Title

Designed Protein Nanoparticles for Cryo-EM and Cellular Imaging

Author

Gee, Morgan

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Designed Protein Nanoparticles for Cryo-EM and Cellular Imaging

A thesis submitted in partial satisfaction of the

requirements for the degree Master of Science

in Biochemistry, Molecular and Structural Biology

by

Morgan Michael Gee

2024

ABSTRACT OF THE THESIS

Designed Protein Nanoparticles for Cryo-EM and Cellular Imaging

by

Morgan Michael Gee

Master of Science in Biochemistry, Molecular and Structural Biology

University of California, Los Angeles, 2024

Professor Todd O. Yeates, Chair

Protein nanoparticles are self-assembling, symmetric, supramolecular protein structures with diverse applications that have significantly furthered the fields of structural biology and molecular biology. Within structural biology, both the size limitation of cryogenic electron microscopy (cryo-EM) and preferred specimen orientation have proven to be outstanding challenges in understanding the protein structure of small protein targets. Similarly, there is a deficit of robust methods for imaging proteins and protein complexes at high resolution within the cellular context. Here, I describe the development of protein nanoparticles, or protein cages, as rigid, symmetric scaffolds capable of resolving the structure of small proteins with cryo-EM and as fluorescent, fiducial biomarkers for targeted cellular imaging with confocal microscopy. This study underscores the profound impact of protein nanoparticle technology and its capacity to further revolutionize biochemical research.

The thesis of Morgan Gee is approved.

Hilary Ann Coller

William M. Gelbart

Todd O. Yeates, Committee Chair

University of California, Los Angeles

2024

Dedicated to my family, friends, and others who have supported me unconditionally.

TABLE OF CONTENTS

LIST OF FIGURES

Acknowledgements

I would like to thank Professor Todd O. Yeates and Dr. Roger Castells-Graells for their continued support and encouragement throughout my undergraduate career. I also thank everyone else in the Yeates lab former and current for their guidance and advice.

Chapter 2 of this thesis includes a version of a published manuscript: Castells-Graells, R., Meador, K., Arbing, M. A., Sawaya, M. R., Gee, M., Cascio, D., Gleave, E., Debreczeni, J. É., Breed, J., Leopold, K., Patel, A., Jahagirdar, D., Lyons, B., Subramaniam, S., Phillips, C. & Yeates, T.O. Cryo-EM structure determination of small therapeutic protein targets at 3 Å-resolution using a rigid imaging scaffold. Proc. Natl. Acad. Sci. United States Am.120, e2305494120(2023). Reproduced with permission. R.C.-G., K.M., S.S., C.P., and T.O.Y. designed research; R.C.-G., K.M., M.A.A., M.R.S., M.G., D.C.,E.G., J.É.D., J.B., K.L., A.P., D.J., B.L., and S.S. performed research; R.C.-G., K.M., M.A.A., M.R.S., K.L., A.P., D.J., B.L.,S.S.,C.P., and T.O.Y. analyzed data; and R.C.-G., K.M., and T.O.Y. wrote the paper.

Chapter 3 of this thesis is a version of a preprint: Gee, M., Atai, K., Coller, H. A., Yeates, T. O., Castells-Graells, R. Designed fluorescent protein cages as fiducial markers for targeted cell imaging. bioRxiv : the preprint server for biology, 2024.02.28.582585. Reproduced with permission.

Chapter 1 – Introduction

1.1 Introduction to protein nanoparticles

The wholly-designed first protein nanoparticles were introduced in a 2001 seminal work in which two naturally self-assembling oligomeric proteins were genetically fused. Their self-assembly properties enabled the two oligomers to form a protein nanohedron, dubbed a protein cage [1]. This initial design specifically relied on the rigid connection between the dimeric M1 matrix protein of influenza virus and the trimeric bromoperoxidase to self-assemble into a symmetric, tetrahedral protein cage. Along with the tetrahedral assembly, the genetic fusion between influenza virus matrix protein M1 and carboxylesterase formed a self-assembling protein filament. The newly designed symmetric, self-assembling protein nanohedra, specifically the protein cage, laid the foundations for new avenues in the development of protein assemblies.

A major breakthrough in protein nanoparticle technology was the elimination of the rigid linker between the two protein oligomers in favor of computationally designing naturally docking subunits [2]. This led to a new generation of designed protein cages that would form from two self-assembling protein oligomers without the need for a genetic linker. Despite this improvement, there still lacked a diversity in the symmetries, sizes, and protein composition of existing protein cages.

Over time, more protein nanoparticles were computationally designed to include icosahedral assemblies and to increase the number of existing tetrahedral and octahedral assemblies [3–7]. These assemblies include the already existing tetrahedral assemblies composed of two trimeric subunits labeled T33 (tetrahedral; trimer subunit 1, trimeric subunit 2), the novel icosahedral assemblies composed of one pentameric subunit and one trimer subunit labeled I53 (icosahedral; pentamer subunit; trimer subunit), and more. Iterations of the T33 and I53 symmetry

and design were eventually further modified and engineered for a range of applications.

1.2 Current applications of protein nanoparticles

Protein nanoparticles have served an integral role in widespread biochemical research, advancing nanotechnology, biotechnology, and medical research. They have broadened structural biology opportunities with cryo-EM, functioned as a basis of molecular medicine with their cargo delivery capabilities, and acted as a nanoparticle immunogen for vaccine development.

The numerous applications of protein nanoparticles have been made feasible through genetically fusing proteins on their external surface. One of their earliest adaptations was their use as imaging scaffolds for cryo-EM by externally presenting engineered DARPins to bind to a targeted protein of interest. With the development of the protein nanoparticle imaging scaffold, the near-atomic structures of small protein targets were successfully determined with cryo-EM, bypassing its inherent size limitation and preferred specimen orientation challenge [8,9]. Their open interior has been theorized to be able to encapsulate and protect drugs and other therapeutics for targeted drug delivery. Using similar binding methodology, protein nanoparticles have been developed for molecular medicine as a means to carry and release cargos of interest or to bind proteins of interest to their surface [10–13]. Owing to properties they share with virus-like particles, which have long been used as vaccines, protein nanoparticles have recently been engineered for the purpose of therapeutics development [14]. By fusing antigen epitopes to the surface rather than binding proteins, protein nanoparticles are being designed to act as immunogens. Following both the conventional two-component nanoparticle as well as newer single-component nanoparticle designs, there has been significant success in eliciting immune responses from mammalian models [15–17]. Exploiting recent advancements, current applications of protein nanoparticles have

significantly impacted numerous disciplines within biochemistry.

Research applications for protein nanoparticles continue to expand as *de novo* nanoparticles are continuously being designed and their full potential as nanoparticle immunogens is still being realized. Here, I describe the development of a protein nanoparticle imaging scaffold for the structure determination of the oncogenic protein KRAS and the design of nanoparticles as fiducial biomarkers for targeted cellular imaging *in vivo*.

1.3 References

(1) Padilla, J. E.; Colovos, C.; Yeates, T. O. Nanohedra: Using Symmetry to Design Self Assembling Protein Cages, Layers, Crystals, and Filaments. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (5), 2217–2221. https://doi.org/10.1073/pnas.041614998.

(2) King, N. P.; Sheffler, W.; Sawaya, M. R.; Vollmar, B. S.; Sumida, J. P.; André, I.; Gonen, T.; Yeates, T. O.; Baker, D. Computational Design of Self-Assembling Protein Nanomaterials with Atomic Level Accuracy. *Science* **2012**, *336* (6085), 1171–1174. https://doi.org/10.1126/science.1219364.

(3) Lai, Y.-T.; Cascio, D.; Yeates, T. O. Structure of a 16-Nm Cage Designed by Using Protein Oligomers. *Science* **2012**, *336* (6085), 1129–1129. https://doi.org/10.1126/science.1219351.

(4) King, N. P.; Bale, J. B.; Sheffler, W.; McNamara, D. E.; Gonen, S.; Gonen, T.; Yeates, T. O.; Baker, D. Accurate Design of Co-Assembling Multi-Component Protein Nanomaterials. *Nature* **2014**, *510* (7503), 103–108. https://doi.org/10.1038/nature13404.

(5) Cristie-David, A. S.; Koldewey, P.; Meinen, B. A.; Bardwell, J. C. A.; Marsh, E. N. G. Elaborating a Coiled-coil-assembled Octahedral Protein Cage with Additional Protein Domains. *Protein Sci.* **2018**, *27* (11), 1893–1900. https://doi.org/10.1002/pro.3497.

(6) Bale, J. B.; Park, R. U.; Liu, Y.; Gonen, S.; Gonen, T.; Cascio, D.; King, N. P.; Yeates, T. O.; Baker, D. Structure of a Designed Tetrahedral Protein Assembly Variant Engineered to Have Improved Soluble Expression. *Protein Sci.* **2015**, *24* (10), 1695–1701. https://doi.org/10.1002/pro.2748.

(7) Bale, J. B.; Gonen, S.; Liu, Y.; Sheffler, W.; Ellis, D.; Thomas, C.; Cascio, D.; Yeates, T. O.; Gonen, T.; King, N. P.; Baker, D. Accurate Design of Megadalton-Scale Two-Component Icosahedral Protein Complexes. *Science* **2016**, *353* (6297), 389–394. https://doi.org/10.1126/science.aaf8818.

(8) Liu, Y.; Huynh, D. T.; Yeates, T. O. A 3.8 Å Resolution Cryo-EM Structure of a Small Protein Bound to an Imaging Scaffold. *Nat. Commun.* **2019**, *10* (1), 1864. https://doi.org/10.1038/s41467-019-09836-0.

(9) Castells-Graells, R.; Meador, K.; Arbing, M. A.; Sawaya, M. R.; Gee, M.; Cascio, D.; Gleave, E.; Debreczeni, J. É.; Breed, J.; Leopold, K.; Patel, A.; Jahagirdar, D.; Lyons, B.; Subramaniam, S.; Phillips, C.; Yeates, T. O. Cryo-EM Structure Determination of Small Therapeutic Protein Targets at 3 Å-Resolution Using a Rigid Imaging Scaffold. *Proc. Natl.*

Acad. Sci. **2023**, *120* (37), e2305494120. https://doi.org/10.1073/pnas.2305494120.
(10) Jain, A.; Singh, S. K.; Arya, S. K.; Kundu, S. C.; Kapoor, S. Protein Nanoparticles: Promising Platforms for Drug Delivery Applications. *ACS Biomater. Sci. Eng.* **2018**, *4* (12), 3939–3961. https://doi.org/10.1021/acsbiomaterials.8b01098.
(11) Miller, J. E.; Srinivasan, Y.; Dharmaraj, N. P.; Liu, A.; Nguyen, P. L.; Taylor, S. D.; Yeates, T. O. Designing Protease-Triggered Protein Cages. *J. Am. Chem. Soc.* **2022**, *144* (28), 12681–12689. https://doi.org/10.1021/jacs.2c02165.
(12) Miller, J. E.; Castells-Graells, R.; Arbing, M. A.; Munoz, A.; Jiang, Y.-X.; Espinoza, C. T.; Nguyen, B.; Moroz, P.; Yeates, T. O. Design of Beta-2 Microglobulin Adsorbent Protein Nanoparticles. *Biomolecules* **2023**, *13* (7), 1122. https://doi.org/10.3390/biom13071122.
(13) Lee, E. J.; Gladkov, N.; Miller, J. E.; Yeates, T. O. Design of Ligand-Operable Protein-Cages That Open Upon Specific Protein Binding. *ACS Synth. Biol.* **2023**. https://doi.org/10.1021/acssynbio.3c00383.
(14) Tariq, H.; Batool, S.; Asif, S.; Ali, M.; Abbasi, B. H. Virus-Like Particles: Revolutionary Platforms for Developing Vaccines Against Emerging Infectious Diseases. *Front. Microbiol.* **2022**, *12*, 790121. https://doi.org/10.3389/fmicb.2021.790121.
(15) Kanekiyo, M.; Ellis, D.; King, N. P. New Vaccine Design and Delivery Technologies. *J. Infect. Dis.* **2019**, *219* (Supplement_1), S88–S96. https://doi.org/10.1093/infdis/jiy745.
(16) Antanasijevic, A.; Ueda, G.; Brouwer, P. J. M.; Copps, J.; Huang, D.; Allen, J. D.; Cottrell, C. A.; Yasmeen, A.; Sewall, L. M.; Bontjer, I.; Ketas, T. J.; Turner, H. L.; Berndsen, Z. T.; Montefiori, D. C.; Klasse, P. J.; Crispin, M.; Nemazee, D.; Moore, J. P.; Sanders, R. W.; King, N. P.; Baker, D.; Ward, A. B. Structural and Functional Evaluation of de Novo-Designed, Two-Component Nanoparticle Carriers for HIV Env Trimer Immunogens. *PLOS Pathog.* **2020**, *16* (8), e1008665. https://doi.org/10.1371/journal.ppat.1008665.
(17) Zhang, Y.-N.; Paynter, J.; Antanasijevic, A.; Allen, J. D.; Eldad, M.; Lee, Y.-Z.; Copps, J.; Newby, M. L.; He, L.; Chavez, D.; Frost, P.; Goodroe, A.; Dutton, J.; Lanford, R.; Chen, C.; Wilson, I. A.; Crispin, M.; Ward, A. B.; Zhu, J. Single-Component Multilayered Self-Assembling Protein Nanoparticles Presenting Glycan-Trimmed Uncleaved Prefusion Optimized Envelope Trimers as HIV-1 Vaccine Candidates. *Nat. Commun.* **2023**, *14* (1), 1985. https://doi.org/10.1038/s41467-023-37742-z.

Chapter 2 – Cryo-EM Imaging Scaffolds Based on Rigid Protein Nanoparticles

The following chapter includes a reprint of a published manuscript from

Proceedings of the National Academy of Sciences

Cryo-EM structure determination of small therapeutic protein targets at 3 Å resolution using a rigid imaging scaffold

120, 37 (2023)

DOI: 10.1073/pnas.2305494120

2.1 Introduction to protein nanoparticle imaging scaffolds

The advent of cryo-EM significantly advanced the field of structural biology as large complexes and proteins unsolvable with other contemporary methods were able to be structurally determined at a high-resolution. Despite the "resolution revolution," cryo-EM suffered due to its inability to resolve structures below a certain molecular weight; the method is theorized to have a physical lower limit of 38-50 kDa [1,2]. In addition, macromolecules suffer from a phenomenon known as preferred specimen orientation when placed on an EM grid, resulting in non-random orientations, a loss of information when imaging, and a concomitant loss of overall resolution [1,3]. To address these problems, protein nanoparticles have been developed as a scaffold for high-resolution cryo-EM imaging.

The structural determination of small protein targets with cryo-EM has been accomplished by employing protein nanoparticles and utilizing their physical properties. In particular, the large size of protein cages and their symmetry have proven to be paramount characteristics in bypassing the size limitation of cryo-EM and the preferred specimen orientation. As the size of the imaging scaffold is slightly above 650 kDa, any cargo protein bound to its exterior will only contribute to the overall molecular weight. This significantly surpasses the suggested 38-50 kDa lower limit, theoretically allowing proteins of any size to be imaged with cryo-EM. Additionally, the symmetry of the scaffolds reduces the number of varied orientations necessary for high resolution imaging to partially circumvent the preferred specimen orientation issue.

To bind and present cargo for imaging, designed ankyrin repeat proteins (DARPins) are genetically fused to enable the non-covalent and specific binding of a target protein to the scaffold [4]. They can be engineered to specifically and strongly interact with a desired protein by modifying its nonconserved loop regions. By conjoining a protein cage subunit and an engineered antiGFP-

DARPin (that binds specifically to GFP), we can present multiple copies of a GFP binding domain on the surface of the imaging scaffold [5]. In the case of a previously developed protein cage known as T33-51, when the cage subunit is genetically fused with antiGFP-DARPins, 12 copies of the DARPin are presented, allowing for up to 12 GFP molecules to bind to each individual protein cage. With this, GFP was structurally determined with cryo-EM at a near-atomic resolution of 3.8 Å.

This protein cage has since been computationally designed to more rigidly interface and present the DARPin domains. A more rigid structure enables higher resolution as flexibility and movement decreases the overall quality of individual micrographs. The improved imaging scaffold was specifically used for the structural determination of the small oncogenic protein KRAS, which has a molecular weight of approximately 19 kDa [6]. These modifications were fundamental for improving the attainable resolution from 3.8 Å to 3 Å for the small target protein. Moreover, the slight variations in the structure of KRAS and its mutant, G12C, were able to be visualized after binding to a commercially available inhibitor drug, AMG510.

Protein nanoparticles have enabled the imaging of small target proteins with cryo-EM, bypassing the lower size limit and the preferred specimen orientation problems. By rigidly fusing DARPins to the external surface of the imaging scaffolds, proteins of interest can non-covalently bind with high specificity and affinity. This opens significantly more opportunities for structural determination of proteins with cryo-EM rather than other contemporary structural biology methods.

