PNAS RESEARCH ARTICLE | BIOPHYSICS AND COMPUTATIONAL BIOLOGY OPEN ACCESS

Improving the secretion of designed protein assemblies through negative design of cryptic transmembrane domains

Jing Yang (John) Wang[a,b,c,1], Alena Khmelinskaia[a,b,d,e,1], William Sheffler[a,b], Marcos C. Miranda[a,b], Aleksandar Antanasijevic[f,g], Andrew J. Borst[a,b], Susana V. Torres[a,b], Chelsea Shu[a,b], Yang Hsia[a,b], Una Nattermann[a,b,h], Daniel Ellis[a,b,c], Carl Walkey[a,b], Maggie Ahlrichs[a,b], Sidney Chan[a,b], Alex Kang[a,b], Hannah Nguyen[a,b], Claire Sydeman[a,b], Banumathi Sankaran[i,j], Mengyu Wu[a,b], Asim K. Bera[a,b], Lauren Carter[a,b], Brooke Fiala[a,b], Michael Murphy[a,b], David Baker[a,b], Andrew B. Ward[f,g], and Neil P. King[a,b,2]

Edited by Daniel W. Kulp, Center for HIV/AIDS Vaccine Immunology and Immunogen Discovery, The Scripps Research Institute, La Jolla, CA; received August 26, 2022; accepted February 3, 2023 by Editorial Board Member William F. DeGrado

Computationally designed protein nanoparticles have recently emerged as a promising platform for the development of new vaccines and biologics. For many applications, secretion of designed nanoparticles from eukaryotic cells would be advantageous, but in practice, they often secrete poorly. Here we show that designed hydrophobic interfaces that drive nanoparticle assembly are often predicted to form cryptic transmembrane domains, suggesting that interaction with the membrane insertion machinery could limit efficient secretion. We develop a general computational protocol, the Degreaser, to design away cryptic transmembrane domains without sacrificing protein stability. The retroactive application of the Degreaser to previously designed nanoparticle components and nanoparticles considerably improves secretion, and modular integration of the Degreaser into design pipelines results in new nanoparticles that secrete as robustly as naturally occurring protein assemblies. Both the Degreaser protocol and the nanoparticles we describe may be broadly useful in biotechnological applications.

biochemistry | protein design | nanoparticles

Significance

Most biopharmaceuticals traverse the eukaryotic secretory pathway during their production as either recombinant proteins or as mRNA-delivered products. As designed proteins more commonly form the basis of new medicines, methods for optimizing their secretion will become increasingly important. Here we present a general computational method for improving protein secretion that focuses on the elimination of cryptic transmembrane domains while maintaining overall protein stability. We show that the method can retrospectively and prospectively improve the secretion of computationally designed protein nanoparticles, a promising platform for the development of new vaccines and therapeutics.

Secreted proteins make up nearly 20% of the human proteome and are the primary medium for intercellular communication in animals (1, 2). Due to their potent and wide-ranging functions, many secreted proteins such as antibodies, hormones, cytokines, and growth factors are of great interest for therapeutic applications. Furthermore, secreted and membrane-anchored proteins from pathogens are common targets for prophylactic or therapeutic interventions in infectious disease. Secretion from eukaryotic cells is required for the recombinant production of many protein biologics, as they often feature secretory pathway-specific post-translational modifications such as proteolytic cleavage (3), glycosylation (4), or disulfide bond formation (5). Understanding and controlling the secretion of a protein of interest is thus mandatory for optimal development of secreted protein technologies.

Several strategies for increasing the yield of secreted proteins have been established. Most approaches have focused on engineering or adapting cell lines for secretion. Chinese hamster ovary (CHO) cells are by far the most widely used for the production of antibodies and other secreted biologics, accounting for 84% of approved monoclonal biopharmaceuticals from 2015 to 2018 (6). Genomic, transcriptomic, and proteomic profiling of this cell line has yielded insights into the adaptations that confer its ability to grow at high density and robustly secrete target proteins and has also identified opportunities for intentional improvements through cell engineering (7–10). For example, overexpression of secretory chaperones or other secretory pathway factors can significantly increase the secreted yield of target proteins (11, 12). Engineering other eukaryotic expression hosts, such as *S. cerevisiae* or *P. pastoris*, has provided further avenues for recombinant expression (13). However, host-directed efforts like these are agnostic to the sequence of the target protein and therefore do not take into account sequence-specific factors that could affect secretion.

An alternative approach that is more customizable and amenable to downstream biological applications, in particular when the host cell cannot be modified, is engineering the sequences of protein biologics themselves. For example, computational or experimental evolution of signal peptides in both yeast and mammalian cells has yielded portable sequences that maximize secretion (14–16). The introduction (17, 18) or removal (19) of potential N-linked glycosylation sites has also been shown to substantially affect protein secretion. Importantly, sequence optimization via rational design or evolutionary methods can maximize production of a protein while still preserving its biological function, such

Competing Interest Statement: N.P.K. is a cofounder, shareholder, paid consultant, and chair of the scientific advisory board of Icosavax, Inc. The King lab has received unrelated sponsored research agreements from Pfizer and GSK. The authors have patent filings to disclose: J.Y.J.W., A.K., U.N., D.E., C.W., Y.H., D.B., W.S., and N.P.K. are coinventors on patent applications related to this work. All other authors declare no competing interests.

This article is a PNAS Direct Submission. D.W.K. is a guest editor invited by the Editorial Board.

Copyright © 2023 the Author(s). Published by PNAS. This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

[1]J.Y.W. and A.K. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: neilking@uw.edu. or neil@ipd.uw.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2214556120/-/DCSupplemental.

Published March 8, 2023.

Downloaded from https://www.pnas.org by "UNIVERSITY OF WASHINGTON LIBRARIES, ARCS - SERIALS" on October 20, 2025 from IP address 205.175.118.15.

as immunogenicity or receptor binding (20, 21). Although powerful, these methods tend to focus on the maximization of secreted yield for single proteins and can be laborious, underscoring the need for general methods applicable to broad classes of secreted proteins. To devise such general methods, it is necessary to understand the effects of common sequence elements on protein secretion and how they can be modulated through design.

Protein translocation across the endoplasmic reticulum (ER) membrane is a key step in the traversal of proteins through the eukaryotic secretory pathway. Transmembrane, secreted, and ER/Golgi-resident proteins typically contain an N-terminal signal peptide that is recognized by the signal recognition particle (SRP) (22–24). The SRP–polypeptide–ribosome complex is then targeted to the ER membrane, across which the nascent polypeptide is translocated in a mostly unfolded state through a protein channel, the Sec translocon. The translocon complex senses the hydrophobicity of segments within a polypeptide to determine its fate: highly hydrophobic segments are partitioned into the ER membrane, while segments of low hydrophobicity are translocated into the ER lumen (25). To predict this biological hydrophobicity sensing, Hessa et al. empirically determined the position-specific amino acid contribution of each residue in a model polypeptide segment to transmembrane insertion potential (dG_{ins}) and generated a theoretical model that can, using sequence alone,

predict the dG_{ins} of any given polypeptide segment, $dG_{ins,pred}$ (26, 27). Unsurprisingly, the model predicts that transmembrane proteins contain segments of low $dG_{ins,pred}$, while secretory proteins are generally devoid of such segments. Interestingly, many proteins that are typically expressed in the prokaryotic or eukaryotic cytoplasm also contain segments of low $dG_{ins,pred}$, as they have not evolved under selective pressure to avoid them. We hypothesize that, when expressed recombinantly, the Sec translocon may interpret segments of low $dG_{ins,pred}$ within proteins targeted for secretion as transmembrane domains, leading to inefficient secretion.

Protein nanoparticles, which in nature evolved to function as containers, reaction chambers, and multivalent display platforms (28–30), have long been a promising biotechnological platform (31–33). Virus-like particles (VLPs) and other naturally occurring protein nanoparticles such as ferritin and lumazine synthase have been engineered or evolved to encapsulate a number of different cargoes, such as inorganic particles, polymers, micelles, enzymes, DNA, and small molecules (34–36). Self-assembling protein complexes have also been used as scaffolds for nanoparticle vaccine design, an application that often requires secretion from eukaryotic cells to obtain natively folded antigens with appropriate post-translational modifications (20, 37, 38). In the last few years, new and accurate computational methods have made possible the design of self-assembling proteins with customized structures (39–44). These methods allow rational exploration of structural and functional space beyond the limited set of protein nanoparticle architectures sampled during evolution (45). Additionally, because they are explicitly designed to adopt low-energy states, designed protein nanoparticles are often hyperstable, a property that can enhance the stability of proteins fused to them (46–48). The recent entry of multiple vaccines based on computationally designed nanoparticles into clinical trials, and the regulatory approval of one of them for COVID-19 (49), highlights the potential technological impact of this promising class of proteins (47, 50, 51).

Naturally occurring and designed protein nanoparticles alike use hydrophobic interactions both in the folding of their subunits as well as in the interactions between subunits that drive supramolecular assembly (40, 52). One approach to protein nanoparticle design computationally docks symmetric protein building blocks into a desired geometry and then designs protein–protein interfaces between them (39, 42). This design strategy has been previously shown to generate robust, well-expressed, stable protein nanoparticles constructed from subunits or pairs of subunits that are most often expressed in the cytoplasm of bacterial cells (39, 41–43, 53, 54). Nonetheless, only a small subset of the components of these designed nanoparticles have been robustly secreted from eukaryotic cells (47, 50, 51).

Here, we develop a method to improve the secretion of computationally designed protein nanoparticles. First, we identify a correlation between the presence of cryptic transmembrane domains and low levels of secreted protein. We then develop a general computational protocol, named the Degreaser, that specifically designs away cryptic transmembrane domains without sacrificing overall structural stability. We demonstrate the ability of the Degreaser to not only retroactively improve the secreted yield of existing nanoparticles and nanoparticle components, but also to avoid the introduction of cryptic transmembrane domains during the design of a new set of robustly secreting designed protein nanoparticles.

