Published in final edited form as:

Structure. 2019 July 02; 27(7): 1148–1155.e3. doi:10.1016/j.str.2019.04.003.

Fusion of DARPin to aldolase enables visualization of small protein by cryoEM

Qing Yao[1,4], **Sara J. Weaver**[2,4], **Jee-Young Mock**[2], and **Grant J. Jensen**[1,3,5,*]

[1]Division of Biology and Biological Engineering, California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125

[2]Division of Chemistry and Chemical Engineering, California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125

[3]Howard Hughes Medical Institute, 1200 E. California Blvd., Pasadena, CA 91125

[5]Lead Contact

Summary

Solving protein structures by single particle cryo-electron microscopy (cryoEM) has become a crucial tool in structural biology. While exciting progress is being made towards the visualization of small macromolecules, the median protein size in both eukaryotes and bacteria is still beyond the reach of cryoEM. To overcome this problem, we implemented a platform strategy where a small protein target was rigidly attached to a large, symmetric base via a selectable adapter. Of our seven designs, the best construct used designed ankyrin repeat protein (DARPin) rigidly fused to tetrameric rabbit muscle aldolase through a helical linker. The DARPin retained its ability to bind its target: green fluorescent protein (GFP). We solved the structure of this complex to 3.0 Å resolution overall, with 5 to 8 Å resolution in the GFP region. As flexibility in the DARPin position limited the overall resolution of the target, we describe strategies to rigidify this element.

Graphical Abstract

[*]Corresponding author: jensen@caltech.edu (GJJ).

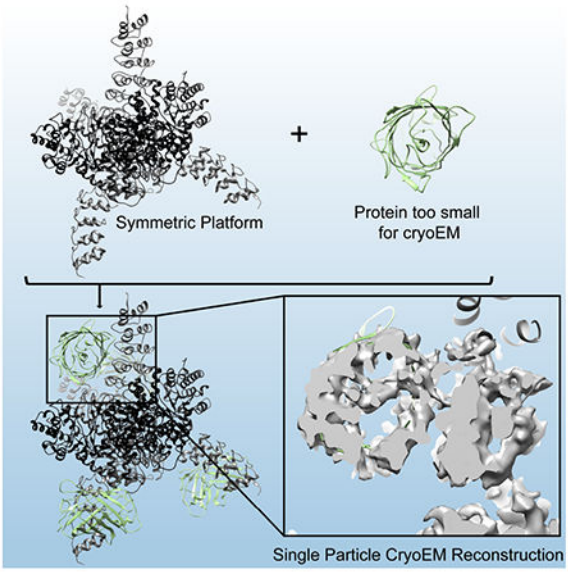
[4]These authors contributed equally to this work

Author Contributions

Conceptualization, Q.Y., S.J.W., G.J.J.; Methodology, Q.Y., S.J.W., J.Y.M; Investigation, Q.Y., S.J.W., J.Y.M; Data Curation, S.J.W.; Writing – Original Draft, S.J.W.; Writing – Review & Editing, S.J.W., Q.Y., G.J.J.; Visualization, S.J.W., Q.Y.; Supervision, G.J.J., Funding Acquisition, G.J.J.

Declaration of interests

The authors declare no competing interests.



eTOC

Proteins <100 kDa are difficult to solve by single particle cryoEM. Here, Yao et al explored seven constructs to display a target protein. Using GFP as a test case, they found a DARPin-aldolase fusion capable of cryoEM visualization of the target. Flexibility in the DARPin position limited the resolution.

Keywords

CryoEM; DARPin; Single particle analysis; Platform; Aldolase

Introduction

Single particle cryoEM can reveal the structures of large macromolecular complexes to near atomic resolution. To solve a protein structure by single particle cryoEM, purified proteins are rapidly frozen in a thin layer of vitreous ice. A transmission electron microscope is used to collect projection images of the protein. Individual proteins are identified in the ice and their orientations are computationally determined. The projection images are then combined to calculate a 3D reconstruction of the protein.

A fundamental challenge in single particle cryoEM is that small proteins do not produce enough contrast in noisy projection images to precisely determine their orientation. Richard Henderson estimated that with ideal images, a 3 Å structure could be reconstructed for a 40 kDa protein (Henderson, 1995). Unfortunately, real electron micrographs are imperfect so this theoretical minimum of macromolecular size has never been reached. The smallest protein to be solved to near atomic resolution so far by cryoEM is hemoglobin (64 kDa) (Khoshouei et al., 2017), but the median protein lengths in both bacteria (27 kDa) and eukaryotes (36 kDa) are about two times smaller (Brocchieri and Karlin, 2005).

Consequently, many proteins in biology are beyond the reach of high-resolution structure determination by single particle cryoEM.

Over the years, several strategies to overcome the size limit problem in single particle cryoEM have been suggested. Two major themes have emerged to increase the target mass and improve its orientation determination. First, the target can be decorated with antibody fragments (Jensen and Kornberg, 1998) (Wu et al., 2012). Second, the target can be rigidly attached to a large platform protein. The platform is typically composed of a base protein and an adapter. The purpose of the base protein is to increase the molecular weight, which facilitates accurate particle picking and precise particle orientation determination. The adapter can be customized (a covalent fusion between the target and the base) or general (a selectable adapter that facilitates non-covalent binding of the target to the platform base). Covalent approaches have utilized direct fusions between the target protein and the base either via a flexible linker adapter (Kratz et al., 1999) or a helical junction adapter (Coscia et al., 2016; Liu et al., 2018a). For the platform to be successful, the adapter must be rigidly attached to the base. The flexible linker adapter was therefore insufficient to determine the structure of the target (Kratz et al., 1999), but the use of a helix-forming peptide linker (Liu et al., 2018a; Padilla et al., 2001) or direct concatenation of two helices (Coscia et al., 2016; Jeong et al., 2016) has shown promise. Most recently, Liu et al. demonstrated that a rigid, continuous α-helix could be formed by linking the terminal α-helices of a designed ankyrin repeat protein (DARPin) and a nanocage subunit through a helix-forming peptide linker (Padilla et al., 2001). Notably, Liu et al were able to show the structure of the 17 kDa DARPin to 3.5 to 5 Å local resolution (Liu et al., 2018a). Unfortunately, these strategies are limited to target proteins with a terminal α-helix, and their implementation requires that the length of the helical junction adapter must be customized for each new target. Utilizing a non-covalent platform strategy with a selectable adapter (like an antibody or a DARPin) has the potential to be generally applicable, regardless of the structure of the target, since the selectable adapter could be raised against any target using phage display, while the invariant nature of the adapter framework region would allow the one-time optimization of a rigid attachment point between it and the base. Along these lines, Liu et al. suggested that their DARPin-nanocage could display a small protein for structure determination by cryoEM (Liu et al., 2018a). A follow up to Liu et al., 2018 was posted to the bioRxiv while this manuscript was under review (Liu et al., 2018b). In the follow up paper, Liu and coworkers reported the fusion of their previously reported nanocage with a DARPin against GFP. They engineered the DARPin to have additional stabilizing mutations designs (Kramer et al., 2010) and achieved a near atomic resolution structure of the target GFP.

Here we report the outcomes of a variety of new designs.