Cryo-EM structure determination of small therapeutic protein targets at 3 Å-resolution using a rigid imaging scaffold

The following chapter is a reprint of a research article from

Proceedings of the National Academy of Sciences

120, 37 (2023)

DOI: 10.1073/pnas.2305494120

PNAS RESEARCH ARTICLE | BIOCHEMISTRY
BIOPHYSICS AND COMPUTATIONAL BIOLOGY OPEN ACCESS

Cryo-EM structure determination of small therapeutic protein targets at 3 Å-resolution using a rigid imaging scaffold

Roger Castells-Graells[a], Kyle Meador[b], Mark A. Arbing[a], Michael R. Sawaya[a], Morgan Gee[b], Duilio Cascio[a], Emma Gleave[c], Judit É. Debreczeni[c], Jason Breed[c], Karoline Leopold[d], Ankoor Patel[d], Dushyant Jahagirdar[d], Bronwyn Lyons[d], Sriram Subramaniam[d,e], Chris Phillips[c], and Todd O. Yeates[a,b,1]

Edited by Edward Egelman, University of Virginia, Charlottesville, VA; received April 5, 2023; accepted July 14, 2023

Cryoelectron microscopy (Cryo-EM) has enabled structural determination of proteins larger than about 50 kDa, including many intractable by any other method, but it has largely failed for smaller proteins. Here, we obtain structures of small proteins by binding them to a rigid molecular scaffold based on a designed protein cage, revealing atomic details at resolutions reaching 2.9 Å. We apply this system to the key cancer signaling protein KRAS (19 kDa in size), obtaining four structures of oncogenic mutational variants by cryo-EM. Importantly, a structure for the key G12C mutant bound to an inhibitor drug (AMG510) reveals significant conformational differences compared to prior data in the crystalline state. The findings highlight the promise of cryo-EM scaffolds for advancing the design of drug molecules against small therapeutic protein targets in cancer and other human diseases.

cryo-EM | small proteins | imaging scaffolds | protein design | cancer drugs

Significance

Cryoelectron microscopy (cryo-EM) is emerging as a major method for elucidating the structures of proteins in atomic detail. A key limitation, however, is that cryo-EM is applicable only to sufficiently large macromolecular complexes. This places a great many important proteins of smaller size, especially those of interest for therapeutic drug development, outside the reach of cryo-EM. We describe a protein engineering effort that overcomes the lower mass limit through the development of a modular imaging scaffold able to rigidly bind and display practically any small protein of interest, greatly increasing its effective mass. We show this technology can be used to visualize molecules, such as a key cancer protein, with important implications for drug design and biomedical research.

Cryoelectron microscopy (cryo-EM) is a rapidly expanding method for determining the atomic structures of large molecular assemblies. It is, however, problematic for determining the structures of small-to-medium-sized protein molecules. A size of about 38 kDa represents a likely theoretical lower limit (1), while about 50 kDa is a practical limit from current work (2). Accordingly, vast numbers of cellular proteins, including many of key therapeutic interest, remain beyond the reach of cryo-EM methods (3).

A potential workaround to the size limitation in cryo-EM is to bind a small protein of interest (the "cargo") to a much larger carrier (the "scaffold") in order to make it large enough to visualize readily. Ideas for scaffolding approaches go back several years (4–6). A key challenge is how to make the binding attachment between the scaffold and the cargo protein sufficiently rigid, as even minor flexibility in the attachment severely compromises the ability to reconstruct a high-resolution image of the bound cargo component. In addition, a general solution to the scaffolding problem calls for modular design, i.e., through the use of a scaffolding component that can be readily diversified to bind any given cargo protein of interest (7–10). Earlier work has explored the use of DARPins as the modular binding domain, genetically fused by way of a continuous alpha helical connection to self-assembling protein cages, to create large symmetric scaffolds for imaging (11–14). Diverse studies have made progress (2, 15–20) (*SI Appendix, Supplementary Text*), but further improvements are needed to develop a facile system for high-resolution cryo-EM of small proteins.

In the present study, we demonstrate a protein design advance that substantially rigidifies a cryo-EM scaffold based on fusion of a DARPin as the modular binding domain to a designed protein cage. Analogous to antibodies, sequence variations in the nonconserved loop regions of a DARPin protein can be selected in the laboratory in order to obtain a variant that binds nearly any protein of interest (21). To demonstrate utility in a critically important area of medicine, we have applied this rigidified cryo-EM scaffolding system to study mutant and drug-bound structures of the key oncogenic protein KRAS, which represents a major target for designing anticancer drugs.

Results and Discussion

Rigidification and Testing of an Imaging Scaffold. A previous cage-scaffold design reached a resolution of about 3.8 Å for the attached cargo protein (11, 12), but residual flexibility made it impossible to reach the higher resolution needed for reliable atomic interpretation (generally about 3 Å or better). In the earlier design, the individual DARPin arms—12 in total emanating from the tetrahedrally symmetric cage—protruded separately from each other, thus suffering from residual flexibility. To make further stabilizing contacts possible, we investigated alternative design choices for a scaffold. A different tetrahedral protein cage

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

Competing interest statement: S.S. is CEO of Gandeeva Therapeutics. T.O.Y. is CEO of AvimerBio. S.S. holds equity in Gandeeva Therapeutics. T.O.Y. holds equity in AvimerBio. R.C.-G., K.M., and T.O.Y. are inventors on a relevant patent application.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: yeates@mbi.ucla.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2305494120/-/DCSupplemental.

Published September 5, 2023.

known as T33-51 (22), when modeled with alpha helical linkers to DARPins, oriented the protruding arms to be in near-contact with each other; three DARPins come together at each of the four vertices of the tetrahedron (Fig. 1). Then, computational interface design methods were used to generate new amino acid sequences at the interfaces formed between three symmetry-related copies of the DARPin (*SI Appendix*, Fig. S1 and *Materials and Methods*). The designed interfaces between protruding DARPins were proposed to confer additional stability to these key binding components of the scaffold (Fig. 1). From 12 candidate sequence designs, five were validated by experimental tests to self-assemble into cage-like structures as intended (*Materials and Methods*).

Before employing the candidate cryo-EM scaffolds to image a protein target of major biological importance, we compared their performance in a test system, using the well-studied superfolder version of the green fluorescent protein (GFP) (23), 26 kDa in size, as the cargo protein. When bound to the imaging scaffold, the overall molecular weight of this complex is 972 kDa. As expected, experimental tests showed that all five scaffold candidates bound to GFP when the DARPin (genetically fused to the cage) was one previously established to bind GFP (*SI Appendix*). Initial cryo-EM datasets were collected on the five candidate scaffolds with GFP bound. Based on data processing of similar numbers of particle images from the five candidates, one design

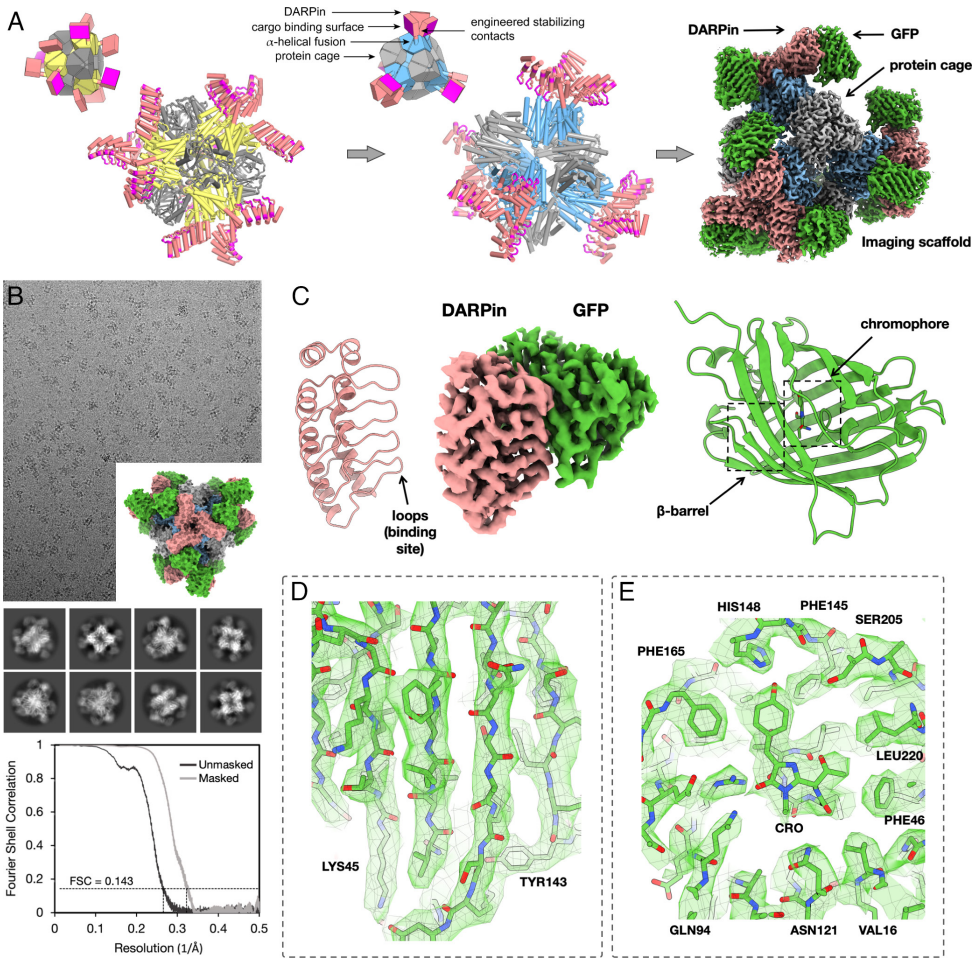

Fig. 1. Rigidified modular cryo-EM imaging scaffolds. (*A*, *Left*) A scheme for a previously described scaffold (11, 12), based on a self-assembling protein cage, displayed protruding DARPin domains as modular binders via continuous alpha helical fusions. The cage subunits bearing the continuous alpha helical fusion are shown in yellow. The other subunit type in this two-component cage is shown in gray. DARPin domains are colored in salmon with their hypervariable binding regions highlighted in magenta. (*A*, *Middle*) A redesigned scaffold based on similar principles, but with protruding DARPin arms disposed to make additional protein–protein contacts with symmetric copies of each other. Designed surface mutations at the newly created interface away from the hypervariable region stabilize the DARPin domain, allowing high-resolution cryo-EM imaging of bound cargo. The *Insets* provide simplified geometric diagrams of the scaffold constructions. (*A*, *Right*) Composite cryo-EM map after focused refinements of GFP bound to a rigidified imaging scaffold. (*B*) Cryo-EM micrograph of the rigidified imaging scaffold bound to GFP (model shown in *Inset*) and 2D classes from selected particles. An FSC plot illustrates agreement between independent half-maps obtained after focused classification and 3D reconstruction, masked around the GFP protein (resolution = 3.1 Å based on a correlation threshold of 0.143). (*C*, *Middle*) A view of the final density map covering the DARPin and its bound GFP protein. Ribbon models of the two components are shown on the sides. (*D* and *E*) Focused views of the density map covering several GFP beta-strands and the GFP chromophore with its surrounding amino acid side chains.

(designated RCG-10; *SI Appendix*) appeared to offer the most rigid presentation of the bound GFP cargo protein. This scaffold was therefore selected for further analysis and cryo-EM data processing. Following data processing from ~877,000 particles obtained from 3,575 cryo-EM movies, a 3-D density map was obtained in which the resolution of the central core of the scaffold was 2.7 Å, with a resolution of 3.1 Å for just the GFP component (Fig. 1 and *SI Appendix*, Figs. S4 and S5). The level of atomic detail is illustrated by the density for the GFP chromophore and side chains from the neighboring amino acid residues (Fig. 1).

In order to assess issues related to coordinate precision and potential perturbances caused by binding to the scaffold, we compared the bound protein structure to crystal structures of GFP in an unbound form. The binding of GFP to the DARPin did not lead to meaningful differences in the backbone, though a different rotamer is seen for a tyrosine residue (Tyr39). The rms deviation for the GFP displayed by the imaging scaffold compared to a crystal structure is 0.59 Å. For data quality and model refinement statistics, see *SI Appendix*, Table S1.

While the significant improvement in resolution of the cargo (compared to the previous, unrigidified scaffold) also reflects various advances in cryo-EM instrumentation and software, analysis of the data shows that the scaffold redesign did lead to a dramatic reduction in the flexibility of the cargo attachment, as anticipated (*SI Appendix*, Fig. S12). The success of the rigidification plan is evident in the pattern of agreement between the atomic model and the cryo-EM density map; the agreement Q-scores decrease steeply with distance from the core-DARPin hinge in the old design but remain nearly uniform in the new design (*SI Appendix*, Fig. S12). Importantly, this supports the hinge as a principal cause of reduced resolution of the cargo in the old design and the reduction in hinge flexibility as a major cause of improvement in the new design.

Additionally, we compared the ability of the deep-learning program ModelAngelo (24) to build de novo atomic models into the cryo-EM density maps. For the earlier 3.8-Å cryo-EM map, the program correctly built only 93 residues (including sidechain atoms) of 156 DARPin residues, a roughly 60% completion for the DARPin. Only 65 of 231 residues could be built for the GFP cargo, corresponding to only 28% completeness. For the new 3.1-Å cryo-EM map, ModelAngelo built all 156 residues of the DARPin domain correctly (100% success), including sidechains. For the GFP cargo, the program built 220 of 231 residues correctly (95% success), including sidechains. The missing residues are in loops (*SI Appendix*, Fig. S13).

Cryo-EM Structures of the Oncogenic KRAS Protein Bound to GDP. For biomedically relevant structural studies, we chose the KRAS protein as a target of high clinical importance. KRAS is a 19-kDa GTPase involved in signal transduction in cell proliferation pathways. KRAS is among the most prevalent human oncogenes, with mutations in KRAS occurring in about 25% of all cancers (25). Some of the most clinically relevant mutations occur at amino acid residues Gly12 and Gly13. Drugs bound to a minor cleft region of the protein near that location are of key pharmaceutical interest, including covalent inhibitors targeting cysteine mutants (i.e., G12C or G13C) (26–29). We therefore undertook a series of structural studies on known KRAS mutants, focusing on the degree of atomic interpretability in 3D density maps obtained using the cryo-EM scaffold described above; a DARPin with loop sequences that bind the GDP-bound form of KRAS was already known from prior work (30, 31), enabling the scaffold to be readily repurposed to image GDP-bound KRAS structures (*Materials and Methods*).

For imaging experiments, we investigated three different sequence variants of KRAS—single site mutants G12V, G12C, and G13C—in their GDP-bound forms. All three KRAS variants were found to bind with good occupancy to our cryo-EM scaffold (presenting the KRAS-specific DARPin). For mutant G13C, ~665,000 particles were obtained from 2,000 cryo-EM movies. Following similar data processing as before, we obtained a 3-D density map showing a resolution of 2.5 Å for the entire particle and 2.9 Å for the KRAS protein (Fig. 2 and *SI Appendix*, Figs. S7 and S8). Among

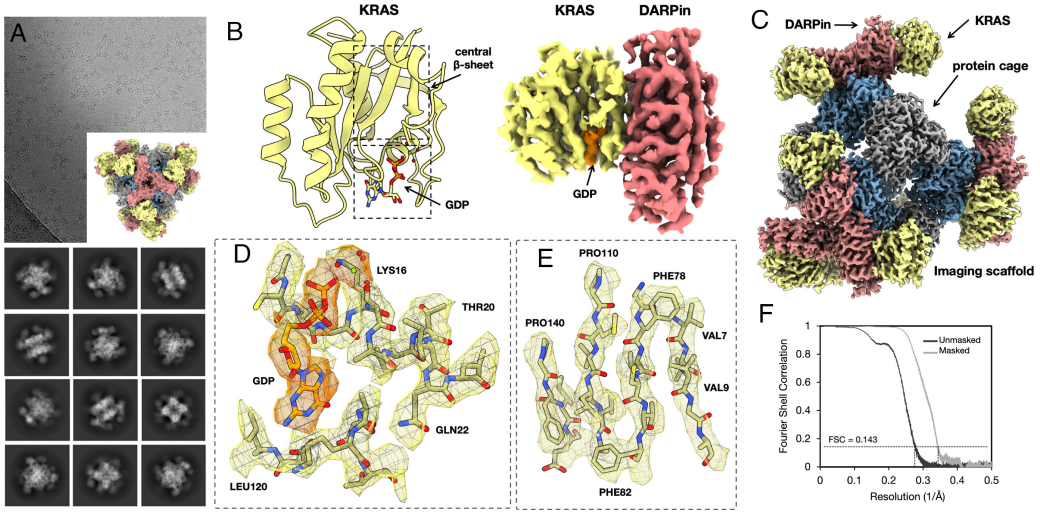

Fig. 2. Cryo-EM structure of KRAS on a rigidified imaging scaffold. (*A*) Cryo-EM micrograph of the rigidified imaging scaffold bound to KRAS (model shown in *Inset*) and 2D classes from the selected particles. (*B*) 3D reconstruction of a density map covering the DARPin and its bound KRAS protein. The GDP ligand is shown in orange. A ribbon model of the KRAS is shown on the left side. (*C*) Composite cryo-EM map after focused refinements of KRAS bound to a rigidified imaging scaffold. (*D* and *E*) Focused views of the density map covering the bound GDP ligand (orange density) and select regions of the KRAS structure. The Mg^{2+} ion is represented by a green sphere. (*F*) An FSC plot illustrates agreement between independent half-maps, obtained after focused classification and 3D reconstruction, masked around the KRAS protein (resolution = 2.9 Å based on a correlation threshold of 0.143).

other metrics of map quality, we assessed the ability of automatic protein model-building software to generate an atomic model for the protein without human intervention. Given the cryo-EM density map and the amino acid sequences for the DARPin and KRAS proteins, ModelAngelo (24) was able to build, de novo, a correct and nearly complete atomic model using default parameters (164 out of 166 residues for KRAS and 150 out of 157 for the DARPin). The amino acid sequence was correctly assigned throughout both KRAS G13C and DARPin molecules. Limited manual fitting was sufficient to join breaks in the chain where the density was weak for mobile loops in the proteins. The success of the modeling exercise shows the utility of the cryo-EM scaffolding approach for an automated structure determination pipeline.