Results

The Degreaser. During the course of our work designing self-assembling protein nanomaterials and using them as scaffolds for heterologous antigen display, we have observed that many [but not all (50, 51)] designed nanoparticles and nanoparticle components secrete from eukaryotic cells at very low levels. We hypothesized that their inefficient secretion may derive from the introduction of long contiguous stretches of hydrophobic amino acids during nanoparticle interface design. These stretches could be interpreted by the Sec translocon as cryptic transmembrane domains, leading to inefficient translocation across the ER membrane and poor protein secretion. Using the model of Hessa et al. (26, 27) and a sliding window of 19 amino acids, we compared the distributions of lowest $dG_{ins,pred}$ per sequence of 489 computationally designed two-component icosahedral nanoparticle scaffolds (53) to those of 309 transmembrane and 585 nontransmembrane proteins in the RCSB Protein Databank (Fig. 1*A*) (55, 56). We identified a $dG_{ins,pred}$ threshold of +2.7 kcal/mol that maximally discriminated between natural transmembrane and nontransmembrane proteins (84.7% and 15.6% < +2.7 kcal/mol, respectively). 61.3% of the protein nanoparticle components contained segments below this threshold, indicating that the designed proteins differ substantially from naturally occurring nontransmembrane proteins in their likelihood to include hydrophobic segments. As a preliminary experimental test of whether cryptic transmembrane domains affect protein secretion, we analyzed the expression of a set of 480 de novo-designed peptide-binding proteins by yeast surface display (57). To reach the cell surface, these proteins must be translocated across the ER membrane and into the lumen. We found that the presence of segments of low $dG_{ins,pred}$ strongly hindered surface display, while proteins lacking such cryptic transmembrane domains were generally efficiently exported to the cell surface (*SI Appendix*, Fig. S1).

The inspection of individual designed nanoparticle components highlighted the impact that interface design can have on $dG_{ins,pred}$. For example, docking naturally occurring pentameric (PDB 2jfb) and computationally designed trimeric (PDB 5hrz) building blocks and designing them to form an icosahedral nanoparticle [I53_dn5; (58)] resulted in two hydrophilic to hydrophobic substitutions in a single helix of the pentamer (Fig. 1*B*). As a result, the $dG_{ins,pred}$ at index 14 (residues 14 to 32) of the pentamer (I53_dn5A) was

Downloaded from https://www.pnas.org by "UNIVERSITY OF WASHINGTON LIBRARIES, ARCS - SERIALS" on October 20, 2025 from IP address 205.175.118.15.

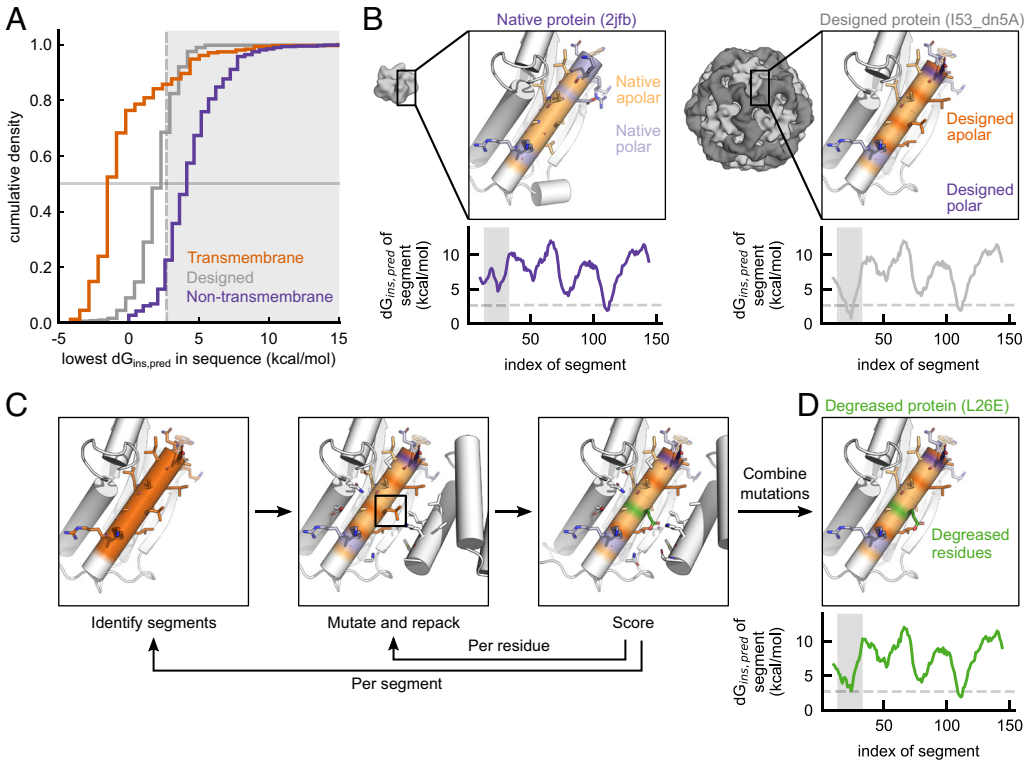

Fig. 1. Graphical depiction of the degreaser protocol. (*A*) Designed proteins have segments of lower dG$_{ins,pred}$ than do nontransmembrane proteins, but higher than those of transmembrane proteins. The shaded region indicates dG$_{ins,pred}$ > +2.7 kcal/mol. (*B*) *Top*, Protein building blocks have surface polar residues (light and dark purple) that are mutated to apolar residues (light and dark orange) during nanoparticle interface design. Alpha helices are depicted as cylinders, and side chains of residues within the identified segment or within 8 A of the designed interface are shown as sticks. *Bottom,* These mutations often decrease the lowest dG$_{ins,pred}$ segment in the sequence (the lowest region of designed protein shaded gray, horizontal dashed line at +2.7 kcal/mol). (*C*) The Degreaser iteratively identifies hydrophobic segments, perturbs each residue within such segments through point mutations, and chooses the optimal variant after neighborhood repacking. (*D*) Degreased proteins feature one or more polar mutations (green) that preserve the originally designed interface while increasing dG$_{ins,pred}$.

reduced from +4.71 kcal/mol to +0.779 kcal/mol, suggesting that this region of the protein may be interpreted as a cryptic transmembrane domain during translocation. Quantification of secreted protein by western blot revealed a nearly 100-fold reduction in secretion of I53_dn5A relative to its naturally occurring counterpart (Fig. 2*A* and *SI Appendix*, Fig. S2). The inspection of several other previously designed nanoparticle components that secrete with varying efficiency from eukaryotic cells supported the relationship between the presence of cryptic transmembrane domains and poor secretion (*SI Appendix*, Fig. S2). In each poorly secreted case, cryptic transmembrane domains resulted from the introduction of a few surface-exposed hydrophobic residues at the designed interface that, combined with existing hydrophobic residues involved in packing within the protein itself, form segments of low transmembrane insertion potential. Together, these observations suggest that the secretion of designed proteins from eukaryotic cells, a prerequisite for several biotechnological applications, may be limited by the presence of cryptic transmembrane domains introduced during design.

To address this issue, we developed the Degreaser, a new design protocol in the Rosetta macromolecular modeling software suite (59, 60) that detects and eliminates cryptic transmembrane domains without disrupting protein structure or stability. We had three primary goals for the protocol: 1) accurate identification of regions of low dG$_{ins,pred}$ and their improvement through minimal sequence perturbations, 2) preservation of the overall integrity of the protein's structure, and 3) compatibility with large-scale protein design protocols. The Degreaser uses an internal implementation of the model of Hessa et al. (26, 27) to identify 19-residue segments of input protein structures that have dG$_{ins,pred}$ < 3.5 kcal/mol. Each residue in these segments is individually mutated to polar amino acids selected from a user-specified set, the mutated residue and nearby residues in three-dimensional space are repacked to attempt to accommodate the mutation, and the overall Rosetta score and dG$_{ins,pred}$ are re-evaluated (Fig. 1*C*). All chains in the input structure are modeled throughout the process, including symmetry-related subunits when appropriate, to capture the potential energetic effects of each mutation in the correct structural context. Mutations are discarded if they fail to increase dG$_{ins,pred}$ by a user-specified amount or are energetically unfavorable (i.e., result in an increase in Rosetta score beyond a user-specified threshold), usually due to steric clashes. Designs that pass these filters (for details, see *SI Appendix, Materials and Methods*) are output as potential candidates for experimental evaluation. Typically, the Degreaser identifies apolar to polar mutations that increase dG$_{ins,pred}$; however, some polar to polar mutations still increase dG$_{ins,pred}$ due to, for example, favorable interactions of positively charged residues near segment boundaries with the lipid membrane. We found that a small number of mutations was often sufficient to substantially raise the dG$_{ins,pred}$ of the segment of interest. For example, the dG$_{ins,pred}$ of residues 14 to 32 in the I53_dn5 pentamer increased from +0.779 to +2.812 kcal/mol with only one mutation to a charged amino acid (Fig. 1*D*).

Improving the Secretion Yield of Existing Designed Protein Nanoparticle Components. As a first test of our protocol, several previously designed two-component nanoparticle proteins that were originally expressed in bacteria were screened for secretion from Expi293F cells (53, 58). Three had no predicted segments of low dG$_{ins,pred}$, needing no Degreaser application, and secreted

Downloaded from https://www.pnas.org by "UNIVERSITY OF WASHINGTON LIBRARIES, ARCS - SERIALS" on October 20, 2025 from IP address 205.175.118.15.