Results

Platform strategy and design

The goal of our study was to design a generally applicable platform to solve small protein structures by single particle cryoEM. We explored several candidate base proteins and selectable adapters (Figure 1A, S1A). We favored bases that were easy to purify and that had

already been solved to high-resolution by single particle cryoEM. We reasoned that oligomeric and symmetric (as a globular protein, or as a helical tube) bases would be best.

As selectable adapters, we first considered antibody fragments (Fabs and scFvs). Fabs have a flexible elbow connecting two immunoglobulin regions, whereas scFvs are made up of one immunoglobulin region (Figure S1B). The Fab elbow could introduce flexibility, so we preferred the smaller, more compact scFv (Figure S1B). However, because the beta sandwich immunoglobulin fold of a scFv could be difficult to rigidly fuse to the surface of a platform base, we identified Staphylococcus Protein A (PrA) as a linker that could bind the invariant region of a scFv (Graille et al., 2000) (PBD 1DEE, Figure S1B). As PrA is a three-helix bundle, we reasoned that it could be rigidly attached to a base via a helical linker. Thus in one of our designs, the C-terminal helix of PrA was fused to the N-terminal helix of the base protein. Since PrA is capable of binding the invariant scFv framework, the base-PrA:scFv interfaces would not need to be redesigned for each new target. Unfortunately, in our biochemical experiments, we observed that the PrA:scFv interaction did not remain stable through a gel filtration column (Figure S1C–D), indicating that the binding affinity was not strong enough for our purposes. Further mutagenesis to the PrA:scFv interface may strengthen the interaction. Regardless, a fundamental concern with this design is that two non-covalent binding interactions are required (PrA:scFv, and scFv:target), which could lead to occupancy issues. As a result, we moved to DARPins as our selectable adapter (Figure 1B).

In our designs, the final alpha helix of the DARPin C-terminal cap (C-cap) was directly fused to the first α-helix of the base (Figure 1C, Methods). All DARPin libraries use a C-cap to stabilize the protein, so we expect it will be straightforward to swap in any DARPin built on the same framework (Figure 1B). In the base-DARPin platform design, only one non-covalent interaction is required (between the DARPin and the target), which results in a more predictable and stable complex. We chose a DARPin that formed a stable complex with GFP with picomolar range binding affinity as a first test case (Brauchle et al., 2014) and screened several base-DARPin candidates.

For all of our DARPin-base constructs, the fusion was designed using the molecular graphics program COOT (Emsley and Cowtan, 2004). On the base protein, a N- or C-terminal alpha helix was identified as the fusion site (see Methods). The base alpha helix and the DARPin terminal alpha helix were aligned in COOT. We assumed that the alpha helicity and the orientation of the fusion helix would not be significantly perturbed by the fusion. For some constructs, a helix-forming peptide linker was added between the base and the DARPin to achieve the desired fusion helix length (Padilla et al., 2001). The possible steric clashes of the designed model were visually examined and the open binding space was maximized. For example, the ferritin-DARPin fusion design was abandoned because adding a DARPin to each ferritin subunit didn't leave enough space for a target to bind (PDB 1EUM, Figure S1A) (Stillman et al., 2001).

Screening base candidates

We performed expression trials for two base-PrA candidates and six base-DARPin candidates (Figure 1D, S1A). These bases included the *E. coli* ribosome, (Noeske et al.,

2015; Shoji et al., 2011) β-galactosidase (β-gal) (Bartesaghi et al., 2015), the vipA/vipB helical tube (Kudryashev et al., 2015), an artificial nanocage based on EPN-01(Votteler et al., 2016), TibC (Yao et al., 2014), and aldolase (Herzik et al., 2017).

Our initial expression trials utilized the PrA/scFv strategy discussed in the previous section with β-gal. Concurrently, we found that the ribosomal protein L29-PrA fusion could be expressed, but we were unable to incorporate it into ΔL29 *E. coli* ribosomes (Shoji et al., 2011).

Because β-gal tetramerization requires the N- and C-termini of each subunit (Ullmann et al., 1967), an internal DARPin insertion was used, flanked by a helix-forming peptide (at the DARPin N-cap) and a flexible linker (at the DARPin C-cap) (Padilla et al., 2001). Biochemically the ß-gal-DARPin platform formed a stable complex with GFP, but no cryoEM density was observed for the DARPin or GFP in our 3 Å reconstruction. This means that the helical linker was flexible relative to the β-gal base. Our design for the EPN-01 based nanocage also inserted the DARPin into the middle of the sequence. The EPN-01 DARPin fusion protein failed the *E. coli* expression test.

We therefore focused on bases with a terminal α-helix that could be rigidly fused to the DARPin. The vipA/vipB, TibC, and aldolase proteins all had long terminal α-helices to facilitate direct fusion (Figure 1A, S1A). In our experiments, the helical tube vipA-DARPin/vipB platform exhibited poor expression in *E. coli*. The purified TibC-DARPin platform formed a stable complex with GFP, but the complex demonstrated aggregation and preferred orientation on plunge frozen grids. In contrast, the DARPin-aldolase platform was well-behaved (Figures S1E–F, S2).

In our DARPin-aldolase platform, the C-terminal α-helix of the DARPin was directly concatenated to the N-terminal α-helix of aldolase (Figure 1C–D). The D2 symmetry of the DARPin-aldolase platform provided extensive space for the target and could potentially accommodate a globular protein of up to 740 kDa without steric clash (Figure 1E–1F, S1 Movie). The purified GFP:DARPin-aldolase complex was stable in a gel filtration column with an apparent 1:1 stoichiometry of DARPin-aldolase to the target (GFP) (Figure S1E–F).

CryoEM analysis of the GFP:DARPin-aldolase complex

To solve the structure of GFP bound to the DARPin-aldolase platform, we collected 1,681 micrographs on a Titan Krios (Figure S3). Because the thin ice forced a slight preferred orientation issue, an additional 1,180 micrographs were collected at 26° tilt (Naydenova and Russo, 2017). High quality micrographs were selected after CTF determination (Figure S2A). The final Relion autopicking round used 2D class references generated earlier, and yielded 841,776 particles from 1,548 micrographs (Figure S3). The particles were converted to cryoSPARC, where 2D classes with strong secondary structure were selected for reconstruction and converted back to Relion. Representative 2D classes from Relion show clear secondary structure (Figure S2D). The GFP:DARPin-aldolase complex reconstruction yielded an overall resolution of 3 Å with C1 symmetry (Figure S3, S2 Movie). Further classification suggested too much conformational heterogeneity to apply D2 symmetry. The aldolase core and the helical linker were resolved to near atomic resolution (Figure 2B–D,

S2 Movie). The DARPin and GFP exhibited a local resolution of 4 to 8 Å, with discontinuous regions of higher resolution of 3.5 Å (Figure 2D, S2 Movie). Although the resolution in the GFP and DARPin portion was not sufficient to build a model or assign sequence *de novo*, the static X-ray structures of GFP and the DARPin could be reliably docked into the map (Figure 2A, S2 Movie). During the design phase we made an approximate mock up of the DARPin-aldolase construct in Coot, but did not use computation methods to relax the structure. Because of this, we cannot quantitatively compare the DARPin-aldolase model from our cryoEM structure to the design. Additionally, at the current resolution, we cannot precisely describe the side chain interactions that facilitate the GFP binding to the DARPin, but the orientation looks similar to that described by the crystal structure of a similar GFP:DARPin interaction (PDB 5MA6) (Hansen et al., 2017).

