All measured differences were significant to $p<0.001$.

With pegRNA packaging confirmed, we chose to move forward with nonspecific packaging of pegRNA due to its simplicity. In order to evolve a library, we need a detectable signal of RNA delivery to start from. However, incubation of up to 50 nM of V4-cp packaging pegRNA did not lead to a detectable signal via the reporter assay. We hypothesized that this was due to the instability of *E. coli* pegRNA, which is not 5' capped and does not have a polyA tail and is therefore susceptible to mammalian nucleases. Delivery of in vitro synthesized pegRNA using the commercial transfection reagent XFect under ideal conditions revealed that there is a 100-fold reduction in frameshifts in pegRNA similar to that produced by *E. coli* (data not shown). Knockout of endogenous nucleases failed to significantly improve this reduction in editing efficiency. Other directions that we were not able to try but may be worth pursuing are to use a eukaryotic host to produce nucleocapsids such as yeast or mammalian cells, which have the advantage of packaging more stable and less immunogenic eukaryotic RNA.

In Vitro Assembly and pegRNA Loading

We hypothesized that packaging of stabilized pegRNA may lead to a detectable signal. This stabilized pegRNA differs from *E. coli*-produced pegRNA in that the first three and last three bases are modified with 2'-O-Methylations and are linked with phosphorothioate bonds. Both these features help resist nuclease degradation. We also hypothesized that the display of targeting domains may lead to increased cellular internalization, and thus a higher concentration of nucleocapsids with membrane-permeabilizing domains in endosomes that will increase the likelihood of RNA delivery.

To test these hypotheses, we constructed ten different assemblies, B1 through B10, with varying targeting and membrane-permeabilizing domains fused to the I53-50-V4 trimeric component (**fig. 5.2**). For targeting domains, we chose an in-lab designed EGFR mini-binder, 'EGFRn', as well as a Transferrin receptor mini-binder, 'TfRn' [35]. These targeting domains were displayed at 25% valency. For endosome permeabilizing domains, we chose sc-pRO-2-flex, pRO-533, and pRLB-540. These domains were displayed at 50% valency. All components were expressed in *E. coli* and purified by IMAC and SEC (**fig. 5.3a**). Components were assembled in vitro at 25°C in the presence of pegRNA at an RNA:nucleocapsid molar ratio of 5:1, and incubated for at least 30 minutes. Assembly components were purified by IMAC and SEC, and verified by SDS-PAGE (**fig. 5.3b**). To confirm assembly, we assessed the size and polydispersity by DLS. All assemblies were monodisperse with a low polydispersity index (PDI) ≤0.2, and had an average diameter around their expected size (**fig. 5.4**). We imaged a subset of these assemblies by negative-stain EM and confirmed the presence of capsid-like structures (**fig 5.5a**). Class averages of two assemblies, B2 and B4, showed density around the corona of the nucleocapsid (**fig 5.5b**). The unresolved corona confirmed the presence of displayed flexibly-linked proteins. Assembly B2, which displays the long pRLB-540 domain at 50% valency, had a particularly blurry periphery. In contrast, assembly B4 had slightly sharper class averages with its smaller EGFRn mini-binder displayed at 25% valency and a smaller pRO-2-sc-flex membrane permeabilizing domain displayed at 50% valency.

We next assessed the extent at which these assemblies package and protect pegRNA. A native gel after RNase A treatment confirmed that all assemblies packaged RNA and most protected against RNase A (**fig 5.6a-b**). Analysis of the area intensity of RNA bands revealed that between 1-3 pegRNAs were packaged per nucleocapsid (**fig 5.6c**). Packaging efficiency varied between 39% - 81% while RNase A protection efficiency varied between 7% - 83%. The RNase A protection efficiency of in vitro assembled V4 was 61%, which is similar to published RT-qPCR data for bicistronic assembly [18].

Live-cell Imaging of Intracellular Localization

To determine whether these nucleocapsids are taken up by cells, two EGFrn- and TfRn-targeted assemblies were labeled with AF647 dye and incubated with cells with an early endosome marker. Live-cell images were taken every 3 minutes over a period of 1 hour starting after 20 minutes of incubation. In both cases, nucleocapsids are seen attaching to the cell surface at the 20 minute time point and being internalized into puncta by the 1 hour time point (**fig. 5.7**). These results confirmed internalization of nucleocapsids, and showed some overlap with early endosome markers which suggests internalization into the endocytic pathway.

We next conducted a similar experiment to determine the localization of packaged RNA, and whether pH-responsive domains change the localization of nucleocapsids. We tested three different assemblies packaging ATTO-550 labeled RNA: 1. untargeted V4 (B1), 2. EGFRn-targeted V4 (B2), and 3. EGFRn-targeted V4 displaying pRO-2-flex (B3). These nucleocapsids were incubated with HeLa cells at a final concentration of 10 nM, with a theoretical maximum of 50 nM of packaged RNA. After incubation over 0.5 hr, 1 hr, and 3 hr time points, cells were fixed, stained with a LAMP2A lysosomal marker, and imaged. Untargeted V4 nucleocapsids were slow to internalize with most localizing to the cell surface, and no visible RNA was detected inside cells (**fig. 5.8a**). In contrast, EGFR-targeted V4 was internalized along with the packaged RNA within an hour into visible puncta. EGFR-targeted V4 mostly colocalized with lysosomes (**fig. 5.8b**). Finally, EGFR-targeted V4 displaying a pH-responsive membrane-permeabilizing domain appeared to not colocalize with lysosomes along with the packaged RNA (**fig. 5.8c**). This data suggests that targeting domains facilitate internalization of nucleocapsids, and that the pH-responsive domains help to halt endosome maturation into lysosomes. Further, RNA appeared to co-localized with cages throughout the 3 hour incubation. However, it is possible that longer time periods may result in the release of the RNA through various cellular processes.

Intracellular Delivery of pegRNA

With cellular targeted internalization confirmed, we then moved to testing these assemblies for the ability to deliver their packaged RNA payload using our reporter system. We incubated cells with 5 nM of each nucleocapsid assembly for 24 hours before replacing the media. Cells were then placed under puromycin selection 48 hours after addition of nucleocapsids to begin selecting for edited cells and amplifying the signal. In HEK293 cells which are known to internalize extracellular proteins through nonspecific uptake, we found that RNase treated V4-pRLB-540 (B2) delivered 19.6-fold more RNA than bare RNase treated V4 alone (B1) (**fig. 5.9a**). This suggests that the pRLB-540 domain dramatically enhanced RNA delivery. To test whether targeting domains enhanced RNA delivery, we switched to HeLa cells which naturally express the EGF- and Tf- receptors and do not have as much of a proclivity for nonspecific uptake as HEK293s as suggested by our localization data. In HeLa cells, we found that EGFR targeted V4-sc-pRO-flex (B4) delivered 5-fold more RNA than just EGFR targeted nucleocapsids (B3) pre-treated with RNase and 31.6-fold more RNA than bare untargeted nucleocapsids (B1) also treated with RNase (**fig. 5.9b**). Similarly, Transferrin-targeted V4-sc-533 (B9) and V4-pRLB-540 (B10) delivered at least 4-fold more RNA than just TfR targeted nucleocapsids (B7) and 12-fold more than untargeted nucleocapsids when not treated with RNase (**fig. 5.9c**). These data show that computationally designed pH-driven membrane-disrupting domains help to overcome the bottleneck of endosomal entrapment, which has been a decades-long hurdle for drug delivery [1-3]. We conjecture that our designed pH-responsive domains disrupt and permeabilize endosomes which prevents them from maturing into lysosomes due to their neutralized pH, thereby preventing the degradation of the nucleocapsids and their payload. It is likely that this endosome permeabilization then enables natural cellular mechanisms such as the ubiquitin-proteasome system (UBS) to cause the release of the RNA payload into the cytosol for complexing with prime editing CRISPR/Cas9.