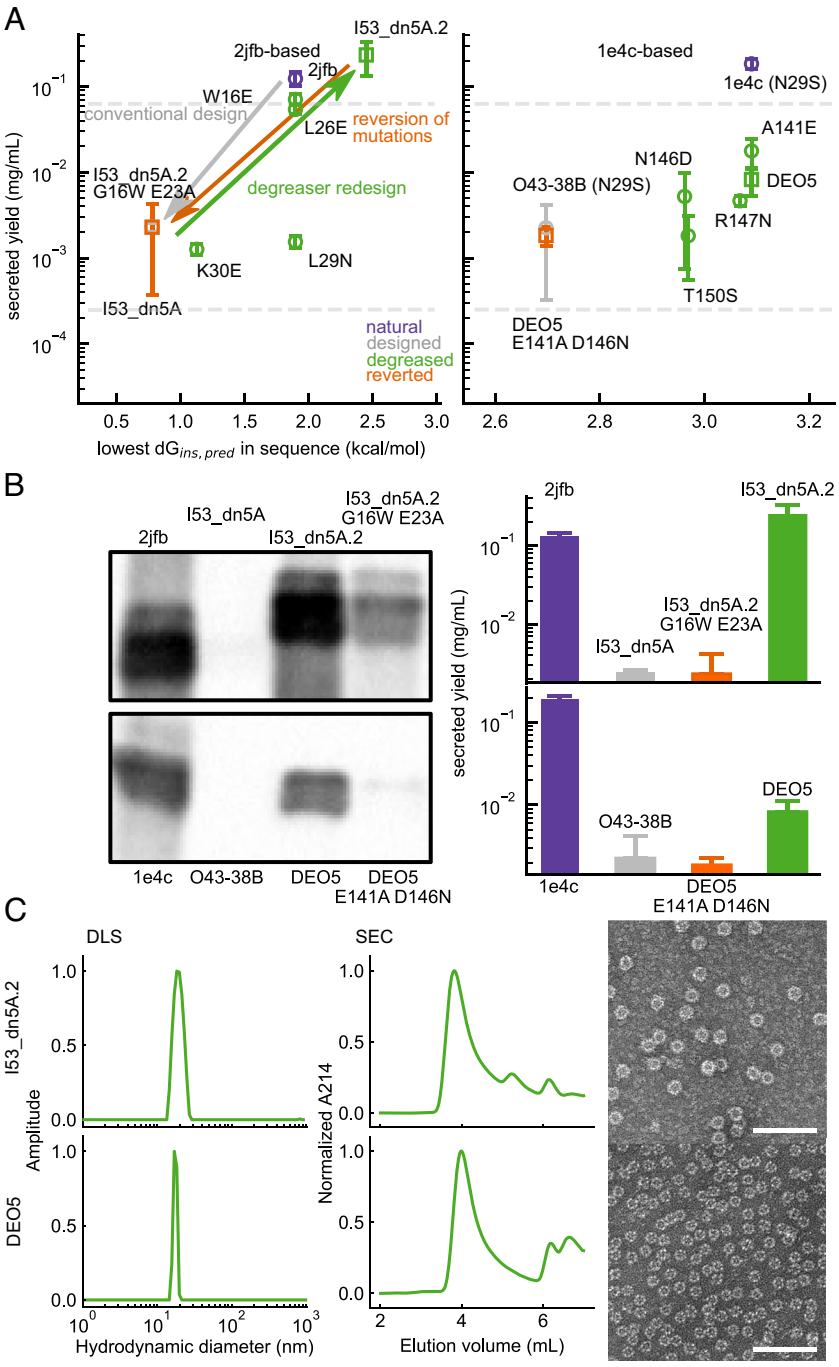

Fig. 2. Retroactive degreasing of protein nanoparticle components improves secretion yield. (*A*) To assess secreted yield of previously designed proteins and their natural counterparts, material from transiently transfected cell cultures was fractionated and analyzed by Western blot. Horizontal dashed lines represent the upper and lower limits of the linear range of a standard curve generated by densitometry. Square markers represent variants that incorporate mutations from parallel redesign for improvement of other phenotypes. (*B*) Representative Western blots of cell culture supernatants of two design trajectories (naturally occurring scaffold, purple; original design, gray; Degreased redesign, green; reversion of Degreaser mutations, orange). (*C*) Characterization of in vitro assemblies of secreted I53_dn5A.2 (*Top* row) and DEO5 (*Bottom* row) with their respective second components I53_dn5B and O43-38A, purified from bacterial cells, reveal nanoparticle sizes and morphologies indistinguishable to those of bacterially-produced I53_dn5 (58) and O43-38 (*SI Appendix*, Fig. S3), respectively. (Scale bar, 100 nm.)

with yields between 3 and 34 μg/mL of cell culture supernatant (*SI Appendix*, Fig. S2). However, seven others yielded between 0.6 and 3 μg/mL, precluding purification and in vitro assembly experiments. In two of the latter cases, the Degreaser was applied to create several single-residue variants intended to improve secretion of I53_dn5A, the pentameric component of the icosahedral nanoparticle I53_dn5 (58), and O43-38B, the tetrameric component of a designed octahedral nanoparticle, O43-38 (*SI Appendix*, Fig. S3*A*), respectively. The two naturally occurring proteins from which these nanoparticle components were derived (including an N29S mutation in 1e4c to knock out a potential N-linked glycosylation site) both secreted at high levels, around 100 μg/mL (Fig. 2*A*). The Degreased variants were predicted to have no significant effects on protein stability, as indicated by the <5% change in Rosetta score compared to the input structures (*SI Appendix*, Fig. S4). In each case, variants with significant improvements in secretion yield were obtained. For example, two of four I53_dn5A pentamer variants tested,

Downloaded from https://www.pnas.org by "UNIVERSITY OF WASHINGTON LIBRARIES, ARCS - SERIALS" on October 20, 2025 from IP address 205.175.118.15.

W16E and L26E, increased $dG_{ins,pred}$ by more than 1.1 kcal/mol and improved secretion yield more than 20-fold (Fig. 2*A*; amino acid sequences of all proteins used in this study are provided in *SI Appendix*, Table S1). For the O43-38B tetramer, two single-residue variants, A141E and R147N, boosted secretion eightfold and twofold, respectively.

To assess the orthogonality and compatibility of the Degreaser with unrelated mutations, some nanoparticle components were further redesigned for other purposes. Because existing methods for designing protein stability, solubility, or expression do not explicitly penalize the introduction of contiguous hydrophobic segments, the variants generated were generally orthogonal to those of the Degreaser. In the case of I53_dn5A, we mutated a pair of cysteine residues retained from the naturally occurring pentameric scaffold to alanine to prevent the formation of off-target disulfide bonds. Ten additional nearby mutations and a three-residue deletion were also included to maintain protein stability. These unrelated mutations were $dG_{ins,pred}$-agnostic and did not substantially change the hydrophobicity of already-hydrophobic segments within the protein (*SI Appendix*, Fig. S5). In these contexts, the Degreaser mutations again positively impacted protein secretion without negatively impacting protein assembly: the stabilized O43-38B variant DEO5 (which incorporated two Degreaser mutations: A141E and N146D) and the I53_dn5A.2 variant (incorporating the A23E Degreaser mutation) had higher secretion yield than their originally designed counterparts (8.2 and 234 μg/mL, respectively; Fig. 2 *A* and *B*). To confirm that the secretion phenotype was dependent on the Degreaser-identified mutations, we reverted these positions to their originally designed identities, which resulted in substantial decreases in $dG_{ins,pred}$ and secretion yield as expected (1.8 and 2.3 μg/mL, respectively; orange data points and bars). Importantly, these nanoparticle components retained their ability to assemble in vitro (Fig. 2*C*), as validated by dynamic light scattering (DLS), size exclusion chromatography (SEC), and negative stain electron microscopy (nsEM) of in vitro assembly reactions containing their respective second components. Taken together, these results demonstrate that the Degreaser can be used to generate nanoparticle component variants with improved secretion that retain their ability to assemble to designed nanoparticle architectures.

Improving the Secretion Yield of Protein Nanoparticles by Degreasing Cryptic Transmembrane Domains. Encouraged by the ability of the Degreaser to improve secretion of the oligomeric components of designed two-component protein nanoparticles, we next evaluated whether it could also improve the secretion of one-component or homomeric nanoparticles that self-assemble during secretion. Several naturally occurring homomeric protein nanoparticles such as lumazine synthase and ferritin have been explored in biotechnological applications such as nanoparticle vaccine design and drug delivery (38, 61). We compared the secretion of these two naturally occurring protein nanoparticles to the computationally designed one-component icosahedral nanoparticle I3-01 (43) and found that they secrete at levels roughly tenfold and 100-fold higher, respectively (Fig. 3*A*). This observation is consistent with the lack of low $dG_{ins,pred}$ segments in lumazine synthase and ferritin, whereas I3-01 has two segments of low $dG_{ins,pred}$, one near the N terminus and a second near the C terminus, both of which include polar to hydrophobic mutations at the designed nanoparticle interface (Fig. 3*B*).