Potential sources of heterogeneity

Because of the 5 to 8 Å local resolution range in the GFP portion of the map (Figure 2D), we suspected that part of the GFP:DARPin-aldolase complex was flexible. Since we observed some preferred orientation in our dataset, we hypothesized that some of this heterogeneity may be related to how the particles interact with the air-water interface. We collected tomograms in the centers of holes (Figure S2E). The particles formed a single layer in a thin sheet of ice that was about as thick as the particles themselves. This suggests that every particle interacts with the air-water interface (Glaeser, 2018). We prepared samples on graphene oxide-coated grids to circumvent this issue, but these samples did not reach high resolution.

DARPin position caused conformational heterogeneity

To better understand the conformational heterogeneity in the data, Relion particle symmetry expansion was used to consider each subunit individually (Figure S3) (Zhou et al., 2015). The symmetry expanded particles were subjected to 3D classification without alignment, a strategy in which the orientation parameters determined in the previous refinement are used to classify the particles into subsets. For this focused classification, a spherical mask that encompassed a single GFP:DARPin region and some of the aldolase subunit was used to increase the signal. The resulting five classes showed reasonable GFP:DARPin conformations (Figure 3A, S3 Movie), but subsequent refinements were still limited to 5 to 6 Å overall, which suggested that additional conformational heterogeneity remained within the subsets. The appearance of each 3D class varied greatly with the threshold level in Chimera, so we have shown the classes at thresholds from 0.02 to 0.01 (Figure 3A, S3 Movie). At the highest threshold (0.02), a hollow GFP barrel with some density for the chromophore is clearly visible in four classes (1, 2, 4 and 5) (Figure 3A, S3 Movie). The majority of the particles (54%) were classified into class 2, which at higher threshold (0.02) appeared to lack a DARPin (Figure 3A, S3 Movie). However, relaxing the Chimera threshold to 0.01 showed sparse density for the GFP:DARPin moiety (Figure 3A and S3 Movie). Class 2 was subjected to an additional round of 3D classification where it revealed several reasonable but lower resolution GFP:DARPin conformations (Figure 3B, S3 Movie). To investigate the heterogeneity in the focused classes, we compared each class to Class 4 (Figure 3C–D). In the different classes, the GFP:DARPin density shows a clear rocking around the Y axis

(Figure 3C) and around the Z axis (Figure 3D) relative to the aldolase base. At this point, we wondered if any these displacements could be attributed to the aldolase subunit. We performed a similar focused classification experiment with a mask around the aldolase subunit and the helical linker, but no rotation or shift was observed in the resulting subsets. Thus, we concluded that some of the displacement we observed likely arose in the C-cap second helix that is fused into the helical linker, and other regions of the DARPin distal to the linker.

Discussion

In this study, we designed and tested a variety of platforms capable of non-covalently binding a small target protein via a selectable adapter for structure determination by single particle cryoEM. In our best construct, we resolved our target protein (GFP) to 5 to 8 Å resolution.

Our DARPin-aldolase platform has several advantages over other strategies. It is simple to express and purify. Aldolase has D2 symmetry and allows attachment of four targets without steric clash. Aldolase can be reconstructed to 2.6 Å resolution with even a 200 keV microscope (Herzik et al., 2017). Because DARPins can be readily generated against a wide range of small protein targets, the attachment of a DARPin to aldolase promises to be a generally applicable strategy. A recent study of the insulin degrading enzyme (IDE) bound to Fabs was able to isolate several IDE conformations using different Fabs (Zhang et al., 2018). It stands to reason that different DARPins could also stabilize different conformations of the target. Because switching DARPins in the platform would be done by straight-forward DNA manipulations, our DARPin-aldolase platform has the potential to resolve a series of conformations of the target protein.

Our biochemistry experiments suggested that the purified GFP:DARPin-aldolase complex was very stable, and clear secondary structure was apparent in the 2D classes, yet heterogeneity remained. Potential sources of heterogeneity include the aldolase core, the angle of the linker helix relative to the aldolase core, and flexibility in the DARPin itself. The interaction of the particles with the air-water interface may also play a role. Since the air-water interface played a role in the heterogeneity observed in the structure reported here, designing a platform that would couch the target inside a cavity may be advantageous. Because the aldolase base and the helical linker region were resolved to near atomic resolution (Figure 2B–C), a large contribution to the heterogeneity we observed likely stemmed from the DARPin C-cap. However, we cannot discount the affect of the other sources listed previously.

The DARPin against GFP used here was from a first generation DARPin library. The C-cap of the first generation DARPins was reported to be less stable than the other repeat modules (Seeger et al., 2013). While the crystal structure contained a well-resolved C-cap, the heterogeneity observed here suggests that it is not yet sufficiently rigid to serve as an attachment point in a cryoEM platform (Figure 3). Recent DARPin phage display libraries contain DARPins with reduced surface entropy and a more stable C-cap sequence (Seeger et al., 2013), however, and additional stabilizing surface interactions could be introduced in

future designs (Kramer et al., 2010) (Interlandi et al., 2008), or even a second attachment point of the DARPin to the base (at both N- and C-terminal caps of DARPin for instance). Together such improvements could allow the DARPin-aldolase platform to reveal the structures of many small proteins to near atomic resolution.

STAR Methods

Contact for Reagent and Resource Sharing—Requests for further information as well as resources and reagents should be directed to and will be fulfilled by the Lead Contact, Grant Jensen (jensen@caltech.edu).

Experimental Model and Subject Details

Bacterial Strains—*E. coli* E. cloni and BL21(DE3) cells were cultured in Luria Broth or Autoinduction medium with appropriate antibiotics.

Method Details

Computational design—Designs were generated by examining the atomic coordinates of the base protein, the selectable adapter, and the target in UCSF Chimera (Pettersen et al., 2004) or COOT (Emsley and Cowtan, 2004), manually adjusting the positions to approximate a fusion protein between the base and the selectable adapter, and assessing the design for potential steric hindrance. Promising designs left enough room for the target protein to bind the selectable adapter. The PDB identifiers of the models used in this process are summarized in the Key Resources Table. The sequences of the constructs tested are reported in DataS1.

The N- and C-termini of base proteins β-galactosidase and the EPN-01 nanocage are involved in multimerization, so the sequence of the selectable adapter was inserted within the base protein sequence. All other constructs utilized a terminal alpha helix to fuse the selectable adapter.

The computational α-helix fusion was generated by manually docking the rabbit muscle aldolase structure (PDB code: 5VY5) and the GFP/DARPin complex (PDB code: 5MA6) (Hansen et al., 2017; Herzik et al., 2017) in COOT. Because the DARPin sequence we selected has not been crystallized (DARPin 3G86.32 (Brauchle et al., 2014)), residues in the 5MA6 PDB model were mutated in COOT to match our sequence. In order to rigidly join the aldolase and DARPin moiety together, we truncated the C-terminal flexible loop on DARPin and N-terminal flexible loop on aldolase, respectively, exposing the two terminal α-helices. The two terminal α-helices were manually concatenated and joined together to form an ideal α-helix using building α-helix tool in UCSF Chimera. The model was inspected for the orientation of DARPin relative to the aldolase, ensuring no steric clash and the providing enough space for target protein attachment. Structural design figures were generated using PyMOL1.8 (https://pymol.org) or UCSF Chimera.

