



In vivo selection of synthetic nucleocapsids for tissue targeting

Audrey Olshefsky^{a,b}, Halli Benasutti^c, Meilyn Sylvestre^{a,1}, Gabriel L. Butterfield^{b,d,2}, Gabriel J. Rocklin^{b,3}, Christian Richardson^{a,b}, Derrick R. Hicks^b, Marc J. Lajoie^{b,4}, Kefan Song^a, Elizabeth Leaf^b, Catherine Treichel^b, Justin Decarreau^b, Sharon Ke^{b,5}, Gargi Kher^{b,6}, Lauren Carter^b, Jeffrey S. Chamberlain^{c,e}, David Baker^{b,c}, Neil P. King^{b,c,7}, and Suzie H. Pun^{a,f,7,8}

Edited by David Tirrell, California Institute of Technology, Pasadena, CA; received April 15, 2023; accepted September 21, 2023

Controlling the biodistribution of protein- and nanoparticle-based therapeutic formulations remains challenging. In vivo library selection is an effective method for identifying constructs that exhibit desired distribution behavior; library variants can be selected based on their ability to localize to the tissue or compartment of interest despite complex physiological challenges. Here, we describe further development of an in vivo library selection platform based on self-assembling protein nanoparticles encapsulating their own mRNA genomes (synthetic nucleocapsids or synNCs). We tested two distinct libraries: a low-diversity library composed of synNC surface mutations (45 variants) and a high-diversity library composed of synNCs displaying miniproteins with binder-like properties (6.2 million variants). While we did not identify any variants from the low-diversity surface library that yielded therapeutically relevant changes in biodistribution, the high-diversity miniprotein display library yielded variants that shifted accumulation toward lungs or muscles in just two rounds of in vivo selection. Our approach should contribute to achieving specific tissue homing patterns and identifying targeting ligands for diseases of interest.

library selection | biopanning | synthetic nucleocapsid | miniprotein | cancer

The emerging drug class of biologics with intracellular targets (e.g., mRNA, siRNA, ribonucleoproteins) is currently limited in clinical translation by a lack of effective systemic delivery systems (1, 2). These biologics often require packaging into delivery systems such as viral vectors or lipid nanoparticles (LNPs) that improve drug stability and mediate cellular entry. However, the default biodistribution of most nanoparticles when delivered systemically skews toward uptake by the mononuclear phagocytic system (MPS) and accumulation in the liver and spleen (3). In vivo library selection is a powerful method to identify delivery vehicles possessing both physicochemical characteristics and molecular ligands that enable organ targeting. For example, in vivo library selection has been successfully employed to alter the vascular targeting profile of lipid nanoparticles (LNPs) by identifying specific surface chemistries that enable therapeutically relevant mRNA delivery to bone marrow, lungs, spleen, liver, and tumor (4–8). In vivo library selection is also amenable to the identification of targeting ligand moieties, wherein millions of potential ligands are simultaneously screened for desired biodistribution behaviors (9). Unlike in vitro library selection, in vivo library selection accounts for both known and unknown physiological complexity and can utilize established animal models that recapitulate disease physiology. Although several in vivo library selection platforms exist, every platform has limitations. Current state-of-the-art library selection platforms are limited by large particle sizes that prevent tissue penetration (e.g., M13 phage display), low stability (e.g., ribosome display and one-bead one-compound display), incompatibility with in vivo selection (one-bead one-compound display), and sensitivity to surface display modifications (e.g., adeno-associated virus display) (9–11). Additional library selection platforms that better represent nanoparticle delivery systems and are amenable to in vivo selection will further enable the discovery of novel targeting ligands for specific cell and tissue types.

In prior work, our group engineered a virus-sized, self-assembling protein nanoparticle that encapsulates its own mRNA genome, I53-50-v4 (12). I53-50-v4 was derived from I53-50, a computationally designed, two-component self-assembling protein nanoparticle with icosahedral symmetry (13). This mRNA-encapsulating version of I53-50, termed a “synthetic nucleocapsid” (synNC), links genotype and phenotype to enable the evolution of new and improved functional properties through diversification and selection. After four stages of evolution, synNC stability in blood was increased by almost 20-fold, and in vivo circulation half-life was increased from less than 5 min to about 4.5 h. We named the resulting synNC “I53-50-v4” or simply “v4,” for the fourth evolved version of I53-50. However, we did not observe specific or preferential association of I53-50-v4 with any major organs. Samples collected either 5 min or 4 h after I53-50-v4 administration showed

Significance

Targeted delivery—getting therapeutics to specific cells or tissues while retaining therapeutic activity—remains one of the greatest challenges and opportunities in drug development. The in vivo library selection platform reported here offers a unique tool to optimize protein nanoparticle and protein minibinder therapeutic performance in a living mammal, a complex physiological environment that cannot yet be comprehensively modeled in silico or in vitro.

Author contributions: A.O., H.B., M.S., G.L.B., G.J.R., C.R., D.R.H., M.J.L., E.L., J.S.C., D.B., N.P.K., and S.H.P. designed research; A.O., H.B., M.S., G.L.B., G.J.R., D.R.H., K.S., E.L., C.T., J.D., S.K., G.K., and L.C. performed research; A.O., H.B., M.S., G.L.B., G.J.R., C.R., and G.K. analyzed data; and A.O., N.P.K., and S.H.P. wrote the paper.

Competing interest statement: N.P.K. is a cofounder, shareholder, paid consultant, and chair of the scientific advisory board of Icosavax, Inc.

This article is a PNAS Direct Submission.

Copyright © 2023 the Author(s). Published by PNAS. This article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

¹Present address: Lynx Biosciences, San Diego, CA 92121.

²Present address: Duke University, Department of Biomedical Engineering, Durham, NC 27710.

³Present address: Northwestern University, Department of Pharmacology, Evanston, IL 60208.

⁴Present address: Outpace Bio, Seattle, WA 98109.

⁵Present address: Rocky Vista University, College of Osteopathic Medicine, Lovins, UT 84738.

⁶Present address: Fred Hutchinson Cancer Center, Seattle, WA 98109.

⁷To whom correspondence may be addressed. Email: neilking@uw.edu or spun@uw.edu.

⁸Present address: Lila Biologics, Seattle, WA 98102.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2306129120/-/DCSupplemental>.

Published November 8, 2023.

that the majority of I53-50-v4 mRNA remained in the blood, while we detected an order of magnitude less I53-50-v4 mRNA in the kidneys, lungs, heart, brain, spleen, or liver compared to the blood. As we were able to significantly enhance circulation time using this selection strategy, we hypothesized that a modified selection strategy could be employed to enrich library variants that preferentially accumulate in specific tissues and yield novel targeting moieties. We sought to address two main questions: First, can a similar library with which circulation half-life was evolved (12) be used to enrich tissue targeting behavior of the synthetic nucleocapsid itself? Second, can the synthetic nucleocapsids be used in a similar manner as phage display, in which binding domains are displayed on the surface of the particles and tested for accumulation in specific tissues? Our approach is similar to the state-of-the-art M13 phage display platform (14). However, since synNCs are an order of magnitude smaller than M13 phage, they might be better able to access different physiological compartments such as tumor microenvironments. Additionally, computationally designed protein nanoparticles can be engineered modularly to display or encapsulate a wide variety of cargoes (13, 15, 16); and can be manufactured at scale (17), so any promising library variant could be directly formulated for applications in therapeutic delivery.

Here, we utilize murine models to evolve two separate synNC libraries toward altered tissue biodistribution. We tested two different libraries in BALB/c mice: a surface library in which point mutations are made on the surface of the synNCs (12) and a miniprotein display library in which the synNCs display a semirandom library of designed miniproteins on their surface (18, 19). While the low-diversity library composed of synNC surface modifications did not yield variants that significantly shifted biodistribution, the high-diversity miniprotein display library successfully shifted synNC biodistribution patterns toward selected tissues including lung and muscle.

Results

Synthetic nucleocapsid libraries can be constructed in a variety of ways. Here, we constructed a surface library with mutated surface residues (Fig. 1 *A, Top*, yellow point mutations), and a miniprotein display library with designed miniproteins displayed on the surface of the nucleocapsids via genetic fusion (Fig. 1 *A, Bottom*, orange miniproteins). Following the delivery of library plasmids into *Escherichia coli* (*E. coli*), the transgene is transcribed and translated

into synthetic nucleocapsid (synNC) protein subunits, which then associate with intracellular mRNA through electrostatic interactions. Since transgene expression is driven by the strong T7 promoter, there is an abundance of self-mRNA inside the *E. coli* cells and synNCs encapsulate their own mRNA genomes (12). For both the surface library and miniprotein library, we injected the libraries into normal mice or 4T1 tumor-bearing mice, allowed the libraries to circulate for the desired amount of time (e.g., 30 min), and extracted RNA from the desired tissues (Fig. 1 *B*). After specifically reverse-transcribing and amplifying synNC mRNA by qRT-PCR, we sequenced the samples and cloned a separate library for the next round of selection. We visualized enrichments of specific sequences in the collected organs using hierarchically clustered heatmaps, from which we identified variants of interest (Fig. 1 *C*).

In Vivo Library Selection of I53-50-v4 Surface Mutant Library.