We used the Degreaser to identify substitutions that could increase the $dG_{ins,pred}$ of both the N- and C-terminal cryptic transmembrane domains of I3-01 (*SI Appendix*, Fig. S5). The single substitution that had the greatest effect on both $dG_{ins,pred}$ and secretion was H35D in the N-terminal region, though K25D also substantially increased secretion yield (Fig. 3*A*). A triple mutation in the C-terminal region (L171Q/S177E/V180N) also increased secreted yield. Combining mutations in both regions resulted in a quadruple mutant (H35D/L171Q/S177E/V180N) that secreted at tenfold higher levels than any of the tested single-residue variants. We again evaluated the utility of Degreaser mutations in the context of orthogonal mutations by introducing the four Degreaser-identified substitutions into a negatively "supercharged" I3-01 variant, I3-01-neg. While the I3-01-neg mutations did not affect the predicted transmembrane insertion potential of I3-01 (*SI Appendix*, Fig. S5), they did substantially improve overall protein expression (*SI Appendix*, Fig. S11*B*), which may explain our observation of improved secreted yield for this variant even in the absence of Degreaser mutations (Fig. 3*A*). Nevertheless, incorporation of the four Degreaser-identified mutations into I3-01-neg further improved its secretion, resulting in a 64-fold increase in secreted yield compared to I3-01 (Fig. 3*C*). This variant, I3-01NS, retained its ability to assemble to the known icosahedral architecture of I3-01: SEC, DLS, and nsEM all indicated monodisperse particles that were indistinguishable from previously published I3-01 data [Fig. 3*D*; (43)]. These results establish that elimination of cryptic transmembrane domains through sequence redesign can be used to improve the secretion of homomeric self-assembling protein nanoparticles, including the I3-01 nanoparticle that has been used as a hyperstable scaffold in several biotechnological applications (46, 48, 62, 63).

De Novo Design of Secretion-Optimized One-Component Protein Assemblies. Given the success of the Degreaser in retroactively improving the secretion of several nanoparticle components and a computationally designed nanoparticle, we next tested its prospective use and compatibility with large-scale design protocols by incorporating it into the design of a set of new one-component nanoparticles intended to secrete robustly from mammalian cells. We used as building blocks a set of 1,094 models of trimeric proteins consisting of de novo helical bundles fused to designed helical repeat proteins as previously described (44). These building blocks were docked as rigid bodies into three target architectures containing three-fold symmetry axes: icosahedral (I3), octahedral (O3), and tetrahedral (T3) (Fig. 4*A*). After docking, residues at interfaces with adjacent building blocks were designed using Rosetta to enable spontaneous self-assembly to the target architecture. Three fully automated design protocols were compared: OG, ND, and DG (Fig. 4*A*). The OG protocol used a conventional, $dG_{ins,pred}$-agnostic protocol and therefore generated designs that had $dG_{ins,pred}$ values both above and below +2.7 kcal/mol. The ND protocol simply applied a post-design filter to the OG design set that rejected any designs with $dG_{ins,pred}$ less than +2.7 kcal/mol. Finally, the DG protocol incorporated the Degreaser after the interface design step and also filtered out any designs with $dG_{ins,pred}$ less than +2.7 kcal/mol. In this benchmark design set, we used the conservative approach of allowing the Degreaser to change at maximum one residue per design, focusing on decreasing the hydrophobicity of the lowest $dG_{ins,pred}$ segment. It is important to note that not all DG designs harbor a Degreaser mutation, as designs that do not have segments of low $dG_{ins,pred}$ pass the Degreaser step and are accepted without modification.

The incorporation of Degreaser-guided design into an otherwise conventional design protocol did not substantially perturb the structural metrics typically used to gauge the quality of designed nanoparticle interfaces. Within the DG design set, 420 of the 1,048 designs (40%) were actually mutated by the Degreaser,

Downloaded from https://www.pnas.org by "UNIVERSITY OF WASHINGTON LIBRARIES, ARCS - SERIALS" on October 20, 2025 from IP address 205.175.118.15.

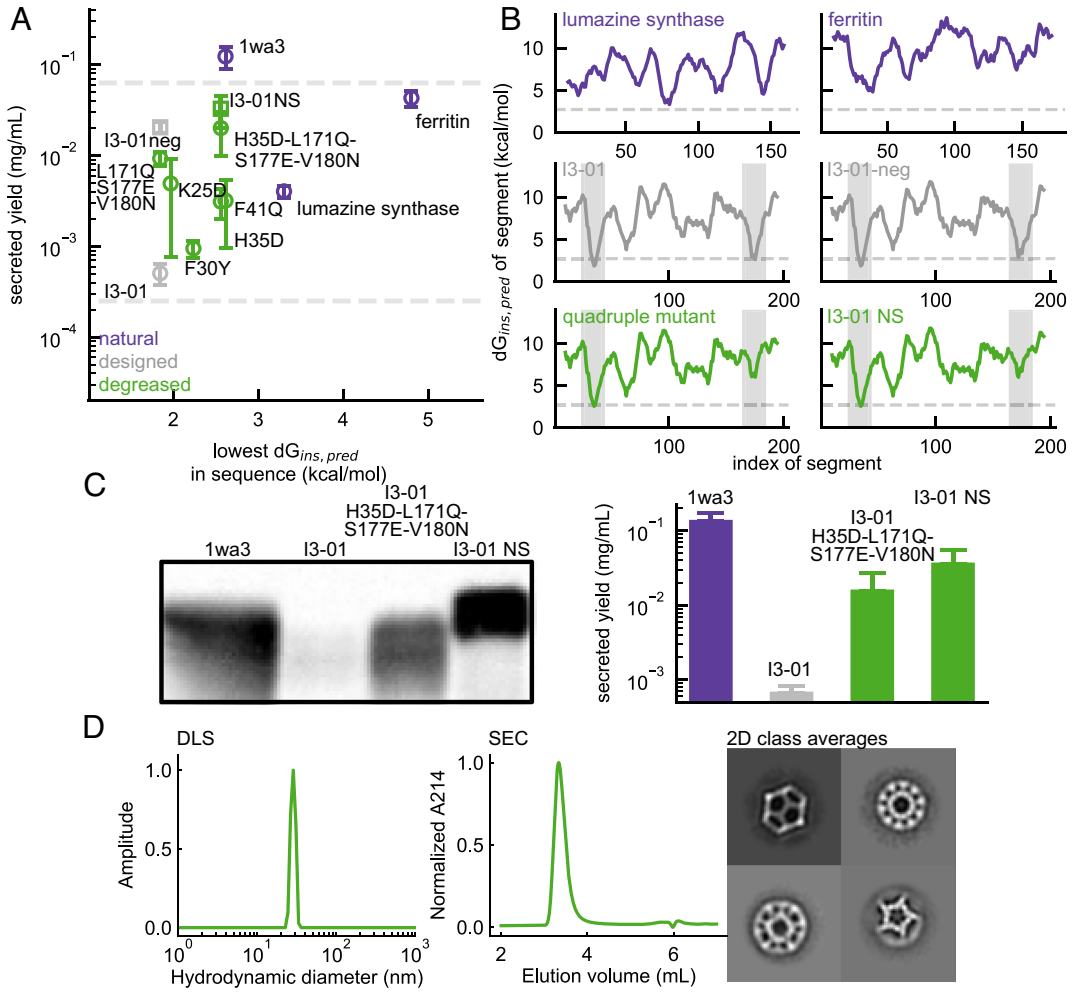

Fig. 3. Retroactive degreasing of a designed protein nanoparticle improves secretion yield. (*A*) I3-01 can be Degreased to boost secretion yield; square markers represent variants that incorporate mutations from orthogonal redesign for improvement of other phenotypes. (*B*) Comparison of the $dG_{ins,pred}$ per segment of wild-type (gray) and best-secreted (green) variants of I3-01 with secreted natural protein assemblies (purple). The lowest regions of wild-type I3-01 are shaded in gray, and the horizontal dashed line denotes the +2.7 kcal/mol threshold. (*C*) *Left*, representative western blot of individual samples across I3-01 designs. *Right*, secreted yield quantification of the same series measured in triplicate. (*D*) Characterization of the Degreased I3-01NS purified from mammalian cell supernatants by DLS, SEC, and nsEM with 2D class averaging confirmed its structure is identical to that of I3-01 expressed in bacteria [Hsia et al. (43)].

while mutations meeting the Degreaser criteria were not identified for 18 designs which were therefore rejected. Notably, there was an increase in $dG_{ins,pred}$ of the Degreaser-guided design models before any sequences were filtered on $dG_{ins,pre}d$ (47% > +2.7 kcal/mol compared to 36% for the OG design models). After filtering, DG designs that were not mutated by the Degreaser had an average $dG_{ins,pred}$ of +3.97 kcal/mol, while those bearing mutations had an average $dG_{ins,pred}$ of +3.38 kcal/mol. Because sequences with originally high $dG_{ins,pred}$ are not mutated, the lower average $dG_{ins,pred}$ of sequences with mutations is due to the low original dGins,pred of those sequences. The Degreaser mutations led to an average increase in $dG_{ins,pred}$ of 1.41 kcal/mol but no significant differences in other design metrics (*SI Appendix*, Fig. S6*A*). A slight shift in the distribution of ddG (the Rosetta-predicted energy of interface formation) was observed, which was expected due to Degreaser-introduced polar residues in what remained predominantly hydrophobic interfaces, accompanied by a small shift in the distribution of the interface shape complementarity [Sc; (64)]. On the other hand, the solvent-accessible surface area buried at the interface (sasa) showed a nearly identical distribution to that of conventional (OG) and $dG_{ins,pred}$-filtered (ND) designs. After filtering on several structural metrics and visual inspection of the top-scoring designs by ddG, we selected 99 KWOCAs (Khmelinskaia-Wang One-Component Assemblies) for experimental characterization. These included 57 OG, 19 ND, and 23 DG designs, which differed in $dG_{ins,pred}$ but not other structural metrics (*SI Appendix*, Fig. S6*B*). Eight of the selected DG designs included mutations introduced by the Degreaser.