Cloning, protein expression, and purification—The cDNA used for all base-selectable adapter fusions and for GFP was synthesized at Integrated DNA Technologies

(IDT, Coralville, Iowa, USA) and cloned into an expression vector as indicated in the Key Resources Table.

The β-galactosidase sequence was derived from PDB 5A1A while the Protein A sequence and the single chain variable fragment (scFv) sequences were derived from PDB 1DEE(Bartesaghi et al., 2015; Graille et al., 2000). Graille et al. reported the crystal structure of *S. aureus* protein A bound to a human IgM Fab. To convert the Fab sequence into a scFv, the variable domains from the Fab immunoglobulin light chain (PDB 1DEE Chain A residues 1 to 106) and the Fab immunoglobulin heavy chain (PDB 1DEE Chain B residues 501 to 621) were connected using a flexible linker (see DataSl for sequences). The β-galactosidase fusion with Protein A was cloned into the pET21a vector with a C-terminal His tag. The scFv antibody against Protein A was cloned into the pET22b expression vector with a C-terminal His tag. The β-gal-PrA fusion and the scFv were each recombinantly expressed in *E. coli* BL21(DE3) cells overnight at 22 °C. Cells were lysed using sonication in a buffer containing 50 mM Tris-HCl pH 8.0 150 mM NaCl and 15 mM imidazole, and purified using Ni-NTA affinity chromatography (Qiagen). About 300 μ g purified β-gal PrA fusion and 200 μ g scFv were mixed together in 1 ml incubation buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl). The mixture was incubated on ice overnight, and was subsequently analyzed by gel filtration using a Superdex 200 10/300 GL column (GE Healthcare) equilibrated with incubation buffer. Fractions for each peak were pooled and analyzed by SDS-PAGE.

The DARPin sequence was derived from DARPin 3G86.32 (Brauchle et al., 2014). The cDNA of GFP and of the DARPin-aldolase fusion were PCR-amplified and inserted into pACYCDuet and pET21b vectors respectively. The resulting GFP protein was untagged, while a C-terminal His-tag was used for the DARPin-aldolase chimeric fusion. GFP and DARPin-aldolase were co-expressed in *E. coli* BL21(DE3) using autoinduction medium with trace elements (Formedium) at 30 °C for overnight. Cells were harvested by centrifugation, lysed with sonication in a buffer containing 50 mM Tris-HCl pH 8.0, 150 mM NaCl and 15 mM imidazole, and then the complex was purified with Ni-NTA affinity chromatography. Gel filtration chromatography (Superdex 200 10/300 GL) was used to isolate the complete GFP:DARPin-aldolase complex. The gel filtration peaks were analyzed by SDS-PAGE. The purified GFP:DARPin-aldolase complex was concentrated to 2.5mg/ml in a buffer containing 25 mM Tris-HCl pH 8.0 and 150 mM NaCl.

CryoEM sample preparation and data collection—Electron microscopy grids were prepared at Scripps Research Institute. Briefly, 3 μL sample of 2.5 mg/ml GFP-DARPin-aldolase complex was applied to a plasma cleaned Au UltraFoil Grid (200 Mesh, R2/2, Quantifoil) in a cold room (4°C, ≥95% relative humidity). The grid was manually blotted with a filter paper (Whatman No.1) for approximately 3 seconds before plunging into liquid ethane using a manual plunger (Herzik et al., 2017). The grids were screened in Talos Arctica 200 kV with Falcon 3 (FEI) direct electron detector for ice thickness and sample distribution. Micrographs of GFP-DARPin-aldolase complex were collected on Titan Krios microscope (FEI) operating and 300 kV with energy filter (Gatan) and equipped with a K2 Summit direct electron detector (Gatan). For untilted data, Serial EM was used for automated EM image acquisition(Mastronarde, 2005). After calculating an efficiency score

from early refinements using cryoEF(Naydenova and Russo, 2017), additional data were collected at 26° using EPU software (FEI). A nominal magnification of 165,000× was used for data collection, corresponding to a pixel size of 0.865 Å at the specimen level, with the defocus ranging from − 1.0 μm to −3.0 μm. Movies were recorded in superresolution mode, with a total dose of ~40 e-/Å^2, fractioned into 20 frames (0° tilt images) or 40 frames (26° tilt images) under the does rate of 8.4 electron per pixel per second. Tilt series were collected using SerialEM on the Titan Krios at a nominal magnification of 165,000×, with a pixel size of 0.865 Å and at −4 μm target defocus (Mastronarde, 2005). A bi-directional tilt scheme was used with a 3° tilt increment. The tilt series began at 0° and covered −45° to +45°. A 0.5 s movie with 5 frames was recorded for each tilt image in K2 super resolution mode using a dose rate of 8 electron per pixel per second.

Single particle cryoEM analysis—A summary of the major steps of data processing is available in Figure S4. Movies were decompressed and gain corrected with IMOD (Kremer et al., 1996). Motion correction was performed using program MotionCor2 (Zheng et al., 2017), and exposure filtered in accordance with relevant radiation damage curves (Grant and Grigorieff, 2015). Micrographs with high CTF Figure of Merit scores and promising maximum resolution (better than 3.6 Å for 0° tilt, and better than 6 Å for 26° tilt) were selected for further processing (1,548 micrographs). Several rounds of autopicking using combinations of different references and manual picking were analyzed to determine optimal settings, and all yielded similar results. In an early processing cycle, to generate 2D class references, a subset of 1,050 micrographs was autopicked with a model of aldolase (PDB: 5vy5) generated to 20 Å resolution in Chimera and subjected to an additional 10 Å low pass filter by Relion auto picking with D2 symmetry and 15° angular sampling. These particles experienced seven iterative rounds of 2D classification, subset selection, and re-extraction to generate the ten 2D classes used as a reference for the final round of autopicking. This yielded 851,776 particles from 1,548 micrographs. Particles were extracted in RELION (Scheres, 2012a, b; (Kimanius et al., 2016); Zivanov et al., 2018) and initial 2D classification was performed in cryoSPARC (Punjani et al., 2017). High quality 2D classes were selected for further processing (442,974 particles). The initial model was *de novo* generated and subsequent 3D refinement were performed using cryoSPARC. The UCSF PyEM package (https://github.com/asarnow/pyem) script was used to convert the cryoSPARC coordinates into Relion (268,905 particles). Duplicate particles were removed and the resulting 236,339 particles were analyzed by 3D refinement, Bayesian Particle Polishing and CTF Refinement in Relion. The data were binned to 1.5 Å/pixel, refined with D2 symmetry, and symmetry expanded. Symmetry expanded particles were used in 3D classification without alignment. All reconstructions were analyzed using the *relion_display* function and USCF Chimera. The coordinate model was built by breaking our initial GFP:DARPin-aldolase PDB model into domains (GFP, DARPin and aldolase subunits) and rigidly docking these individual protein structures into the EM map using Chimera. Once the orientations were identified, the model was then fit and adjusted manually in USCF Chimera and Coot. The figures were generated using UCSF Chimera, and local resolution and final Fourier shell correlation were calculated using ResMap (Kucukelbir et al., 2014) and cryoSPARC.