As imaged here by cryo-EM, the KRAS protein matches closely to known structures of KRAS-GDP reported in previous X-ray crystallography studies (30, 31). Our refined structure of the G13C mutant overlaps with a previous X-ray crystal structure with an rms deviation of only 0.5 Å over protein backbone atoms.

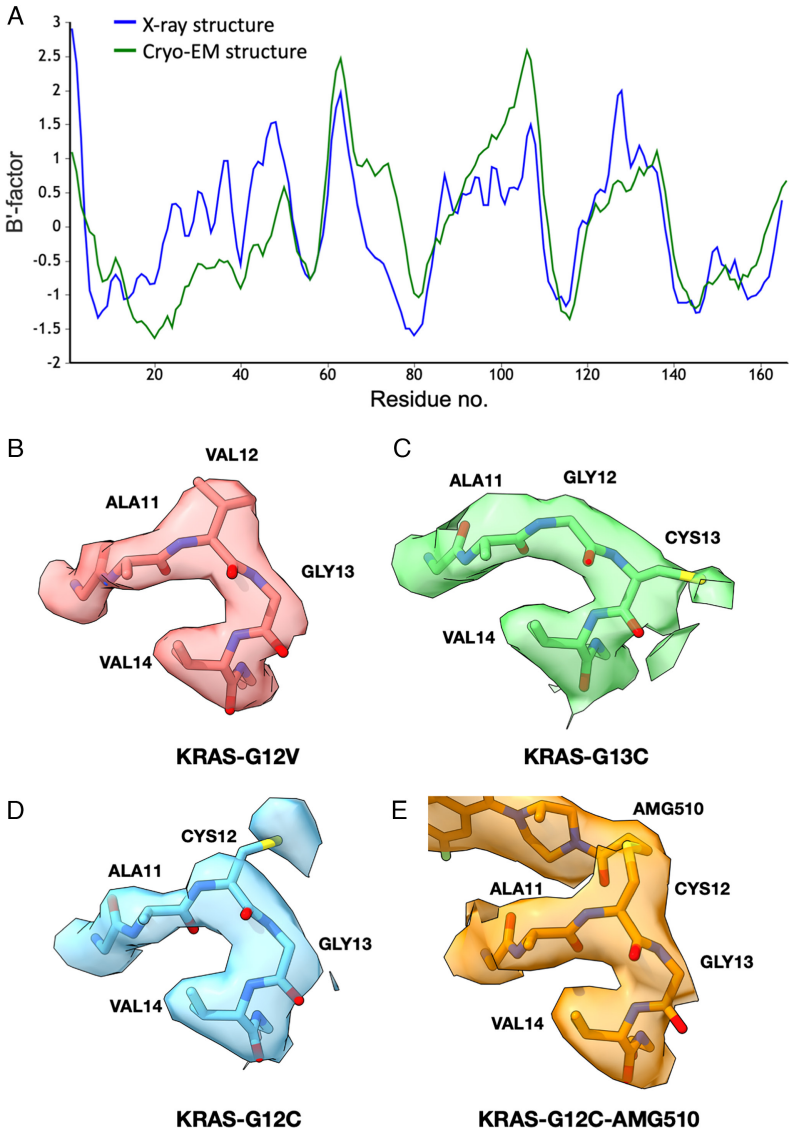

Fig. 3. Structural and dynamical interpretability of cryo-EM maps of KRAS and single-site mutants. (*A*) A plot of refined B-factors—a measure of flexibility or dynamic mobility—for the KRAS structure. Agreement is evident between the X-ray crystal structure (pdb 5o2s) and the cryo-EM structure, which was built and refined de novo (after setting B-factors to a uniform starting value of 20 Å^2). The B-factors are averaged over individual amino acid residues and smoothed over a three-residue window, then normalized for direct comparison using the BANΔIT toolkit (32). The calculated correlation coefficient is 0.65. (*B*–*E*) Cryo-EM density maps around the single site mutations for KRAS G12V, G13C, G12C, and G12C bound to AMG510. A higher-than-average mobility of Cys12 is also reported by X-ray crystallography (pdb6oim).

The region around the bound GDP cofactor further emphasizes the atomic interpretability (Fig. 2). A Mg^{2+} ion bound near the terminal GDP phosphate group is also clearly visible. An interpretation of protein flexibility and dynamics from the cryo-EM map also agreed well with prior data, as revealed by an analysis of B-factors (or atomic displacement parameters). When examined across the length of the KRAS protein sequence, the correlation coefficient was 0.65 for the atomic structure obtained by cryo-EM compared to an earlier structure reported by X-ray crystallography (Fig. 3*A*). This highlights that the resolution and map quality obtained by cryo-EM are high enough to provide detailed atomic interpretation as well as potentially important information about conformational flexibility.

Structures of additional KRAS mutants provided further opportunities to evaluate atomic interpretability. Following similar protocols as for the G13C mutant, for the G12V mutant, we obtained a final map reconstruction with a resolution of 2.4 Å for the entire particle and 3.1 Å around the KRAS protein (*Materials and Methods*). For the G12C mutant, the resolution was 2.2 Å for the entire particle and 3.0 Å around the KRAS protein (*Materials and Methods*). The maps and refined KRAS structures were all closely comparable, with significant differences in the maps occurring only at the mutated amino acid side chains, as anticipated (Fig. 3). As an assessment of coordinate precision, the rms deviation between the two most closely related cryo-EM structures (the G12V and G12C mutants) was 0.58 Å; this is slightly less than the differences when compared to previously reported X-ray crystal structure, which are between 0.73 and 1.1 Å (*SI Appendix*, Table S2).

Conformational Variations and Drug Binding to KRAS G12C. A minor or "cryptic" cleft in the KRAS protein around residues 12 and 13 has been a site of intense focus for drug design efforts (27–29). Substantial protein conformational changes occur in that region upon drug binding; energetic and structural differences caused by drug binding stabilize the KRAS protein in its inactive form, which binds preferentially to GDP. Understanding the conformational and energetic landscape of the KRAS protein in this binding cleft region is expected to advance the discovery of new cancer drugs. Among drugs targeting clinically important KRAS mutations are a subset that form covalent bonds to cysteine mutants in that site.

As a test of our cryo-EM scaffold for analyzing KRAS drug binding, we determined the structure of the KRAS G12C mutant bound to the covalent inhibitor drug AMG510 [also known as sotorasib; (33)]. Following similar data processing protocols as before, from a set of 69,949 particle images obtained from 2072 cryo-EM movies, we obtained a density map with a resolution of 2.2 Å for the entire particle and 3.2 Å around the KRAS protein bound to AMG510. The map revealed significant conformation changes in the KRAS G12C mutant protein upon binding the AMG510 inhibitor compared to the G12C structure without drug bound. This was anticipated based on prior X-ray crystal structures showing conformational changes in this key region upon drug binding (28, 34–37). Most notable, however, is that the AMG510-bound structure we obtained by cryo-EM differs in the drug-binding region from the structure of the same complex reported earlier by X-ray crystallography protein structure database (PDB 6oim). The nominal resolution in the cryo-EM map is lower than that reported for the X-ray crystal structure (1.65 Å) (33), but the density is sufficiently well resolved to derive a conformation for bound AMG510 that is different from that observed in the crystallographic structure (*SI Appendix*, Fig. S9), especially at the covalent attachment point (residue 12) and the loop residues 60-GQEEYSAM-67 (Fig. 4). The torsion angle at the covalent bond between Cys12 and the drug molecule AMG510 differs by about 100° in the cryo-EM model from the conformation reported in a crystallographic model of the same drug complex (Fig. 4). A movement of ~ 2.7 Å is evident in regions of the drug molecule around the isopropyl pyridyl group, distal from the point of covalent attachment to Cys 12. We assessed the confidence in our modeling of the AMG510 drug molecule in a test in which we refined atomic models separately into density maps produced using two independent half-datasets. For the drug molecule, the differences between the independent models were only 0.1 to 0.3 Å. This is considerably smaller than the coordinate differences observed in comparison to the reported X-ray structure, which exceeded 2 Å, supporting the conclusion that meaningful differences are being revealed between the reported X-ray and cryo-EM conformations for drug binding (*SI Appendix*, Fig. S14).

Motivated by differences observed in the drug-binding pocket of the KRAS G12C mutant, we surveyed the PDB for examples of KRAS G12C bound to other inhibitors or drug molecules. An analysis of a set of 12 such structures (pdb 7a47, 6pgp, 6pgo, 8dnj, 8dnk, 8dni, 7a1y, 5v9o, 5v9l, 4lv6, 4luc, and 4lyh), all elucidated by X-ray crystallography, highlights a substantial degree of conformational variability for the KRAS protein in the binding region. Some of this variation is clearly the result of differences in the chemical structures of the various bound drugs. But there are unexpected patterns. Interestingly, whereas the cryo-EM structure reported here for the AMG510 drug

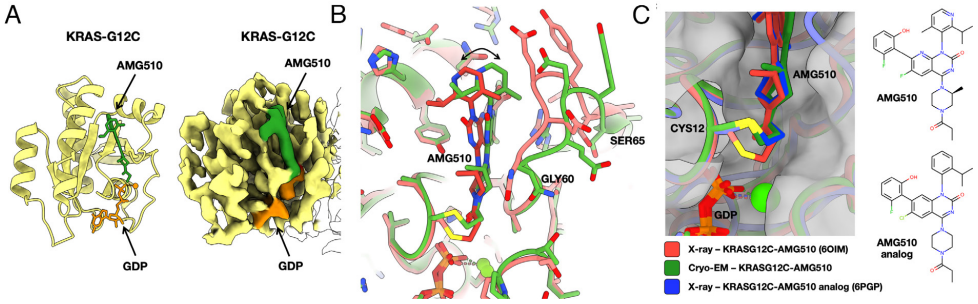

Fig. 4. Cryo-EM structure of KRAS G12C bound to AMG510. (*A*) A refined atomic model (*A*, *Left*) and a cryo-EM density map (*A*, *Right*) covering the KRAS protein, with the AMG510 drug molecule bound. The GDP ligand is shown in orange, and the AMG510 drug is in green. (*B*) Comparison between the cryo-EM structure and a prior X-ray crystal structure of KRAS G12C bound to AMG510. (*C*) Conformational variation at the covalent bond between Cys12 and the AMG510 and an AMG510 analog in X-ray and cryo-EM structures. At the thioether attachment, the cryo-EM model resembles an X-ray crystal structure of a complex with an AMG510 analog.

complex differs from a prior X-ray crystal structure of the identical complex (as discussed above), it matches more closely to an alternative X-ray crystal structure of a complex with a slightly different AMG510 analog (Fig. 4*C*). In particular, we note that the covalent attachment geometry for AMG 510 derived by cryo-EM occurs as well in the context of different drug bound complexes of KRAS G12C.

The findings on AMG510 binding suggest a substantial range of apparently low-energy conformations for the drug molecules and surrounding segments of the protein. The particular conformation observed appears to be affected at least in part by other molecular interactions. In the X-ray crystal structure, the drug-binding region (residues 62 to 73) is at a crystal packing interface (*SI Appendix*, Fig. S15*A*); conformational changes imposed by crystallographic molecular packing have long been studied and proven useful in uncovering conformational states involved in molecular function such as catalysis (38). Likewise, it is notable that in the cryo-EM structure, residue Met 67 is in contact with one of the DARPin domains protruding from the scaffold (*SI Appendix*, Fig. S15*B*). The observed variation across structures provides potentially useful insight into the conformational landscape for drug binding.

Conclusions

These initial structural findings serve as a starting point for deeper explorations of KRAS, and other small therapeutic protein targets, by cryo-EM scaffolding methods. Two immediate messages emerge. The first concerns feasibility. The rigidified scaffold described here provides a number of advantageous properties for cryo-EM structure determination—size, symmetry, and modular binding—making it suitable for future applications to many important systems. Second, the observation of conformational variability in drug binding emphasizes that cryo-EM approaches are likely to offer alternative structural views and distinct atomic frameworks for drug design efforts across broad areas of medicine.

Materials and Methods

Conformational Sampling of Rigidified Scaffolds. The N-terminal helix of DARP14–3G124Mut5 (12) was spatially aligned to the C-terminal helix of each subunit from the T33-51 cage (22). Using local programs, superpositions were performed between the first five helical residues of the DARPin to five residue windows from the terminal helical region of the protein cage, with different choices for the alignment segment from the protein cage. Following superposition, each conformation was evaluated for detrimental, overlapping collisions, and potentially favorable contacts in the fully assembled symmetric environment using local programs as well as visual inspection. Promising conformations–those where multiple protruding DARPin arms came into close proximity–were subjected to further conformational exploration by allowing for minor helix flexing. Modeling of allowable deviations from ideal alpha helix geometry was based on natural deviations observed in a large set of alpha helices extracted from high-resolution crystal structures.

Interface Design Calculation. All calculations were performed in the context of tetrahedral symmetry. For each sampled alignment and helical bend conformation, the resulting pose was relaxed into the REF2015 score function (39) using the FastRelax mover (40). Then, residues in the aligned helical fusion as well as any residues located in cage subunits or other DARPins (excluding variable loop regions) within 8 Å of the aligned DARPin were marked as designable. Further, all residues within 8 Å of designable residues were designated as packable. Sequence design trajectories were performed with a coordinate constraint applied to backbone atoms using Rosetta FastDesign with the InterfaceDesign2019 protocol (41) and REF2015 score function. We collected interface design metrics to quantify the resulting design success as compared to native interfaces (42). After analysis of the global design pool, we removed entire poses from consideration where the average design trajectory had a measured shape complementarity below 0.6, leaving eight viable poses for sampling sequence variations. Next, we ranked the design trajectories from each passing pose by applying a linear weighting scheme to the normalized metrics from each pose. These consisted of favoring fewer buried unsatisfied hydrogen bonds, lower interface energy (between complexed and unbound forms), higher interface shape complementarity, and lower interface solvation energy. Each normalized metric was equally weighted and summed to rank each trajectory. Finally, by examining the sequence diversity of the top candidates from each pose, we removed redundant sequence mutation patterns and selected 12 individual designs for characterization.

Protein Production. The sequences of the imaging scaffolds used in this paper are listed below. DNA fragments carrying the designed imaging scaffold sequences were synthesized (Integrated DNA Technologies and Twist Bioscience) and separately cloned into the vectors pET-22b (subunitB–DARPin) or pSAM (subunitA) (gifted from Jumi Shin, Addgene plasmid #45174; http://n2t.net/addgene:45174; RRID:Addgene_45174). The superfolder GFP V206A (sfGFP V206A) vector was previously described (12). DNA manipulations were carried out in *Escherichia coli* XL2 cells (Agilent). The proteins were expressed in *E. coli* BL21(DE3) cells (New England Biolabs) in Terrific Broth at 18 °C overnight with 0.5 mM IPTG induction at an OD_{600} of 1.0.

Upon collection of the cells, pellets were resuspended in buffer (50 mM Tris, 300 mM NaCl, 20 mM imidazole, pH 8.0) supplemented with benzonase nuclease, 1 mM PMSF, EDTA-free protease inhibitor cocktail (Thermo Scientific) and 0.1% LDAO and lysed using an EmulsiFlex C3 homogenizer (Avestin). The cell lysate was cleared by centrifugation at 20,000 × g for 20 min at 4 °C; the resulting supernatant was recovered and centrifuged at 10,000 × g for 10 min at 4 °C and then loaded onto a HisTrap column (GE Healthcare) pre-equilibrated with the same resuspension buffer. The imaging scaffold was eluted with a linear gradient to 300 mM imidazole. Upon elution, 5 mM EDTA and 5 mM BME were added immediately for designs 5, 8, 10, 13, and 14. The eluted proteins were concentrated using Amicon Ultra-15 100-kDa molecular weight cutoff for the imaging scaffold and 3-kDa molecular weight cutoff for the GFP protein. The concentrated proteins were further purified by size exclusion chromatography using a Superose six Increase column, eluted with 20 mM Tris pH 8.0, 100 mM NaCl, 5 mM BME, 5 mM EDTA for designs 5, 8, 10, 13, and 14 and 20 mM Tris, pH 8.0, and 100 mM NaCl for design 33. Chromatography fractions were analyzed by SDS–PAGE and negative stain EM for the presence of the imaging scaffold. KRAS G12V and KRAS G13C proteins were prepared as previously described by Kettle et al. (43).