In an attempt to modulate the biodistribution of I53-50-v4, we first used a similar surface library that was used to evolve longer circulation half-life in our previous work (Figs. 1 and 2*A*) (12). Here, we combinatorially mutated charged and polar surface residues (aspartic acid, arginine, glutamic acid, lysine, and glutamine) to other charged and polar amino acids as allowed by codon degeneracies (*SI Appendix, Table S1* and Fig. 2*B* middle structure). Instead of selecting for variants with longer circulation half-lives in blood, we enriched synNC RNA recovered from major organs (brain, heart, kidneys, liver, lungs, and spleen) over two rounds of selection. We primarily detected synNC library mRNA in the blood, liver, and spleen (*SI Appendix, Fig. S1* and Fig. 2*C*). After sequencing the synNC mRNA from each collected tissue, we calculated $\log_{10}(\text{enrichment})$ scores for each sequence compared to v4 (Eq. 1 in *Materials and Methods*, Fig. 2*D*). To ensure sufficient sequencing coverage, we analyzed each mutation individually. Of 45 possible single-mutation variants, we selected two variants representative of different biodistribution patterns. The selected variants were representative of two different biodistribution patterns. First, v4_D100K, a negative-to-positive mutation on the pentamer subunit, exhibited positive $\log_{10}(\text{enrichment})$ between 0.2 and 0.7 in major organs and negative enrichment in blood and brain [Fig. 2*B* (left structure), Fig. 2*D* (left heatmap)]. Second, v4_E243D, a negative-to-negative mutation on the trimer subunit, exhibited positive $\log_{10}(\text{enrichment})$ in blood and all tissues collected, albeit low in magnitude (between 0.1 and 0.2) [Fig. 2*B* (right structure), Fig. 2*D* (right heatmap)].

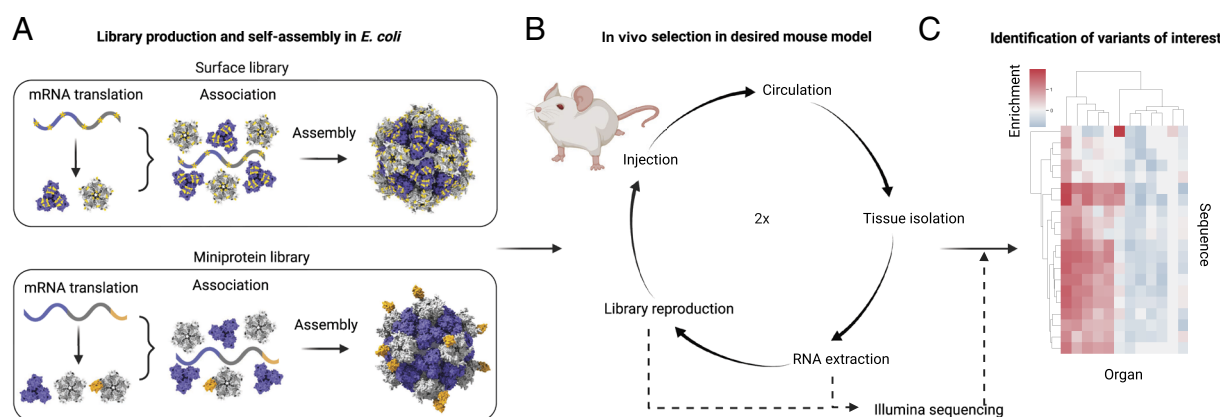


Fig. 1. Experimental process for in vivo library selection based on synthetic nucleocapsids. (*A*) Library production and self-assembly inside *E. coli*. *Top*: surface library with yellow point mutations. *Bottom*: Miniprotein library with orange miniproteins. (*B*) In vivo selection scheme. (*C*) Example heatmap format used to identify variants of interest.

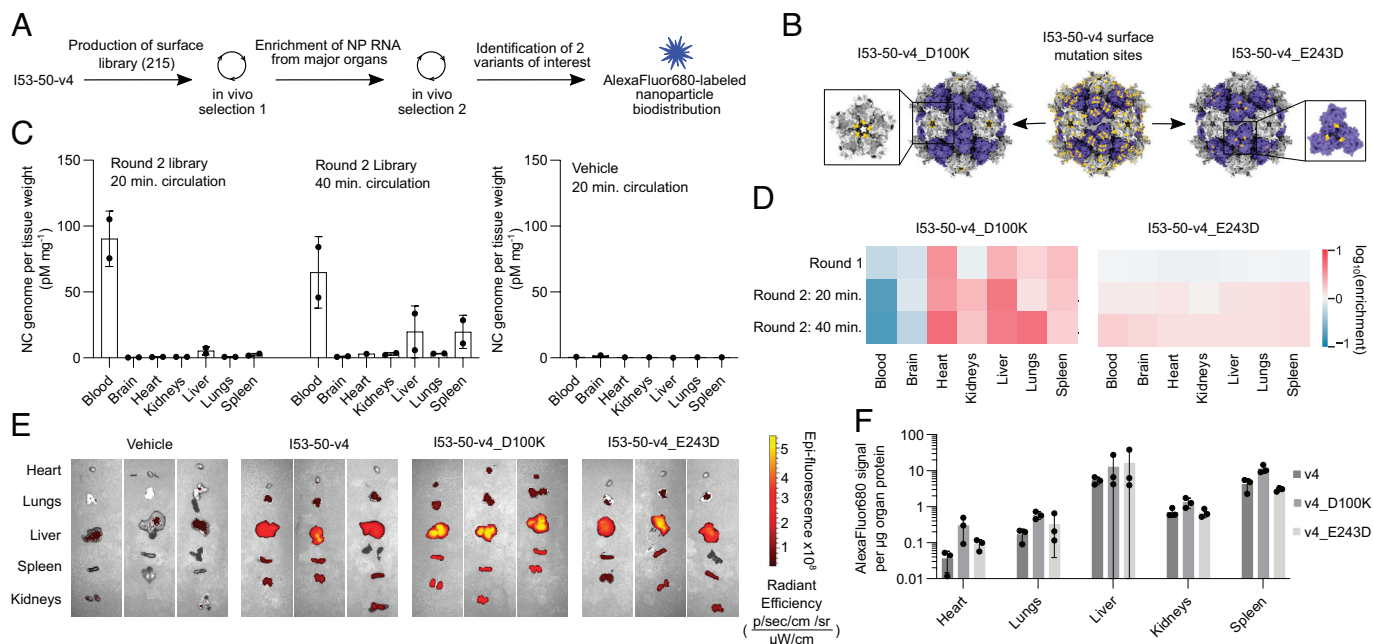


Fig. 2. In vivo library selection and biodistribution of I53-50-v4 surface library. (A) Experimental trajectory of surface library selection and fluorescence biodistribution studies. SynNC library mRNA was obtained from organs (brain, heart, kidneys, liver, lungs, and spleen) and amplified for in vivo selection round 2. SynNC mRNA-seq was used to identify two variants of interest (v4_D100K and v4_E243D) that were produced as individual synNCs and fluorescently labeled for biodistribution studies. (B) Model of locations of variants of interest (*Left* and *Right*) selected from the I53-50-v4 surface library (*Middle*). The D100K mutation resides near the small central pore of the pentamer and the E243D mutation resides near the threefold axis of the trimer. (C) Bulk biodistribution of the i.v.-injected surface library from in vivo selection round 2 after 20 or 40 min of circulation (*Left*) and the vehicle control (*Right*), measured by RT-qPCR. The majority of library mRNA was detected in the blood at 20 min, and in the blood, liver, and spleen at 40 min. (D) Enrichment of the two variants of interest compared to v4 (Eq. 1). v4_D100K shows positive enrichment in all tissues except blood and brain. v4_E243D shows slight positive enrichment in all tissues. (E) SynNC-Alexa Fluor 680 signal in the whole organs of mice injected with library variants of interest or vehicle control shows statistically insignificant increase in organ accumulation by v4_D100K compared to v4 and v4_E243D and similar tissue accumulation of v4 and v4_E243D. Radiant efficiency color scale: min = 2.8×10^7 , max = 5.5×10^9 . (F) BCA-fluorescence assay on homogenized organs from (E) to measure the AF680 fluorescence per μg organ protein in major organs of interest. C: left, $n = 2$ mice per group; right, $n = 1$ mouse; error bars depict SEM. F: $n = 3$ mice per group; error bars depict SDs. No statistically significant differences were observed in (F) by one-way ANOVA in each different organ.

We next evaluated the biodistribution in normal mice of fluorescently labeled v4_D100K and v4_E243D as an orthogonal method to mRNA sequencing (Fig. 2E and F). Instead of administering self-mRNA-encapsulating synNCs to mice, we administered synNCs in which the exterior residues were conserved but the interior residues were mutated to both ablate nucleic acid encapsulation and introduce a unique cysteine handle for conjugation to Alexa Fluor 680 C2-maleimide (AF680) (SI Appendix, Fig. S2). These modified synNCs are herein referred to as protein nanoparticles, as they do not encapsulate their mRNA genomes, and denoted “v4v0-Cys-AF680,” which stands for v4 exterior (version 4: surface evolved for increased circulation half-life), v0 interior (version 0: non-mRNA-encapsulating), and a unique internal cysteine (Cys) on the trimer component to which Alexa Fluor 680-C2-Maleimide was conjugated (AF680). To compare the biodistribution of synNCs, we injected 1 nanomole (nmole) of asymmetric unit of each sample into mice intravenously (the asymmetric unit comprises one subunit from the pentamer and one subunit from the trimer, with 60 asymmetric units per nanoparticle, organized into 12 pentamers and 20 trimers), including a nontargeted v4v0-Cys-AF680 control and a vehicle control, perfused with PBS to reduce background signal, imaged the whole organs for qualitative analysis, and quantified AF680 fluorescence per microgram (μg) protein via BCA-fluorescence assay in each homogenized organ sample (Fig. 2E and F). We observed similar overall biodistribution patterns in the original v4v0-Cys-AF680 synNC and v4v0-Cys-AF680_E243D and increased but nonspecific tissue accumulation by v4v0-Cys-AF680_D100K. We did not observe any statistically significant differences in organ

accumulation but note the congruence between enrichment values (Fig. 2D) and fluorescence signal per μg protein (Fig. 2F). While we did not identify any variants that shifted biodistribution in a therapeutically relevant manner, this experiment validated our platform for organ enrichment and laid the groundwork for our miniprotein display library.