This work demonstrates the first example of intracellular RNA delivery via a completely synthetic biological carrier made purely of computationally designed components and serves as a benchmark for all future designed protein assemblies. These RNA-delivery vehicles have great potential for development into a targeted therapeutics platform. There are many directions from which to build on these advances. One immediate direction is evolving highly potent RNA delivery as well as other desired functionality using a high-throughput library based screen with our mammalian reporter system, where nucleocapsid libraries are assembled in eukaryotic hosts such as yeast or mammalian cells. Other exciting directions include moving to mouse and nonhuman primate models for further evolution *in vivo*. These delivery vehicles make significant headway towards the development of next generation medicines, and offer an alternative to the top-down approach of VLPs through a modular bottom-up design approach that can be tailor-made to the specific needs of each application.

5.2 FIGURES

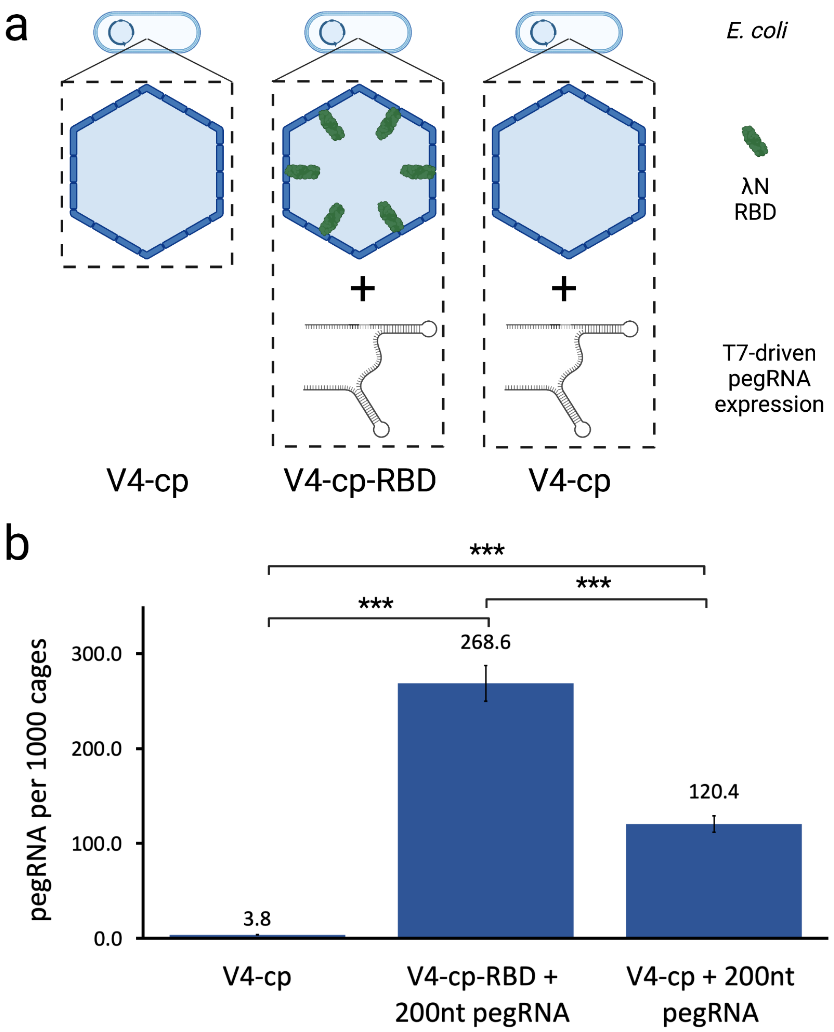
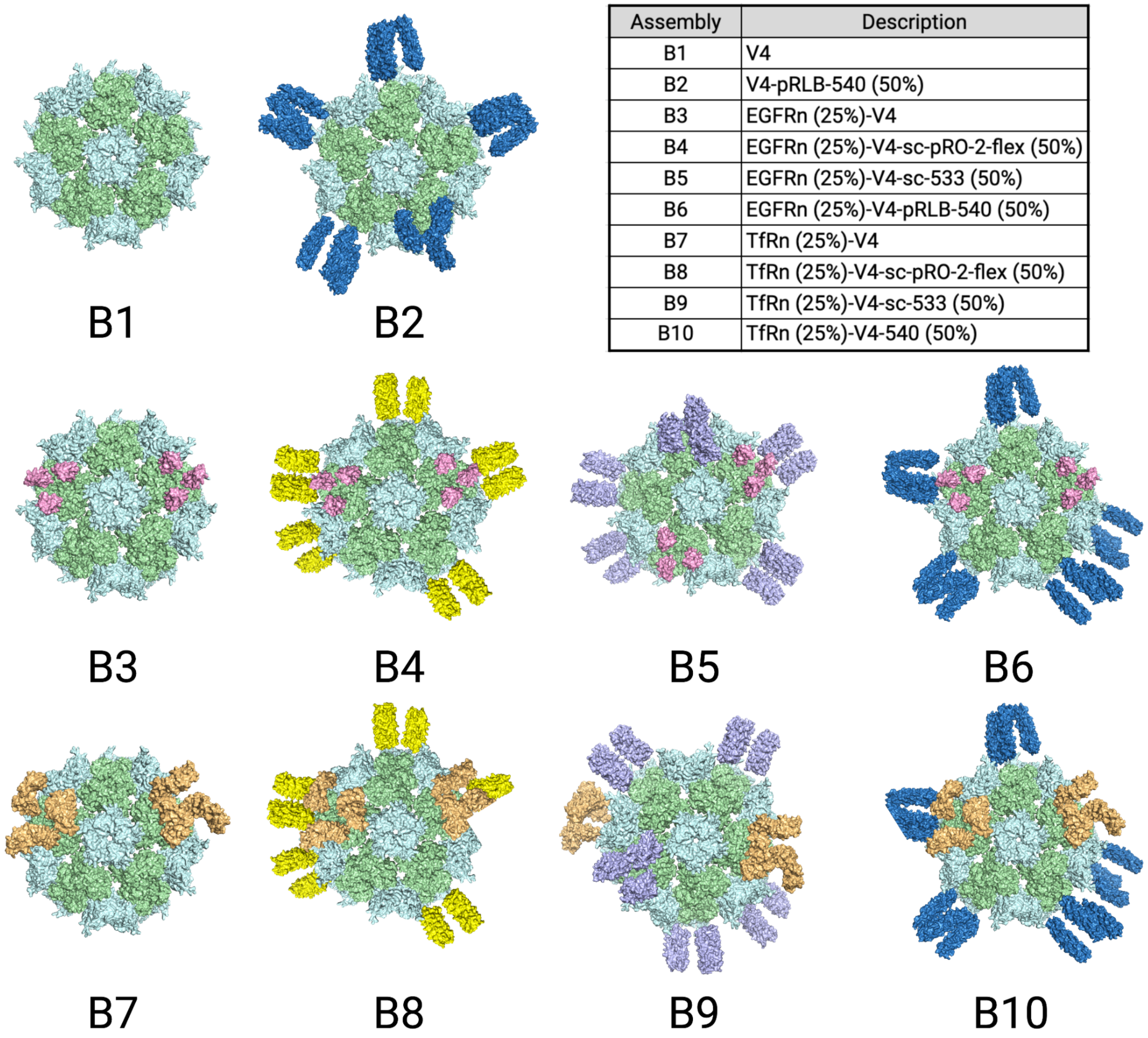


Figure 5.1. Engineering I53-50-V4 for specific and nonspecific packaging of pegRNA