Tomography—Frames were motion corrected using MotionCor2 and reconstructed in IMOD. Because the grids lacked gold fiducial markers, patch tracking was used to align the tilt series. The weighted back projection method was used to reconstruct tilt series into tomograms.

Quantification and Statistical Analysis

Statistical analyses were performed within published software as described in *Single particle cryoEM analysis* method section.

Data and Software Availability

Data deposition—The cryoEM density map and coordinates of the GFP:DARPin-aldolase complex have been deposited in the Electron Microscopy Data Bank (EMDB) and the Protein Data Bank (PDB) with access codes: EMD-9277 and PDB 6MWQ.

Software availability—This study did not produce additional software. The software used here (summarized in Key Resources Table) has been published and is publically available. Software was accessed from developer websites, or from the SBGrid Consortium(Morin et al., 2013).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Dr. Mark Herzik Jr. and Mengyu Wu at The Scripps Research Institute for help with sample preparation. We also thank Dr. Songye Chen, Dr. Andrey Malyutin, Dr. Rebecca Voorhees, and Dr. Bil Clemons at Caltech for technical assistance. We are also grateful to all members of the Jensen laboratory for discussion and technical assistance. In particular, we thank Dr. Lauren Ann Metskas and Dr. Ariana Peck. This work was supported by funds from NIH NIGMS P50 082545. CryoEM work was performed in the Caltech Beckman Institute Resource Center for Transmission Electron Microscopy.

References

Bartesaghi A, Merk A, Banerjee S, Matthies D, Wu X, Milne JL, and Subramaniam S (2015). 2.2 A resolution cryo-EM structure of beta-galactosidase in complex with a cell-permeant inhibitor. Science 348, 1147–1151. [PubMed: 25953817]

Brauchle M, Hansen S, Caussinus E, Lenard A, Ochoa-Espinosa A, Scholz O, Sprecher G, Pluckthun A, and Affolter M (2014). Protein interference applications in cellular and developmental biology using DARPins that recognize GFP and mCherry. Biol Open 3, 1252–1261. [PubMed: 25416061]

Brocchieri L, and Karlin S (2005). Protein length in eukaryotic and prokaryotic proteomes. Nucleic Acids Res 33, 3390–3400. [PubMed: 15951512]

Coscia F, Estrozi LF, Hans F, Malet H, Noirclerc-Savoye M, Schoehn G, and Petosa C (2016). Fusion to a homo-oligomeric scaffold allows cryo-EM analysis of a small protein. Sci Rep 6 30909. [PubMed: 27485862]

Emsley P, and Cowtan K (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60, 2126–2132. [PubMed: 15572765]

Glaeser RM (2018). Proteins, Interfaces, and Cryo-Em Grids. Current opinion in colloid & interface science 34, 1–8. [PubMed: 29867291]

Graille M, Stura EA, Corper AL, Sutton BJ, Taussig MJ, Charbonnier JB, and Silverman GJ (2000). Crystal structure of a Staphylococcus aureus protein A domain complexed with the Fab fragment of

a human IgM antibody: structural basis for recognition of B-cell receptors and superantigen activity. Proc Natl Acad Sci U S A 97, 5399–5404. [PubMed: 10805799]

Grant T, and Grigorieff N (2015). Measuring the optimal exposure for single particle cryo-EM using a 2.6 A reconstruction of rotavirus VP6. Elife 4, e06980. [PubMed: 26023829]

Hansen S, Stuber JC, Ernst P, Koch A, Bojar D, Batyuk A, and Pluckthun A (2017). Design and applications of a clamp for Green Fluorescent Protein with picomolar affinity. Sci Rep 7, 16292. [PubMed: 29176615]

Henderson R (1995). The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. Q Rev Biophys 28, 171–193. [PubMed: 7568675]

Herzik MA Jr., Wu M, and Lander GC (2017). Achieving better-than-3-A resolution by single-particle cryo-EM at 200 keV. Nat Methods 14, 1075–1078. [PubMed: 28991891]

Interlandi G, Wetzel SK, Settanni G, Pluckthun A, and Caflisch A (2008). Characterization and further stabilization of designed ankyrin repeat proteins by combining molecular dynamics simulations and experiments. J Mol Biol 375, 837–854. [PubMed: 18048057]

Jensen GJ, and Kornberg RD (1998). Single-particle selection and alignment with heavy atom cluster-antibody conjugates. Proc Natl Acad Sci U S A 95, 9262–9267. [PubMed: 9689068]

Jeong WH, Lee H, Song DH, Eom JH, Kim SC, Lee HS, Lee H, and Lee JO (2016). Connecting two proteins using a fusion alpha helix stabilized by a chemical cross linker. Nat Commun 7, 11031. [PubMed: 26980593]

Khoshouei M, Radjainia M, Baumeister W, and Danev R (2017). Cryo-EM structure of haemoglobin at 3.2 A determined with the Volta phase plate. Nat Commun 8, 16099. [PubMed: 28665412]

Kimanius D, Forsberg BO, Scheres SH, and Lindahl E (2016). Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. Elife 5.

Kramer MA, Wetzel SK, Pluckthun A, Mittl PR, and Grutter MG (2010). Structural determinants for improved stability of designed ankyrin repeat proteins with a redesigned C-capping module. J Mol Biol 404, 381–391. [PubMed: 20851127]

Kratz PA, Bottcher B, and Nassal M (1999). 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. [PubMed: 10051569]

Kremer JR, Mastronarde DN, and McIntosh JR (1996). Computer visualization of three-dimensional image data using IMOD. J Struct Biol 116, 71–76. [PubMed: 8742726]

Kucukelbir A, Sigworth FJ, and Tagare HD (2014). Quantifying the local resolution of cryo-EM density maps. Nat Methods 11, 63–65. [PubMed: 24213166]

Kudryashev M, Wang RY, Brackmann M, Scherer S, Maier T, Baker D, DiMaio F, Stahlberg H, Egelman EH, and Basler M (2015). Structure of the type VI secretion system contractile sheath. Cell 160, 952–962. [PubMed: 25723169]

Liu Y, Gonen S, Gonen T, and Yeates TO (2018a). 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. [PubMed: 29507202]

Liu Y, Huynh D, and Yeates T (2018b). A 3.8 Å Resolution Cryo-EM Structure of a Small Protein Bound to a Modular Imaging Scaffold. bioRxiv.

Mastronarde DN (2005). Automated electron microscope tomography using robust prediction of specimen movements. J Struct Biol 152, 36–51. [PubMed: 16182563]

Morin A, Eisenbraun B, Key J, Sanschagrin PC, Timony MA, Ottaviano M, and Sliz P (2013). Collaboration gets the most out of software. Elife 2, e01456. [PubMed: 24040512]

Naydenova K, and Russo CJ (2017). Measuring the effects of particle orientation to improve the efficiency of electron cryomicroscopy. Nat Commun 8, 629. [PubMed: 28931821]

Noeske J, Wasserman MR, Terry DS, Altman RB, Blanchard SC, and Cate JH (2015). High-resolution structure of the Escherichia coli ribosome. Nature structural & molecular biology 22, 336–341.