We first expressed the KWOCAs in the cytoplasm of *E. coli* to determine which ones successfully assembled to the intended structures regardless of secretion from mammalian cells. All KWOCAs but one yielded sufficient protein in the soluble fraction of clarified *E. coli* lysates for purification and characterization (Fig. 4*B*). During SEC purification, 22 of 99 KWOCAs yielded a peak with an elution volume corresponding to a protein complex larger than that expected for a trimer, but smaller than unbounded aggregates (*SI Appendix*, Fig. S7*A*). Most of these also showed peaks at elution volumes corresponding to unassembled trimeric protein. DLS of fractions from the early peaks indicated that the majority of these 22 designs formed monodisperse assemblies (*SI Appendix*, Fig. S7*B*), and nsEM confirmed that 13 assembled to homogeneous nanoparticle structures (*SI Appendix*, Fig. S8). The proportion of designs confirmed to adopt homogeneous structures was identical between conventionally designed proteins (OG+ND) and

Downloaded from https://www.pnas.org by "UNIVERSITY OF WASHINGTON LIBRARIES, ARCS - SERIALS" on October 20, 2025 from IP address 205.175.118.15.

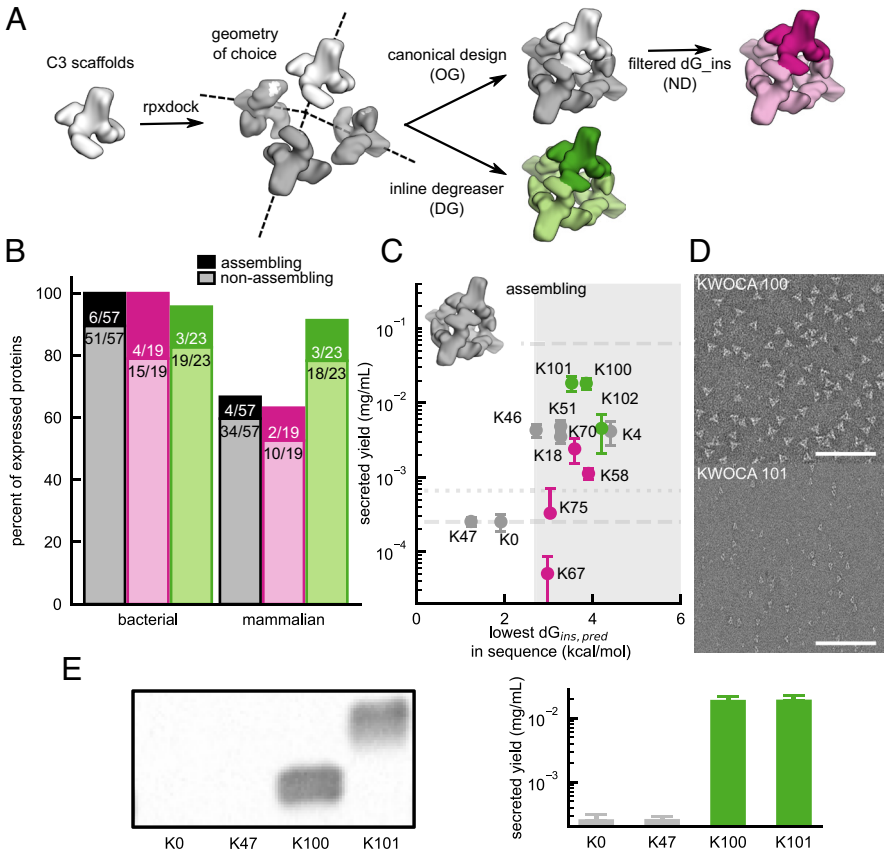

Fig. 4. Incorporation of the degreaser prospectively during design to generate de novo designed secreted protein assemblies. (*A*) Trimeric building blocks were docked into a desired geometry: tetrahedral, octahedral, or icosahedral. For the KWOCAs, designs were run independently for DG and OG sets, while ND designs were selected from a filtered subset of all OG designs. (*B*) Expression and secretion characterization of KWOCAs shows the benefit of the Degreaser on secreted yield (positive expression in mammalian cells determined as greater secretion than I3-01). Assemblies validated by nsEM are highlighted in darker color. (*C*) nsEM-verified assembling (*SI Appendix*, Fig. S9) secreted proteins partitioned into OG, ND, and DG groups show the enhanced secreted yield of DG designs. (*D*) Constructs purified from mammalian material assembled into well-defined particles, indistinguishable from those expressed in bacteria (*SI Appendix*, Fig. S9). (Scale bar, 100 nm.) (*E*) *Left*, representative western blot of KWOCAs with lowest $dG_{ins,pred}$ and secretion yield (K0 and K47) and highest secretion yield (K100 and K101). *Right*, quantification of secreted yield, measured in triplicate.

Degreaser-designed candidates [10/76 (13%) vs. 3/23 (13%), respectively; Fig. 4*B*]. Crystallization of several KWOCAs resulted in high-resolution structures of trimeric building blocks of one nonassembling KWOCA (KWOCA 39; *SI Appendix*, Fig. S9) and four assembling KWOCAS, including one confirmed by nsEM (KWOCAs 60, 65, 73, and 102) (*SI Appendix*, Fig. S10*A*). The crystal structures matched closely the monomeric subunit of the design models (1.0-2.2 A Ca rmsd), with larger deviations observed at the level of the trimers (2.2-4.0 A Ca rmsd). The differences suggest that multiple aspects of the designed proteins, such as the helical bundle interface, flexibility within the designed helical repeat domains, and flexibility at the junction of the helical repeat domains and trimeric helical bundles, contribute subtle structural deviations that propagate across the trimer, which may prevent formation of the target architecture in the crystal.

We next evaluated secretion of the KWOCAs from transfected Expi293F cells by measuring the levels of myc-tagged protein in clarified harvest fluid by western blot (*SI Appendix*, *Methods*). A majority of the KWOCAs (72%) secreted with greater yield than I3-01, our benchmark modestly secreted protein nanoparticle (Fig. 4 *B* and *C*). The apparent higher success in secretion within the DG KWOCA set (87% compared to 67% and 63% in the OG and ND sets, respectively) supports the utility of the Degreaser in predicting and improving protein secretion. For further analysis, we separated the experimentally characterized KWOCAs into two categories: nonassembling proteins (*SI Appendix*, Figs. S9 and S11) and confirmed assemblies (*SI Appendix*, Figs. S8, S10, and S11). The nonassembling proteins show only a weak trend of higher secretion yield with higher $dG_{ins,pred}$ (*SI Appendix*, Fig. S11*A*), though this is confounded by the varying overall expression levels of these proteins (*SI Appendix*, Fig. S11*B*). Although we did not observe consistent differences in secretion yield among the OG, ND, and DG sets of nonassembling proteins, inline application of the Degreaser (DG) tended to provide a greater benefit to secreted yield than filtering on $dG_{ins,pred}$ after conventional design (ND) (*SI Appendix*, Fig. S11*A*). Of the 13 EM-validated assembling designs (*SI Appendix*, Fig. S8), nine secrete at higher levels than the original I3-01 design (Fig. 4*C*), and the highest secreted yield (KWOCA 101) was within twofold of the highest secreting redesigned variant of I3-01 (I3-01NS; Fig. 3*C*). The characterization of eight of these KWOCAs by SEC, DLS, and nsEM revealed that each was indistinguishable from its bacterially produced counterpart (Fig. 4*D* and *SI Appendix*, Figs. S7 and S8). Only two of the verified assemblies (KWOCAs 0 and 47, from the OG design set) have segments of low $dG_{ins,pred}$ (<+2.7 kcal/mol; Fig. 4*C*). Those two assemblies were secreted at very low levels, while inline application of the Degreaser led to the highest yield of secreted nanoparticles (KWOCAs 100 and 101; Fig. 4*E*). Together, these data indicate that the Degreaser can be applied to improve secreted yield from mammalian cells while maintaining a similar success rate in design outcome.

Downloaded from https://www.pnas.org by "UNIVERSITY OF WASHINGTON LIBRARIES, ARCS - SERIALS" on October 20, 2025 from IP address 205.175.118.15.

Comparison of several pairs of closely related designs yielded additional insights into secretion determinants. For example, KWOCA 51 and 101, which form closely related tetrahedral assemblies, used the same input scaffold for design and differ by only two residues. However, KWOCA 101 has a higher lowest $dG_{ins,pred}$ and

secreted with a roughly fourfold greater yield than KWOCA 51 (Fig. 4*C*), highlighting how small changes in protein sequence can lead to considerable changes in secreted yield. Also related are the Degreased KWOCA 100 and the conventionally designed KWOCA 46, both confirmed assemblies (*SI Appendix*, Fig. S9), in which a one-residue difference led to a +1.13 kcal/mol change in $dG_{ins,pred}$ and a fivefold increase in secretion. In both of these cases, the conventional design pipeline resulted in assemblies that secrete poorly and would require retrospective application of the Degreaser. Two other pairs of designs suggested that assembly state may affect secreted yield. The nonassembling KWOCA 88 differs from the octahedral KWOCA 4 by only two residues, but KWOCA 88 secretes with a roughly fourfold higher yield (Fig. 4*C* and *SI Appendix*, Fig. S11*a*). Finally, even though KWOCA 73 differs from KWOCA 41 by only two residues, the former showed higher-order material by SEC and DLS, whereas the latter did not (*SI Appendix*, Fig. S8), and KWOCA 73 secretes at about half the yield of KWOCA 41 even though its lowest dGins,pred value is much higher. Thus, although there appears to be a general secretion penalty for self-assembling proteins, these data further support that in-line use of the Degreaser during design can improve secreted nanoparticle yield.