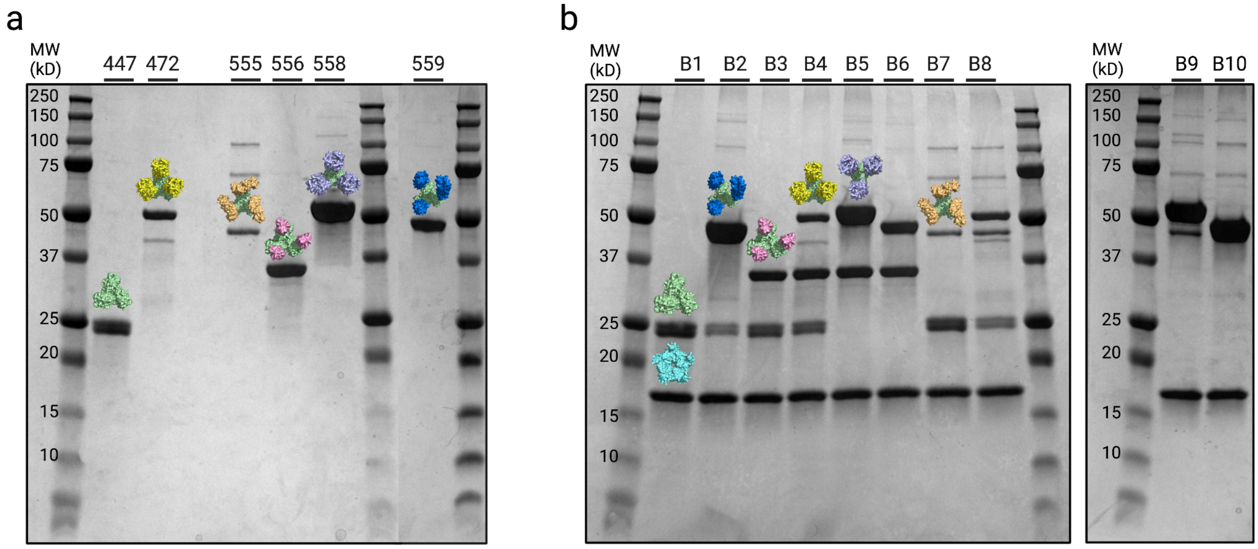
a. Depiction of three different nucleocapsid constructs used to compare specific and nonspecific packaging of pegRNA. **b.** Co-expression of nucleocapsid components with T7-driven pegRNA resulted in enriched pegRNA loading into the nucleocapsids. V4-cp-RBD with specific packaging domains performed about 2-fold better than nonspecific packaging.



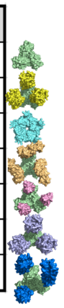
Assembly	Description
B1	V4
B2	V4-pRLB-540 (50%)
B3	EGFRn (25%)-V4
B4	EGFRn (25%)-V4-sc-pRO-2-flex (50%)
B5	EGFRn (25%)-V4-sc-533 (50%)
B6	EGFRn (25%)-V4-pRLB-540 (50%)
B7	TfRn (25%)-V4
B8	TfRn (25%)-V4-sc-pRO-2-flex (50%)
B9	TfRn (25%)-V4-sc-533 (50%)
B10	TfRn (25%)-V4-540 (50%)

Figure 5.2. Design models of targeted I53-50-V4 nucleocapsid assemblies B1-B10 packaging pegRNA for RNA delivery

a. Design models for assemblies B1-B10 displaying various targeting and pH-responsive membrane-permeabilizing domains. pRLB-540 is shown in dark blue. EGFRn minibinder is light pink. sc-pRO-2-flex is pale yellow, pRO-533 is pale violet, , and TfRn minibinder is pale orange.



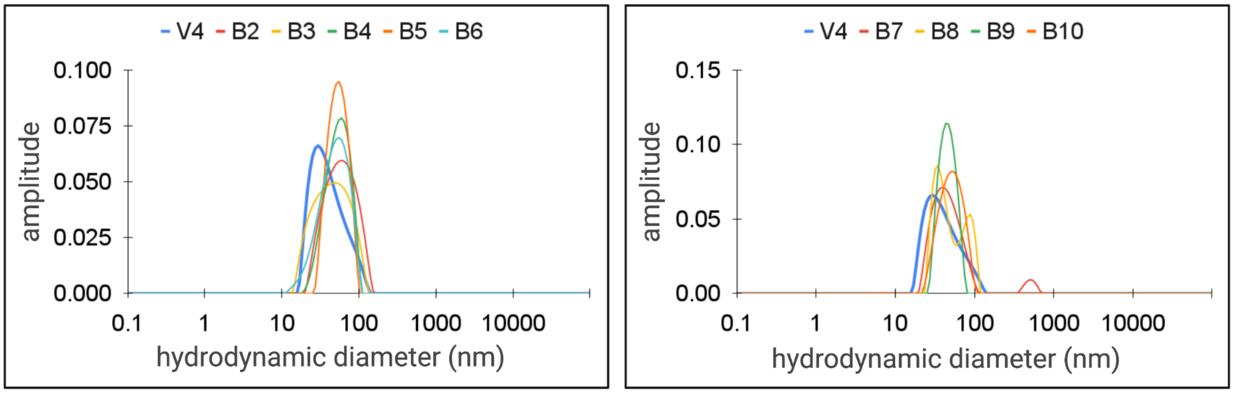
ID	Descripton	kD
447	Trimer	23.7
472	sc-pROI56V-flex-Trimer	52.9
473	Pentamer	18
555	TfRn-Trimer	50.1
556	EGFRn-Trimer	33.2
558	533-pIVgraft-sc-pROI56V-Trimer	53.5
559	540-Trimer	51.6



Assembly	Description
B1	V4
B2	V4-pRLB-540 (50%)
B3	EGFRn (25%)-V4
B4	EGFRn (25%)-V4-sc-pRO-2-flex (50%)
B5	EGFRn (25%)-V4-sc-533 (50%)
B6	EGFRn (25%)-V4-pRLB-540 (50%)
B7	TfRn (25%)-V4
B8	TfRn (25%)-V4-sc-pRO-2-flex (50%)
B9	TfRn (25%)-V4-sc-533 (50%)
B10	TfRn (25%)-V4-540 (50%)

Figure 5.3. SDS-Page of purified components of I53-50-V4 components displaying various targeting and membrane permeabilizing domains

a. Nucleocapsid components after purification by SEC. **b.** Nucleocapsid components after in vitro assembly.



Assembly	Sample	Z-Ave. Dia. (nm)	PDI	Pk 1 Mode Dia. (nm)
B1	V4	36.23	0.16	30.57
B2	V4-pRLB-540 (75%)	48.46	0.156	56.32
B3	EGFRn (25%) -V4	36.06	0.204	48.33
B4	EGFRn (25%)-V4-sc-pRO-2-flex (50%)	41.92	0.197	56.44
B5	EGFRn (25%)-V4-sc-533 (50%)	52.45	0.073	57.38
B6	EGFRn (25%)-V4-pRLB-540 (50%)	43.99	0.178	57.41
B7	TfRn (25%)-V4	39.89	0.174	38.28
B8	TfRn (25%)-V4-sc-pRO-2-flex (50%)	41.94	0.145	32.79
B9	TfR (25%)-V4-sc-533 (50%)	41.58	0.074	41.36
B10	TfR (25%)-V4-540 (50%)	48.63	0.136	53.05

Figure 5.4. Dynamic light scattering of targeted nucleocapsid assemblies packaging pegRNA

a. DLS of assemblies confirms monodisperse nucleocapsids with low PDI ≤0.2 at their expected sizes.