Padilla JE, Colovos C, and Yeates TO (2001). Nanohedra: using symmetry to design self assembling protein cages, layers, crystals, and filaments. Proc Natl Acad Sci U S A 98, 2217–2221. [PubMed: 11226219]

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, and Ferrin TE (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25, 1605–1612. [PubMed: 15264254]

Punjani A, Rubinstein JL, Fleet DJ, and Brubaker MA (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat Methods 14, 290–296. [PubMed: 28165473]

Seeger MA, Zbinden R, Flutsch A, Gutte PG, Engeler S, Roschitzki-Voser H, and Grutter MG (2013). Design, construction, and characterization of a second-generation DARP in library with reduced hydrophobicity. Protein Sci 22, 1239–1257. [PubMed: 23868333]

Shoji S, Dambacher CM, Shajani Z, Williamson JR, and Schultz PG (2011). Systematic chromosomal deletion of bacterial ribosomal protein genes. J Mol Biol 413, 751–761. [PubMed: 21945294]

Stillman TJ, Hempstead PD, Artymiuk PJ, Andrews SC, Hudson AJ, Treffry A, Guest JR, and Harrison PM (2001). The high-resolution X-ray crystallographic structure of the ferritin (EcFtnA) of Escherichia coli; comparison with human H ferritin (HuHF) and the structures of the Fe(3+) and Zn(2+) derivatives. J Mol Biol 307, 587–603. [PubMed: 11254384]

Ullmann A, Jacob F, and Monod J (1967). Characterization by in vitro complementation of a peptide corresponding to an operator-proximal segment of the beta-galactosidase structural gene of Escherichia coli. J Mol Biol 24, 339–343. [PubMed: 5339877]

Votteler J, Ogohara C, Yi S, Hsia Y, Nattermann U, Belnap DM, King NP, and Sundquist WI (2016). Designed proteins induce the formation of nanocage-containing extracellular vesicles. Nature 540, 292–295. [PubMed: 27919066]

Wu S, Avila-Sakar A, Kim J, Booth DS, Greenberg CH, Rossi A, Liao M, Li X, Alian A, Griner SL, et al. (2012). Fabs enable single particle cryoEM studies of small proteins. Structure 20, 582–592. [PubMed: 22483106]

Yao Q, Lu Q, Wan X, Song F, Xu Y, Hu M, Zamyatina A, Liu X, Huang N, Zhu P et al. (2014). A structural mechanism for bacterial autotransporter glycosylation by a dodecameric heptosyltransferase family. Elife 3.

Zhang Z, Liang WG, Bailey LJ, Tan YZ, Wei H, Wang A, Farcasanu M, Woods VA, McCord LA, Lee D et al. (2018). Ensemble cryoEM elucidates the mechanism of insulin capture and degradation by human insulin degrading enzyme. Elife 7.

Zheng SQ, Palovcak E, Armache JP, Verba KA, Cheng Y, and Agard DA (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat Methods 14, 331–332. [PubMed: 28250466]

Zhou M, Li Y, Hu Q, Bai XC, Huang W, Yan C, Scheres SH, and Shi Y (2015). Atomic structure of the apoptosome: mechanism of cytochrome c- and dATP-mediated activation of Apaf-1. Genes Dev 29, 2349–2361. [PubMed: 26543158]

Highlights

- Exploration of constructs to display small proteins for single particle cryoEM
- New fusion protein allows cryoEM visualization of <100 kDa proteins, like GFP
- 3-Å structure of GFP:DARPin-aldolase platform with 5 to 8-Å in GFP:DARPin region
- Higher resolution of the target currently limited by DARPin position flexibility