In Vivo Selection of an I53-50-v4 Miniprotein Display Library.

After observing only minor enrichment of the I53-50-v4 surface library variants in major organs, we hypothesized that displaying miniproteins on the surface of the synNC could better enable specific organ accumulation. The Baker group had previously designed millions of miniproteins intended to fold into stable structures and bind specific target proteins. For our purposes, we viewed these miniproteins as simply small, stable domains with surface patches that have physicochemical properties typical of protein-protein interfaces (20, 21). The intended structures of these designed miniproteins are either trihelical bundles or a mix of alpha helices and beta sheets, composed of 30 to 50 amino acids—small enough to be synthesized by high-throughput oligonucleotide microarray technologies. To make a large miniprotein display library based on synNCs, we genetically fused a random subset of designed miniprotein libraries to either the N terminus of the v4 trimer subunit at 100% display valency or the C terminus of the v4 pentamer subunit at about 20 to 40% valency by using a programmed ribosomal frameshifting B sequence (prfB), herein denoted v4-trimer display library or v4-pentamer display library, respectively (Fig. 3A–C). The prfB sequence shifts the open

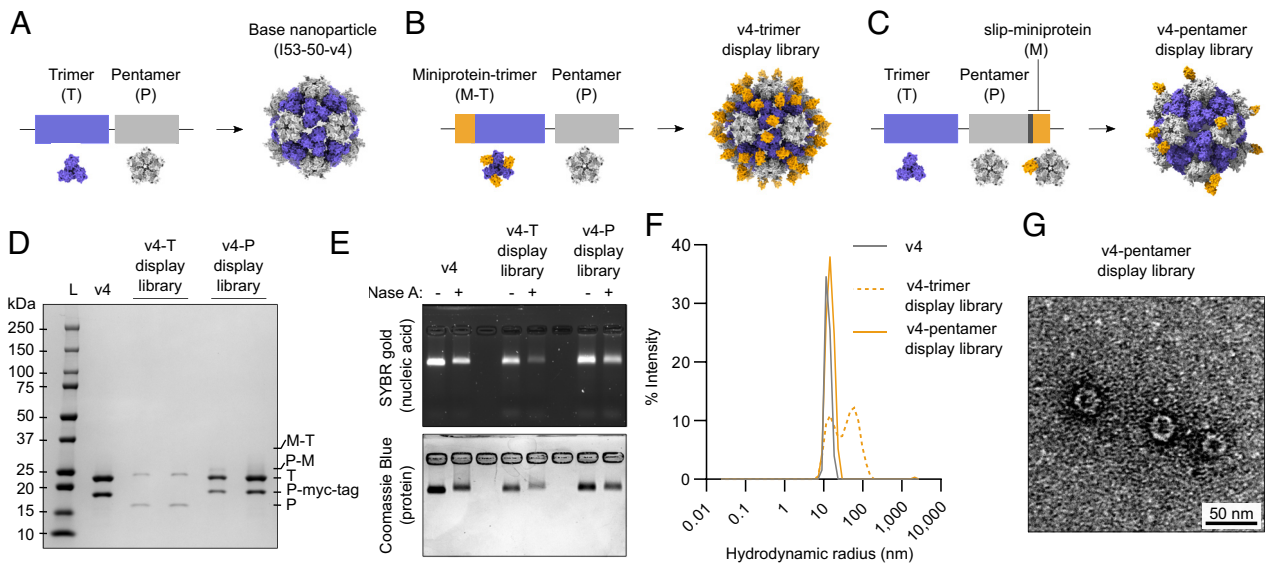


Fig. 3. Biochemical characterization of synthetic nucleocapsid miniprotein display library. (A–C) Model of genetic constructs of base nanoparticle (A) and the nanoparticle displaying miniproteins on the N terminus of the trimer (B, v4-trimer display library) or C terminus of the pentamer (C, v4-pentamer display library). (D) Reducing SDS-PAGE of libraries compared to unmodified synNC (v4) shows lack of trimer-miniprotein fusions (“M-T”) but the presence of pentamer-miniprotein fusions (“P-M”) at a variety of sizes as demonstrated by the smeared P-M band. The reduced valency of fused miniproteins is evident by the ratio of P-myc-tag band intensity to P-M smear intensity. These data agree with the individual library variant characterization in *SI Appendix, Fig. S3*. (E) Evaluation of nuclease resistance. SynNC libraries were incubated in the presence or absence of RNase H and then analyzed for nucleic acid and protein migration on the same native agarose gel (SYBR gold nucleic acid stain, *Top*; Coomassie Blue protein stain, *Bottom*). All three proteins comigrate with mRNA in the presence and absence of RNase A, suggesting mRNA encapsulation and protection. However, the v4-trimer display library shows qualitatively lighter bands after RNase A treatment, suggesting instability. (F) The hydrodynamic radii of the synthetic nucleocapsids were analyzed by dynamic light scattering (DLS) (I53-50-v4: gray, v4-trimer display library: dotted orange line, v4-pentamer display library: solid orange line). The v4-pentamer display library exhibited a slightly larger hydrodynamic radius than v4 alone, as expected. The v4-trimer display library exhibited aggregation, further pointing to instability of this library format. (G) Representative negative stain transmission electron micrograph of v4-pentamer display library shows nanoparticles of expected sizes and shapes. (Scale bar: 50 nm.)

reading frame +1 to avoid a stop codon and instead introduce a flexible polypeptide linker leading into the subsequent encoded miniprotein (22, 23). The v4-trimer display library resulted in very low synNC yield and significant aggregation by dynamic light scattering (DLS), while the v4-pentamer display library exhibited high yield, mRNA encapsulation and protection, and monodisperse nanoparticle sizes by DLS and negative stain transmission electron microscopy (nsEM) (Fig. 3 D–G). We noted that the exact molecular weights of the library display subunits were difficult to determine by SDS-PAGE (Fig. 3D), likely due to the range of sizes of miniproteins being displayed. While de novo designed miniproteins are often 30 to 50 amino acids (aa) long, we observed a broad range of miniprotein sizes by Illumina sequencing (~10 to 120 aa long with a mode of 44 aa). Oligonucleotide microarray synthesis error often results in stop codon insertions and deletions, altering the total length of the miniprotein open reading frame (24). As it is generally assumed that truncated miniproteins lack the key structural elements required to fold into the desired tertiary structures, we only considered miniproteins of the expected size range of 30 to 50 aa in downstream sequencing analyses. DLS confirmed a larger hydrodynamic radius of the v4-pentamer display library compared to v4 alone (Fig. 3F), and purification of individual library clones confirmed the presence of miniproteins fused to the pentamers at reduced valency (*SI Appendix, Fig. S3*). We also observed increased susceptibility to RNase degradation of encapsulated mRNA by the v4-pentamer library compared to the original I53-50-v4 synthetic nucleocapsid (“v4”), but encapsulation and protection were efficient enough to sequence the mRNA of individual library variants (Fig. 3E and *SI Appendix, Fig. S2*). The resulting synthetic nucleocapsid (synNC) library, the v4-pentamer display library, comprised 6.2×10^6 different synNC variants that each displayed a different miniprotein fused to the surface via the pentamer (Fig. 3C).

In a nearly identical manner as described in the surface library section, we tested the miniprotein display library in BALB/c mice (Fig. 1). However, this time we utilized both normal mice and 4T1 tumor-bearing mice to test the library platform in the context of solid tumors (25, 26). 4T1 tumors are a murine model for triple-negative breast cancer. Here, we were especially interested in miniprotein enrichment in tumors, lungs, and muscles, as these tissues are important targets for cancer therapeutics, respiratory illness therapeutics, and muscular dystrophy gene therapies, respectively. We performed three stages of library selection (rounds 1A, 1B, and 2), reducing the library size at each step from 6.2×10^6 unique variants to 1.1×10^5 to 8.4×10^3 , ultimately selecting eight variants of interest with unique organ enrichment patterns (Fig. 4A). To ensure the integrity of encapsulated synNC mRNA, we elected to collect mouse organs after allowing the library to circulate for either 5 or 30 min. We observed similar synNC mRNA yields at both time points and thus used 30-min circulation times in all following in vivo studies (*SI Appendix, Fig. S4 A–C*). We performed qRT-PCR on the bulk synNC library genomes collected from major tissues at each round of selection and detected lower blood accumulation and higher liver and spleen accumulation compared to the surface library (Fig. 4 B and C and *SI Appendix, Fig. S4*). This was expected, as the display of targeting domains on the surface of synNCs shields part of the protein nanoparticle surface that was previously evolved for a longer circulation half-life (12). Additionally, we estimated a shorter circulation half-life for the bulk miniprotein display library compared to I53-50-v4 by qRT-PCR (about 44 min compared to 4.5 h, respectively) (*SI Appendix, Fig. S4D*).