The DNA sequence encoding wild-type KRAS (1 to 169) was synthesized (Genscript) and cloned into a pET28 vector with an N-terminal 6xHis tag followed by a TEV site. The G12C mutation was introduced using site-directed mutagenesis and confirmed by sequencing. Protein was expressed in BL21(DE3) cells in LB at 16 °C overnight, following induction at OD_{600} of 0.7 with 0.5 mM IPTG. After harvesting, cell pellets were resuspended in purification buffer (20 mM HEPES, pH 7.4, 300 mM NaCl, 0.5 mM TCEP, and 5 mM $MgCl_2$) supplemented with 1x EDTA-free protease inhibitor cocktail and 400 units benzonase and lysed by sonication. Cleared lysate was loaded onto a 1-mL HisTrap column (Cytiva), washed with 20 CV purification buffer +25 mM Imidazole, and eluted using an imidazole gradient to 500 mM Imidazole. Peak fractions were pooled, concentrated, and loaded onto a Superdex 75 Increase size-exclusion column in SEC buffer (purification buffer excluding $MgCl_2$). For AMG510-bound protein, KRAS G12C was incubated with AMG510 at a 2:1 molar ratio for 30 min and subjected to size-exclusion chromatography (Superose 6 Increase). Peak fractions yielded a mixture of AMG510-bound and free KRAS G12C (see *SI Appendix*, Fig. S10, first lane).

Either KRAS G12C or KRAS G12C-AMG510 was mixed with the imaging scaffold at a 2:1 molar ratio, incubated on ice for 5 min, and complex formation was confirmed through size-exclusion chromatography (Superose 6 Increase).

Negative Stain EM. The concentration of a 3.5-μL sample of fresh Superose six Increase eluent was adjusted to ~100 μg/mL, applied to glow-discharged Formvar/Carbon 400 mesh Cu grids (Ted Pella Inc) for 1 min and blotted to remove excess liquid. After a wash with filtered MilliQ water, the grid was stained with 2% uranyl acetate for 1 min. Images were taken on a Tecnai T12, a T20, a TF20, and a Talos F200C.

Cryo-EM Data Collection. Concentrated imaging scaffolds (1 to 10 mg/mL) were mixed with the GFP cargo or KRAS G13C/KRAS G12V/ KRAS G12C/KRAS

G12C-AMG510 to a molar ratio of 1:2 and diluted to a final concentration of 0.5 to 0.7 mg/mL. The final buffer composition was 20 mM Tris, pH 8.0, and 100 mM NaCl.

Quantifoil 300 mesh R2/2 copper grids were glow discharged for 30 s at 15 mA using a PELCO easiGLow (Ted Pella). A 1.8- to 3.5-μL volume of sample was applied to the grid at a temperature of 10 or 18 °C at ~100% relative humidity, followed by blotting and vitrification into liquid ethane using a Vitrobot Mark IV Thermo Fisher Scientific. Cryo-EM data were collected on an FEI Titan Krios cryoelectron microscope equipped with a Gatan K3 Summit direct electron detector and on a Titan Krios G4 cryoelectron microscope (Thermo Fisher Scientific) equipped with a Falcon4 direct electron detector in electron event registration mode. With the Gatan K3 Summit detector, movies were recorded with Leginon (44) and SerialEM (45) at a nominal magnification of 81,000× (calibrated pixel size of 1.1 Å per pixel) for designs 5, 8, 10, 13, 14, 33 (G13C) datasets and at a nominal magnification of 105,000× (calibrated pixel size of 0.856 Å per pixel) for design 33 (G12V) dataset, over a defocus range of −1.0 to −2.2 μm. With the Falcon4 detector, movies were recorded with the EPU automated acquisition software at a nominal magnification of 155,000× (calibrated pixel size of 0.5 Å per pixel), for design 33 (G12C and G12C-AMG510) datasets, over a target defocus range of −1.00 μm to −2.25 μm with increment steps of 0.25 μm and a total dose of 40 $e^{-}/Å^{2}$.

Fourier shell correlation (FSC) calculations are summarized in *SI Appendix*, Fig. S11. Plots showing dependence of resolution on the number of particles are shown in *SI Appendix*, Fig. S16.

Cryo-EM Data Processing and Model Building. Motion correction, CTF estimation, particle picking, 2D classification, and further data processing were performed with cryoSPARC v.3.2 (46). An initial set of particles was automatically picked using a blob-picker protocol. The extracted particles were 2D classified after which an ab initio reconstruction was generated. This reconstruction was then used for the 3D refinements enforcing T symmetry. The 3D structure was used to generate 2D projections of the particles and then used to repick the particles from the images using a template picker. The picked particles were extracted from the micrographs and went through 3D refinements enforcing T symmetry. The symmetry was then expanded, followed by further focused 3D classification without alignments and focused refinements using a mask encompassing the density for one DARPin and one cargo protein, GFP or KRAS, respectively. The best-resolved classes from the focused 3D classification were focused refined (C1 symmetry) performing local angular searches with the fulcrum at the center of mass of the mask. For the GFP imaging scaffold, we obtained an overall resolution of 2.7 Å for the entire particle and a resolution of 3.1 Å over the GFP protein, based on an FSC threshold of 0.143. For the KRAS G13C imaging scaffold, we obtained an overall resolution of 2.5 Å for the entire particle, and the resolution over the KRAS protein was 2.9 Å. We performed automatic de novo atomic model building into our KRAS G13C cryo-EM density using the program ModelAngelo (24) in the $COSMIC^{2}$ platform (47). The structure of GFP was built de novo using the automated chain tracing program, Buccaneer (48). The other three structures reported here were built starting from atomic models of close homologs, as noted in *SI Appendix*, Table S1. Manual adjustments to the models were performed using Coot (49), and automated refinement was performed using Phenix (50). Figures were prepared using ChimeraX (51, 52) and PyMOL (Version 2.0 Schrödinger, LLC)

Refinement into Half-Maps. We used refinement against independent half-maps (reconstructed from independent half-datasets) as an assessment of coordinate precision for the bound AMG510 drug molecule. Prior to independent real-space refinement, the molecules were subjected to computational simulated annealing–heating to 1,000 K and slow cooling to 300 K–in the program Phenix.

FSC Calculation. FSC plots were generated using the *mtriage* tool of Phenix (53). Each refined model and final map were submitted to *mtriage* along with two half-maps. Masked curves correspond to the use of a smoothed mask to perform FSC calculation only around the model (54).

Retrospective Test of Scaffold Structure Predictability by AI Methods. Given the important interplay between protein sequence design and protein structure prediction, we considered whether a leading machine learning algorithm, AlphaFold2 (55), would correctly predict the structure of our designed scaffold based on amino acid sequence. Such a success would argue that an unguided algorithm might have reached the same (or a similar) design result. A key element of the present scaffold design is the association of a homomeric protein trimer–based on a protein chain comprising a cage subunit fused to a DARPin–in such a fashion that stabilizing interactions occur between three copies of the DARPin; the trimer is mainly held together by association of the cage subunit component. When applied to our designed protein sequence, and specifying three chains to be associated, the AlphaFold2 program did not faithfully recapitulate the key stabilizing features between DARPins that were critical in rigidifying the scaffold to enable high-resolution imaging, and which were validated by cryo-EM. For example, residue ARG 254 was engineered to make a stabilizing interaction with residue ASP 181 from an adjacent DARPin. In our cryo-EM structure, those two residues come into atomic contact, as intended. In contrast, prediction by AlphaFold2 leaves those two residues ~15 Å apart, which is well beyond interaction distance. We furthermore attempted to use AlphaFold2 to computationally assemble the entire 24 subunit ($a_{12}b_{12}$) scaffold architecture given just the amino acid sequence information. That computational exercise did not assemble the cage subunits into a correct tetrahedral assembly. These results emphasize the importance in the present work of expert human input in the overall design strategy.

Data, Materials, and Software Availability. The structures of the imaging scaffolds and the protein targets, and their associated atomic coordinates, have been deposited into the Electron Microscopy Data Bank (EMDB) and the Protein Data Bank (PDB) with EMDB accession codes EMD-29700 (56), EMD-29713 (57), EMD-29715 (58), EMD-29718 (59), EMD-29719 (60), and EMD-29720 (61) and PDB accession codes 8G3K (62), 8G42 (63), 8G47 (64), 8G4E (65), 8G4F (66), and 8G4H (67), respectively. The sequences of the protein designs are included in *SI Appendix*.

ACKNOWLEDGMENTS. We thank David Strugatsky and Peng Ge for assistance in cryo-EM data collection acquired at the Electron Imaging Center for Nanomachines at the University of California, Los Angeles California for NanoSystems Institute, and Alison Berezuk for assistance with cryo-EM data collection carried out at University of British Columbia, Vancouver. We also thank Kevin Cannon, Ivo Atanasov, and Wong Hoi Hui for training in cryo-EM. We thank Yi Xiao Jiang, Tom Dendooven, Jack Bravo, and Yuval Mazor for helpful discussions about cryo-EM data processing, and Tom Ceska, Matt Lucas, and Lee Freiburger for helpful KRAS-related discussions. We thank Chris Garcia and Nathanael Caveney for discussions regarding cryo-EM scaffolding, and Alex Lisker for computing support. This work was supported by NIH grant R01GM129854 (T.O.Y.). Additional resources for sample preparation and electron microscopy screening were supported by DOE grant DE-FC02-02ER63421.

Author affiliations: [a]Department of Energy, Institute for Genomics and Proteomics, University of California, Los Angeles, CA 90095; [b]Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095; [c]Discovery Sciences, R&D, AstraZeneca, Cambridge CB2 0AA, United Kingdom; [d]Gandeeva Therapeutics, Inc., Burnaby, British Columbia V5C 6N5, Canada; and [e]Department of Biochemistry and Molecular Biology, The University of British Columbia, Vancouver, BC V6T 1Z3, Canada

1. R. Henderson, The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Q. Rev. Biophys.* **28**, 171–193 (1995).
2. M. A. Herzik, M. Wu, G. C. Lander, High-resolution structure determination of sub-100 kDa complexes using conventional cryo-EM. *Nat. Commun.* **10**, 1032 (2019).
3. T. O. Yeates, M. P. Agdanowski, Y. Liu, Development of imaging scaffolds for cryo-electron microscopy. *Curr. Opin. Struct. Biol.* **60**, 142–149 (2020).
4. F. Coscia *et al.*, Fusion to a homo-oligomeric scaffold allows cryo-EM analysis of a small protein. *Sci. Rep.* **6**, 30909 (2016).
5. P. A. Kratz, B. Böttcher, M. Nassal, Native display of complete foreign protein domains on the surface of hepatitis B virus capsids. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 1915–1920 (1999).
6. T. G. Martin *et al.*, Design of a molecular support for cryo-EM structure determination. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E7456–E7463 (2016).
7. C. McMahon *et al.*, Yeast surface display platform for rapid discovery of conformationally selective nanobodies. *Nat. Struct. Mol. Biol.* **25**, 289–296 (2018).
8. M. S. Morrison, T. Wang, A. Raguram, C. Hemez, D. R. Liu, Disulfide-compatible phage-assisted continuous evolution in the periplasmic space. *Nat. Commun.* **12**, 5959 (2021).
9. H. K. Binz *et al.*, High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat. Biotechnol.* **22**, 575–582 (2004).
10. S. Rothenberger *et al.*, The trispecific DARPin ensovibep inhibits diverse SARS-CoV-2 variants. *Nat. Biotechnol.* **40**, 1845–1854 (2022).