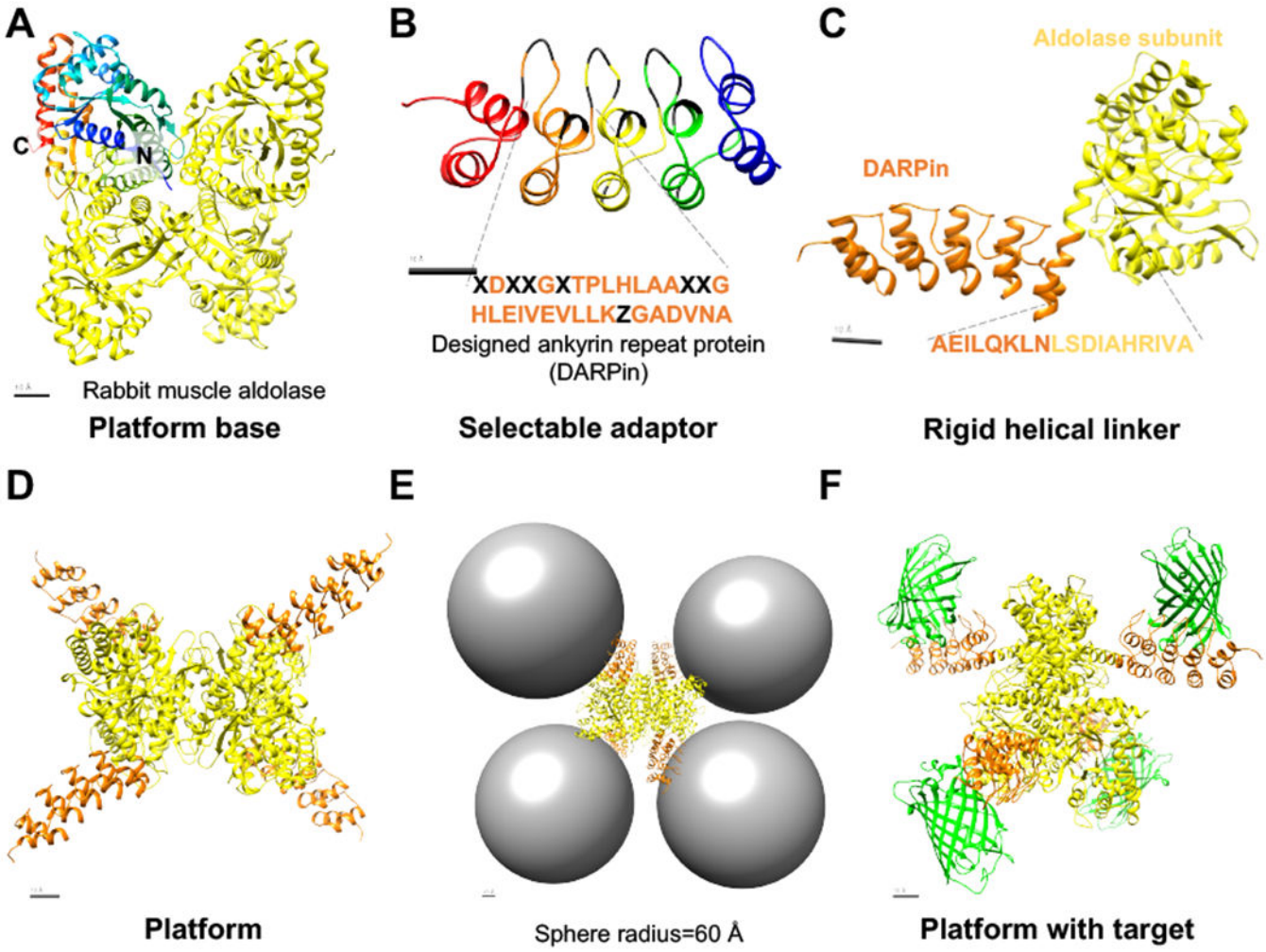


Figure 1. The Design of the platform

(A) The platform base was homotetrameric rabbit muscle aldolase (PDB ID code 5VY5). One subunit was depicted with rainbow coloring and N and C labels to indicate the orientation of the monomer chain. The other three identical subunits are shown in yellow. **(B)** The selectable adapter was a Designed Ankyrin Repeat Protein (DARPin) (PDB 5MA6). Shown below is a close-up view of the repetitive motif of DARPin with its amino acid sequence (orange). Using a phage display library, DARPins can be generated against a protein target. The selectable residues are depicted in black as X (any amino acid except cysteine or proline) or Z (amino acids asparagine, histidine or tyrosine) (Brauchle et al., 2014). **(C)** The final helix of the C-terminal cap of the DARPin (orange) was directly fused to the first alpha helix of aldolase (yellow) to form the platform subunit. **(D)** The D2 symmetry of the DARPin-aldolase fusion demonstrates ample space for target binding. **(E)** Spheres (radius=60 Å) were drawn in the position where each DARPin binds its target. A globular protein of up to 740 kDa could be accommodated on the DARPin-aldolase platform without steric clash. **(F)** The model of the DARPin-aldolase platform in complex with GFP (green) is shown. See also Figure S1 and Movie S1.

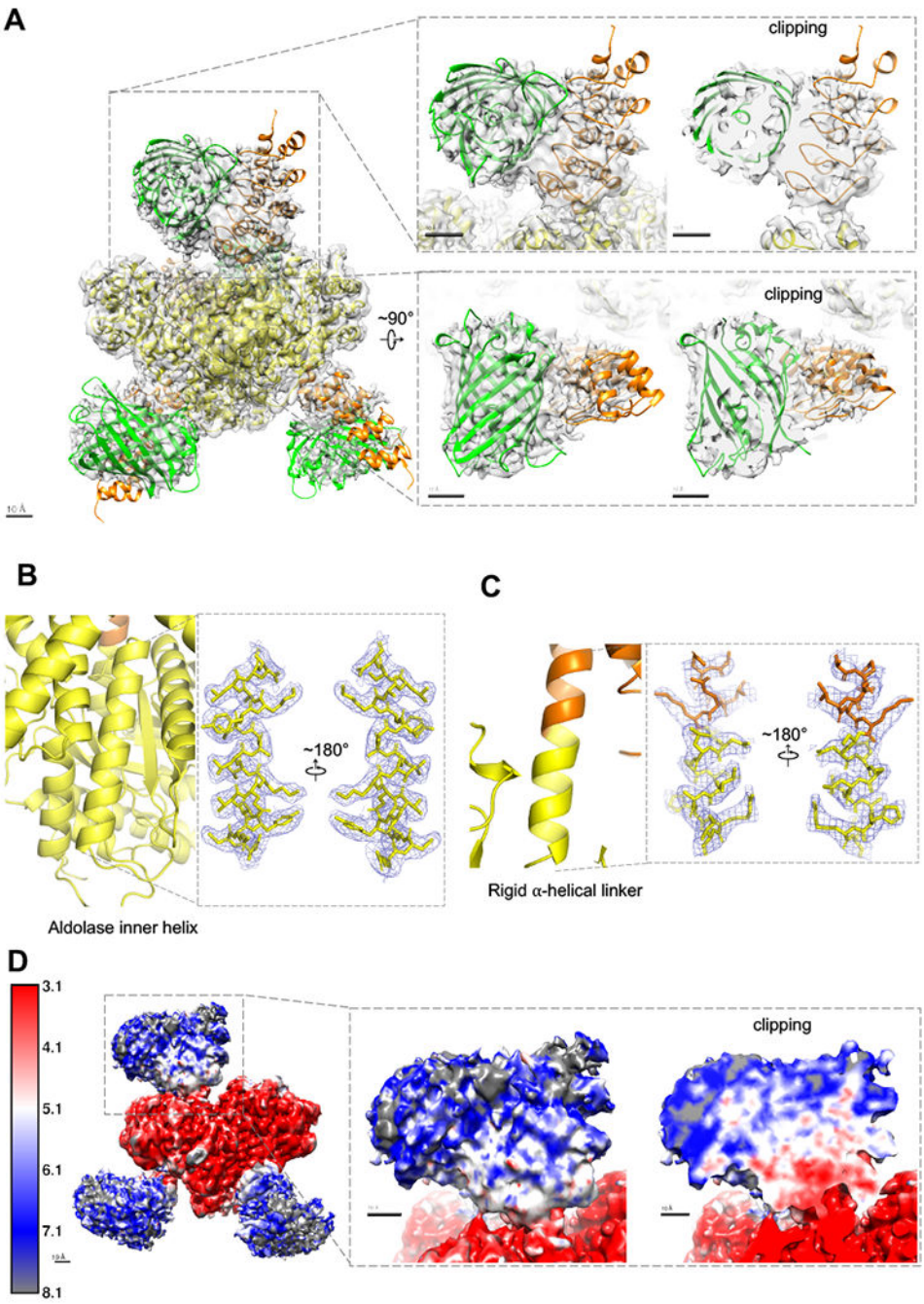


Figure 2. The cryoEM structure of the DARPin-aldolase platform in complex with the target GFP

(A) Surface of the sharpened 3 Å C1 reconstruction of DARPin-aldolase platform in complex with GFP. The crystal structures of GFP (green), the DARPin (orange) and aldolase (yellow) were docked into the cryoEM density. Two expanded views of the best resolved subunit are shown on the right. The expanded views in the dashed line boxes are shown from the top or side, and clipped halfway to indicate the quality of the fit. The cryoEM map is sharpened with Relion PostProcess. Chimera threshold, 0.005. Scale bars, 10 Å. **(B)** Ribbon

diagram (left) and cryoEM density (right, blue mesh, zoned 1.8 Å within atoms) of an internal aldolase helix (residues Arg369 to Asp387). **(C)** Ribbon diagram (left) and cryoEM density (right, blue mesh, zoned 1.8 Å within atoms) of the helical linker (residues Ala176 to Ile191) between the DARPin (orange, residues Ala176 to Lys181) and aldolase (yellow, residues Leu182 to Ile191). **(D)** ResMap local resolution estimate of the unsharpened cryoEM density (left) and of the best subunit (right). The expanded view in the dashed line box is shown from the side (left) and halfway into the GFP:DARPin density with clipping (right). The gradient color scale from red (3.1 Å) to grey (8.1 Å) was generated with ResMap. Chimera threshold 0.003. Scale bars, 10 Å. See also Figure S2 and Movie S2.