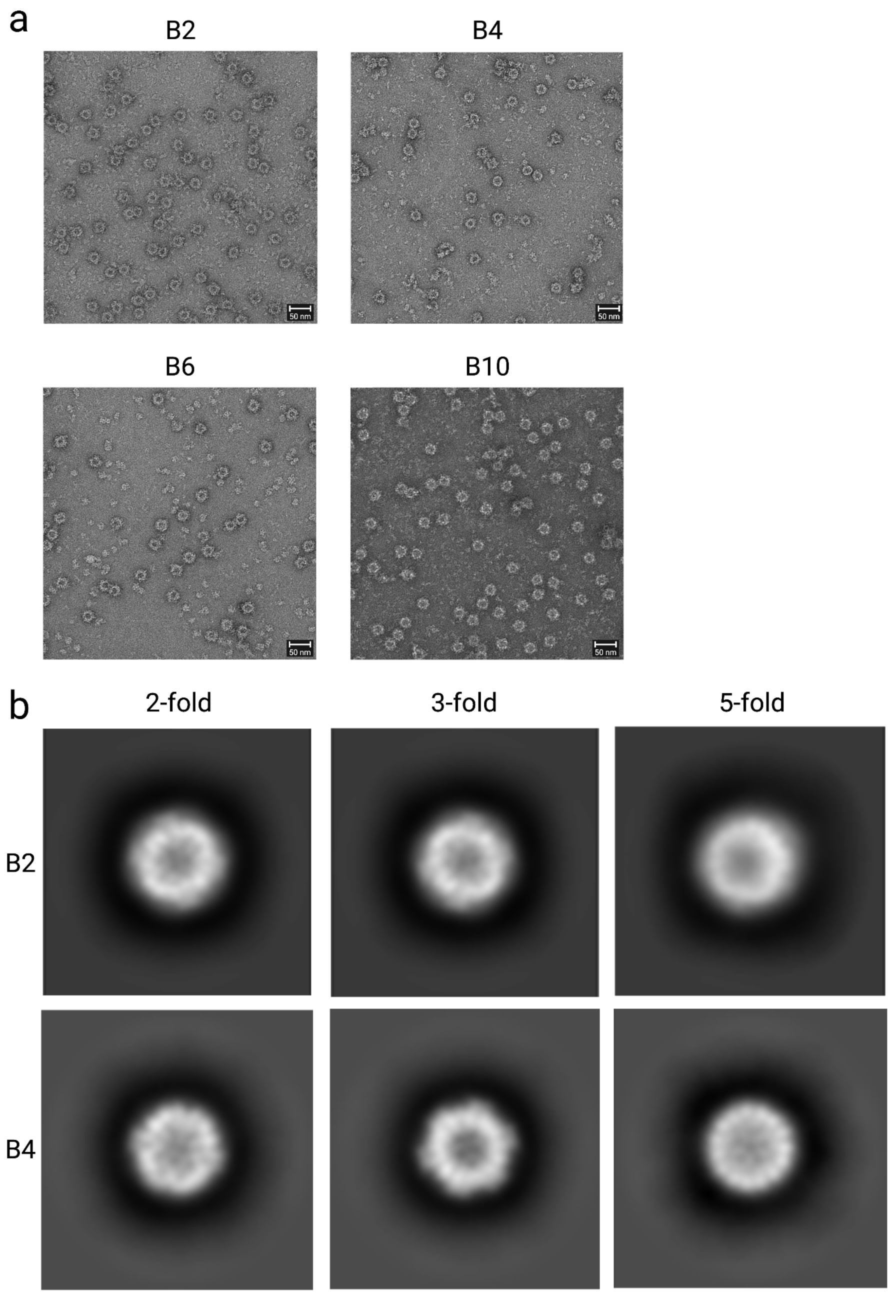
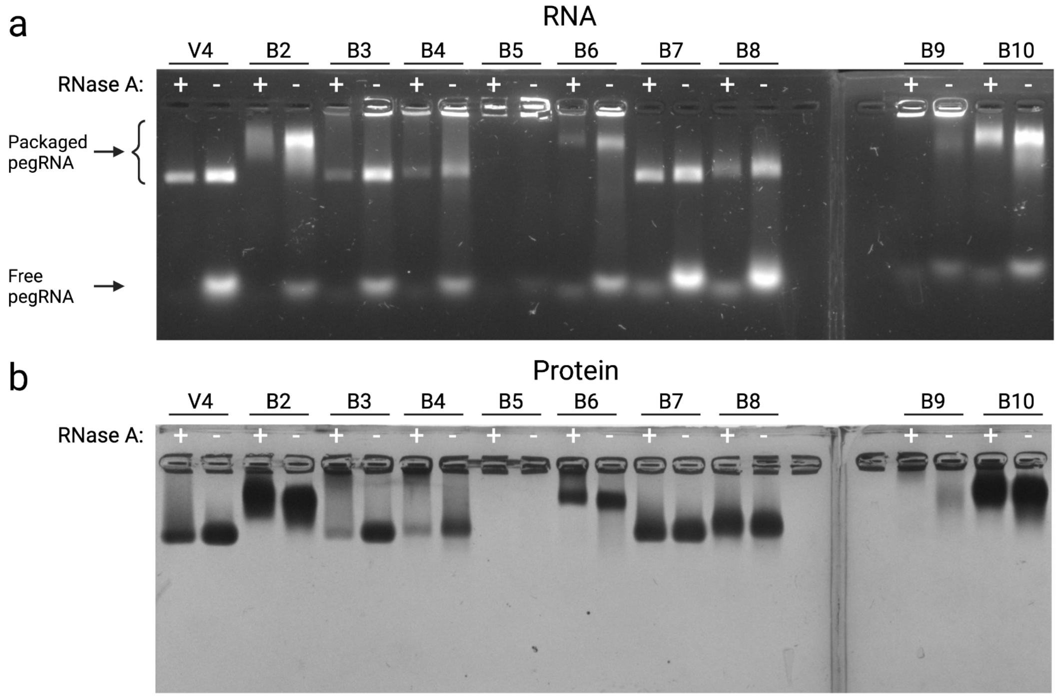


Figure 5.5. Negative-stain transmission electron microscopy and class averages

a. Negative-stain electron micrographs of assemblies B2, B4, B6, and B10 confirming intact nanoparticles. **b.** 2D class averages of B2 and B4 show a blurry corona, where flexibly linked domains are displayed.



c

Assembly	Description	Packaging Efficiency[1]	RNase Protection Efficiency[2]	Protected pegRNA/Cage[3]
B1	V4	48%	61%	1.4
B2	V4-pRLB-540 (50%)	81%	21%	0.8
B3	EGFRn (25%)-V4	63%	24%	0.8
B4	EGFRn (25%)-V4-sc-pRO-2-flex (50%)	60%	58%	1.7
B5	EGFRn (25%)-V4-sc-533 (50%)	-	-	-
B6	EGFRn (25%)-V4-pRLB-540 (50%)	58%	27%	0.8
B7	TfRn (25%)-V4	43%	85%	1.9
B8	TfRn (25%)-V4-sc-pRO-2-flex (50%)	39%	51%	1.0
B9	TfRn (25%)-V4-sc-533 (50%)	58%	7%	0.2
B10	TfRn (25%)-V4-540 (50%)	75%	83%	3.1

1. Area of RNase treated band divided by area of untreated band
2. Area of packaged RNA band divided by area of unpackaged band (untreated)
3. Calculated by multiplying: (5 pegRNA/cage molar ratio) * (protection efficiency) * (packaging efficiency)

Figure 5.6. pegRNA packaging and RNase protection efficiencies of nucleocapsid assemblies B1-B10

a, b. Native agarose gel confirming RNA packaging as visualized by the co-migration of protein and RNA. Assemblies protected against RNase A at different extents. **c.** ImageJ analysis of packaged RNA efficiency, and RNase A protection efficiency based on band area intensity.