11. Y. Liu, S. Gonen, T. Gonen, T. O. Yeates, Near-atomic cryo-EM imaging of a small protein displayed on a designed scaffolding system. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 3362–3367 (2018).
12. Y. Liu, D. T. Huynh, T. O. Yeates, A 3.8 Å resolution cryo-EM structure of a small protein bound to an imaging scaffold. *Nat. Commun.* **10**, 1864 (2019).
13. Q. Yao, S. J. Weaver, J.-Y. Mock, G. J. Jensen, Fusion of DARPin to aldolase enables visualization of small protein by Cryo-EM. *Structure* **27**, 1148–1155.e3 (2019).
14. I. Vulovic *et al.*, Generation of ordered protein assemblies using rigid three-body fusion. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2015037118 (2021).
15. T. Uchański *et al.*, Megabodies expand the nanobody toolkit for protein structure determination by single-particle cryo-EM. *Nat. Methods* **18**, 60–68 (2021).
16. R. J. Cater *et al.*, Structural basis of omega-3 fatty acid transport across the blood–brain barrier. *Nature* **595**, 315–319 (2021).
17. X. Fan *et al.*, Single particle cryo-EM reconstruction of 52 kDa streptavidin at 3.2 angstrom resolution. *Nat. Commun.* **10**, 2386 (2019).
18. J. S. Bloch *et al.*, Development of a universal nanobody-binding Fab module for fiducial-assisted cryo-EM studies of membrane proteins. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2115435118 (2021).
19. X. Wu, T. A. Rapoport, Cryo-EM structure determination of small proteins by nanobody-binding scaffolds (Legobodies). *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2115001118 (2021).
20. K. Zhang *et al.*, Cryo-EM, protein engineering, and simulation enable the development of peptide therapeutics against acute myeloid leukemia. *ACS Cent. Sci.* **8**, 214–222 (2022).
21. Y. L. Boersma, A. Plückthun, DARPins and other repeat protein scaffolds: Advances in engineering and applications. *Curr. Opin. Biotechnol.* **22**, 849–857 (2011).
22. K. A. Cannon *et al.*, Design and structure of two new protein cages illustrate successes and ongoing challenges in protein engineering. *Protein Sci.* **29**, 919–929 (2020).
23. J.-D. Pédelacq, S. Cabantous, T. Tran, T. C. Terwilliger, G. S. Waldo, Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24**, 79–88 (2006).
24. K. Jamali, D. Kimanius, S. H. W. Scheres, A graph neural network approach to automated model building in cryo-EM maps. arXiv [Preprint] (2022). https://doi.org/10.48550/arXiv.2210.00006 (Accessed 7 February 2023).
25. S. Li, A. Balmain, C. M. Counter, A model for RAS mutation patterns in cancers: Finding the sweet spot. *Nat. Rev. Cancer* **18**, 767–777 (2018).
26. L. Huang, Z. Guo, F. Wang, L. Fu, KRAS mutation: From undruggable to druggable in cancer. *Signal Transduct. Target. Ther.* **6**, 1–20 (2021).
27. A. Mullard, Cracking KRAS. *Nat. Rev. Drug Discov.* **18**, 887–891 (2019).
28. J. M. Ostrem, U. Peters, M. L. Sos, J. A. Wells, K. M. Shokat, K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. *Nature* **503**, 548–551 (2013).
29. J. M. L. Ostrem, K. M. Shokat, Targeting KRAS G12C with covalent inhibitors. *Annu. Rev. Cancer Biol.* **6**, 49–64 (2022).
30. S. Guillard *et al.*, Structural and functional characterization of a DARPin which inhibits Ras nucleotide exchange. *Nat. Commun.* **8**, 16111 (2017).
31. N. Bery *et al.*, KRAS-specific inhibition using a DARPin binding to a site in the allosteric lobe. *Nat. Commun.* **10**, 2607 (2019).
32. F. Barthels, T. Schirmeister, C. Kersten, BAN∆IT: B'-Factor analysis for drug design and structural biology. *Mol. Inform.* **40**, 2000144 (2021).
33. J. Canon *et al.*, The clinical KRAS(G12C) inhibitor AMG 510 drives anti-tumour immunity. *Nature* **575**, 217–223 (2019).
34. M. Mathieu *et al.*, KRAS G12C fragment screening renders new binding pockets. *Small GTPases* **13**, 225–238 (2022).
35. B. A. Lanman *et al.*, Discovery of a covalent inhibitor of KRASG12C (AMG 510) for the treatment of solid tumors. *J. Med. Chem.* **63**, 52–65 (2020).
36. M. Zeng *et al.*, Potent and selective covalent quinazoline inhibitors of KRAS G12C. *Cell Chem. Biol.* **24**, 1005–1016.e3 (2017).
37. K. Zhu *et al.*, Modeling receptor flexibility in the structure-based design of KRASG12C inhibitors. *J. Comput. Aided Mol. Des.* **36**, 591–604 (2022).
38. M. Sawaya, J. Kraut, Loop and subdomain movements in the mechanism of Escherichia coli dihydrofolate reductase: Crystallographic evidence. *Biochemistry* **36**, 586–603 (1997).
39. R. F. Alford *et al.*, The rosetta all-atom energy function for macromolecular modeling and design. *J. Chem. Theory Comput.* **13**, 3031–3048 (2017).
40. L. G. Nivón, R. Moretti, D. Baker, A pareto-optimal refinement method for protein design scaffolds. *PLoS One* **8**, e59004 (2013).
41. J. B. Maguire *et al.*, Perturbing the energy landscape for improved packing during computational protein design. *Proteins* **89**, 436–449 (2021).
42. J. Janin, R. P. Bahadur, P. Chakrabarti, Protein–protein interaction and quaternary structure. *Q. Rev. Biophys.* **41**, 133–180 (2008).
43. J. G. Kettle *et al.*, Structure-based design and pharmacokinetic optimization of covalent allosteric inhibitors of the mutant GTPase KRASG12C. *J. Med. Chem.* **63**, 4468–4483 (2020).
44. C. Suloway *et al.*, Automated molecular microscopy: The new Leginon system. *J. Struct. Biol.* **151**, 41–60 (2005).
45. D. N. Mastronarde, Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
46. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
47. M. A. Cianfrocco, M. Wong-Barnum, C. Youn, R. Wagner, A. Leschziner, "COSMIC2: A science gateway for cryo-electron microscopy structure determination" in *Proceedings of the Practice and Experience in Advanced Research Computing 2017 on Sustainability, Success and Impact, PEARC17* (Association for Computing Machinery, 2017), pp. 1–5.
48. K. Cowtan, The buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1002–1011 (2006).
49. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
50. D. Liebschner *et al.*, Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* **75**, 861–877 (2019).
51. T. D. Goddard *et al.*, UCSF ChimeraX: Meeting modern challenges in visualization and analysis: UCSF ChimeraX visualization system. *Protein Sci.* **27**, 14–25 (2018).
52. E. F. Pettersen *et al.*, UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
53. P. V. Afonine *et al.*, New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallogr. D Struct. Biol.* **74**, 814–840 (2018).
54. G. Pintilie, D.-H. Chen, C. A. Haase-Pettingell, J. A. King, W. Chiu, Resolution and probabilistic models of components in CryoEM maps of mature P22 bacteriophage. *Biophys. J.* **110**, 827–839 (2016).
55. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
56. R. Castells-Graells, M. R. Sawaya, T. O. Yeates, EMD-29700, Cryo-EM imaging scaffold subunits A and B used to display KRAS G12C complex with GDP. Electron Microscopy Database. https://www.ebi.ac.uk/emdb/EMD-29700. Deposited 8 February 2023.
57. R. Castells-Graells, M. R. Sawaya, T. O. Yeates, EMD-29713, KRAS G12C complex with GDP imaged on a cryo-EM imaging scaffold. Electron Microscopy Database. https://www.ebi.ac.uk/emdb/EMD-29713. Deposited 8 February 2023.
58. R. Castells-Graells, M. R. Sawaya, T. O. Yeates, EMD-29715, KRAS G12C complex with GDP and AMG 510 imaged on a cryo-EM imaging scaffold. Electron Microscopy Database. https://www.ebi.ac.uk/emdb/EMD-29715. Deposited 8 February 2023.
59. R. Castells-Graells, M. R. Sawaya, T. O. Yeates, EMD-29718, Green Fluorescence Protein imaged on a cryo-EM imaging scaffold. Electron Microscopy Database. https://www.ebi.ac.uk/emdb/EMD-29718. Deposited 9 February 2023.
60. R. Castells-Graells, M. R. Sawaya, T. O. Yeates, EMD-29719, KRAS G12V complex with GDP imaged on a cryo-EM imaging scaffold. Electron Microscopy Database. https://www.ebi.ac.uk/emdb/EMD-29719. Deposited 9 February 2023.
61. R. Castells-Graells, M. R. Sawaya, T. O. Yeates, EMD-29720, KRAS G13C complex with GDP imaged on a cryo-EM imaging scaffold. Electron Microscopy Database. https://www.ebi.ac.uk/emdb/EMD-29720. Deposited 9 February 2023.
62. R. Castells-Graells, M. R. Sawaya, T. O. Yeates, 8G3K, Cryo-EM imaging scaffold subunits A and B used to display KRAS G12C complex with GDP. Protein Data Bank. https://www.rcsb.org/structure/8G3K. Deposited 8 February 2023.
63. R. Castells-Graells, M. R. Sawaya, T. O. Yeates, 8G42, KRAS G12C complex with GDP imaged on a cryo-EM imaging scaffold. Protein Data Bank. https://www.rcsb.org/structure/8G42. Deposited 8 February 2023.
64. R. Castells-Graells, M. R. Sawaya, T. O. Yeates, 8G47, KRAS G12C complex with GDP and AMG 510 imaged on a cryo-EM imaging scaffold. Protein Data Bank. https://www.rcsb.org/structure/8G47. Deposited 8 February 2023.
65. R. Castells-Graells, M. R. Sawaya, T. O. Yeates, 8G4E, Green Fluorescence Protein imaged on a cryo-EM imaging scaffold. Protein Data Bank. https://www.rcsb.org/structure/8G4E. Deposited 9 February 2023.
66. R. Castells-Graells, M. R. Sawaya, T. O. Yeates, 8G4F, KRAS G12V complex with GDP imaged on a cryo-EM imaging scaffold. Protein Data Bank. https://www.rcsb.org/structure/8G4F. Deposited 9 February 2023.
67. R. Castells-Graells, M. R. Sawaya, T. O. Yeates, 8G4H, KRAS G13C complex with GDP imaged on a cryo-EM imaging scaffold. Protein Data Bank. https://www.rcsb.org/structure/8G4H. Deposited 9 February 2023.

2.3 References

(1) Glaeser, R. M. How Good Can Single-Particle Cryo-EM Become? What Remains Before It Approaches Its Physical Limits? *Annu. Rev. Biophys.* **2019**, *48* (1), 45–61. https://doi.org/10.1146/annurev-biophys-070317-032828.

(2) Yeates, T. O.; Agdanowski, M. P.; Liu, Y. Development of Imaging Scaffolds for Cryo-Electron Microscopy. *Curr. Opin. Struct. Biol.* **2020**, *60*, 142–149. https://doi.org/10.1016/j.sbi.2020.01.012.

(3) Glaeser, R. M. How Good Can Cryo-EM Become? *Nat. Methods* **2016**, *13* (1), 28–32. https://doi.org/10.1038/nmeth.3695.

(4) Liu, Y.; Gonen, S.; Gonen, T.; Yeates, T. O. Near-Atomic Cryo-EM Imaging of a Small Protein Displayed on a Designed Scaffolding System. *Proc. Natl. Acad. Sci.* **2018**, *115* (13), 3362–3367. https://doi.org/10.1073/pnas.1718825115.

(5) Liu, Y.; Huynh, D. T.; Yeates, T. O. A 3.8 Å Resolution Cryo-EM Structure of a Small Protein Bound to an Imaging Scaffold. *Nat. Commun.* **2019**, *10* (1), 1864. https://doi.org/10.1038/s41467-019-09836-0.

(6) Castells-Graells, R.; Meador, K.; Arbing, M. A.; Sawaya, M. R.; Gee, M.; Cascio, D.; Gleave, E.; Debreczeni, J. É.; Breed, J.; Leopold, K.; Patel, A.; Jahagirdar, D.; Lyons, B.; Subramaniam, S.; Phillips, C.; Yeates, T. O. Cryo-EM Structure Determination of Small Therapeutic Protein Targets at 3 Å-Resolution Using a Rigid Imaging Scaffold. *Proc. Natl. Acad. Sci.* **2023**, *120* (37), e2305494120. https://doi.org/10.1073/pnas.2305494120.

Chapter 3 – Designed fluorescent protein cages as fiducial markers for targeted cell imaging

The following is a version of a preprint from:

bioRxiv : the preprint server for biology

2024.02.28.582585

Designed fluorescent protein cages as fiducial markers for targeted cell imaging

Morgan Gee[1,*], Kaiser Atai[2,3,4,*], Hilary A. Coller[3,4,5], Todd O. Yeates[1,5,6], Roger Castells-Graells[6]

[1]Department of Chemistry and Biochemistry, University of California, Los Angeles, CA, USA 90095

[2]Molecular Biology Interdepartmental Doctoral Program, University of California, Los Angeles, Los Angeles, CA, USA 90095

[3]Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, Los Angeles, CA, USA 90095

[4]Department of Biological Chemistry, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA 90095

[5]Molecular Biology Institute, University of California, Los Angeles, Los Angeles, CA, USA 90095

[6]UCLA-DOE Institute for Genomics and Proteomics, Los Angeles, CA, USA 90095

* Equal contributions

e-mail contact: rcastellsg@ucla.edu

Keywords: Protein design, cell imaging, nanobiology, protein cage, nanoparticles, fluorescent, fiducial markers, DARPin

ABSTRACT

Understanding how proteins function within their cellular environments is essential for cellular biology and biomedical research. However, current imaging techniques exhibit limitations, particularly in the study of small complexes and individual proteins within cells. Previously, protein cages have been employed as imaging scaffolds to study purified small proteins using cryo-electron microscopy (cryo-EM). Here we demonstrate an approach to deliver designed protein cages – endowed with fluorescence and targeted binding properties – into cells, thereby serving as fiducial markers for cellular imaging. We used protein cages with anti-GFP DARPin domains to target a mitochondrial protein (MFN1) expressed in mammalian cells, which was genetically fused to GFP. We demonstrate that the protein cages can penetrate cells, are directed to specific subcellular locations, and are detectable with confocal microscopy. This innovation represents a milestone in developing tools for in-depth cellular exploration, especially in conjunction with methods such as cryo-correlative light and electron microscopy (cryo-CLEM).

INTRODUCTION

Visualizing the spatial and temporal organization of proteins inside cells is essential for understanding their biological functions and their role in cellular processes [1]. *In vivo* imaging of proteins has played a crucial role in rapidly developing therapeutics, specifically in areas of research such as cancer biology and drug discovery [2–5]. While fluorescence microscopy is a widely used technique for protein imaging, its application is limited by the need for specific protein labeling and by the limited resolution and photophysical properties of fluorescent proteins [6,7]. Cryo-electron tomography (cryo-ET) offers higher resolution, but it is currently constrained to large assemblies due to the difficulty in identification of the molecule of interest in the crowded and complex environment of the cell [8–10].

To address these limitations, we have developed fluorescent designed protein cages that enable targeted imaging of proteins inside cells by acting as identifiable markers. Protein cages are hollow, nanoscale structures that can be designed to spontaneously self-assemble from multiple copies of modular protein subunits [11–17]. The resulting cage structures have well-defined sizes and shapes and can be engineered to display functional proteins on their surfaces and work as imaging scaffolds for cryo-electron microscopy (cryo-EM) [18–26].

The protein cages presented in this study are genetically fused to a designed ankyrin repeat protein (DARPin) [27]. Similar to antibody domains, DARPins have been exploited in recent work as a modular system for obtaining sequence variants that bind with high affinity to diverse proteins of interest; the DARPin in the present study binds to GFP [28]. By using a DARPin as the targeting moiety, we can image proteins of interest with high specificity without the need for direct genetic manipulation of the target protein or the chemical labeling of cells. Protein cages also offer other prospective advantages over traditional protein imaging methods, including high stability, low cytotoxicity, and biocompatibility [29]. They can be targeted to bind to any protein of interest, enabling imaging of previously inaccessible proteins. Additionally, our protein cages have a larger size compared to conventional fluorescent proteins, enabling improved localization accuracy.

In addition to fluorescence microscopy, our designed protein cages show potential in applications in high-resolution cryo-correlative light and electron microscopy (cryo-CLEM) [30]. As we successfully visualized the colocalization of the protein cages with their target protein in the cellular context with confocal microscopy, future applications in cryo-fluorescence microscopy (cryo-FLM) will allow for their use as a marker in cryo-CLEM. With the previous work on

developing protein cages for cryo-EM, protein cages are effective macromolecular structures for both aspects of cryo-FLM and cryo-EM. Moreover, the protein cages are designed to be large enough to be identifiable by the data processing software used to analyze tomograms obtained by electron microscopy, which could enable precise localization and 3D reconstruction of the target protein [8,9,31]. This further expands on the ability to use protein cages for cryo-ET as an additional alternative to cryo-EM. As many current probes for cryo-CLEM are inorganic, there has been recent interest in developing biomolecules such as nucleic acids [32,33] (van den Dries, K., 2022; Silvester, E., 2021). The protein cages would better maintain the native cellular environment compared to that of inorganic markers and would serve as a modular protein system to study proteins with cryo-CLEM. With their potential for cryo-CLEM and the targeting capabilities conferred by their DARPins, the protein cages address two prominent methods for improved sample identification [34].

We describe the development and characterization of protein cages designed for targeted imaging of proteins inside cells **(Fig. 1)**. We demonstrate that our protein cages can be delivered into mammalian cells, can specifically bind to proteins of interest in the cytosol and in organelles, and can be imaged using confocal microscopy. Overall, the development of designed protein cages for intracellular delivery offers a promising approach for studying proteins in their native cellular context and opens up new avenues to utilize methods like cryo-CLEM as a biomarker[23,25,35–43].

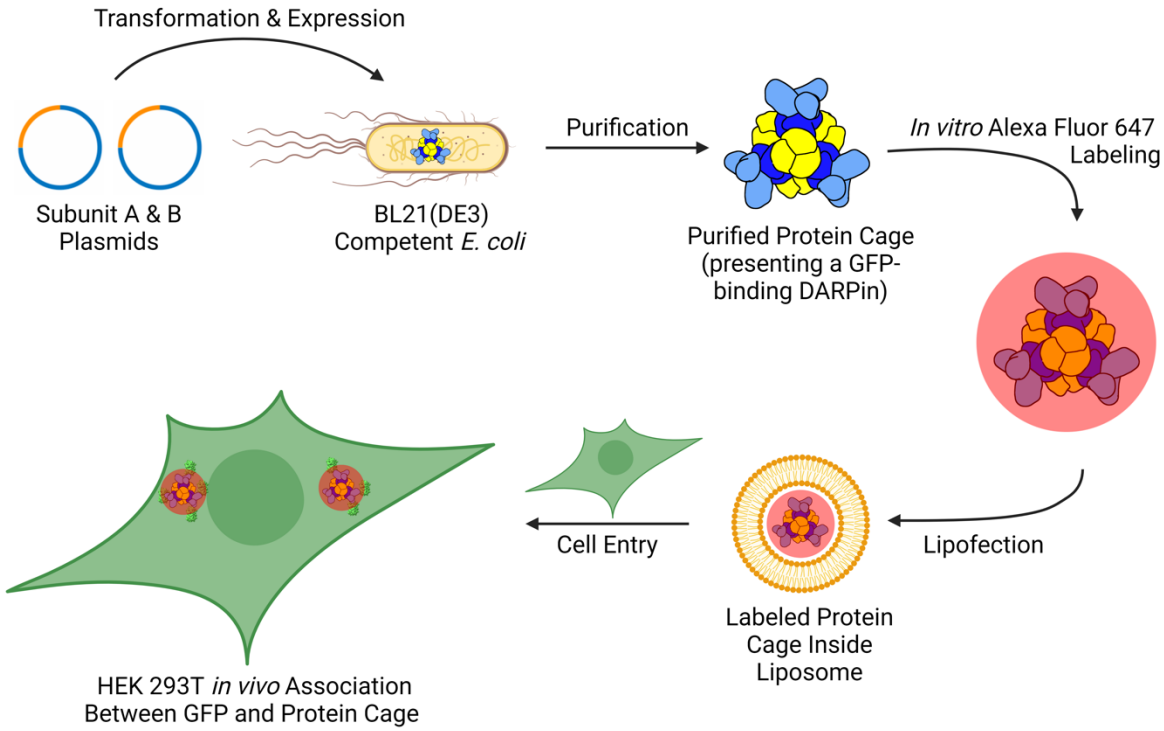

Figure 1. Workflow for the production, characterization, and lipofection of fluorescently labeled protein cages into HEK-293T cells. Plasmids containing the genes for the two subunits of the protein cage were transformed into *E. coli* cells followed by protein expression. The protein cages present a GFP-binding DARPin on their surface. Following the spontaneous, *in vivo* self-assembly of the protein cage, they were purified and fluorescently labeled with Alexa Fluor 647, a bright far-red fluorescent dye. These underwent lipofection to be transported into HEK-239T cells, integrating within liposomes as a payload for cell imaging. After an incubation period, colocalization of Alexa Fluor 647-labeled protein cages and endogenously expressed GFP within the HEK-239T cells occurred, which was visualized using confocal microscopy.