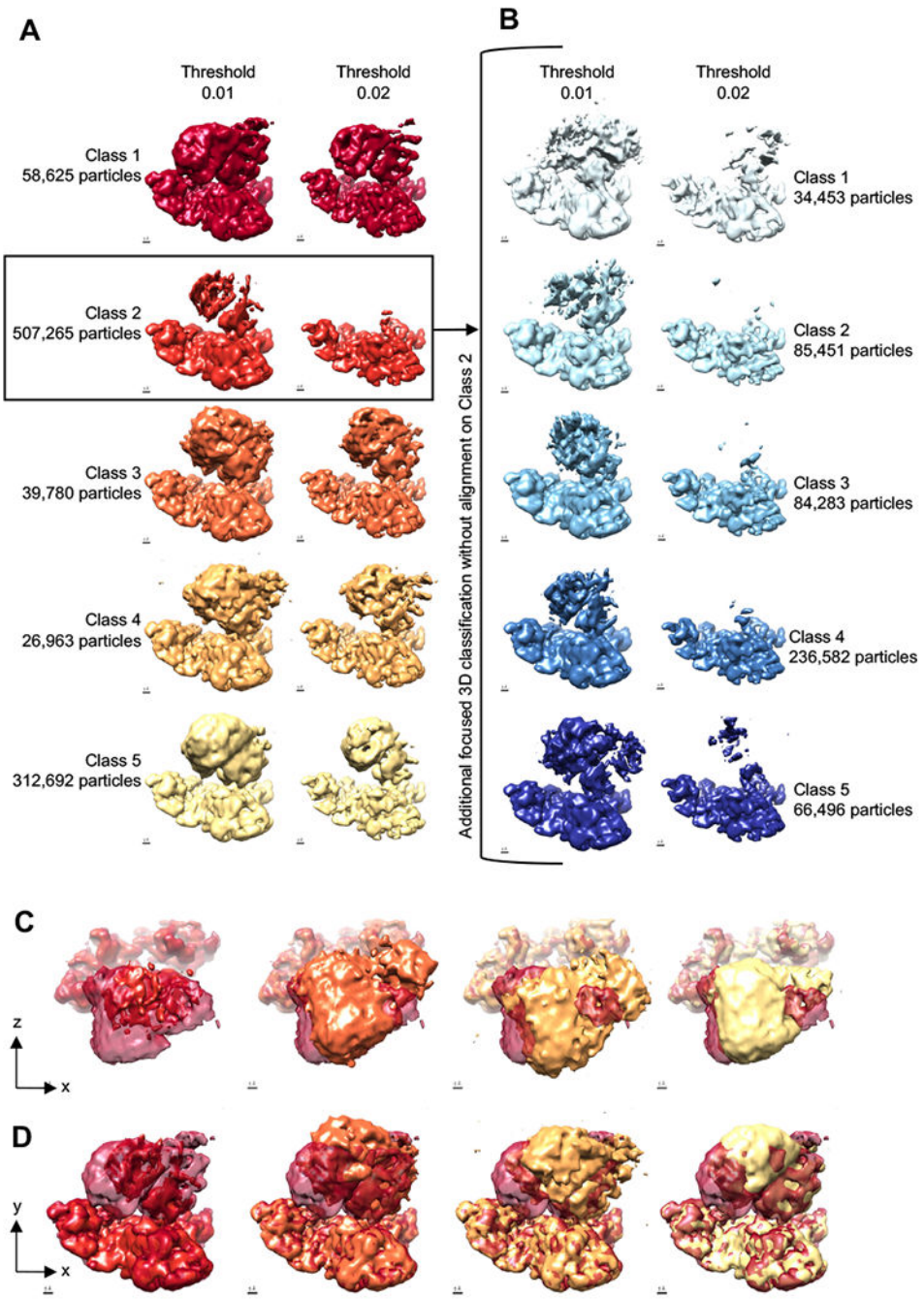


Figure 3. Symmetry expanded 3D classification of the GFP:DARPin region of the density

(A) 3D classification without alignment of the symmetry expanded particles with five classes. The number of particles per class is indicated to the left of each class. Each class was viewed at two Chimera thresholds (0.01, 0.02) to facilitate direct comparisons. Scale bar, 5 Å. **(B)** An additional round of 3D classification without alignment was performed on Class 2 from (B). Scale bars, 5 Å. **(C)** The classes in (A) were each compared with Class 1 (dark red) to show the displacement between classes. The XZ plane is shown and the Y axis is perpendicular to the page. Class 1 was clearly displaced relative to the other classes.

Chimera threshold, 0.01. Scale bar, 5 Å. **(D)** The comparison from (C) is viewed looking down the Z axis. The Chimera threshold was 0.01. Scale bar, 5 Å. See also Figure S3 and Movie S3.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
E. coli E. cloni 10G Chemically Competent cells	Lucigen	Cat #60106-2
E. coli BL21(DE3)	Lucigen	Cat #60401-3
Deposited Data		
Cryo-EM map of GFP:DARPin-aldolase structure	This paper	EMD-9277
Coordinates of GFP:DARPin-aldolase structure	This paper	PDB 6MWQ
Coordinates of rabbit muscle aldolase	(Herzik et al., 2017)	PDB 5VY5
Coordinates of GFP/DARPin complex	(Hansen et al., 2017)	PDB 5MA6
Coordinates of *S. aureus* protein A bound to a human IgM Fab	(Graille et al., 2000)	PDB 1DEE
Coordinates of *E. coli* ribosome	(Noeske et al., 2015)	PDB 4YBB
Coordinates of *E. coli* Beta-galactosidase	(Bartesaghi et al., 2015)	PDB 5A1A
Coordinates of *V. cholerae* vipA/vipB helical tube	(Kudryashev et al., 2015)	PDB 3J9G
Coordinates of EPN-01 nanocage	(Votteler et al., 2016)	PDB 5KP9
Coordinates of *E. coli* TibC	(Yao et al., 2014)	PDB 4RAP
Coordinates of *E. coli* Ferritin	(Stillman et al., 2001)	PBD 1EUM
Recombinant DNA		
pACYCDuet-GFP	This paper	n/a
pET21b-His-DARPin-aldolase	This paper	n/a
pET21a-Beta-galactosidase-ProteinA	This paper	n/a
pET22b-scFv	This paper	n/a
pET21a-Beta-galactosidase-DARPin	This paper	n/a
Software and Algorithms		
IMOD	(Kremer et al., 1996)	https://bio3d.colorado.edu/imod/
Relion	(Scheres, 2012b; 2012a; Zivanov et al., 2018)	https://www2.mrc-lmb.cam.ac.uk/relion/index.php/Main_Page
cryoSPARC	(Punjani et al., 2017)	https://cryosparc.com/
PyMOL	(DeLano, 2002)	https://pymol.org/2/
COOT	(Emsley and Cowtan, 2004)	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
Chimera	(Pettersen et al., 2004)	https://www.cgl.ucsf.edu/chimera/
MotionCor2	(Zheng et al., 2017)	http://msg.ucsf.edu/em/software/motioncor2.html
CtfFind4	(Rohou and Grigorieff, 2015)	http://grigoriefflab.janelia.org/ctffind4
cryoEF	(Naydenova and Russo, 2017)	https://www.mrc-lmb.cam.ac.uk/crusso/cryoEF/
SerialEM	(Mastronarde, 2005)	http://bio3d.colorado.edu/SerialEM/

REAGENT or RESOURCE	SOURCE	IDENTIFIER
UCSF PyEM Package	n/a	https://github.com/asarnow/pyem
SBGrid Software Library	(Morin et al., 2013)	https://sbgrid.org/
Other		
UltrAufoil® R2/2 holey gold film on Au 200 mesh grid	Quantifoil, Großlöbichau, Germany	Cat #687-200-AU