We next sequenced the synNC RNA recovered from the blood, heart, lungs, liver, kidneys, and spleen from all normal mice and tumors isolated from tumor-bearing mice. We calculated the $\log_{10}(\text{enrichment})$ values of sequences identified from

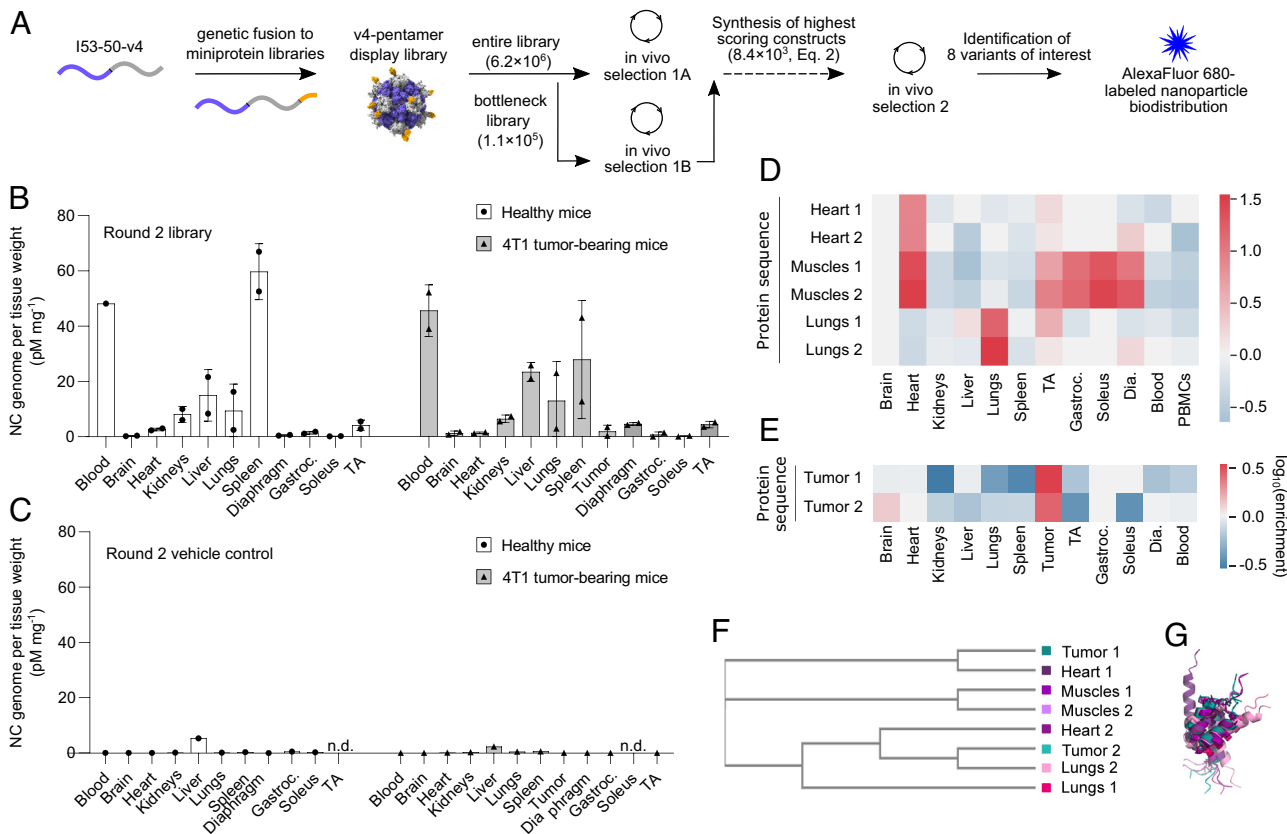


Fig. 4. Biodistribution and sequencing analysis of miniprotein display library. (A) Experimental trajectory of library selection and fluorescence biodistribution studies. SynNCs produced in *E. coli* were used in two rounds of in vivo library selection. In round 1A, the entire library was used. In round 1B, a smaller ("bottleneck") library was injected to achieve higher sequencing coverage. After rounds 1A and 1B, enriched miniprotein sequences in the heart or tumor were selected and synthesized as a separate library solely containing these sequences (Eq. 2). After in vivo selection round 2, eight variants of interest with preferential enrichment in the lungs, heart, heart and muscles, or tumor were selected for individual biodistribution assessment. (B and C) Biodistribution of bulk synNC mRNA recovered from major organs by qRT-PCR for the library (B) and vehicle control (C), reported as picomolar (pM) nucleocapsid (NC) genome per mg of tissue. (D and E) Enrichment of sequences of interest in tissues of interest [heart only, heart and all other muscles tested (Muscles), lungs, and tumor] in normal mice (D) and tumor-bearing mice (E). (F) Phylogenetic tree from Multiple Sequence Comparison by Log-Expectation (MUSCLE) of miniproteins identified in D and E. (G) Superimposition of miniprotein structures predicted by ColabFold shows distinct conformations, despite similar protein sequences. The color scale matches the phylogenetic tree labels in (F). B: n = 2 mice per group; error bars depict SEM. C: n = 1 mouse per group. "n.d.": no data.

these organs compared to the input library that was administered to the mice and selected 10,000 sequences that were preferentially enriched in the heart or tumor (Eq. 2 in *Materials and Methods*). We resynthesized a library for the third round of in vivo selection, which contained 8,407 of the 10,000 selected variants after synNC production and purification. We then performed mRNA-Seq on all aforementioned organs in addition to major muscle groups (diaphragm, gastrocnemius, soleus, and tibialis anterior) and selected two variants enriched in each target organ or organ group for further characterization: heart only, heart and major muscle groups, lungs, or 4T1 tumors from tumor-bearing mice (Fig. 4 D and E and Eq. 1). It has previously been demonstrated that the incorporation of free cysteines in peptides increases tumor accumulation by forming disulfide bonds to serum albumin rather than a specific receptor-ligand interaction; therefore, we only selected miniprotein sequences without cysteines (*SI Appendix, Table S3*) (27). Although some of the eight selected miniproteins have homologous sequences, AlphaFold2 predictions revealed unique 3D conformations, surface hydrophobic networks, and surface charge distributions (Figs. 4 F and G and 5 and *SI Appendix, Table S3*) (28, 29). We were encouraged to see that each of the eight variants, with the exception of the second tumor-enriched miniprotein (Fig. 5 I), featured hydrophobic surface patches that are characteristic of high-affinity minibinders (30).

We then tested the biodistribution of fluorescently labeled nanoparticles displaying each of the eight selected miniproteins. Using the same method as for the surface library variants of interest (Fig. 1 F and G), we individually produced and purified the protein nanoparticles, as well as a bare nanoparticle (i.e., nontargeted) control, each with AF680 covalently conjugated to the interior surface of the nanoparticle to enable fluorescence-based quantification (*SI Appendix, Figs. S2 and S5*). We also produced two fluorescent nanoparticles displaying miniprotein sequences that were enriched in the spleen but found that the observed spleen enrichment was likely the result of free mRNA escaping unstable synNC variants (*SI Appendix, Figs. S5 N, O, S, and T and S6*). As with the surface library variants, we examined fluorescence biodistribution by both qualitative whole organ IVIS imaging (*SI Appendix, Fig. S7*) and a quantitative BCA-fluorescence assay on homogenized organs (Fig. 6 and *SI Appendix, Fig. S8*). We measured the AF680 fluorescence per μg organ protein in each sample and normalized against the summed values across the major tissues in each mouse to enable relative comparisons between samples (heart, lungs, liver, kidneys, and spleen). The same data are plotted as raw values in *SI Appendix, Fig. S8*, in which skeletal muscle tissues were only processed from mice treated with Heart 1, Heart 2, Muscles 1, and Muscles 2 variants. We observed 5.7-fold and 4.8-fold higher relative normalized accumulation of the Lungs 1 and Lungs 2 variants in the lungs compared to the nontargeted

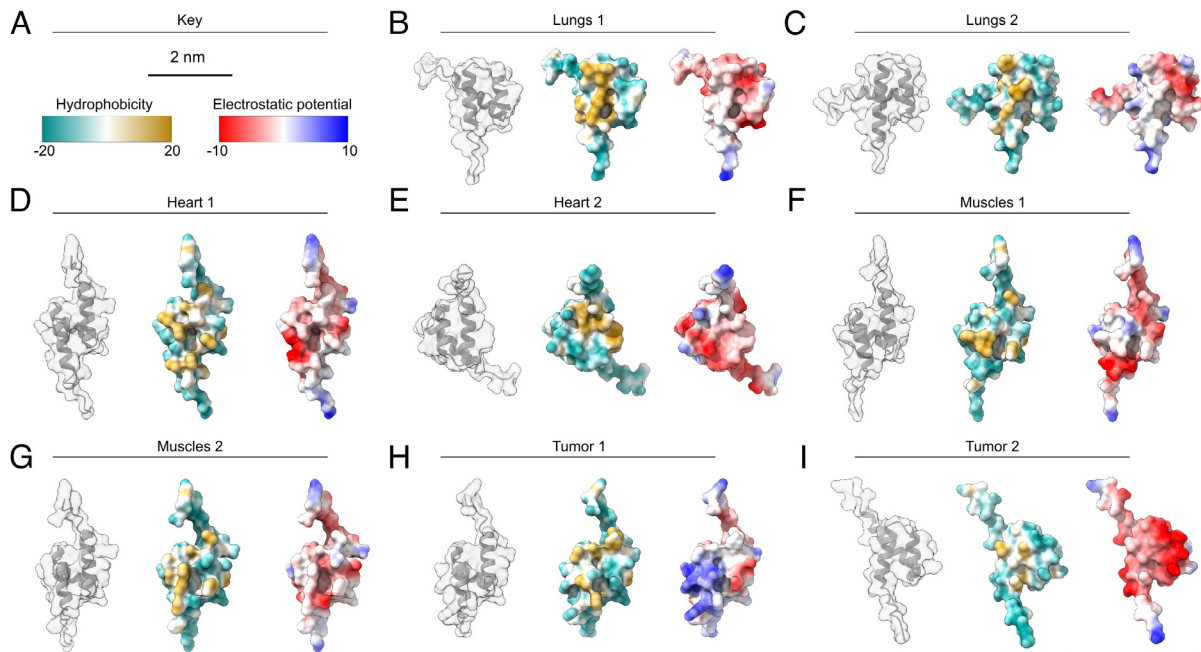


Fig. 5. Enriched miniproteins have distinct topologies, surface hydrophobic networks, and surface charge distributions. (A) Key. (B–I) AlphaFold2 predictions (28) of eight selected miniprotein structures enriched in lungs, heart, muscles, or tumor. The secondary structure, surface hydrophobicity, and surface electrostatic potentials of each miniprotein were visualized in UCSF ChimeraX (29).