RESULTS

Characterization of Fluorescently Labeled Designed Protein Cages

As a basis for our nanoparticle imaging, we used a designed 12 nm tetrahedral protein cage ($A_{12}B_{12}$ subunit stoichiometry) with a genetically fused outward-facing DARPin that binds to green fluorescent protein (GFP), previously described in Castells-Graells *et al.*, 2023. We transformed the A and B subunit genes of the protein cage into *Escherichia coli* BL21(DE3) cells. We purified the protein cages from cell lysate using Ni-NTA affinity chromatography based on a polyhistidine tail on the A subunit. To be able to subsequently label the protein cages with Alexa Fluor 647 (AF647), we had to modify existing protocols used for protein cage purification to avoid amine-based buffers, as the fluorophore needs to react with lysine residues of the protein to form covalent amide bonds (see Methods).

Following Ni-NTA affinity chromatography, we conducted size exclusion chromatography (SEC) and identified the presence of three differently sized protein populations, where the first peak represents the whole protein cage, the second peak represents trimer complexes, and the third peak represents the monomeric subunits **(Fig. 2a)**.

The SEC plots demonstrate that there is little to no naturally occurring excitation at 650 nm, the absorbance maximum for AF647, exhibited by the protein cage subunits, allowing us to visualize successful labeling with fluorophore using both SEC and confocal microscopy **(Fig. 2a)**. SDS-PAGE analysis confirmed the presence of subunit A (19.3 kDa) and subunit B (35.2 kDa).

After the initial SEC purification, we labeled the purified protein cages with AF647 and performed another SEC purification to isolate the labeled complex **(Fig. 2b)**. We visualized the co-elution of the assembled protein cage and AF647 with two peaks at 280 nm and 650 nm, on the same fraction, for the protein cage and the fluorophore, respectively. This confirmed attachment of the AF647 and the protein cage. SDS-PAGE analysis confirmed the presence of the two distinct cage subunit types.

We also wanted to ensure that the assembled protein cages could properly associate with sfGFP, through binding to the genetically fused DARPin in the novel amine-free environment [26,44]. After *in vitro* incubation of purified sfGFP and purified protein cage, we purified the sample with SEC to remove any disassembled cages and any unbound sfGFP **(Fig. 2c).** sfGFP exhibits maximum absorption at 490 nm, allowing us to visualize complex formation between the sfGFP and assembled protein cage with co-elution of two peaks at 280 nm and 490 nm. Additionally, we noticed that there is no noticeable excitation at 650 nm in the absence of chemical fluorescent

labeling, suggesting that there is no significant excitation caused by the complex formation of protein cage and sfGFP that could interfere with detection of AF647. SDS-PAGE analysis confirmed the presence of the two subunits from the protein cage as well as sfGFP (26.6 kDa).

Binding of sfGFP to the protein cages does not have a detrimental effect on their assembly [19], however it was not known if the covalent bonding of AF647 could have any significant deleterious effects. We used negative stain electron microscopy to visualize the purified AF647-labeled protein cages and confirm proper cage assembly in the presence of chemical labelling **(Fig. 2d)**.

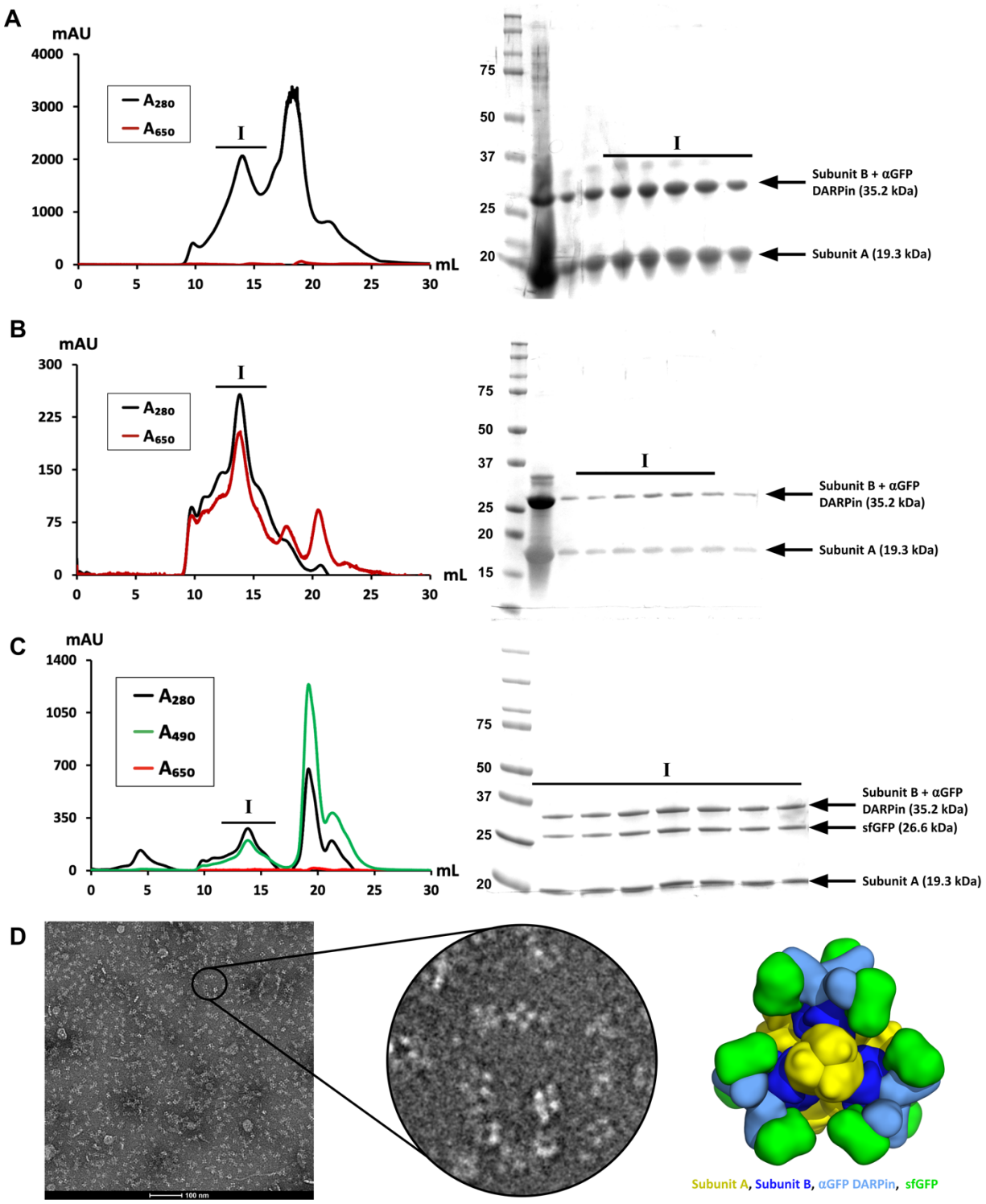

Figure 2. Characterization of fluorescently labeled protein cages. **(a)** SEC profile of unlabeled protein cage. Excitation at 650 nm was measured to determine naturally occurring absorbance that could interfere with detection of AF647. The SDS-PAGE band at 19.3 kDa supports the presence of subunit A, and the band at 35.2 kDa supports the presence of subunit B. **(b)** SEC profile of AF647-labeled cage. At peak I, there is

significant absorbance at both 280 nm and 650 nm, representing the co-elution and the complex formation of assembled protein cage and AF647. The SDS-PAGE band at 19.3 kDa supports the presence of subunit A, and the band at 35.2 kDa supports the presence of subunit B. **(c)** SEC profile of sfGFP-bound cage. At peak I, there is significant absorbance at both 280 nm and 490 nm, representing the co-elution and the complex formation of assembled protein cage and sfGFP. The SDS-PAGE band at 19.3 kDa supports the presence of subunit A, the band at 26.6 kDa supports the presence of sfGFP, and the band at 35.2 kDa supports the presence of subunit B. **(d)** Negative stain electron microscopy images of AF647-labeled protein cages. Imaged properly assembled AF647-labeled protein cages could be visualized in a relatively monodisperse environment. An enhanced image of various protein cages can be compared to a 3D-reconstruction of a native protein cage.

Internalization of Designed Fluorescent Protein Cages Into HEK-293T Cells

After the protein cages were successfully labeled with both AF647 and sfGFP, they were prepared for cellular experiments. For the first set, we wanted to determine if the protein cages could be properly internalized and, if so, whether the internalized protein cages were cytotoxic. To determine the capacity of protein cages to enter the cells, we used a lipofection-based approach using Lipofectamine CRISPRmax, which has been shown to effectively lipofect ribonucleoprotein complexes into mammalian cells [45]. We predicted that lipofection would allow the cages to be efficiently packaged into liposomes and released into the cytoplasm. As an independent mode for tracking the cages, we used their DARPin domains to bind them to as many as twelve GFP molecules each *in vitro*, prior to incorporation into HEK-293T cells. With this approach, any green fluorescent puncta we observe intracellularly would be indicative of GFP-bound cages entering the cell. We tested a range of protein cage concentrations (from 2 μg/mL to 250 μg/mL) with incubation for 4 hours **(Fig. 3a)**. After 4 hours of lipofection, we detected green fluorescent puncta in all the tested conditions, demonstrating that even at low concentrations the cages can be delivered into cells quickly and efficiently. As we increased the concentration of cages during lipofection, we found that there was a concomitant increase of intracellular green puncta, suggesting that the lipofection of these cages is tunable. Internalization of protein cages did not result in noticeable cytotoxicity.

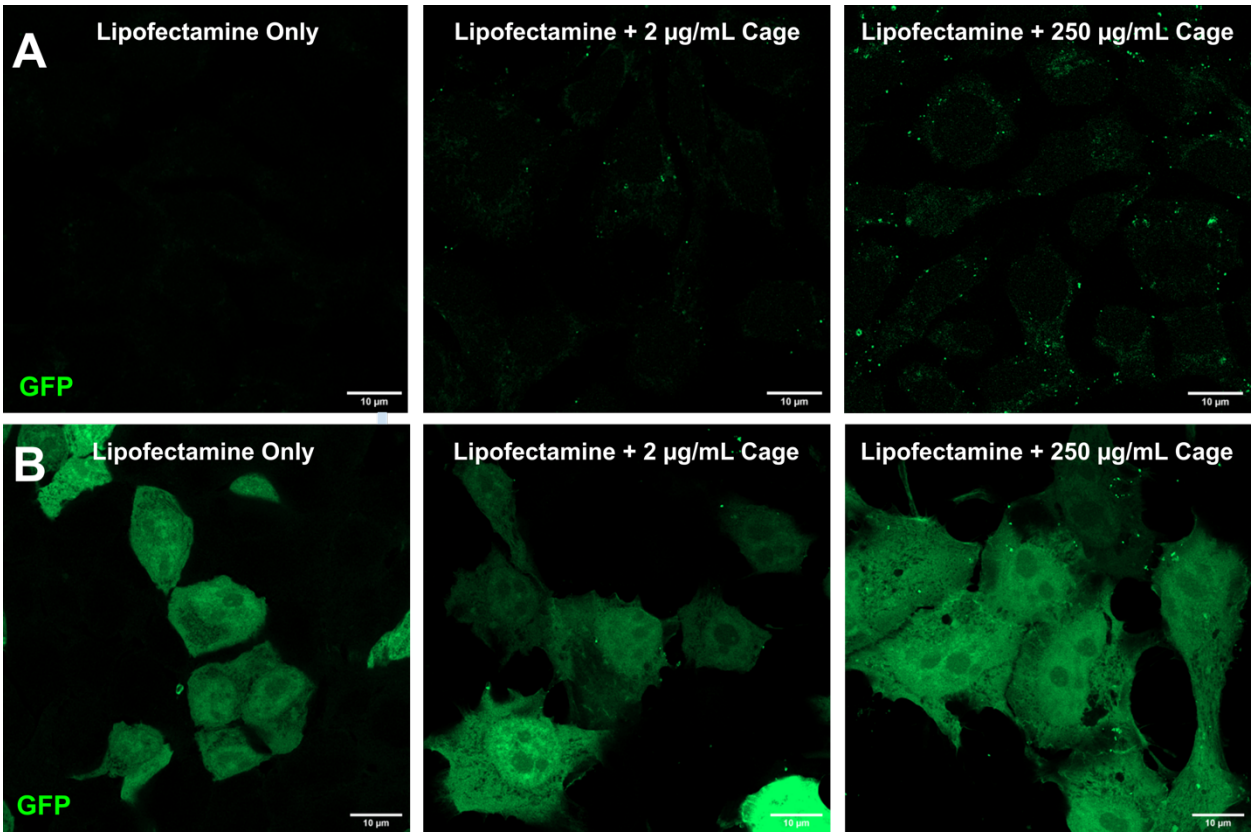

Figure 3. Visualization of green puncta as a result of internalized designed protein cages. **(a)** After incubation of HEK-293T cells with varying concentrations of sfGFP-bound protein cages, green puncta can be visualized. For the condition with only Lipofectamine and no protein cages, there are no visible green puncta, which confirms that the laser/filter used does not detect any fluorescence without protein cages. As the concentration of the protein cages increases, so does the number of green puncta, suggesting that there is a positive correlation between protein cages and number of green puncta. **(b)** After incubation of GFP-expressing HEK-293T cells with varying concentrations of "naked" protein cages, green puncta can be visualized. This suggests that the protein cages can escape the liposomes from the lipofection process and interact with cytoplasmic proteins.

Internalization of Designed Protein Cages into HEK-293T Cells and Binding to Endogenously Expressed GFP

After establishing that our designed protein cages can penetrate cells without any noticeable deleterious effects on the protein cages or on the cells, we wanted to determine if they are able to enter the cytoplasm. We theorized that if the cages are able to readily enter the cytoplasm, the exposed DARPins on the cages should nucleate any endogenously expressed GFP that is in the cytoplasm and produce bright green fluorescent puncta. Note, importantly, that each molecular cage binds (up to) 12 GFP proteins, making it feasible to detect individual

particles (which would typically not be possible for a single fluorescent protein). Alternatively, if the protein cages are unable to enter the cytoplasm and instead remain inside the liposomes, we would not visualize green puncta. To test this, we used the same lipofection approach as before, but instead used cages that were not loaded with GFP and therefore had DARPins that were accessible and unbound. We then lipofected these “naked” cages into HEK-293T cells that express cytoplasmic GFP **(Fig. 3b)**. After an incubation period, we were able to clearly visualize green puncta across all conditions with lipofected protein cages, while there was a clear absence of green puncta in the lipofectamine-only control. This demonstrates that even at low concentrations, protein cages can exit the liposome formed during the lipofection process and enter the cytoplasm to freely interact with any cytoplasmic protein to which they might be targeted (i.e. by their DARPin domain).

Internalization of Designed Alexa Fluor 647-labeled Protein Cages into HEK-293T Cells with Endogenous GFP

To confirm that the green puncta observed upon cage entry represent GFP bound to the cages, we performed colocalization experiments to visualize AF647-labeled protein cages and endogenously expressed GFP. Because AF647 has a maximum excitation of 650 nm and a maximum emission at 671 nm, we expect to see a clear difference in spectral emission compared to GFP. For these experiments, we followed the prior lipofection method but instead used cages that were fluorescently labeled with AF647. When transfecting the protein cages into HEK-293T cells expressing cytoplasmic GFP, we can clearly visualize AF647-positive puncta in the far-red spectral region, representing the fluorescent protein cages. We similarly see fluorescent puncta in the GFP-emission spectra, but only when cells are presented with the protein cages. Without the presence of the protein cages, the cytoplasmically expressed GFP and subsequent green fluorescent signal remains diffuse **(Fig. 4a-b)**. By merging the images, we can see the overlap of AF647 and green fluorescent puncta, demonstrating colocalization of the protein cages and endogenous GFP. This suggests that protein cages induce the formation of the green puncta by aggregating up to 12 endogenously expressed GFP molecules per protein cage **(Fig. 4b)**.