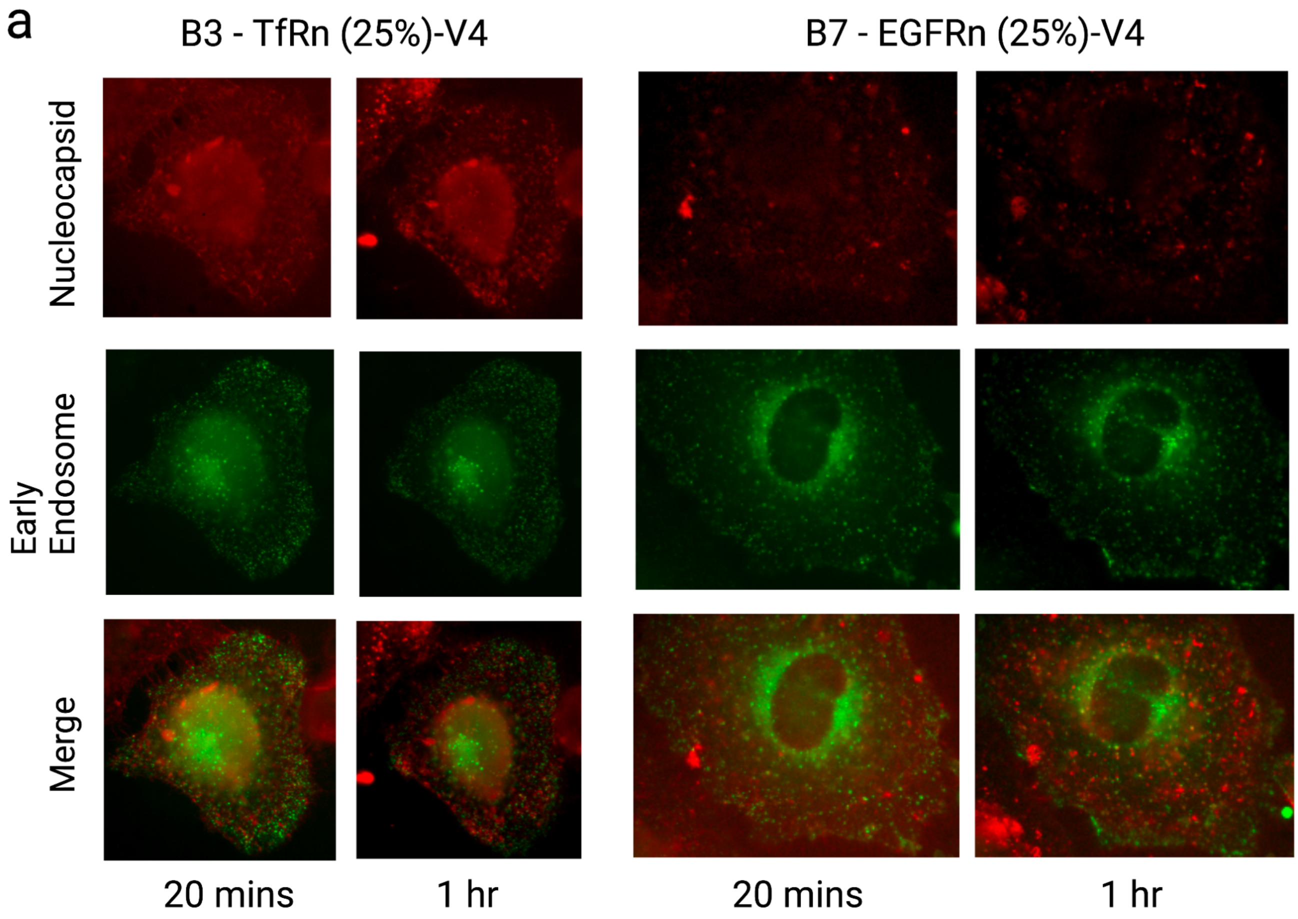
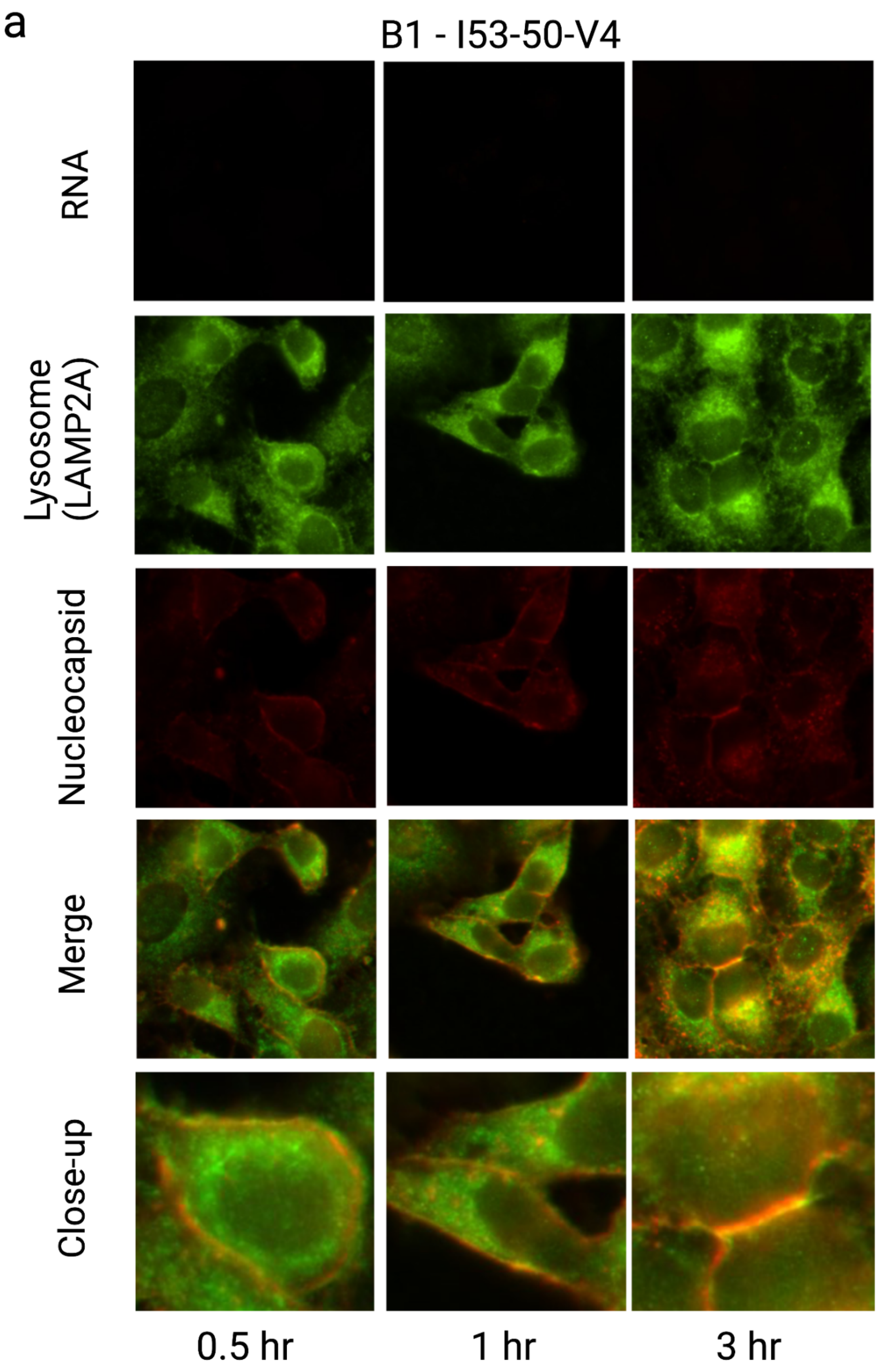
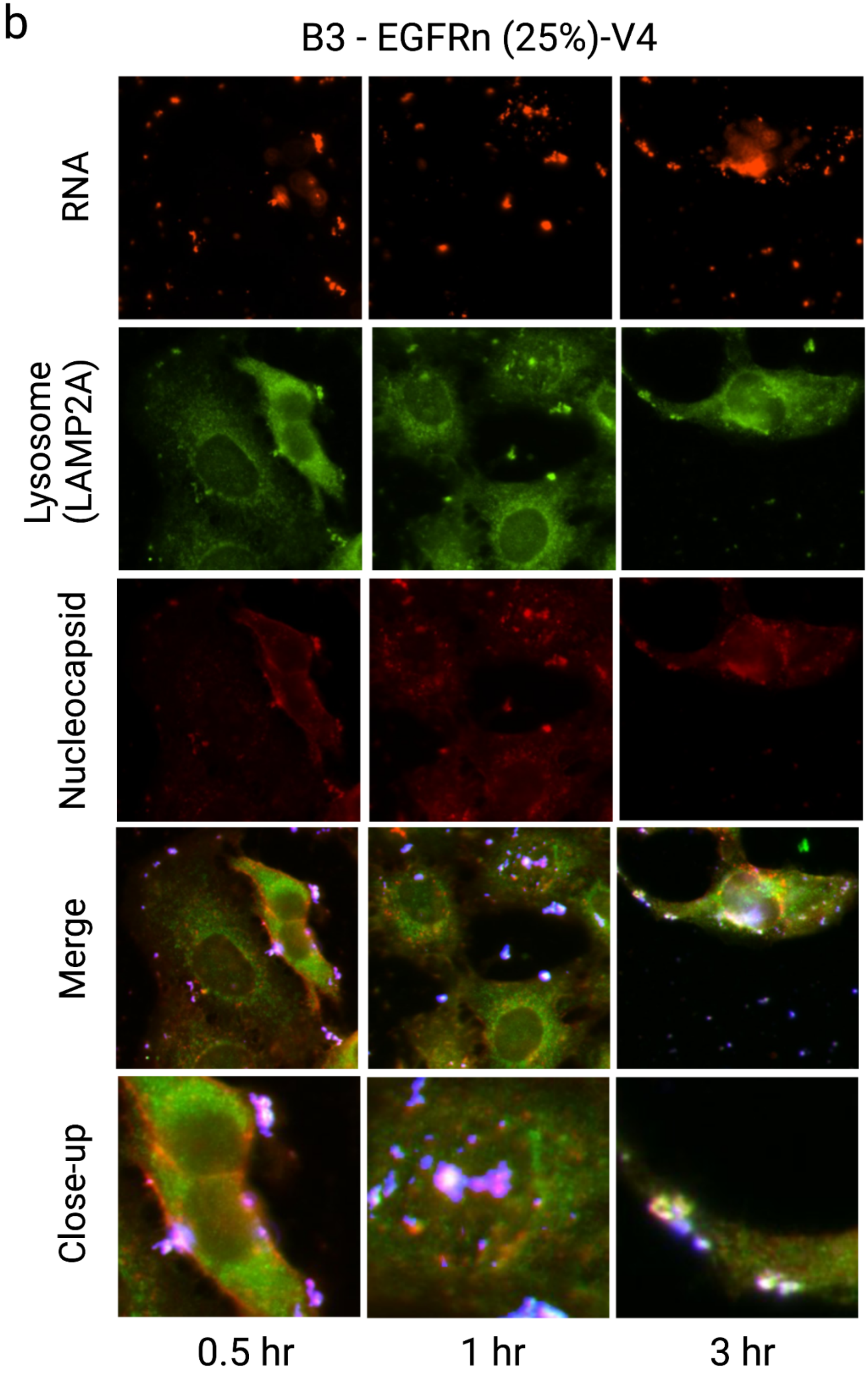