control, respectively ($P \leq 0.01$) (Fig. 6 C and I); 5.7-fold and 3.0-fold higher relative accumulation of the Muscles 1 and Muscles 2 variants in heart tissue compared to the nontargeted control, respectively ($P \leq 0.01$) (Fig. 6 E and I); and did not observe a statistically significant increase in accumulation of any other

variant. Tissue cryosectioning and immunofluorescence of nontargeted, Lungs 1, and Muscles 1 nanoparticles in the heart and lungs qualitatively agreed with the biodistribution data in Fig. 6. We observed very little nontargeted nanoparticle signal in the heart and lungs (SI Appendix, Fig. S9 A and B). Importantly, the

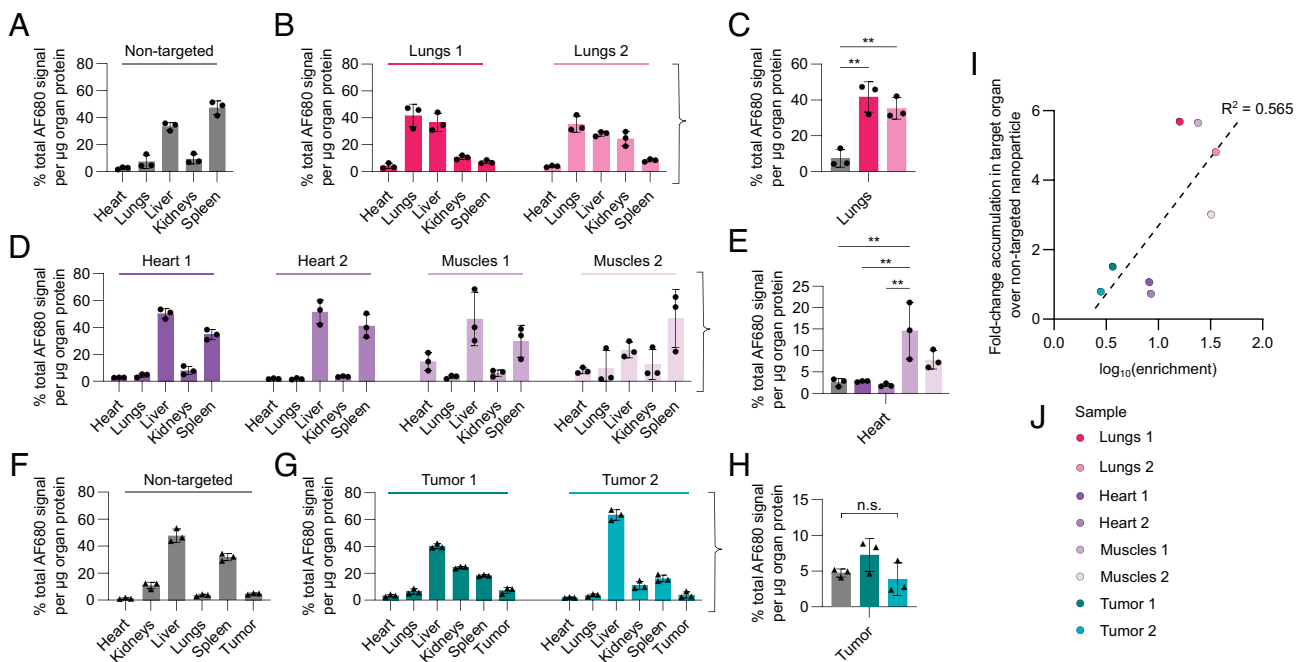


Fig. 6. Normalized fluorescence biodistribution of variants of interest. Each graph shows the normalized fluorescence signal per μg homogenized organ protein from mice injected with AF680-labeled nanoparticles (% total fluorescence per μg organ protein). (A) Nontargeted nanoparticle in normal mice; (B and C) Lungs 1 and 2 library variants in all organs (B) and compared to nontargeted nanoparticle in normal mice in the lungs only (C); (D and E) Heart 1, Heart 2, Muscles 1, and Muscles 2 library variants in all organs (D) and compared to nontargeted nanoparticle in normal mice in the heart only (E); (F) Nontargeted nanoparticle-AF680 in 4T1 tumor-bearing mice; (G and H) Tumor 1 and 2 library variants in all organs (G) and compared to nontargeted nanoparticle in the tumor only (H). In A–H, circles depict individual values from normal mice. Triangles depict individual values from 4T1 tumor-bearing mice. $n = 3$ mice per group. Error bars depict SDs. Statistics: Ordinary one-way ANOVA with Tukey's correction for multiple comparisons. $**P \leq 0.01$; ns: not statistically significant. Only statistically significant comparisons are indicated in C and E. (I) $\text{Log}_{10}(\text{enrichment})$ of specific sequences in target organs from mRNA-Seq (Fig. 4 D and E and Eq. 2) versus fold-change normalized accumulation in target organs over nontargeted nanoparticles by fluorescence biodistribution show positive correlation. The dashed line is a linear fit to the data. (J) Key.

Lungs 1 signal is diffuse in the lungs but not heart (*SI Appendix, Fig. S9 C and D*), and the Muscles 1 signal is diffuse in the heart but not lungs (*SI Appendix, Fig. S9 E and F*).

We suspected that the minor increased relative accumulation of Tumor 1 in 4T1 tumors could have been due to nonspecific accumulation enabled by the large, positively charged patch on the surface of the miniprotein (Figs. 5H and 6H). Additionally, since in vivo phage display often yields targeting domains that bind tumor vasculature (31, 32), we suspected that Tumor 1 could have bound tumor vasculature instead of the 4T1 tumor cells. By tissue cryosectioning and immunofluorescence, we observed Tumor 1 signal exclusively in the vascular endothelium of the tumor, and we did not observe signal from nontargeted or Muscles 1 nanoparticle samples in tumor (*SI Appendix, Fig. S9 G–I*). However, we did not observe binding of Tumor 1 to 4T1 tumor cells or to murine brain endothelial (BEND3) cells in vitro (*SI Appendix, Fig. S10*). To examine distribution- and elimination-related effects on organ accumulation, we measured the serum circulation half-lives of AF680-labeled nontargeted, Lungs 1, Muscles 1, and Tumor 1 nanoparticles (*SI Appendix, Fig. S11 and Table S3*). Based on a two-compartment pharmacokinetic model, all samples exhibited distribution half-lives less than 30 min except M1-AF680, which exhibited a distribution half-life of 1.26 h. We also noted increased relative target organ accumulation by fluorescent nanoparticles displaying organ-matched miniproteins compared to nanoparticles displaying non-organ-matched miniproteins. For example, the Lungs 1 and Lungs 2 samples were the only variants to exhibit significantly increased accumulation in the lungs, and Muscles 1 and Muscles 2 were the only variants to exhibit significantly increased relative accumulation in heart tissue (Fig. 6). Last, we observed a slight positive correlation between enrichment score and fold-change of both nonrelative and relative organ accumulation compared to nontargeted nanoparticles (Fig. 6I and *SI Appendix, Fig. S8M*).

Discussion

Here, we employ in vivo library selection of synthetic nucleocapsids to identify variants with altered tissue biodistribution and to select for specific tissue accumulation. In vivo library selection is a powerful method for targeting domain identification because millions of candidates can be simultaneously injected and evaluated in vivo, and those exhibiting the desired behavior can be identified by linking genotype to phenotype (e.g., biodistribution). Unlike in vitro library selection approaches, which are typically based on a subset of the functional requirements required to achieve a desired outcome (e.g., receptor binding), in vivo library selection subjects the libraries to the complete set of known and unknown functional requirements and physiological barriers. We are still unable to comprehensively recapitulate and account for these physiological barriers during in silico design or in vitro testing.

The stringency of in vivo selection together with the semirational library design (i.e., randomly selected miniproteins that were computationally designed to fold into binder-like structures) enabled the identification of well-performing variants in just two rounds of selection. Other in vivo library selection techniques like peptide phage display have also been shown to yield promising results in just two to three rounds of in vivo selection (33, 34). In comparison, in vitro selection techniques like SELEX, in which random oligonucleotide libraries are screened against a target, typically require 5 to 10 rounds of selection to identify aptamers that bind specific cell populations or 10 to 15 rounds of selection to identify aptamers that bind specific proteins (35–37). Notably, incorporating more

rational design into both library design and in vivo selection strategy could further decrease the number of rounds of selection needed to achieve a desired result and further improve the quality of the molecules identified. For example, designing more complex surface libraries or miniprotein binder libraries against specific targets is now possible using cutting-edge machine learning-based design methods and would enable the in vivo library selection strategy to be tailored to those targets (27, 38).