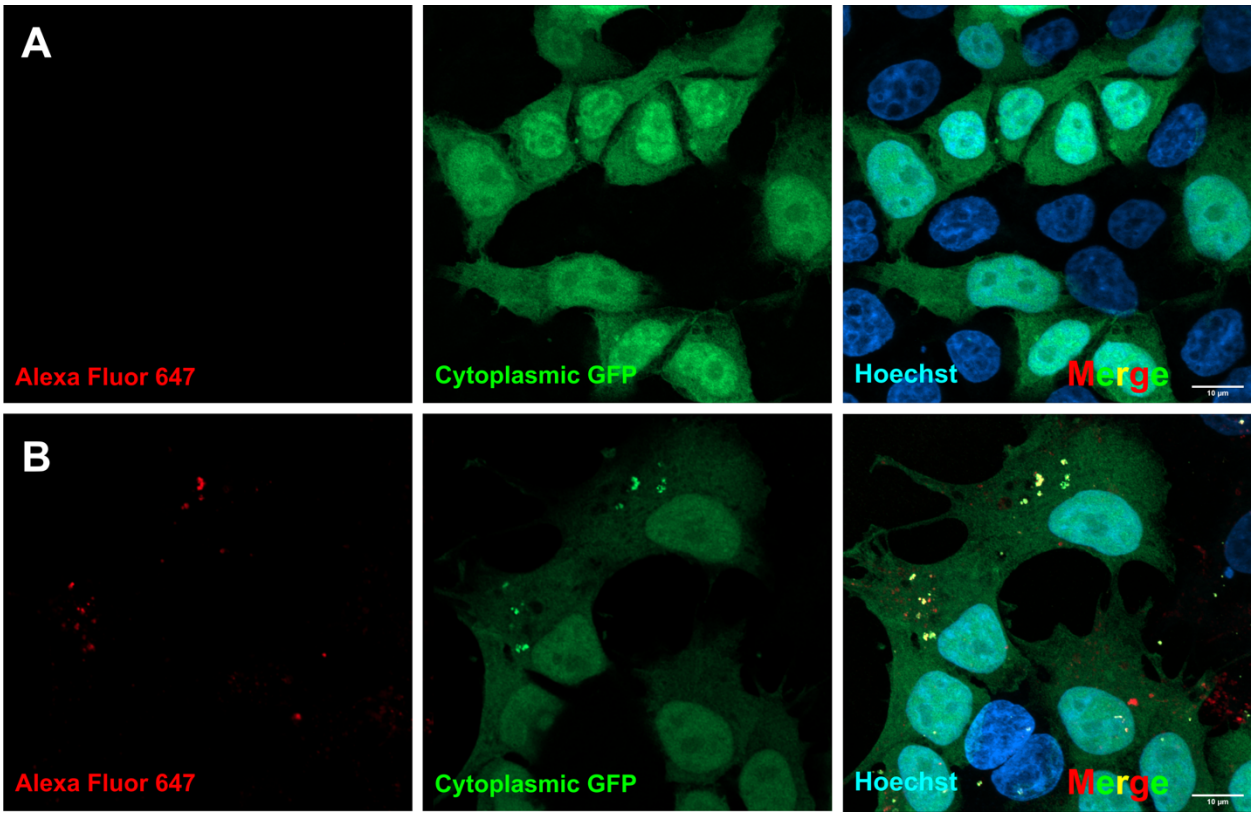

Figure 4. Visualization of colocalization of designed AF647-labeled protein cages and endogenous GFP. **(a)** Control experiment of cytoplasmic GFP without AF647-labeled protein cage. AF647, representing the fluorescently labeled protein cage, and cytoplasmic GFP can both be visualized using confocal microscopy. AF647 was not visualized by the confocal microscope due to the lack of protein cage. Without the protein cage, green puncta were unable to form and only diffuse cytoplasmic GFP can be visualized. **(b)** Colocalization of AF647-labeled protein cage and cytoplasmic GFP. Visualization of the AF647-labeled protein cage can be seen, resulting in the formation of green puncta. By merging the two images, colocalization of the protein cage and cytoplasmic GFP can be seen.

Targeted imaging of specific proteins in live cells with designed protein cages

Our data suggests that fluorescently labeled protein cages can enter cells and target cytoplasmic proteins. We wanted to further investigate whether these cages can diffuse through different parts of the cell and have the potential to be targeted to specific cell compartments. To test this, we expressed a transmembrane mitochondrial protein, MFN1, fused to GFP. These proteins are fused to GFP at the N-terminus such that the GFP is presented as a cytoplasmic domain. Expression of MFN1-GFP in HEK-293T cells presents bright green puncta throughout the cytoplasm, in agreement with the localization and abundance of mitochondria present in HEK-293T cells. Next, we hypothesized that cages, once released into the cytoplasm, will diffuse until they recognize and bind the exposed GFP molecule and thus localize to MFN1 proteins on

the surface of the mitochondria. Similarly, these cages are labeled with AF647 to enhance their visualization and localization in the cells.

After performing lipofection and introducing these labeled cages into the organelle-tagged cells, we found bright fluorescent puncta that colocalize between the green and far-red channels (**Fig. 5a-b**). This finding suggests that protein cages can be targeted to membrane-bound proteins belonging to specific cellular compartments.

To validate that the interaction is indeed specific to GFP and the DARPin domains on the protein cages, we also created HEK-293T cells expressing mNeonGreen (mNG) fused to MFN1. mNeonGreen is comparable in size and spectral properties to GFP but has almost no sequence similarity so we did not expect it to interact with the DARPins on the protein cages.

We found that, like MFN1-GFP, cells expressing MFN1-mNG also exhibit bright, green puncta reflective of mitochondrial distribution in the cell. However, when we transfected protein cages into HEK-293T cells expressing mNG-tagged MFN1, colocalization between the MFN1-mNG and the AF647 was not observed, indicating that cages labeled with the non-cognate DARPin (against GFP instead of mNG) were not targeted to the MFN1 protein or mitochondria (**Fig. 5c-d).** Furthermore, the similar spectral properties of mNeonGreen and GFP provide a useful control to rule out autofluorescence or crosstalk between the far-red and green fluorescent channels as a complicating factor.

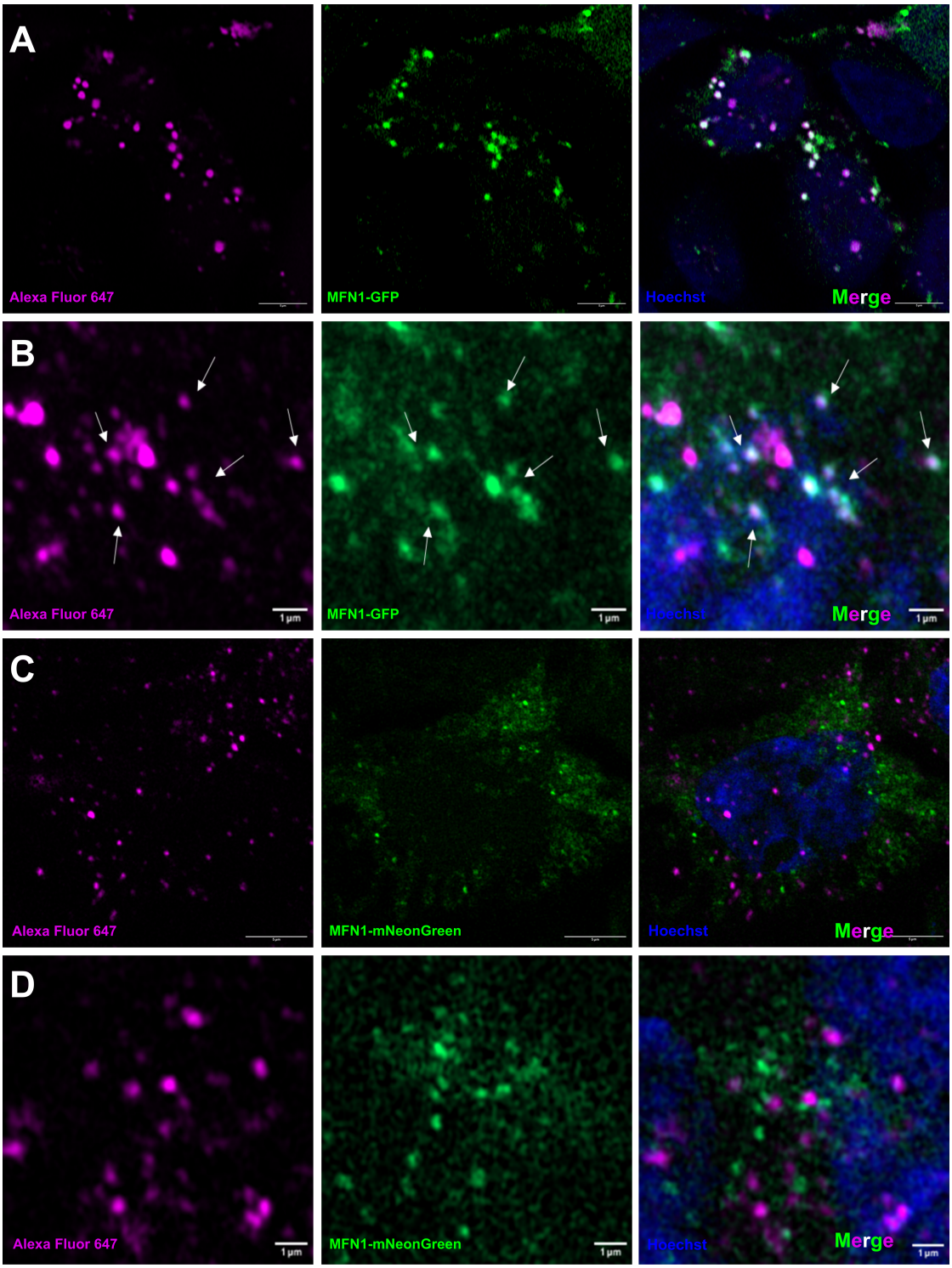

Figure 5. GFP-targeted cages are specific to GFP and do not display off-target binding to other fluorescent proteins. **(a-b)** Colocalization of AF647-labeled protein cages and the mitochondrial protein MFN1(labeled by GFP). By imaging AF647 emission (from the fluorescently labeled protein cage) and GFP emission (representing MFN1-GFP), and merging the two images, colocalization of the protein cage and GFP-tagged proteins can be visualized. **(c-d)** Lack of colocalization of AF647-labeled protein cages and MFN1-mNG. As the DARPins are unable to interact with mNG, the protein cages do not colocalize with MFN1.

DISCUSSION AND CONCLUSION

The central aim of this work was to develop innovative protein-based fiducial markers for cell imaging, with potential applications in molecular, cell, and structural biology research. We established that our large, designed, polyvalent protein cages could be internalized within HEK-293T cells and be readily visualized by fluorescence microscopy. The inclusion of DARPins in our protein cage design significantly enhances their modularity, allowing them to be targeted to specific cellular proteins or organelles. Using the mitochondrial transmembrane protein MFN1 as a first example, we showed that the location of this protein in the cell could be discerned. In this case study, the protein of interest (MFN1) was fused to GFP so that the initial protein cage construct (bearing an anti-GFP DARPin) could be used directly. For broader applications, including where genetic modification of a protein of interest is not possible or convenient, a suitably selected DARPin would be fused to the protein cage to direct it to the cellular protein of interest, in its native form.

The prospective avenues for further development are myriad. The large size of the protein cages permit for continued development for cryo-ET, while the successful visualization of protein cages *in vivo* suggests strong potential for cryo-FLM. Future research will focus on broader applications in mammalian cells along with more advanced imaging readouts in combination with cryo-CLEM.

MATERIALS AND METHODS

Protein Production:

The protein cage and the superfolder GFP V206A (sfGFP) were previously described [19,26]. The proteins were expressed overnight at 18°C in Terrific Broth (Thermo Scientific) and 0.4% v/v glycerol with 0.5 mM isopropyl ß-D-1-thiogalactopyranoside (IPTG) induction at an $O.D._{600nm}$ of 0.6-1.0. The cells were harvested by centrifugation at 4,000 xg for 20 minutes at 4°C.

For purification, cell pellets were resuspended in an amine-free extraction buffer [10 mM KCl, 220 mM NaCl, 50 mM $NaHCO_3$, 20 mM Imidazole pH 8.0] in addition to benzonase nuclease (EMD Millipore), 1 mM PMSF, 1X EDTA-free protease inhibitor cocktail (Thermo Scientific), and 0.1% LDAO. The cells were lysed using an EmulsiFlex C3 homogenizer (Avestin), and the cell lysate was centrifuged at 20,000 xg for 20 minutes at 4°C. The remaining supernatant was centrifuged again at 5,000 x g for 20 minutes at 4°C and equilibrated for 1 hour in 3 mL of Ni-NTA resin (Qiagen) that was pre-equilibrated in the extraction buffer. Using a gravity column, the protein cage was eluted with elution buffer [10 mM KCl, 220 mM NaCl, 50 mM $NaHCO_3$, 300 mM Imidazole pH 8.0] following a step elution model. Fractions containing the eluted protein cage were concentrated using an Amicon Ultra-15 100-kDa molecular weight cutoff centrifugal filter (MilliporeSigma), which were then purified with SEC. Using a Superose 6 Increase 10/300 GL column (Cytiva), the proteins were eluted with SEC buffer [5 mM KCl, 44 mM $NaHCO_3$, 110 mM NaCl]. The chromatography fractions were analyzed using SDS-PAGE and negative stain electron microscopy to ensure proper assembly.

Protein Labeling with Alexa Fluor 647 and sfGFP V206A:

Protein cages were labeled with AF647 Antibody Labeling Kits (Invitrogen). Over 500 µg of protein cage was incubated per reaction for 1 hour at 25°C with periodic inversion, followed by an overnight incubation at 4°C [46,47]. For sfGFP binding to the protein cage, a molar ratio of 2:1 of protein cage to sfGFP was mixed and incubated on ice for 5 minutes. Complex formation for both labeling systems was confirmed and isolated through SEC with a Superose 6 Increase 10/300 GL column.

Negative stain electron microscopy:

The concentration of a 3.5 µl sample of fresh SEC eluent was adjusted to ~100 µg/ml, applied to glow-discharged Formvar/Carbon 400 mesh Cu grids (Ted Pella Inc) for one minute and blotted to remove excess liquid. After a wash with filtered MilliQ water, the grid was stained with 2% uranyl acetate for one minute. Images were taken on a Tecnai T12 electron microscope.

Cell transformation and lipofection:

HEK-293T cells were grown under standard culture conditions at 5% CO2 with DMEM containing 10% fetal bovine serum. The cells were transfected with plasmids containing either GFP, MFN1-GFP, or MFN1-mNeonGreen, downstream of a CMV promoter to mark the cytoplasm or mitochondria, respectively.

Protein cages were packed into liposomes using Lipofectamine CRISPRmax Cas9 transfection reagents (Invitrogen). Protein cages were incubated with Lipofectamine CRISPRmax transfection reagents for 30 minutes in opti-MEM then added dropwise to cells. The lipofection mix was left on the cells for 4 hours, or 12 hours for MFN1 targeting experiments, before cells were fixed and processed for microscopy.

Confocal imaging:

HEK-293T cells were grown on 18 mm diameter circular coverslips in a 12-well dish in preparation for visualization. Coverslips were fixed with 4% PFA for 15 minutes then washed with PBS. Coverslips were then stained with Hoechst 33342 at a concentration of 1 µg/mL in PBS for 15 minutes then washed with PBS. Coverslips were then mounted with ProLong Glass Antifade Mountant (Thermo Scientific) on microscope slides. The mountant was left to cure overnight and confocal microscopy was performed the following day.

Super-resolution images were acquired on a Zeiss LSM 980 with Airyscan 2 using a 63x objective lens and appropriate lasers/filters for GFP, Hoechst, and AF647. Final images were analyzed using ImageJ [48].

Acknowledgments

We thank Michael Stowell for suggestions and advice on cell targeting and imaging. We thank Nika Gladkov and Michael Sawaya for helping with the preparation of the cartoon picturing the protein cage. We thank Shannon Chau for helping with the preparation of the graphical abstract.

Author Conflict Statement

The authors declare no competing financial interests.

Funding Support

This work was supported by the U. S. Department of Energy Office of Science award DE-FC02-02ER63421, NIH/NCI R01 CA221296-01A1 to HAC, Muscle Biology T32 2T32AR065972-06 to KA, and a Whitcome Predoctoral Fellowship to KA.

REFERENCES

(1) Hsiao, Y.-T.; Tsai, C.-N.; Chen, T.-H.; Hsieh, C.-L. Label-Free Dynamic Imaging of Chromatin in Live Cell Nuclei by High-Speed Scattering-Based Interference Microscopy. *ACS Nano* **2022**, *16* (2), 2774–2788. https://doi.org/10.1021/acsnano.1c09748.

(2) Condeelis, J.; Weissleder, R. In Vivo Imaging in Cancer. *Cold Spring Harb. Perspect. Biol.* **2010**, *2* (12), a003848. https://doi.org/10.1101/cshperspect.a003848.

(3) Hoffman, R. M. Strategies for In Vivo Imaging Using Fluorescent Proteins. *J. Cell. Biochem.* **2017**, *118* (9), 2571–2580. https://doi.org/10.1002/jcb.25677.

(4) Sun, N.; Malide, D.; Liu, J.; Rovira, I. I.; Combs, C. A.; Finkel, T. A Fluorescence-Based Imaging Method to Measure in Vitro and in Vivo Mitophagy Using Mt-Keima. *Nat. Protoc.* **2017**, *12* (8), 1576–1587. https://doi.org/10.1038/nprot.2017.060.

(5) Kim, J.; Ahn, S. B.; Hong, S.; Kim, K.; Ko, E. H.; Jo, I. J.; Chang, J.; Kim, M.; Lee, W.; Lee, H. Intracellular Dynamics-Resolved Label-Free Scattering Reveals Real-Time Metabolism of Single Bacteria. *Nano Lett.* **2023**, *23* (17), 8225–8232. https://doi.org/10.1021/acs.nanolett.3c02370.

(6) Díaz, M.; Malacrida, L. Advanced Fluorescence Microscopy Methods to Study Dynamics of Fluorescent Proteins In Vivo. In *Fluorescent Proteins: Methods and Protocols*; Sharma, M., Ed.; Methods in Molecular Biology; Springer US: New York, NY, 2023; pp 53–74. https://doi.org/10.1007/978-1-0716-2667-2_3.