Figure 5.7. Live cell imaging of targeted nucleocapsids internalization over time

a. Live cell imaging of AF647-labeled TfR and EGFR targeted assemblies at 20 min and 1 hr time points. In both cases nucleocapsids can be seen attaching to the surface and being internalized. Nucleocapsids overlapped with early endosomes, suggesting tracking into the endocytic pathway.


a
B1 - I53-50-V4
RNA
Lysosome (LAMP2A)
Nucleocapsid
Merge
Close-up
0.5 hr
1 hr
3 hr


b
B3 - EGFRn (25%)-V4
RNA
Lysosome (LAMP2A)
Nucleocapsid
Merge
Close-up
0.5 hr
1 hr
3 hr

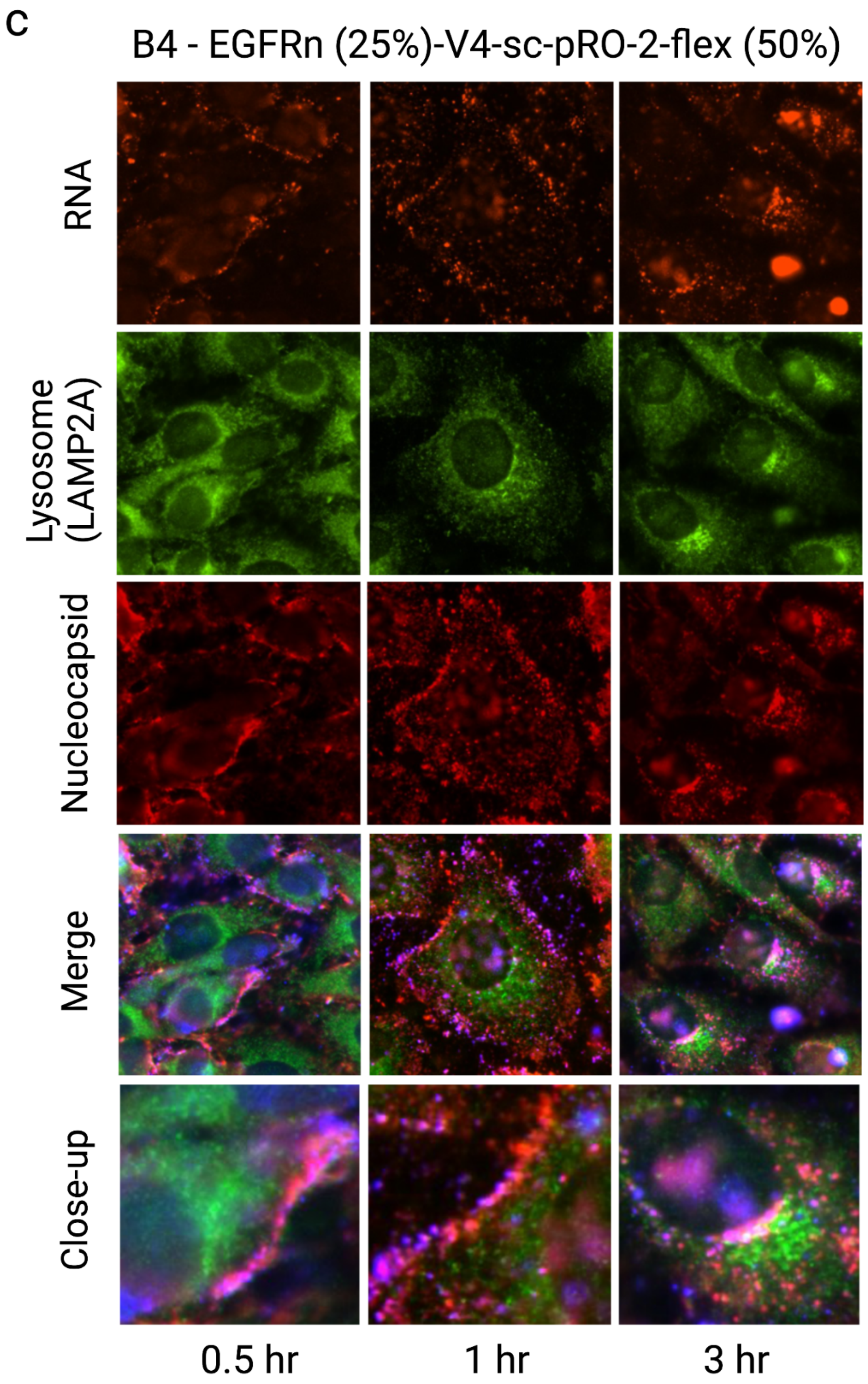
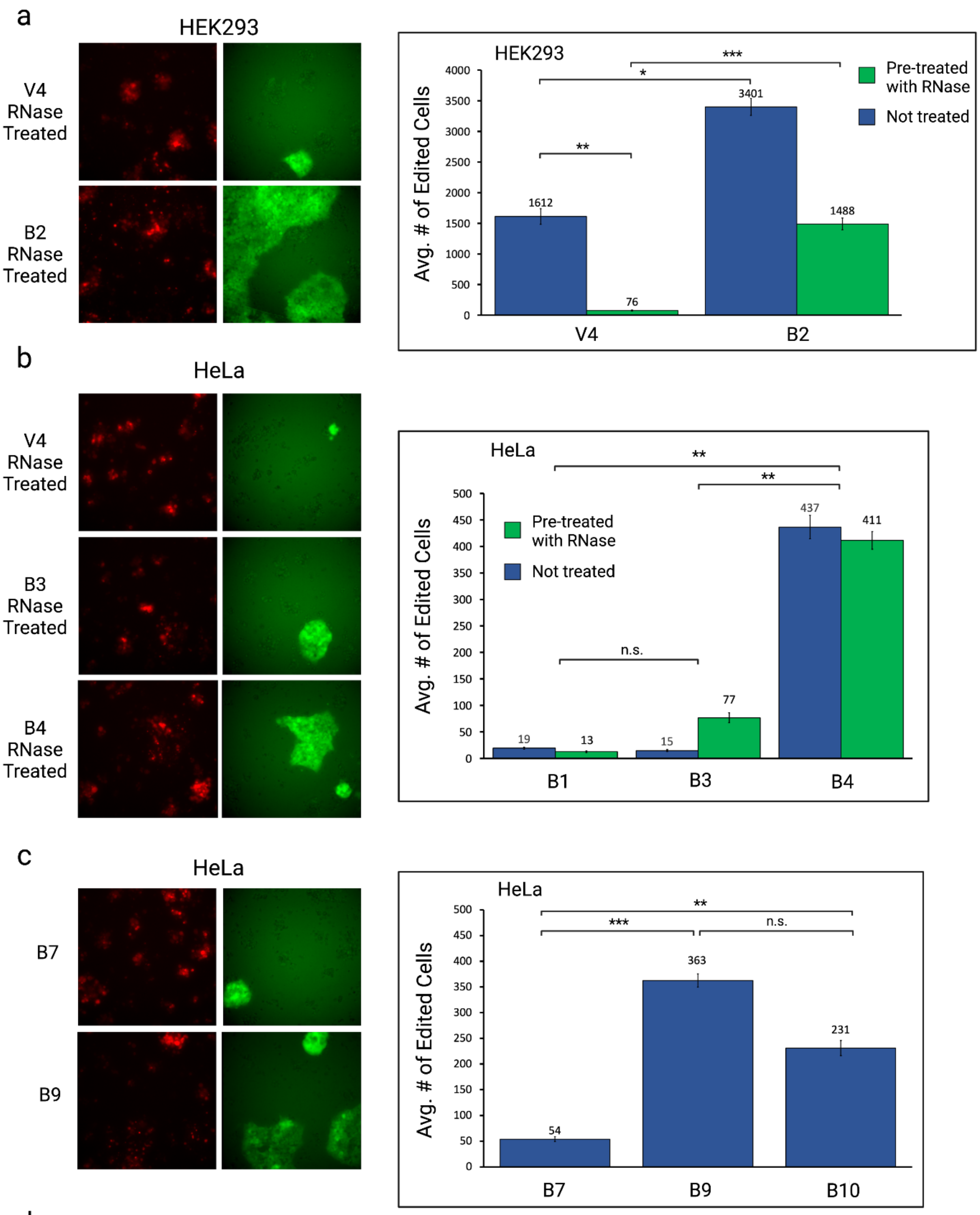