The ability to sequence library mRNA from multiple tissues enabled simultaneous positive selection for accumulation in tissues of interest and negative selection for accumulation in off-target tissues. However, this selection scheme is not immune to the pervasive hurdle of distinguishing false positives from true positives that recognize the target through a specific molecular interaction. Four of the eight selected variants exhibited statistically significant shifts in biodistribution and target organ accumulation compared to nontargeted nanoparticles (Lungs 1, Lungs 2, Muscles 1, and Muscles 2), and one library variant showed a minor, nonstatistically significant increase in relative 4T1 tumor accumulation compared to nontargeted nanoparticles (Tumor 1). Tissue cryosectioning and immunofluorescence of the Lungs 1, Muscles 1, and Tumor 1 variants agreed with these biodistribution trends, where Lungs 1 and Muscles 1 variants exhibit diffuse signal throughout the target tissue and the Tumor 1 variant shows a low magnitude of signal lining the tumor vasculature (*SI Appendix, Fig. S9*). Consistent with previous results (12) and the surface library results (Fig. 2D–F), none of the variants with \log_{10} (enrichment) scores below 1.0 exhibited a statistically significant increase in fluorescent nanoparticle accumulation in the target tissue(s), while three of the four variants with \log_{10} (enrichment) scores greater than 1.0 did (Fig. 6I). However, we have yet to demonstrate whether these shifts in biodistribution are caused by specific receptor-ligand interactions; other undesired or nonspecific phenomena could be responsible. Further experiments including additional organ cryosectioning and immunofluorescence imaging are needed to understand the mechanisms of target organ accumulation and to identify molecular targets the miniproteins might be binding.

Importantly, the ability to detect congruence between separate readouts may have helped avoid the selection of false positives in some cases. Both variants that were selected for preferential enrichment in all five of the muscles—heart, diaphragm, gastrocnemius, soleus, and tibialis anterior—were mutually corroborating (Muscles 1 and Muscles 2, Fig. 4D). Indeed, both variants successfully showed increased accumulation in the target tissues by fluorescence biodistribution (Fig. 6D and E and *SI Appendix, Fig. S8 D–I*). Additionally, the ability to rapidly predict the 3D conformation of miniproteins of interest with reasonably high confidence, confirming that they were likely folding into tertiary protein structures with binder-like properties, aided our selection of variants to further examine (Fig. 5) (28, 39).

Our results support the growing body of research highlighting the power of in vivo library selection for identifying nanoparticles with desired physiological and functional properties (4, 5, 12). Not only could miniprotein binders of interest from this study be modularly applied to other delivery platforms to potentially confer altered tissue tropism [e.g., the Muscles 2 miniprotein could be displayed on lentivirus or AAV via SpyCatcher-SpyTag (40, 41)], synNCs displaying miniproteins of interest could also be formulated as delivery vehicles themselves by encapsulating therapeutic molecules instead of self-mRNA or fluorophores (1, 42). Additionally, miniproteins could be optimized for desired in vivo behaviors using this platform, then extracted and used on their own in a manner similar to antibody-based therapeutics (43). These

methods could also be applied to virus-like particles or protein nanoparticles of different sizes, shapes, and surface chemistries that could be better suited to specific delivery applications (38, 44–49). In conclusion, the synthetic nucleocapsid library selection platform described in this work can be implemented in various ways to advance the discovery and optimization of protein-based therapeutics.

Materials and Methods

\log_{10} (enrichment) score of synNC surface library variant sequence s in tissue t compared to v_4 sequence v from tissue t containing up to x unique sequences.

$$\log_{10}(\text{enrichment})_{s,t} = \left[\frac{\text{representation of } s \text{ in } t}{\text{representation of } v \text{ in } t} \right] = \left[\frac{\left(\sum_s \text{reads in } t / \sum_0^x \text{reads in } t \right)}{\left(\sum_v \text{reads in } t / \sum_0^x \text{reads in } t \right)} \right] = \left[\frac{\left(\sum_s \text{reads in } t \right)}{\left(\sum_v \text{reads in } t \right)} \right]. \quad [1]$$

\log_{10} (enrichment) score of miniprotein sequence m in tissue t containing up to x unique sequences compared to input library dose i containing up to y unique sequences.

$$\log_{10}(\text{enrichment})_{m,t} = \left[\frac{\text{representation of } m \text{ in } t}{\text{representation of } m \text{ in } i} \right] = \left[\frac{\left(\sum_m \text{reads in } t / \sum_0^x \text{reads in } t \right)}{\left(\sum_m \text{reads in } i / \sum_0^y \text{reads in } i \right)} \right]. \quad [2]$$

AF680 signal per μg tissue protein normalized to the sum of each organ for each individual mouse.

$$\begin{aligned} & \% \text{ total AF680 signal per } \mu\text{g tissue protein } t \text{ from mouse with } x \text{ number of tissues} \\ & = \left[\frac{\text{AF680 RFU in } t}{(\mu\text{g protein in } t)} \right] / \left[\sum_{i=0}^x \frac{\text{AF680 RFU in } i}{(\mu\text{g protein in } i)} \right]. \quad [3] \end{aligned}$$

Cloning Synthetic Nucleocapsid Libraries.

Surface library. The I53-50-v4 synthetic nucleocapsid (synNC) genome was PCR-amplified as previously described (12). Oligonucleotides encoding for the degenerate codons described in *SI Appendix, Table S1* were utilized to construct the surface library via PCR-based site-directed mutagenesis.

Miniprotein library. The I53-50-v4 synthetic nucleocapsid (synNC) genome was PCR-amplified as previously described (12). Randomly chosen (double-blind) libraries of miniprotein designs were obtained from the Baker Lab oligonucleotide chip stocks, synthesized by Twist, IDT, CustomArray, or Agilent. Miniprotein sequences were qPCR-amplified by primers that included overlap sites with the synNC genome (*SI Appendix, Table S2*). Assembly PCR was performed to genetically fuse the miniproteins to the synthetic nucleocapsids. Assemblies were cloned into pet29b(+) protein production vectors via Gibson Assembly (NEB E2611L), electroporated into DH10beta cells (NEB C3019H), and grown at 37 °C on agar-kanamycin plates overnight. Plates were scraped, plasmid DNA was isolated, plasmid DNA was electroporated into BL21 Star (DE3) (ThermoFisher C601003) or Lemo21(DE3) (NEB C2528J) producer cells, and cells were grown at 37 °C on agar-kanamycin plates overnight. Fresh chemically competent cells were expanded and prepared for electroporation per the manufacturer's instructions in Milli-Q water. Bacterial lawns were collected from the plates and used to inoculate large producer cultures for protein production. At both plating steps, bacteria were serially diluted and spotted onto separate agar plates to calculate the approximate library size and coverage. Library coverages of 10-fold to 100-fold were obtained in all selection steps except miniprotein library selection round 1A.

Protein Production and Fluorescent Labeling. Proteins were produced and purified as previously described (12). Briefly, proteins were expressed in *E. coli* via IPTG induction or autoinduction (50), cell pellets were homogenized and microfluidized, and the supernatant was subjected to immobilized metal affinity chromatography (IMAC) at pH 8.0. Triton X-114 (Sigma Aldrich X114-500ML, CAS 9036-19-5) was used to remove lipopolysaccharides (endotoxin). Samples were incubated with 10 $\mu\text{g}/\text{mL}$ RNase H (ThermoFisher 18021071) for 10 min at room temperature to remove free RNA. Immediately following RNase treatment, size-exclusion chromatography (SEC) was performed on a Superose 6 Increase 10/300 GL column (Cytiva 29-0915-96) on a GE AKTA Pure with PBS or HEPES buffer at pH 7.4. Samples were stored at 4 °C and sterile filtered before use. For fluorescent labeling, samples were labeled with fivefold to 10-fold molar excess of Alexa Fluor 488 C5 maleimide (ThermoFisher A10254) or Alexa Fluor 680 C2

maleimide (ThermoFisher A20344) at 4 °C overnight or at room temperature for 2 h prior to dialysis to remove free fluorophore, quenching with 10-fold molar excess dithiothreitol (DTT, ThermoFisher P2325), purification on a PD-10 desalting column (Cytiva 17085101), and purification by SEC.

Biochemical Characterization. Proteins were characterized as previously described (12). Briefly, the molecular weight was analyzed by SDS-PAGE, nucleic acid and protein comigration and RNase resistance was analyzed by native agarose gel electrophoresis, and polydispersity and size were measured by dynamic light scattering (DLS) and negative stain transmission electron microscopy (nsEM). Protein concentrations were measured by either Cubit (Invitrogen Q33212), UV-Vis spectroscopy, or Bradford assay (Thermo Scientific 23200). Molecular protein

masses were measured by intact mass spectra obtained via reverse-phase LC/MS on an Agilent G6230B TOF on an AdvanceBio RP-Desalting column and deconvoluted with BioConfirm using a total entropy algorithm. Endotoxin was measured (Charles River Endosafe LAL Endotoxin Cartridges, PTS201F) (*SI Appendix, Table S5*).

In Vivo Library Circulation, Fluorescence Biodistribution, and Organ Recovery.

All animal studies were approved by the Institutional Animal Care and Use Committee at the University of Washington and were conducted in accordance with use and regulations and OAW standards. Female BALB/c mice (BALB/cAnNCrI, Inbred, strain code 028) were purchased from Charles River Laboratories at 6 to 8 wk of age. To establish tumor-bearing mice, 9- to 10-wk-old mice were inoculated subcutaneously near the lower right nipple with 1 to 2 million 4T1 cells, and tumors were allowed to grow to about 100 to 200 mm^3 in size, approximately 1 to 2 wk before the start of the study. Mice were intravenously (IV) (retro-orbitally) injected (ROI) with 150 μL of sample (0.5 to 4 nanomoles protein nanoparticle asymmetric unit). After the designated circulation time, mice were killed with Avertin overdose and cardiac puncture. Blood was collected from the vena cava into EDTA-lined tubes. PBS perfusions through the left ventricle were performed. Organs were collected and immediately snap-frozen. For fluorescent imaging, organs were imaged with Xenogen IVIS before snap-freezing.

RNA isolation from snap-frozen tissues was performed within 1 wk of the in vivo study. Frozen tissues were ground in liquid nitrogen with a mortar and pestle over dry ice to remain frozen. A subset of the ground tissue was weighed, dissolved in TRIzol, sonicated with a Sonic Dismembrator 60 (Fisher Scientific), and stored at -80 °C until RNA extraction. We experienced great difficulty homogenizing the muscles with the hand-held homogenizer. Snap-freezing and grinding the frozen muscles with a mortar and pestle, as we employed to recover RNA from muscles and organs during the selection steps, could be a more reliable method to measure accumulation in muscles moving forward.