We obtained single-particle cryo-EM density maps of two highly secreted assemblies, KWOCA 51 and KWOCA 4, to evaluate our design protocol in more detail. DLS and SEC indicated that both designs assemble into monodisperse nanoparticles, with KWOCA 51 forming a smaller particle than KWOCA 4 (~19 and 26 nm hydrodynamic diameter, respectively) as expected by design (~17 and 32 nm, respectively) (Fig. 5*A*). Comparing calculated to experimental SAXS profiles further revealed that KWOCA 51 homogeneously assembles into the intended tetrahedral geometry, while KWOCA 4 significantly deviates from the design model (Fig. 5*B*). Indeed, a single-particle cryo-EM density map of KWOCA 51 at 5.1 A resolution closely matched the design model, and relaxing the model into the density led to only minor deviations within each subunit that mainly reflect slight structural flexibility of the helical repeat domain (Fig. 5*C* and *SI Appendix*, Fig. S10*B*). In contrast, a cryo-EM map of KWOCA 4 at 6.6 A resolution revealed that the protein does not form the computationally designed icosahedral assembly, instead identifying an octahedral nanoparticle as the only species present in the assembly fraction from SEC (Fig. 5*D* and *SI Appendix*, Fig. S10 *B* and *C*). Accordingly, a SAXS profile calculated from a cryo-EM model obtained by fitting and relaxing trimeric building blocks into the density closely matched the experimental data (*SI Appendix*, Fig. S10*D*). Interestingly, only minor structural deviations within the trimeric building blocks were observed when comparing the computational design model to the relaxed cryo-EM model, indicating that the off-target assembly must be due to differences in the computationally designed interface between the trimers. Indeed, the angle between two contiguous subunits in the cryo-EM model is rotated by 27°, resulting in a deviation of 18 A of the contiguous subunit compared to the design model (Fig. 5*E* and *SI Appendix*, Fig. S10*C*). This rotation is further accompanied by a 3 A transverse translation of the center of mass of the designed interface past the C2 symmetry

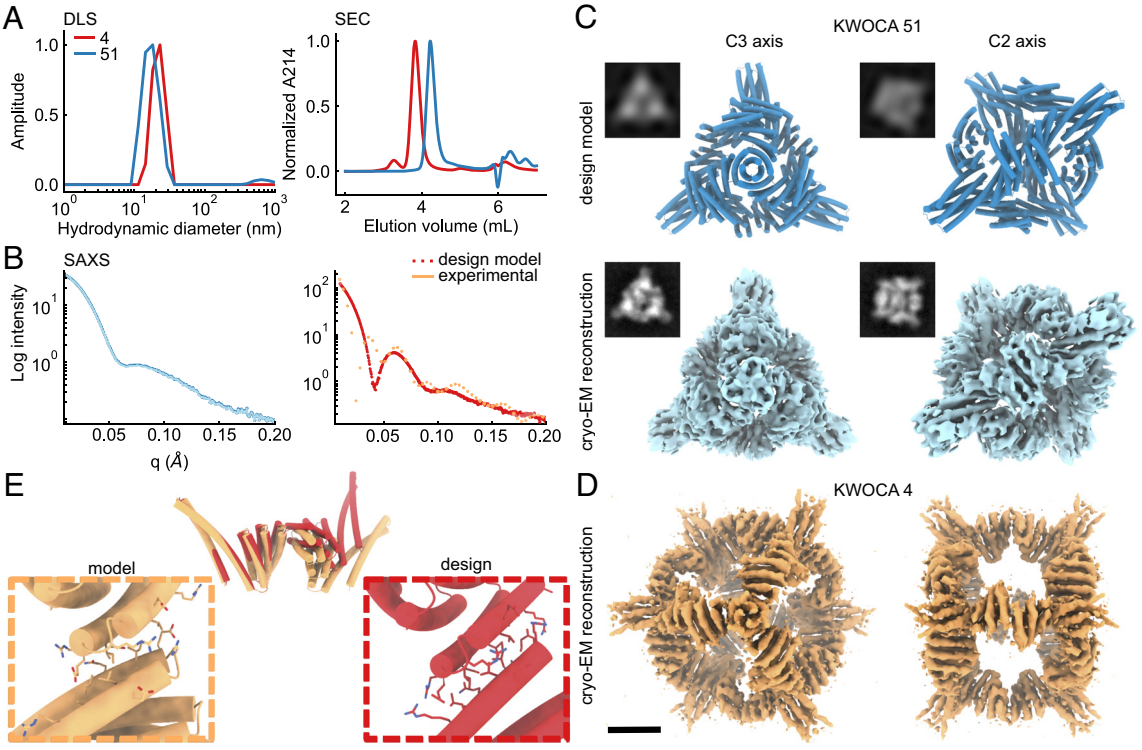

Fig. 5. Structural characterization of KWOCA 4 and KWOCA 51. (*A*) DLS, SEC traces and (*B*) SAXS profiles of KWOCA 51 (blue) and KWOCA 4 (red/orange). (*C*) Design model and cryo-EM density map of KWOCA 51. (*D*) Cryo-EM density map of KWOCA 4. (*E*) Overlay of two KWOCA 4 subunits across the designed nanoparticle interface, highlighting the interface contact angle difference between the design model and the best-fitting cryo-EM model. Theoretical SAXS profiles calculated from the design models (dotted darker lines) are overlaid with the experimentally obtained SAXS profiles (*B*). The fits to the theoretical SAXS profiles calculated to the best-fitting cryo-EM models can be found in *SI Appendix*, Fig. S10*D*. (Scale bar, 5 nm.) (*C* and *D*).

Downloaded from https://www.pnas.org by "UNIVERSITY OF WASHINGTON LIBRARIES, ARCS - SERIALS" on October 20, 2025 from IP address 205.175.118.15.

axis, suggesting that the residues on the periphery of the originally designed interface were loosely packed and only weakly contributing to the interface energy. To our knowledge, this is the first report of a de novo computationally designed protein nanoparticle that forms a well-defined architecture distinct from the one intended. Nevertheless, the two cryo-EM density maps together establish that KWOCAs 4 and 51, which secrete from mammalian culture at higher levels than lumazine synthase and I3-01, form monodisperse *de novo*-designed nanoparticles with some conformational flexibility, as judged by the relatively low resolutions of the reconstructed cryo-EM maps.

Discussion

Computational protein design methodologies are advancing rapidly, enabling access to previously unexplored spaces in protein structure and function (45, 65). In addition to increasing our fundamental understanding of proteins, these advances have made commercial application of computationally designed proteins a reality. For example, computationally designed cytokine mimetics (66), enzymes for gluten degradation (67), and nanoparticle vaccines (47, 51, 68) have recently advanced to clinical trials, and a designed nanoparticle vaccine for COVID-19 (50) was recently approved for use in South Korea. As designed proteins become increasingly useful, methods for optimizing various phenotypes other than structure and stability become more important. Recent examples include the ability to design heterodimers while maintaining the solubility of the individual components (69) and methods for modulating the brightness and chromophore specificity of de novo mini fluorescent proteins (70, 71). To enable their modular implementation in a wide variety of design applications, design methods like these must be able to optimize such phenotypes without sacrificing structural stability. The Degreaser was explicitly constructed to be modular-as showcased by our redesign of existing proteins as well as our application of the Degreaser in-line during the design of new secretable protein assemblies-while preserving structural stability and integrity. These features in principle enable its application to any protein. Furthermore, application of the Degreaser in-line during design is minimally invasive: It only mutates proteins that require elimination of cryptic transmembrane domains, and it identifies the minimal sufficient perturbation. As we showed during KWOCA design, this approach allows in-line implementation of the Degreaser that should eliminate the requirement for retroactive redesign of poorly secreting proteins.

More broadly, any method for improving the yield of recombinant biologics is valuable. For example, the decades of effort invested in optimizing and industrializing the production of monoclonal antibodies now underpins the biologics industry (72). Methods like the Degreaser that encode improved yield or performance in the sequence of the molecule itself are especially desirable, as they make the improvements "automatic": They do not require other actions like the use of specialized cell culture media or co-transfection of chaperones. Numerous examples of enzyme and antigen redesign to improve yield highlight the utility of this approach (19, 73–80). Methods that encode improvements genetically are of increasing importance now that genetic delivery of biologics has become clinical reality with the recent licensure of several AAV-based gene therapies and mRNA vaccines (81, 82). Furthermore, secreted protein nanoparticle immunogens are beginning to be explored as a strategy for improving the potency of mRNA vaccines (83–85). The Degreaser and the new highly secretable KWOCAs we describe here could pave the way towards the design of mRNA-launched nanoparticle vaccines with atomic level accuracy. This approach will enable structural and functional optimization of the nanoparticle scaffolds in ways that are not possible when relying on naturally occurring scaffolds. For example, we previously showed that the *de novo* design of nanoparticle scaffolds enabled precise optimization of termini placement for genetic fusion of functional domains (58).

Several features in our data suggest potential future directions for the Degreaser. First, to test the suitability of the Degreaser as part of a "set it and forget it" large-scale design protocol, we used the conservative strategy of only allowing one mutation maximum in the DG KWOCA designs. However, the results we obtained retroactively degreasing I3-01 and the individual scaffolds of two-component nanoparticles indicate that more aggressive use of the Degreaser (i.e., enabling multiple mutations) could lead to larger improvements in the secretion of specific proteins, and could lead to a higher success rate in designing highly secretable proteins while still preserving the intended structure and function. Second, although we have shown the presence of cryptic transmembrane domains significantly influences protein secretion, there are other factors that affect secretion and should be considered when designing secreted proteins (19, 73). For example, a recent high-throughput study of the secretion of protein fragments in yeast identified several sequence-based features that could be incorporated in future secretion models (86). Finally, here we applied the Degreaser to homomeric nanoparticles and the oligomeric components of two-component nanoparticles due to the many potential applications of highly secretable designed protein nanoparticles. Future application of the Degreaser to two-component nanoparticles, other designed proteins, and naturally occurring proteins that have not evolved to secrete at high levels could be used to improve the secretion of many different potential biologics and enable new applications of designed proteins.

Methods Summary

Details for methods can be found in the Supporting Information Appendix. The Degreaser was written in Rosetta-compatible C++, and design protocols that incorporated it were written with RosettaScripts. Designed protein nanoparticles and their components were expressed recombinantly as secreted constructs in HEK Expi293F cells to determine their secreted yields as well as for purification. For proteins with low secreted yields, constructs were also expressed with bacterial culture and subsequently purified. Biochemical characterization (DLS, nsEM) was carried out for constructs with assembly signal by preparative SEC. The characterization by cryo-EM was carried out for selected secreted assemblies.