c
B4 - EGFRn (25%)-V4-sc-pRO-2-flex (50%)
RNA
Lysosome (LAMP2A)
Nucleocapsid
Merge
Close-up
0.5 hr
1 hr
3 hr

Figure 5.8. Localization of nucleocapsids and RNA over time

Assemblies labeled with AF657 dye and encapsulating ATTO550-labeled RNA were incubated with HeLa cells for 30 mins, 1 hr, and 3 hours before being fixed and imaged. **a.** I53-50-V4 labeled with AF647 was slow to internalize, with most nucleocapsids remaining on the cell surface at the 3 hour mark. No detectable RNA is seen. **b.** EGFR-targeted V4 is internalized into puncta by the 1 hour mark along with its packaged RNA, and both co-localize with lysosomes. **c.** EGFR-targeted V4 displaying sc-pRO-2-I56V-flex is robustly internalized with its packaged RNA, and does not colocalize with lysosomes.



d

Assembly	Description
B2	V4-pRLB-540 (50%)
B3	EGFRn (25%)-V4
B4	EGFRn (25%)-V4-sc-pRO-2-flex (50%)
B7	TfRn (25%)-V4
B9	TfRn (25%)-V4-sc-533 (50%)
B10	TfRn (25%)-V4-540 (50%)

Figure 5.9. Delivery of functional pegRNA via synthetic nucleocapsids

a. RNase A treated V4 displaying pRLB-540 delivered ~20-fold more functional pegRNA than V4 alone in HEK293 cells. **b.** RNase A EGFR-targeted V4 displaying sc-pRO-2-I56V-flex delivered ~5-fold more RNA than just EGFR-targeted V4, and 31-fold more than untargeted V4. **c.** TfR-targeted nucleocapsids displaying pRO-533 delivered 6.7-fold more RNA than just TfR-targeted V4.

5.3 METHODS

Purification of nucleocapsid components

Capsid components were purified in the following Tris buffer: 25 mM Tris, 1M NaCl, 0.75% CHAPS, with the addition of imidazole for IMAC (wash: 40 mM, elution: 300 mM). They were purified on a Superose 200 Increase 10/300 column, equilibrated with the described Tris buffer without imidazole. 5 mM TCEP was added as needed for constructs containing cysteines, specifically the trimers 447 and 472 in Figure 5.3.

Fluorophore conjugation

Versions of the trimer, 447 and 472 in Figure 5.3, had a single free cysteine for maleimide conjugation of AF647 C2 maleimide dye (ThermoFisher). Dye was incubated at a 10:1 molar ratio for 2 hours with shaking at 200 rpm, according to manufacturer’s instructions. Labeled protein was then separated from free dye via three rounds of dialysis at 25 °C with a 20K MWCO dialysis cassette (ThermoFisher).

In vitro assembly

Components were mixed at their desired ratios in the presence of pegRNA in the following order: trimer-components first, then RNA, then pentamer components. The assembly buffer was

50 mM Tris pH 8, 150 mM NaCl, 100 mM L-arginine, 5% glycerol. Assemblies were incubated at 25 °C for at least 30 minutes.

Live-cell imaging

Labeled nucleocapsids were incubated with HeLa cells on a 18-well glass Ibidi slide at a final concentration of 1 nM, and incubated for 15 minutes at 37 C with 5% CO2. The media was then replaced with fresh phenol-free DMEM containing 10% FBS and 1% penicillin-streptomycin. Cells were imaged every 3-5 minutes over several hours using the IN Cell Analyzer 2000.

Dynamic light scattering

Constructs were assessed by DLS ~1 hour after assembly. See Methods section of Chapter 2.

RNA delivery quantification using the reporter assay

HEK293 or HeLa cells were plated on a Corning 96-well plate pre-treated with Poly-L-Lysine (MPBio) at 15k cells/well. 24 hours after plating, cells were transfected with the PE2 and PE3b plasmids using Lipofectamine 3000 (ThermoFisher) according to manufacturer's instructions. 48 hours after plating, the media was replenished with 50% phenol-free DMEM containing 10% FBS and 1% penicillin-streptomycin and 50% OptiMEM. Assemblies were then added in duplicates to a final concentration of 10 nM, or 1 pmol total per well. Cells were incubated for 12-16 hours before the media was replaced with fresh phenol-free DMEM containing 10% FBS and 1% penicillin-streptomycin. 48-72 hours after addition of the assemblies, the the media was replaced with fresh media containing 5 ug/mL puromycin. Cells were monitored for mNeon expression using a Leica benchtop microscope as well as an IN Cell 2000 Analyzer. Cells were then imaged every 1-2 days once frameshifted cells became visible. 37 total images were collected per replicate or 74 per condition, at a 20X magnification using the IN Cell 2000 Analyzer.