To measure AF680 fluorescence per μg tissue protein, tissues were thawed and suspended in RIPA buffer (Thermo 89901) supplemented with DNase (Thermo 90083) before homogenizing with OmniTissue Homogenizer (TH). Homogenized tissues were spun down and lysate relative fluorescence units at 679 nm were read on a plate reader (infinite M200PRO plate reader, Tecan). In parallel, Pierce BCA Protein Assay Kits were utilized according to the manufacturer's instructions to quantify the amount of tissue protein per sample (ThermoFisher 23225). AF680 signal per μg organ (or tissue) protein was calculated by dividing the AF680 RFU

by the μg organ protein calculated by the BCA assay. These values were then normalized (Eq. 3 in *Materials and Methods*).

To measure serum circulation half-lives, 0.5 nmoles of AF680-labeled nanoparticle ASU was I.V. R.O.I.-injected into normal BALB/c mice. Each sample was administered to two groups of mice ($n = 3$ mice per group). Blood was collected from group A at 5 min, 2 h, and 24 h. Blood was collected from group B at 30 min, 6 h, and 24 h. Blood was immediately centrifuged, serum was aspirated, and samples were frozen on dry ice. Samples were stored at -80°C for up to 4 d prior to measuring fluorescence with an infinite M200PRO plate reader (Tecan). Two-phase exponential decay was performed in GraphPad Prism 9.

RNA Extraction, Reverse Transcription, qPCR, and Illumina Sequencing. Samples were stored in TRIzol (Invitrogen 15596026) at -80°C . RNA was purified from samples using RNeasy Kits (Qiagen 74104 or 74106) according to the manufacturer's instructions. RNA and DNA concentrations were measured with NanoDrop or Qubit (Invitrogen Q32852, Q32854). Reverse transcription (PrimeScript 1st strand cDNA Synthesis Kit, Takara 6110A), quantitative PCR (KAPA HiFi HotStart Ready Mix, Roche; SYBR Green, Invitrogen S7563), and sequencing on the Illumina MiSeq (Illumina MiSeq Reagent Kit v3, MS-102-3003; 5% PhiX Control v3, FC-110-3001) were performed as previously described and/or according to manufacturer's instructions unless otherwise noted (12).

The skpp-132-R reverse primer was used for reverse transcription instead of the primers and 6mers provided by the Takara PrimeScript 1st strand cDNA Synthesis Kit (5'-CATACTGTTGTTGCTAGGC-3'). The skpp-132-F forward primer was used to amplify synNC genomes with qPCR (5'-TAGGATTACTGCTCGTGAC-3'). The skpp-offset-R reverse primer was used to amplify synNC genomes and to later amplify miniprotein sequences out of synNC genomes with qPCR (5'-GTTGCTAGGCTCAGTGATGG-3'). A separate forward primer was used in combination with skpp-132-R to amplify the miniprotein sequences out of synNC genomes for sequencing (forward primer named c _ lib _ fwd, 5'-GTATCTTGACGGCTCCGGT-3').

Sequencing Analysis. Sequencing analysis was performed by aligning the MiSeq output files with PEAR (a fast and accurate Illumina Paired-End read mergeR) and using custom Python scripts (51). Library diversity was calculated by summing the number of unique variants observed in the sequencing results for each library, from which variants of interest were chosen. Scripts and processed sequencing data are available in the following GitHub repository: https://github.com/aolshesky/in_vivo_miniprotein_selection.git (52). Raw sequencing data are available in the NIH Sequence Read Archive, BioProject accession number [PRJNA994506](https://www.ncbi.nlm.nih.gov/PRJNA994506) and [PRJNA999343](https://www.ncbi.nlm.nih.gov/PRJNA999343) (53, 54).

Flow Cytometry Binding Study. 4T1 and BEND3 cells were separately cultured in RPMI 1640 with 10% FBS (Gibco 11875093) at 37°C at 5% CO_2 . Cells were dissociated with StemPro Accutase (Gibco A1110501). Cells were stained with ZombieViolet viability stain (BioLegend 423113) in HEPES buffer (20 mM HEPES, 150 mM NaCl, pH 7.4). Fluorescent nanoparticles were incubated with cells in HEPES buffer with 1% w/v BSA for 30 min on ice. Cells were washed three times with HEPES-1% BSA prior to examination on an Attune NxT Flow Cytometer (Invitrogen). Data were analyzed using FlowJo.

Tissue Cryosectioning and Immunofluorescence. Nanoparticles and mice were prepared and administered as described in the In vivo library circulation, fluorescence biodistribution, and organ recovery methods section. After a 30-min circulation time, mice were perfused with 20 mL PBS followed by 10 mL 4% PFA. Organs were removed and transferred to 5 mL 4% PFA and stored at 4°C overnight, protected from light. The next day, organs were transferred to 15% sucrose w/v in TBS (20 mM Tris, 150 mM NaCl, pH 7.4) and stored at 4°C for approximately 24 h, protected from light. Tissues were then transferred to

30% sucrose w/v in TBS (20 mM Tris, 150 mM NaCl, pH 7.4) and stored at 4°C for approximately 24 to 48 h, protected from light. Tissues were placed in OCT Compound, equilibrated for several minutes, and then frozen on dry ice before storage at -80°C for approximately 1 to 2 wk. Tissues were sectioned into $10\text{-}\mu\text{m}$ slices onto positively charged slides using a freezing microtome, allowed to dry for approximately 1 h at room temperature in the dark, and then stored at -80°C until immunofluorescence staining.

In preparation for immunofluorescence staining, OCT residue was removed from the slides, and tissues were incubated in 4% PFA in PBS for 10 min at room temperature. Samples were blocked and permeabilized for 1 h at room temperature [2% donkey serum (DKS), 2% bovine serum albumin (BSA), 0.3% Triton X in TBS]. Samples were then incubated with primary stain overnight at 4°C , rocking [2% DKS, 2% BSA, 0.3% Triton X in TBS, goat anti-CD31 antibody at 1:500 dilution (Novus Biological, AF3628) and rat anti-LYVE-1 antibody at 1:40 dilution (R&D Systems, MAB2125-100)]. Samples were washed three times for 30 min each with TBS 0.1% Tween-20. Samples were incubated with secondary stain for 1 h at room temperature [2% DKS, 0.1% Tween-20 in TBS, anti-goat antibody-AF488 at 1:500 dilution (Jackson Labs 705-545-003) and anti-rat antibody-AF568 at 1:400 dilution (Abcam, ab175475)]. Finally, samples were washed three times, allowed to dry completely, and covered with polyvinyl alcohol and glass coverslips. Samples were allowed to dry at least overnight before imaging.

Samples were imaged on a Leica SP8X confocal microscope, and images were processed in Fiji (55).

Figures and Statistical Analysis. Figures were made using Inkscape, GraphPad Prism, PyMOL, UCSF ChimeraX, BioRender, Python, and FlowJo. Statistical analyses were performed in GraphPad Prism 9.

Data, Materials, and Software Availability. Scripts and processed sequencing data, Raw sequencing data have been deposited in Github; NIH Sequence Read Archive (https://github.com/aolshesky/in_vivo_miniprotein_selection.git; [PRJNA994506](https://www.ncbi.nlm.nih.gov/PRJNA994506) and [PRJNA999343](https://www.ncbi.nlm.nih.gov/PRJNA999343)) (52–54). All other data are included in the manuscript and/or *SI Appendix*.

ACKNOWLEDGMENTS. We gratefully acknowledge Dr. Drew Sellers, Nataly Kacherovsky, Dr. Heather H. Gustafson, Dr. Karla Herpoldt, Brooke Fiala, Dr. Inna Goreschnik, Dr. David Peeler, Dr. Cassie Bryan, Dr. Franziska Seeger, Dr. Ian Cardle, Dr. Gary Liu, Dr. Ivan Anishchanka, Dr. Ratika Krishnamurthy, Dr. Barry Stoddard, Dr. Erin Yang, Kandise A. VanWormer, and Luki Goldschmidt for technical and scientific advice, feedback on the manuscript, lab and administrative support, and insightful conversations. We would also like to thank Xinting Li, Mila Lamb, and Stephen Rettie for assistance with mass spectrometry and the W.M. Keck Center for Advanced Studies in Neural Signaling (NIH grant S10 OD016240) and the help of Nathaniel Peters for confocal microscopy experiments. This work was supported by the National Cancer Institute (NCI) grants 1R21CA232430 (N.P.K. and S.H.P.), 1R01CA257563 (S.H.P.), and T32CA080416 (A.O.), Defense Threat Reduction Agency Grant HDTRA1-18-1-0001 (D.B. and N.P.K.), the Bill and Melinda Gates Foundation INV-010680 (D.B. and N.P.K.), the Audacious Project at the Institute for Protein Design (N.P.K. and D.B.), and National Institute of Allergy and Infectious Disease (NIAID) grant U54AI170856 (N.P.K.).