Data, Materials, and Software Availability. Data deposition, atomic coordinates, and structure factors have been deposited in the Protein Data Bank, http://www.rcsb.org (PDB ID 8FBI, 8FBJ, 8FBK, 8FBN and 8FBO). Cryo-EM maps were deposited in the Electron Microscopy Data Bank (EMD-28862 and EMD-28929). All study data are included in the article and/or *SI Appendix*.

ACKNOWLEDGMENTS. We thank Bill Anderson, Hannah L. Turner, and Jean-Christophe Ducom (The Scripps Research Institute) for their assistance with electron microscopy experiments; Thomas Schlichthaerle for related fluorescence microscopy experiments, Ratika Krishnamurty for program management, and members of the King laboratory for comments on the manuscript. This work was supported by the Bill & Melinda Gates Foundation (OPP1156262 to D.B. and N.P.K., OPP1115782 to A.B.W.), a generous gift from the Audacious Project at the Institute for Protein Design, the NIH (P50 AI150464 and1P01AI167966 to N.P.K.), the Defense Threat Reduction Agency (HDTRA1-18-1-0001 to D.B. and N.P.K.) and the Collaboration for AIDS Vaccine Discovery (INV-002916 to A.B.W.).

Downloaded from https://www.pnas.org by "UNIVERSITY OF WASHINGTON LIBRARIES, ARCS - SERIALS" on October 20, 2025 from IP address 205.175.118.15.

A.A. is supported by amfAR Mathilde Krim Fellowship in Biomedical Research #110182-69-RKVA. U.N. was supported in part by a Public Health Service National Research Service Award (T32GM007270) from National Institute of General Medical Sciences. Small-angle X-ray scattering experiments were conducted at the Advanced Light Source (ALS) a national user facility operated by Lawrence Berkeley National Laboratory on behalf of the Department of Energy, Office of Basic Energy Sciences, through the Integrated Diffraction Analysis Technologies program, supported by Department of Energy Office of Biological and Environmental Research. Additional support comes from the NIH project ALS-ENABLE (P30 GM124169) and a High-End Instrumentation Grant S10OD018483. X-ray diffraction data were collected at the Advanced Photon Source Northeastern Collaborative Access Team beamlines, which are funded by the National Institute of General Medical Sciences from the NIH (P30 GM124165), and ALS beamline 8.2.2/8.2.1 at Lawrence Berkeley National Laboratory. The Berkeley Center for Structural Biology is supported in part by the NIH, National Institute of General Medical Sciences, and the Howard Hughes Medical Institute. Molecular graphics and analysis performed with University of California San Francisco (UCSF) Chimera and UCSF ChimeraX, developed by the Resource for Biocomputing, Visualization, and Informatics at UCSF with support from NIH P41-GM103311, NIH R01-GM129325 and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases.

Author affiliations: [a]Department of Biochemistry, University of Washington, Seattle, WA 98195; [b]Institute for Protein Design, University of Washington, Seattle, WA 98195; [c]Graduate Program in Molecular and Cellular Biology, University of Washington, Seattle, WA 98195; [d]Transdisciplinary Research Area "Building Blocks of Matter and Fundamental Interactions", University of Bonn, 53113 Bonn, Germany; [e]Life and Medical Sciences Institute, University of Bonn, 53121 Bonn, Germany; [f]Department of Integrative, Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA 92037; [g]Scripps Consortium for HIV/AIDS Vaccine Development, The Scripps Research Institute, La Jolla, CA 92037; [h]Graduate Program in Biological Physics, Structure and Design, University of Washington, Seattle, WA 98195; [i]Molecular Biophysics and Integrated Bioimaging, Lawrence Berkeley Laboratory, Berkeley, CA 94720; and [j]Berkeley Center for Structural Biology, Lawrence Berkeley Laboratory, Berkeley, CA 94720

Author contributions: J.Y.W., A.K., and N.P.K. designed research; J.Y.W., A.K., M.C.M., A.A., A.J.B, S.V.T., C. Shu, M.A., S.C., A.K., H.N., C. Sydeman, B.S., M.W., and A.K.B performed research; J.Y.W., A.K., W.S., Y.H., U.N., D.E., and C.W. contributed new reagents/analytic tools; J.Y.W., A.K., M.C.M., A.A., A.J.B., S.V.T., C. Shu, A.K., H.N., A.K.B., L.C., B.F., M.M., D.B., A.B.W., and N.P.K. analyzed data; L.C., B.F., M.M., D.B., A.B.W., and N.P.K. supervised research; and J.Y.W., A.K., W.S., M.C.M., A.A., A.J.B., S.V.T., A.K.B., and N.P.K. wrote the paper.

Downloaded from https://www.pnas.org by "UNIVERSITY OF WASHINGTON LIBRARIES, ARCS - SERIALS" on October 20, 2025 from IP address 205.175.118.15.

1. M. Uhlen *et al.*, The human secretome. *Sci. Signal.* **12**, eaaz0274 (2019).
2. H. Farhan, C. Rabouille, Signalling to and from the secretory pathway. *J. Cell Sci.* **124**, 171–180 (2011).
3. E. Braun, D. Sauter, Furin-mediated protein processing in infectious diseases and cancer. *Clin. Transl. Immunol.* **8**, e1073 (2019).
4. K. Ohtsubo, J. D. Marth, Glycosylation in cellular mechanisms of health and disease. *Cell* **126**, 855–867 (2006).
5. K. D. Wittrup, Disulfide bond formation and eukaryotic secretory productivity. *Curr. Opin. Biotechnol.* **6**, 203–208 (1995).
6. G. Walsh, Biopharmaceutical benchmarks 2018. *Nat. Biotechnol.* **36**, 1136–1145 (2018).
7. D. Baycin-Hizal *et al.*, Proteomic analysis of Chinese hamster ovary cells. *J. Proteome Res.* **11**, 5265–5276 (2012).
8. O. Rupp *et al.*, A reference genome of the Chinese hamster based on a hybrid assembly strategy. *Biotechnol. Bioeng.* **115**, 2087–2100 (2018).
9. C.-C. Kuo *et al.*, The emerging role of systems biology for engineering protein production in CHO cells. *Curr. Opin. Biotechnol.* **51**, 64–69 (2018).
10. X. Xu *et al.*, The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nat. Biotechnol.* **29**, 735–741 (2011).
11. S. Xiao, J. Shiloach, M. J. Betenbaugh, Engineering cells to improve protein expression. *Curr. Opin. Struct. Biol.* **26**, 32–38 (2014).
12. V. Le Fourn, P.-A. Girod, M. Buceta, A. Regamey, N. Mermod, CHO cell engineering to prevent polypeptide aggregation and improve therapeutic protein secretion. *Metab. Eng.* **21**, 91–102 (2014).
13. A. M. V. Gomes, T. S. Carmo, L. S. Carvalho, F. M. Bahia, N. S. Parachin, Comparison of yeasts as hosts for recombinant protein production. *Microorganisms* **6** (2018).
14. J. A. Rakestraw, S. L. Sazinsky, A. Piatesi, E. Antipov, K. D. Wittrup, Directed evolution of a secretory leader for the improved expression of heterologous proteins and full-length antibodies in Saccharomyces cerevisiae. *Biotechnol. Bioeng.* **103**, 1192–1201 (2009).
15. G. Guler-Gane *et al.*, Overcoming the refractory expression of secreted recombinant proteins in mammalian cells through modification of the signal peptide and adjacent amino acids. *PLoS One* **11**, e0155340 (2016).
16. L. Kober, C. Zehe, J. Bode, Optimized signal peptides for the development of high expressing CHO cell lines. *Biotechnol. Bioeng.* **110**, 1164–1173 (2013).
17. C. M. Sagt *et al.*, Introduction of an N-glycosylation site increases secretion of heterologous proteins in yeasts. *Appl. Environ. Microbiol.* **66**, 4940–4944 (2000).
18. Y. Liu, A. Nguyen, R. L. Wolfert, S. Zhuo, Enhancing the secretion of recombinant proteins by engineering N-glycosylation sites. *Biotechnol. Prog.* **25**, 1468–1475 (2009).
19. N. C. Dalvie *et al.*, Molecular engineering improves antigen quality and enables integrated manufacturing of a trivalent subunit vaccine candidate for rotavirus. *Microb. Cell Fact.* **20**, 94 (2021).
20. J. Jardine *et al.*, Rational HIV immunogen design to target specific germline B cell receptors. *Science* **340**, 711–716 (2013).
21. G. Rojas *et al.*, Directed evolution of super-secreted variants from phage-displayed human Interleukin-2. *Sci. Rep.* **9**, 800 (2019).
22. D. Akopian, K. Shen, X. Zhang, S.-O. Shan, Signal recognition particle: An essential protein-targeting machine. *Annu. Rev. Biochem.* **82**, 693–721 (2013).
23. G. von Heijne, The signal peptide. *J. Membr. Biol.* **115**, 195–201 (1990).
24. P. F. Egea, R. M. Stroud, P. Walter, Targeting proteins to membranes: Structure of the signal recognition particle. *Curr. Opin. Struct. Biol.* **15**, 213–220 (2005).
25. T. A. Rapoport, Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature* **450**, 663–669 (2007).
26. T. Hessa *et al.*, Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **433**, 377–381 (2005).
27. T. Hessa *et al.*, Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* **450**, 1026–1030 (2007).
28. N. G. A. Abrescia, D. H. Bamford, J. M. Grimes, D. I. Stuart, Structure unifies the viral universe. *Annu. Rev. Biochem.* **81**, 795–822 (2012).
29. T. A. Bobik, B. P. Lehman, T. O. Yeates, Bacterial microcompartments: Widespread prokaryotic organelles for isolation and optimization of metabolic pathways. *Mol. Microbiol.* **98**, 193–207 (2015).
30. D. S. Goodsell, A. J. Olson, Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 105–153 (2000).
31. S. Zhang, Fabrication of novel biomaterials through molecular self-assembly. *Nat. Biotechnol.* **21**, 1171–1178 (2003).
32. J. E. Padilla, C. Colovos, T. O. Yeates, Nanohedra: Using symmetry to design self assembling protein cages, layers, crystals, and filaments. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 2217–2221 (2001).
33. T. Douglas, M. Young, Viruses: Making friends with old foes. *Science* **312**, 873–875 (2006).
34. M. V. de Ruiter *et al.*, Polymorphic assembly of virus-capsid proteins around DNA and the cellular uptake of the resulting particles. *J. Control. Release* **307**, 342–354 (2019).
35. T. G. W. Edwardson, S. Tetter, D. Hilvert, Two-tier supramolecular encapsulation of small molecules in a protein cage. *Nat. Commun.* **11**, 5410 (2020).
36. S. Tetter *et al.*, Evolution of a virus-like architecture and packaging mechanism in a repurposed bacterial protein. *Science* **372**, 1220–1224 (2021).
37. M. Kanekiyo *et al.*, Self-assembling influenza nanoparticle vaccines elicit broadly neutralizing H1N1 antibodies. *Nature* **499**, 102–106 (2013).
38. J. López-Sagaseta, E. Malito, R. Rappuoli, M. J. Bottomley, Self-assembling protein nanoparticles in the design of vaccines. *Comput. Struct. Biotechnol. J.* **14**, 58–68 (2016).
39. N. P. King *et al.*, Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* **336**, 1171–1174 (2012).
40. A. Khmelinskaia, A. Wargacki, N. P. King, Structure-based design of novel polyhedral protein nanomaterials. *Curr. Opin. Microbiol.* **61**, 51–57 (2021).
41. G. Ueda *et al.*, Tailored design of protein nanoparticle scaffolds for multivalent presentation of viral glycoprotein antigens. *Elife* **9**, e57659 (2020).
42. N. P. King *et al.*, Accurate design of co-assembling multi-component protein nanomaterials. *Nature* **510**, 103–108 (2014).
43. Y. Hsia *et al.*, Design of a hyperstable 60-subunit protein dodecahedron. [corrected]. *Nature* **535**, 136–139 (2016).
44. Y. Hsia *et al.*, Design of multi-scale protein complexes by hierarchical building block fusion. *Nat. Commun.* **12**, 2294 (2021).
45. P.-S. Huang, S. E. Boyken, D. Baker, The coming of age of de novo protein design. *Nature* **537**, 320–327 (2016).
46. T. U. J. Bruun, A.-M.C. Andersson, S. J. Draper, M. Howarth, Engineering a rugged nanoscaffold to enhance plug-and-display vaccination. *ACS Nano* **12**, 8855–8866 (2018).
47. 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).
48. L. He *et al.*, Single-component, self-assembling, protein nanoparticles presenting the receptor binding domain and stabilized spike as SARS-CoV-2 vaccine candidates. *Sci. Adv.* **7** (2021).
49. J. Y. Song *et al.*, Safety and immunogenicity of a SARS-CoV-2 recombinant protein nanoparticle vaccine (GBP510) adjuvanted with AS03: A randomised, placebo-controlled, observer-blinded phase 1/2 trial. *EClinicalMedicine* **51**, 101569 (2022).
50. A. C. Walls *et al.*, Elicitation of potent neutralizing antibody responses by designed protein nanoparticle vaccines for SARS-CoV-2. *Cell* **183**, 1367–1382.e17 (2020).
51. S. Boyoglu-Barnum *et al.*, Quadrivalent influenza nanoparticle vaccines induce broad protection. *Nature* **592**, 623–628 (2021).
52. J. Janin, R. P. Bahadur, P. Chakrabarti, Protein-protein interaction and quaternary structure. *Q. Rev. Biophys.* **41**, 133–180 (2008).
53. J. B. Bale *et al.*, Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science* **353**, 389–394 (2016).
54. A. Sahasrabuddhe *et al.*, Confirmation of intersubunit connectivity and topology of designed protein complexes by native MS. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 1268–1273 (2018).
55. S. K. Burley *et al.*, RCSB protein data bank: Powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Res.* **49**, D437–D451 (2021).
56. UniProt Consortium, UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).
57. E. T. Boder, K. D. Wittrup, Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* **15**, 553–557 (1997).