(7) Wiedenmann, J.; Oswald, F.; Nienhaus, G. U. Fluorescent Proteins for Live Cell Imaging: Opportunities, Limitations, and Challenges. *IUBMB Life* **2009**, *61* (11), 1029–1042. https://doi.org/10.1002/iub.256.

(8) Fung, H. K. H.; Hayashi, Y.; Salo, V. T.; Babenko, A.; Zagoriy, I.; Brunner, A.; Ellenberg, J.; Müller, C. W.; Cuylen-Haering, S.; Mahamid, J. Genetically Encoded Multimeric Tags for Subcellular Protein Localization in Cryo-EM. *Nat. Methods* **2023**, *20* (12), 1900–1908. https://doi.org/10.1038/s41592-023-02053-0.

(9) Turk, M.; Baumeister, W. The Promise and the Challenges of Cryo-Electron Tomography. *FEBS Lett.* **2020**, *594* (20), 3243–3261. https://doi.org/10.1002/1873-3468.13948.

(10) Hylton, R. K.; Swulius, M. T. Challenges and Triumphs in Cryo-Electron Tomography. *iScience* **2021**, *24* (9), 102959. https://doi.org/10.1016/j.isci.2021.102959.
(11) Yeates, T. O. Geometric Principles for Designing Highly Symmetric Self-Assembling Protein Nanomaterials. *Annu. Rev. Biophys.* **2017**, *46* (1), 23–42. https://doi.org/10.1146/annurev-biophys-070816-033928.
(12) Yeates, T. O.; Liu, Y.; Laniado, J. The Design of Symmetric Protein Nanomaterials Comes of Age in Theory and Practice. *Curr. Opin. Struct. Biol.* **2016**, *39*, 134–143. https://doi.org/10.1016/j.sbi.2016.07.003.
(13) Padilla, J. E.; Colovos, C.; Yeates, T. O. Nanohedra: Using Symmetry to Design Self Assembling Protein Cages, Layers, Crystals, and Filaments. *Proc. Natl. Acad. Sci.* **2001**, *98* (5), 2217–2221. https://doi.org/10.1073/pnas.041614998.
(14) Haas, R. J. de; Brunette, N.; Goodson, A.; Dauparas, J.; Yi, S. Y.; Yang, E. C.; Dowling, Q.; Nguyen, H.; Kang, A.; Bera, A. K.; Sankaran, B.; Vries, R. de; Baker, D.; King, N. P. Rapid and Automated Design of Two-Component Protein Nanomaterials Using ProteinMPNN. bioRxiv August 4, 2023, p 2023.08.04.551935. https://doi.org/10.1101/2023.08.04.551935.
(15) Meador, K.; Castells-Graells, R.; Aguirre, R.; Sawaya, M. R.; Arbing, M. A.; Sherman, T.; Senarathne, C.; Yeates, T. O. A Suite of Designed Protein Cages Using Machine Learning Algorithms and Protein Fragment-Based Protocols. bioRxiv October 9, 2023, p 2023.10.09.561468. https://doi.org/10.1101/2023.10.09.561468.
(16) Edwardson, T. G. W.; Hilvert, D. Virus-Inspired Function in Engineered Protein Cages. *J. Am. Chem. Soc.* **2019**, *141* (24), 9432–9443. https://doi.org/10.1021/jacs.9b03705.
(17) Castells-Graells, R.; Laniado, J.; Meador, K.; Yeates, T. O. The Design of Self-Assembling Protein Cages and Other Reticular Protein Materials. In *Chemistry Challenges of the 21st Century*; WORLD SCIENTIFIC, 2023; pp 207–212. https://doi.org/10.1142/9789811282324_0022.
(18) Miller, J. E.; Castells-Graells, R.; Arbing, M. A.; Munoz, A.; Jiang, Y.-X.; Espinoza, C. T.; Nguyen, B.; Moroz, P.; Yeates, T. O. Design of Beta-2 Microglobulin Adsorbent Protein Nanoparticles. *Biomolecules* **2023**, *13* (7), 1122. https://doi.org/10.3390/biom13071122.
(19) Castells-Graells, R.; Meador, K.; Arbing, M. A.; Sawaya, M. R.; Gee, M.; Cascio, D.; Gleave, E.; Debreczeni, J. É.; Breed, J.; Leopold, K.; Patel, A.; Jahagirdar, D.; Lyons, B.; Subramaniam, S.; Phillips, C.; Yeates, T. O. Cryo-EM Structure Determination of Small Therapeutic Protein Targets at 3 Å-Resolution Using a Rigid Imaging Scaffold. *Proc. Natl. Acad. Sci.* **2023**, *120* (37), e2305494120. https://doi.org/10.1073/pnas.2305494120.
(20) McConnell, S. A.; Cannon, K. A.; Morgan, C.; McAllister, R.; Amer, B. R.; Clubb, R. T.; Yeates, T. O. Designed Protein Cages as Scaffolds for Building Multienzyme Materials. *ACS Synth. Biol.* **2020**, *9* (2), 381–391. https://doi.org/10.1021/acssynbio.9b00407.
(21) Gladkov, N.; Scott, E. A.; Meador, K.; Lee, E. J.; Laganowsky, A. D.; Yeates, T. O.; Castells-Graells, R. Design of a Symmetry-Broken Tetrahedral Protein Cage by a Method of Internal Steric Occlusion. bioRxiv November 9, 2023, p 2023.11.08.566319. https://doi.org/10.1101/2023.11.08.566319.
(22) Lee, E. J.; Gladkov, N.; Miller, J. E.; Yeates, T. O. Design of Ligand-Operable Protein-Cages That Open Upon Specific Protein Binding. *ACS Synth. Biol.* **2024**, *13* (1), 157–167. https://doi.org/10.1021/acssynbio.3c00383.
(23) Naskalska, A.; Borzęcka-Solarz, K.; Różycki, J.; Stupka, I.; Bochenek, M.; Pyza, E.; Heddle, J. G. Artificial Protein Cage Delivers Active Protein Cargos to the Cell Interior. *Biomacromolecules* **2021**, *22* (10), 4146–4154. https://doi.org/10.1021/acs.biomac.1c00630.
(24) Divine, R.; Dang, H. V.; Ueda, G.; Fallas, J. A.; Vulovic, I.; Sheffler, W.; Saini, S.; Zhao, Y. T.; Raj, I. X.; Morawski, P. A.; Jennewein, M. F.; Homad, L. J.; Wan, Y.-H.; Tooley, M. R.; Seeger, F.; Etemadi, A.; Fahning, M. L.; Lazarovits, J.; Roederer, A.; Walls, A. C.; Stewart, L.; Mazloomi, M.; King, N. P.; Campbell, D. J.; McGuire, A. T.; Stamatatos, L.; Ruohola-Baker, H.; Mathieu, J.; Veesler, D.; Baker, D. Designed Proteins Assemble Antibodies into Modular

Nanocages. *Science* **2021**, *372* (6537), eabd9994. https://doi.org/10.1126/science.abd9994.
(25) Dashti, N. H.; Abidin, R. S.; Sainsbury, F. Programmable In Vitro Coencapsidation of Guest Proteins for Intracellular Delivery by Virus-like Particles. *ACS Nano* **2018**, *12* (5), 4615–4623. https://doi.org/10.1021/acsnano.8b01059.
(26) Liu, Y.; Huynh, D. T.; Yeates, T. O. A 3.8 Å Resolution Cryo-EM Structure of a Small Protein Bound to an Imaging Scaffold. *Nat. Commun.* **2019**, *10* (1), 1864. https://doi.org/10.1038/s41467-019-09836-0.
(27) Binz, H. K.; Amstutz, P.; Kohl, A.; Stumpp, M. T.; Briand, C.; Forrer, P.; Grütter, M. G.; Plückthun, A. High-Affinity Binders Selected from Designed Ankyrin Repeat Protein Libraries. *Nat. Biotechnol.* **2004**, *22* (5), 575–582. https://doi.org/10.1038/nbt962.
(28) Brauchle, M.; Hansen, S.; Caussinus, E.; Lenard, A.; Ochoa-Espinosa, A.; Scholz, O.; Sprecher, S. G.; Plückthun, A.; Affolter, M. Protein Interference Applications in Cellular and Developmental Biology Using DARPins That Recognize GFP and mCherry. *Biol. Open* **2014**, *3* (12), 1252–1261. https://doi.org/10.1242/bio.201410041.
(29) Edwardson, T. G. W.; Levasseur, M. D.; Tetter, S.; Steinauer, A.; Hori, M.; Hilvert, D. Protein Cages: From Fundamentals to Advanced Applications. *Chem. Rev.* **2022**, *122* (9), 9145–9197. https://doi.org/10.1021/acs.chemrev.1c00877.
(30) Tuijtel, M. W.; Koster, A. J.; Jakobs, S.; Faas, F. G. A.; Sharp, T. H. Correlative Cryo Super-Resolution Light and Electron Microscopy on Mammalian Cells Using Fluorescent Proteins. *Sci. Rep.* **2019**, *9* (1), 1369. https://doi.org/10.1038/s41598-018-37728-8.
(31) Herzik, M. A.; Wu, M.; Lander, G. C. High-Resolution Structure Determination of Sub-100 kDa Complexes Using Conventional Cryo-EM. *Nat. Commun.* **2019**, *10* (1), 1032. https://doi.org/10.1038/s41467-019-08991-8.
(32) Silvester, E.; Vollmer, B.; Pražák, V.; Vasishtan, D.; Machala, E. A.; Whittle, C.; Black, S.; Bath, J.; Turberfield, A. J.; Grünewald, K.; Baker, L. A. DNA Origami Signposts for Identifying Proteins on Cell Membranes by Electron Cryotomography. *Cell* **2021**, *184* (4), 1110-1121.e16. https://doi.org/10.1016/j.cell.2021.01.033.
(33) Van Den Dries, K.; Fransen, J.; Cambi, A. Fluorescence CLEM in Biology: Historic Developments and Current Super-resolution Applications. *FEBS Lett.* **2022**, *596* (19), 2486–2496. https://doi.org/10.1002/1873-3468.14421.
(34) Nogales, E.; Mahamid, J. Bridging Structural and Cell Biology with Cryo-Electron Microscopy. *Nature* **2024**, *628* (8006), 47–56. https://doi.org/10.1038/s41586-024-07198-2.
(35) Brillault, L.; Jutras, P. V.; Dashti, N.; Thuenemann, E. C.; Morgan, G.; Lomonossoff, G. P.; Landsberg, M. J.; Sainsbury, F. Engineering Recombinant Virus-like Nanoparticles from Plants for Cellular Delivery. *ACS Nano* **2017**, *11* (4), 3476–3484. https://doi.org/10.1021/acsnano.6b07747.
(36) Steinmetz, N. F.; Lim, S.; Sainsbury, F. Protein Cages and Virus-like Particles: From Fundamental Insight to Biomimetic Therapeutics. *Biomater. Sci.* **2020**, *8* (10), 2771–2777. https://doi.org/10.1039/D0BM00159G.
(37) Sainsbury, F. Virus-like Nanoparticles: Emerging Tools for Targeted Cancer Diagnostics and Therapeutics. *Ther. Deliv.* **2017**, *8* (12), 1019–1021. https://doi.org/10.4155/tde-2017-0098.
(38) Wang, Y.; Uchida, M.; Waghwani, H. K.; Douglas, T. Synthetic Virus-like Particles for Glutathione Biosynthesis. *ACS Synth. Biol.* **2020**, *9* (12), 3298–3310. https://doi.org/10.1021/acssynbio.0c00368.
(39) Chan, S. K.; Steinmetz, N. F. microRNA-181a Silencing by Antisense Oligonucleotides Delivered by Virus-like Particles. *J. Mater. Chem. B* **2023**, *11* (4), 816–825. https://doi.org/10.1039/D2TB02199D.
(40) Chung, Y. H.; Ortega-Rivera, O. A.; Volckaert, B. A.; Jung, E.; Zhao, Z.; Steinmetz, N. F. Viral Nanoparticle Vaccines against S100A9 Reduce Lung Tumor Seeding and Metastasis. *Proc. Natl. Acad. Sci.* **2023**, *120* (43), e2221859120. https://doi.org/10.1073/pnas.2221859120.
(41) Antanasijevic, A.; Ueda, G.; Brouwer, P. J. M.; Copps, J.; Huang, D.; Allen, J. D.; Cottrell, C.

A.; Yasmeen, A.; Sewall, L. M.; Bontjer, I.; Ketas, T. J.; Turner, H. L.; Berndsen, Z. T.; Montefiori, D. C.; Klasse, P. J.; Crispin, M.; Nemazee, D.; Moore, J. P.; Sanders, R. W.; King, N. P.; Baker, D.; Ward, A. B. Structural and Functional Evaluation of de Novo-Designed, Two-Component Nanoparticle Carriers for HIV Env Trimer Immunogens. *PLOS Pathog.* **2020**, *16* (8), e1008665. https://doi.org/10.1371/journal.ppat.1008665.

(42) Esquirol, L.; McNeale, D.; Douglas, T.; Vickers, C. E.; Sainsbury, F. Rapid Assembly and Prototyping of Biocatalytic Virus-like Particle Nanoreactors. *ACS Synth. Biol.* **2022**, *11* (8), 2709–2718. https://doi.org/10.1021/acssynbio.2c00117.

(43) Wang, J.; Li, Y.; Nie, G. Multifunctional Biomolecule Nanostructures for Cancer Therapy. *Nat. Rev. Mater.* **2021**, *6* (9), 766–783. https://doi.org/10.1038/s41578-021-00315-x.

(44) Pédelacq, J.-D.; Cabantous, S.; Tran, T.; Terwilliger, T. C.; Waldo, G. S. Engineering and Characterization of a Superfolder Green Fluorescent Protein. *Nat. Biotechnol.* **2006**, *24* (1), 79–88. https://doi.org/10.1038/nbt1172.

(45) Cardarelli, F.; Digiacomo, L.; Marchini, C.; Amici, A.; Salomone, F.; Fiume, G.; Rossetta, A.; Gratton, E.; Pozzi, D.; Caracciolo, G. The Intracellular Trafficking Mechanism of Lipofectamine-Based Transfection Reagents and Its Implication for Gene Delivery. *Sci. Rep.* **2016**, *6* (1), 25879. https://doi.org/10.1038/srep25879.

(46) Berlier, J. E.; Rothe, A.; Buller, G.; Bradford, J.; Gray, D. R.; Filanoski, B. J.; Telford, W. G.; Yue, S.; Liu, J.; Cheung, C.-Y.; Chang, W.; Hirsch, J. D.; Beechem Rosaria P. Haugland, J. M.; Haugland, R. P. Quantitative Comparison of Long-Wavelength Alexa Fluor Dyes to Cy Dyes: Fluorescence of the Dyes and Their Bioconjugates. *J. Histochem. Cytochem.* **2003**, *51* (12), 1699–1712. https://doi.org/10.1177/002215540305101214.

(47) Panchuk-Voloshina, N.; Haugland, R. P.; Bishop-Stewart, J.; Bhalgat, M. K.; Millard, P. J.; Mao, F.; Leung, W.-Y.; Haugland, R. P. Alexa Dyes, a Series of New Fluorescent Dyes That Yield Exceptionally Bright, Photostable Conjugates. *J. Histochem. Cytochem.* **1999**, *47* (9), 1179–1188. https://doi.org/10.1177/002215549904700910.

(48) Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W. NIH Image to ImageJ: 25 Years of Image Analysis. *Nat. Methods* **2012**, *9* (7), 671–675. https://doi.org/10.1038/nmeth.2089.

Chapter 4 – Conclusions and Discussion

In this thesis, I delineate the applications of protein nanoparticles in structural, cell, and molecular biology research. Protein nanoparticles were used early on as imaging scaffolds for cryo-EM for the structural determination of small proteins such as GFP. By improving the rigidity of the protein cage by interfacing the external DARPin protein domains, we increased the attainable resolution of a target cargo protein to 3 Å from the previous 3.8 Å. The rigidified imaging scaffold was used to resolve both the structures of key oncogenic protein KRAS and mutational variants bound to a cancer inhibitor drug. In further work, we developed protein nanoparticles as fluorescent, fiducial biomarkers for cellular imaging of target proteins *in vivo*. After cellular uptake of an Alexa Fluor 647-labeled protein cage in HEK-293T cells through lipofection, we successfully visualized its colocalization with both endogenous cytoplasmic GFP and MFN1-GFP.

Protein nanoparticles continue to advance structural biology research and have the potential to revolutionize adjacent fields such as oncology. Applications of imaging scaffolds are nearly limitless as DARPins can be engineered to study the expansive set of therapeutic protein targets with cryo-EM. Current results of protein nanoparticles as *in vivo* biomarkers suggest opportunities for cryo-ET, cryo-FLM, and cryo-CLEM within the cellular context. Other future directions include their applications as immunogens and vehicles for targeted drug delivery.