Image analysis

Conducted as described in the Methods section of Chapter 4. Collected images were analyzed using a CellProfiler protocol normalized to the imaged controls of each specific experiment. An ANOVA test was performed on all conditions and t-test with equal variance was performed between conditions.

REFERENCES

1. Pei D, Buyanova M. Overcoming endosomal entrapment in drug delivery. Bioconjugate Chem. 2019;30(2):273-283.

2. Smith SA, Selby LI, Johnston APR, Such GK. The endosomal escape of nanoparticles: toward more efficient cellular delivery. Bioconjugate Chem. 2019;30(2):263-272.

3. Stewart MP, Sharei A, Ding X, Sahay G, Langer R, Jensen KF. In vitro and ex vivo strategies for intracellular delivery. Nature. 2016;538(7624):183-192.

4. Endosomal escape pathways for delivery of biologicals. Journal of Controlled Release. 2011;151(3):220-228.

5. Li C, Samulski RJ. Engineering adeno-associated virus vectors for gene therapy. Nat Rev Genet. 2020;21(4):255-272.

6. Banskota S, Raguram A, Suh S, et al. Engineered virus-like particles for efficient in vivo delivery of therapeutic proteins. Cell. 2022;185(2):250-265.e16.

7. Björklund T, Davidsson M. Next-generation gene therapy for parkinson’s disease using engineered viral vectors. J Parkinsons Dis. 11(Suppl 2):S209-S217.

8. Tang R, Xu Z. Gene therapy: a double-edged sword with great powers. Mol Cell Biochem. 2020;474(1):73-81.

9. Wang D, Tai PWL, Gao G. Adeno-associated virus vector as a platform for gene therapy delivery. Nat Rev Drug Discov. 2019;18(5):358-378.

10. Daussy CF, Wodrich H. “Repair me if you can”: membrane damage, response, and control from the viral perspective. Cells. 2020;9(9):2042.

11. Staring J, Raaben M, Brummelkamp TR. Viral escape from endosomes and host detection at a glance. Journal of Cell Science. 2018;131(15):jcs216259.

12. Dimitrov DS. Virus entry: molecular mechanisms and biomedical applications. Nat Rev Microbiol. 2004;2(2):109-122.

13. Nooraei S, Bahrulolum H, Hoseini ZS, et al. Virus-like particles: preparation, immunogenicity and their roles as nanovaccines and drug nanocarriers. J Nanobiotechnology. 2021;19:59.

14. Comas-Garcia M, Colunga-Saucedo M, Rosales-Mendoza S. The role of virus-like particles in medical biotechnology. Mol Pharmaceutics. 2020;17(12):4407-4420.

15. Le DT, Müller KM. In vitro assembly of virus-like particles and their applications. Life (Basel). 2021;11(4):334.

16. Hill BD, Zak A, Khera E, Wen F. Engineering virus-like particles for antigen and drug delivery. CPPS. 2017;19(1).

17. Bale JB, Gonen S, Liu Y, et al. Accurate design of megadalton-scale two-component icosahedral protein complexes. Science. 2016;353(6297):389-394.

18. Butterfield GL, Lajoie MJ, Gustafson HH, et al. Evolution of a designed protein assembly encapsulating its own RNA genome. Nature. 2017;552(7685):415-420.

19. Boyken SE, Benhaim MA, Busch F, et al. De novo design of tunable, pH-driven conformational changes. Science. 2019;364(6441):658-664.

20. Jimah JR, Schlesinger PH, Tolia NH. Liposome disruption assay to examine lytic properties of biomolecules. Bio Protoc. 2017;7(15):e2433.

21. Boyken SE, Chen Z, Groves B, et al. De novo design of protein homo-oligomers with modular hydrogen bond network-mediated specificity. Science. 2016;352(6286):680-687.

22. Kuhlman B, Baker D. Native protein sequences are close to optimal for their structures. Proc Natl Acad Sci U S A. 2000;97(19):10383-10388.

23. Leaver-Fay A, Tyka M, Lewis SM, et al. Rosetta3: an object-oriented software suite for the simulation and design of macromolecules. Methods Enzymol. 2011;487:545-574.

24. Huang PS, Oberdorfer G, Xu C, et al. High thermodynamic stability of parametrically designed helical bundles. Science. 2014;346(6208):481-485.

25. Wiethoff CM, Wodrich H, Gerace L, Nemerow GR. Adenovirus protein vi mediates membrane disruption following capsid disassembly. J Virol. 2005;79(4):1992-2000.

26. Martinez R, Schellenberger P, Vasishtan D, et al. The amphipathic helix of adenovirus capsid protein vi contributes to penton release and postentry sorting. J Virol. 2014;89(4):2121-2135.

27. Maier O, Galan DL, Wodrich H, Wiethoff CM. An n-terminal domain of adenovirus protein vi fragments membranes by inducing positive membrane curvature. Virology. 2010;402(1):11-19.

28. Lönn P, Kacsinta AD, Cui XS, et al. Enhancing endosomal escape for intracellular delivery of macromolecular biologic therapeutics. Sci Rep. 2016;6(1):32301.

29. Verdurmen WPR, Mazlami M, Plückthun A. A quantitative comparison of cytosolic delivery via different protein uptake systems. Sci Rep. 2017;7(1):13194.

30. Kilchrist KV, Dimobi SC, Jackson MA, et al. Gal8 visualization of endosome disruption predicts carrier-mediated biologic drug intracellular bioavailability. ACS Nano. 2019;13(2):1136-1152.

31. Smith SA, Selby LI, Johnston APR, Such GK. The endosomal escape of nanoparticles: toward more efficient cellular delivery. Bioconjugate Chem. 2019;30(2):263-272.

32. Anzalone AV, Randolph PB, Davis JR, et al. Search-and-replace genome editing without double-strand breaks or donor DNA. Nature. 2019;576(7785):149-157.

33. Baron-Benhamou J, Gehring NH, Kulozik AE, Hentze MW. Using the λn peptide to tether proteins to rnas. In: Schoenberg DR, ed. MRNA Processing and Metabolism: Methods and Protocols. Methods in Molecular BiologyTM. Humana Press; 2004:135-153.

34. Cocozaki AI, Ghattas IR, Smith CA. The rna-binding domain of bacteriophage p22 n protein is highly mutable, and a single mutation relaxes specificity toward λ. J Bacteriol. 2008;190(23):7699-7708.

35. Sahtoe DD, Coscia A, Mustafaoglu N, et al. Transferrin receptor targeting by de novo sheet extension. Proc Natl Acad Sci U S A. 2021;118(17):e2021569118.

36. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-589.

37. Stirling DR, Swain-Bowden MJ, Lucas AM, Carpenter AE, Cimini BA, Goodman A (2021). CellProfiler 4: improvements in speed, utility and usability. BMC Bioinformatics, 22 (1), 433.

38. Crick FHC. The Fourier transform of a coiled-coil. Acta Cryst. 1953;6(8):685-689.

39. Grigoryan G, DeGrado WF. Probing designability via a generalized model of helical bundle geometry. J Mol Biol. 2011;405(4):1079-1100.

40. Sanders CL, Curran JF. Genetic analysis of the E site during RF2 programmed frameshifting. RNA. 2007;13(9):1483-1491.

41. Curran JF. Analysis of effects of tRNA:message stability on frameshift frequency at the Escherichia coli RF2 programmed frameshift site. Nucleic Acids Res. 1993;21(8):1837-1843.