Author affiliations: ^aDepartment of Bioengineering, University of Washington, Seattle, WA 98195; ^bInstitute for Protein Design, University of Washington, Seattle, WA 98195; ^cDepartment of Biochemistry, University of Washington, Seattle, WA 98195; ^dDepartment of Molecular and Cellular Biology, University of Washington, Seattle, WA 98195; ^eDepartment of Neurology, University of Washington, Seattle, WA 98195; and ^fMolecular Engineering and Sciences Institute, University of Washington, Seattle, WA 98195

1. A. Olshefsky, C. Richardson, S. H. Pun, N. P. King, Engineering self-assembling protein nanoparticles for therapeutic delivery. *Bioconjug. Chem.* **33**, 2018–2034 (2022).
2. M. J. Mitchell *et al.*, Engineering precision nanoparticles for drug delivery. *Nat. Rev. Drug Discov.* **20**, 101–124 (2021).
3. E. Blanco, H. Shen, M. Ferrari, Principles of nanoparticle design for overcoming biological barriers to drug delivery. *Nat. Biotechnol.* **33**, 941–951 (2015).
4. C. D. Sago *et al.*, Nanoparticles that deliver RNA to bone marrow identified by in vivo directed evolution. *J. Am. Chem. Soc.* **140**, 17095–17105 (2018).

5. Q. Cheng *et al.*, Selective organ targeting (SORT) nanoparticles for tissue-specific mRNA delivery and CRISPR-Cas gene editing. *Nat. Nanotechnol.* **15**, 313–320 (2020).
6. M. P. Lokugamage *et al.*, Optimization of lipid nanoparticles for the delivery of nebulized therapeutic mRNA to the lungs. *Nat. Biomed. Eng.* **5**, 1059–1068 (2021).
7. M. Qiu *et al.*, Lung-selective mRNA delivery of synthetic lipid nanoparticles for the treatment of pulmonary lymphangioleiomyomatosis. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2116271119 (2022), 10.1073/pnas.2116271119.

8. S. G. Huayamares *et al.*, High-throughput screens identify a lipid nanoparticle that preferentially delivers mRNA to human tumors in vivo. *J. Control. Release* **257**, 394–403 (2023), 10.1016/j.jconrel.2023.04.005.
9. H. H. Gustafson, A. Olshefsky, M. Sylvestre, D. L. Sellers, S. H. Pun, Current state of in vivo panning technologies: Designing specificity and affinity into the future of drug targeting. *Adv. Drug Deliv. Rev.* **130**, 39–49 (2018).
10. K. S. Lam *et al.*, A new type of synthetic peptide library for identifying ligand-binding activity. *Nature* **354**, 82–84 (1991).
11. R. C. Münch *et al.*, Displaying high-affinity ligands on adeno-associated viral vectors enables tumor cell-specific and safe gene transfer. *Mol. Ther.* **21**, 109–118 (2013).
12. G. L. Butterfield *et al.*, Evolution of a designed protein assembly encapsulating its own RNA genome. *Nature* **552**, 415–420 (2017).
13. J. B. Bale *et al.*, Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science* **353**, 389–394 (2016).
14. K. Hertveldt, T. Beliën, G. Volckaert, General M13 phage display: M13 phage display in identification and characterization of protein-protein interactions. *Methods Mol. Biol.* **502**, 321–339 (2009).
15. T. G. W. Edwardson, T. Mori, D. Hilvert, Rational engineering of a designed protein cage for siRNA delivery. *J. Am. Chem. Soc.* **140**, 10439–10442 (2018).
16. J. Marcandalli *et al.*, Induction of potent neutralizing antibody responses by a designed protein nanoparticle vaccine for respiratory syncytial virus. *Cell* **176**, 1420–1431.e17 (2019).
17. D. Martínez-Cano *et al.*, Process development of a SARS-CoV-2 nanoparticle vaccine. *Process Biochem.* **129**, 241–256 (2023), 10.1016/j.procbio.2023.03.014.
18. H. Benasutti, "Development of a novel gene therapy & investigation of synthetic gene therapy delivery systems", *University of Washington ProQuest Dissertations Publishing* (2023), 30310348.
19. A. Olshefsky, "Modular engineering and in vivo library selection of tissue-targeting protein nanoparticles", *University of Washington ProQuest Dissertations Publishing* (2023), 30420295.
20. A. Chevalier *et al.*, Massively parallel de novo protein design for targeted therapeutics. *Nature* **550**, 74–79 (2017).
21. G. J. Rocklin *et al.*, Global analysis of protein folding using massively parallel design, synthesis, and testing. *Science* **357**, 168–175 (2017).
22. P. V. Baranov, R. F. Gesteland, J. F. Atkins, Release factor 2 frameshifting sites in different bacteria. *EMBO Rep.* **3**, 373–377 (2002).
23. C. L. Sanders, J. F. Curran, Genetic analysis of the E site during RF2 programmed frameshifting. *RNA* **13**, 1483–1491 (2007).
24. E. LeProust, Agilent's microarray platform: how high-fidelity DNA synthesis maximizes the dynamic range of gene expression measurements. agilent.com. https://www.agilent.com/library/applications/5989-9159en_lo.pdf. Accessed 14 July 2023.
25. B. A. Pulaski, S. Ostrand-Rosenberg, Mouse 4T1 breast tumor model. *Curr. Protoc. Immunol.*, in press.
26. B. Schrörs *et al.*, Multi-omics characterization of the 4T1 murine mammary gland tumor model. *Front. Oncol.* **10**, 1195 (2020).
27. J. Tubiana *et al.*, Funneling modulatory peptide design with generative models: Discovery and characterization of disruptors of calcineurin protein-protein interactions. *PLoS Comput. Biol.* **19**, e1010874 (2023).
28. M. Mirdita *et al.*, ColabFold: Making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
29. T. D. Goddard *et al.*, UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).
30. L. Cao *et al.*, Design of protein-binding proteins from the target structure alone. *Nature* **605**, 551–560 (2022).
31. E. Ruoslahti, Targeting tumor vasculature with homing peptides from phage display. *Semin. Cancer Biol.* **10**, 435–442 (2000).
32. A. J. Zurita, W. Arap, R. Pasqualini, Mapping tumor vascular diversity by screening phage display libraries. *J. Control. Release* **91**, 183–186 (2003).
33. D. L. Sellers *et al.*, Targeted axonal import (Taxi) peptide delivers functional proteins into spinal cord motor neurons after peripheral administration. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 2514–2519 (2016).
34. R. Pasqualini, E. Ruoslahti, Organ targeting in vivo using phage display peptide libraries. *Nature* **380**, 364–366 (1996).
35. M. Sylvestre *et al.*, Identification of a DNA aptamer that binds to human monocytes and macrophages. *Bioconjug. Chem.* **31**, 1899–1907 (2020).
36. N. Kacherovsky *et al.*, Traceless aptamer-mediated isolation of CD8+ T cells for chimeric antigen receptor T-cell therapy. *Nat. Biomed. Eng.* **3**, 783–795 (2019).
37. L. F. Yang *et al.*, Aptamer sandwich lateral flow assay (AptaFlow) for antibody-free SARS-CoV-2 detection. *Anal. Chem.* **94**, 7278–7285 (2022).
38. J. L. Watson *et al.*, Broadly applicable and accurate protein design by integrating structure prediction networks and diffusion generative models. *bioRxiv* [Preprint] (2022). <https://doi.org/10.1101/2022.12.09.519842>. Accessed 14 July 2023.
39. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
40. N. Kasaraneni, A. M. Chamoun-Emanuelli, G. Wright, Z. Chen, Retargeting lentiviruses via SpyCatcher-SpyTag chemistry for gene delivery into specific cell types. *MBio* **8**, e01860-17 (2017), 10.1128/mBio.01860-17.
41. A. Muik *et al.*, Covalent coupling of high-affinity ligands to the surface of viral vector particles by protein trans-splicing mediates cell type-specific gene transfer. *Biomaterials* **144**, 84–94 (2017).
42. E. N. Baker, C. J. Squire, P. G. Young, Self-generated covalent cross-links in the cell-surface adhesins of Gram-positive bacteria. *Biochem. Soc. Trans.* **43**, 787–794 (2015).
43. P. D. Senter, E. L. Sievers, The discovery and development of brentuximab vedotin for use in relapsed Hodgkin lymphoma and systemic anaplastic large cell lymphoma. *Nat. Biotechnol.* **30**, 631–637 (2012).
44. S. Kang *et al.*, Implementation of p22 viral capsids as nanoplatfoms. *Biomacromolecules* **11**, 2804–2809 (2010).
45. J. O. Jeon *et al.*, Designed nanocage displaying ligand-specific Peptide bunches for high affinity and biological activity. *ACS Nano* **7**, 7462–7471 (2013).
46. G. Ueda *et al.*, Tailored design of protein nanoparticle scaffolds for multivalent presentation of viral glycoprotein antigens. *Elife* **9**, e57659 (2020), 10.7554/eLife.57659.
47. S. Tetter *et al.*, Evolution of a virus-like architecture and packaging mechanism in a repurposed bacterial protein. *Science* **372**, 1220–1224 (2021).
48. A. Van de Steen *et al.*, Bioengineering bacterial encapsulin nanocompartments as targeted drug delivery system. *Synth. Syst. Biotechnol.* **6**, 231–241 (2021).
49. J. Dauparas *et al.*, Robust deep learning-based protein sequence design using ProteinMPNN. *Science* **378**, 49–56 (2022).
50. F. W. Studier, Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
51. J. Zhang, K. Kobert, T. Flouri, A. Stamatakis, PEAR: A fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**, 614–620 (2014).
52. A. Olshefsky, G. Kher, in_vivo_miniprotein_selection. Github. https://github.com/aolshefsky/in_vivo_miniprotein_selection. Deposited 13 July 2023.
53. A. Olshefsky, G. Butterfield, PRJNA999343. National Institutes of Health Sequence Read Archive. <https://www.ncbi.nlm.nih.gov/sra/PRJNA999343>. Deposited 28 July 2023.
54. A. Olshefsky, PRJNA994506. National Institutes of Health Sequence Read Archive. <https://www.ncbi.nlm.nih.gov/sra/PRJNA994506>. Deposited 14 July 2023.
55. J. Schindelin *et al.*, Fiji: An open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).