58. G. Ueda *et al.*, Tailored design of protein nanoparticle scaffolds for multivalent presentation of viral glycoprotein antigens. *Elife* **9**, e57659 (2020).
59. A. Leaver-Fay *et al.*, ROSETTA3: An object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487**, 545–574 (2011).
60. J. K. Leman *et al.*, Macromolecular modeling and design in Rosetta: Recent methods and frameworks. *Nat. Methods* **17**, 665–680 (2020).
61. F. Palombarini, E. Di Fabio, A. Boffi, A. Macone, A. Bonamore, Ferritin nanocages for protein delivery to tumor cells. *Molecules* **25**, 825 (2020).
62. S. Rajoo, P. Vallotton, E. Onischenko, K. Weis, Stoichiometry and compositional plasticity of the yeast nuclear pore complex revealed by quantitative fluorescence microscopy. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E3969–E3977 (2018).
63. A. A. Cohen *et al.*, Mosaic nanoparticles elicit cross-reactive immune responses to zoonotic coronaviruses in mice. *Science* **371**, 735–741 (2021).
64. M. C. Lawrence, P. M. Colman, Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946–950 (1993).
65. M. Baek, D. Baker, Deep learning and protein structure modeling. *Nat. Methods* **19**, 13–14 (2022).
66. D.-A. Silva *et al.*, De novo design of potent and selective mimics of IL-2 and IL-15. *Nature* **565**, 186–191 (2019).
67. S. R. Gordon *et al.*, Computational design of an a-gliadin peptidase. *J. Am. Chem. Soc.* **134**, 20513–20520 (2012).
68. J. G. Jardine *et al.*, HIV-1 VACCINES. Priming a broadly neutralizing antibody response to HIV-1 using a germline-targeting immunogen. *Science* **349**, 156–161 (2015).
69. D. D. Sahtoe *et al.*, Reconfigurable asymmetric protein assemblies through implicit negative design. *Science* **375**, eabj7662 (2022).
70. J. Dou *et al.*, De novo design of a fluorescence-activating β-barrel. *Nature* **561**, 485–491 (2018).
71. J. C. Klima *et al.*, Incorporation of sensing modalities into de novo designed fluorescence-activating proteins. *Nat. Commun.* **12**, 856 (2021).
72. B. Kelley, Industrialization of mAb production technology: The bioprocessing industry at a crossroads. *MAbs* **1**, 443–452 (2009).
73. N. C. Dalvie *et al.*, Engineered SARS-CoV-2 receptor binding domain improves manufacturability in yeast and immunogenicity in mice. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2106845118 (2021).
74. D. Ellis *et al.*, Stabilization of the SARS-CoV-2 spike receptor-binding domain using deep mutational scanning and structure-based design. *Front. Immunol.* **12**, 710263 (2021).
75. C.-L. Hsieh *et al.*, Structure-based design of prefusion-stabilized SARS-CoV-2 spikes. *Science* **369**, 1501–1505 (2020).
76. R. Rawi *et al.*, Automated design by structure-based stabilization and consensus repair to achieve prefusion-closed envelope trimers in a wide variety of HIV strains. *Cell Rep.* **33**, 108432 (2020).
77. J. Pallesen *et al.*, Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E7348–E7357 (2017).
78. S. K. Malladi *et al.*, One-step sequence and structure-guided optimization of HIV-1 envelope gp140. *Curr Res Struct Biol* **2**, 45–55 (2020).
79. Y. Peleg *et al.*, Community-wide experimental evaluation of the PROSS stability-design method. *J. Mol. Biol.* **433**, 166964 (2021).
80. I. Campeotto *et al.*, One-step design of a stable variant of the malaria invasion protein RH5 for use as a vaccine immunogen. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 998–1002 (2017).
81. D. Wang, P. W. L. Tai, G. Gao, Adeno-associated virus vector as a platform for gene therapy delivery. *Nat. Rev. Drug Discov.* **18**, 358–378 (2019).
82. N. Chaudhary, D. Weissman, K. A. Whitehead, mRNA vaccines for infectious diseases: Principles, delivery and clinical translation. *Nat. Rev. Drug Discov.* **20**, 817–838 (2021).
83. K. M. Konrath *et al.*, Nucleic acid delivery of immune-focused SARS-CoV-2 nanoparticles drives rapid and potent immunogenicity capable of single-dose protection. *Cell Rep.* **38**, 110318 (2022).
84. Z. Xu *et al.*, In vivo assembly of nanoparticles achieved through synergy of structure-based protein engineering and synthetic DNA generates enhanced adaptive immunity. *Adv. Sci.* **7**, 1902802 (2020).
85. Z. Mu *et al.*, Ability of nucleoside-modified mRNA to encode HIV-1 envelope trimer nanoparticles. bioRxiv [Preprint] (2021), https://doi.org/10.1101/2021.08.09.455714 (Accessed 9 August 2021).
86. M. Boone *et al.*, Massively parallel interrogation of protein fragment secretability using SECRiFY reveals features influencing secretory system transit. *Nat. Commun.* **12**, 6414 (2021).

Downloaded from https://www.pnas.org by "UNIVERSITY OF WASHINGTON LIBRARIES, ARCS - SERIALS" on October 20, 2025 from IP address 205.175.118.15.