Design of Orthogonal Far-Red, Orange and Green Fluorophore-binding Proteins for Multiplex Imaging

Long Tran[1,6], Shajesh Sharma[2,6,#], Steffen Klein[3,#], David Juergens[4,5,6,#], Justin Decarreau[6,#], Bingxu Liu[4,6], Yujia Wang[4,6], Asim K. Bera[6], Alex Kang[6], Jon Woods[6], Emily Joyce[6], Dionne K Vafeados[6], Nicole Roullier[6], Wei Chen[4,6], Gyu Rie Lee[4,6], Julia Mahamid[3,7], Luke D. Lavis[8], Linna An[4,6,*], David Baker[4,6,9,*]

Address:

1. Department of Chemical Engineering, University of Washington, Seattle, WA, USA
2. Department of Bioengineering, University of Washington, Seattle, WA, USA
3. Molecular Systems Biology Unit, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany
4. Department of Biochemistry, University of Washington, Seattle, WA, USA
5. Graduate Program in Molecular Engineering, University of Washington, Seattle, WA, USA.
6. Institute for Protein Design, University of Washington, Seattle, WA, USA.
7. Cell Biology and Biophysics Unit, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany.
8. Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, Virginia, USA
9. Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA.

# These authors contributed equally as second authors

* Co-corresponding authors: linnaan@uw.edu;dabaker@uw.edu

Abstract

Fluorescent proteins and small molecule dyes have complementary strengths for biological imaging: the former are genetically manipulatable enabling tagging of specific proteins and detection of protein interactions, while the latter have greater photostability and brightness but are difficult to target. To combine these strengths, we used de novo protein design to generate binders to three bright, stable, cell-permeable dyes spanning the visible spectrum: JF657 (far red), JF596 (orange-red) and JF494 (green). For each dye, we obtain nanomolar binders with weak or no binding to the other two dyes; the accuracy of the design approach is confirmed by a crystal structure of one binder which is very close to the design model. Fusion of the JF567, JF596 and JF494 binders to three different targets followed by staining with the three dyes simultaneously enables multiplex imaging. We further expand functionality by incorporating an active site carrying out nucleophilic aromatic substitution to form a covalent linkage with the dye, and developing split versions which reconstitute fluorescence at subcellular locations where both halves are present, enabling both protein-protein interaction detection and chemically

induced dimerization with fluorescence reporting. Our designs combine the advantages of fluorescent proteins and small molecule dyes and should be broadly useful for cellular imaging.

Introduction

Biological imaging was transformed by the ability to specifically tag cellular components with genetically encoded fluorescent proteins. However, fluorescent proteins have limited brightness and photostability which constrain the resolution, duration, and complexity of imaging experiments(*1*). Small-molecule dyes, by contrast, offer far greater brightness and photostability(*2*–*4*). Advanced synthetic fluorophores such as the Janelia Fluor® (JF) dyes are notable for their exceptional quantum yield, photostability, and cell permeability across a range of emission wavelengths(*2*, *3*). Analogous to fluorescent proteins, JF dye usage in cellular imaging requires a targeting mechanism to the proteins of interest. Current approaches primarily utilize fusion tags, including HaloTag(*5*) and SNAP-tag(*6*), which form covalent bonds with modified versions of the dye. While effective, these tagging systems are relatively large (297aa), and generally not orthogonal – a single tag reacts with any compatible dye – allowing only a single color at a time, thus limiting multiplexing capability(*7*).

We reasoned that a set of small, designed proteins that each bind to a single JF dye, and that collectively span the color spectrum, could provide a general solution to the multiplexed imaging problem. This has not been possible to date due to the lack of natural proteins that specifically bind to the JF dyes, which have quite similar molecular structures(*8*). Recent advances in machine learning (ML)-based de novo protein design(*9*, *10*) could enable a solution to this challenge. We set out to create a new class of JF dye-binding proteins that are genetically encodable, small in size, each with high affinity and specificity for a single bright dye, and evaluate their use in multiplex cellular imaging.

Designing and properties of fluorophore-binding proteins

We selected as design targets three JF dyes with well-separated excitation and emission spectra, JF494, JF596, and JF657, that span the visible spectrum(*2*) **(Fig 1** and **Extended Fig 1)**. These compounds share a common rhodamine-derived core, with modifications that increase the brightness and tune the fluorescence excitation and emission maxima(*2*). Designing a set of binders specific for each of these compounds is challenging because of their high overall structural similarity. We reasoned that a design strategy that generated protein structures with extensive shape complementarity to an input ligand structure could achieve such specificity as even small changes in the chemical structure of the ligand could introduce clashes or eliminate favorable contacts, compromising binding in both cases(*9*, *11*).

To obtain such extensive shape complementarity, we chose to design pseudocyclic binders, which achieve high affinity and specificity by encircling the target ligand, and have the added advantage of being readily split into two domains; we anticipated downstream uses of split versions as chemically inducible dimerizers with a wide range of applications. Pseudocyclic proteins have been generated using hallucination approaches, but not directly in the presence

of a target ligand; for greater efficiency and control we sought to adapt the generative AI method CA-RFdiffusion for this task(*10*). Pseudocycles have approximate internal N-fold symmetry, where N is a parameter selected at inference time along with symmetry unit sequence length L. The repeating units surround a central cavity that constitutes the ligand binding site.

We generated pseudocyclic binders using CA-RFdiffusion(*10*). Given a choice of symmetry (N) and symmetry unit length (L, the full protein length will be LxN), at each denoising step (*i.e.* protein generation step) we symmetrize the RoseTTAfold pair representation by collecting all L by L length blocks of features corresponding to a unique subunit-subunit interaction, aggregating them into a single block (*via* a mean or a max down the collected stack) and re-tiling this back into the source block locations **(see Methods)**. We encode the ligand structure via the CA-RFdiffusion motif template input, and omit symmetry operations in any ligand-ligand or ligand-protein blocks in the pairwise representation. We found that this method effectively generated monomeric or multimeric pseudocyclic structures around a variety of small-molecule targets **(Extended Fig 2**, **see Methods)**. This RFdiffusion-based method has the advantage over our previous hallucination-based method(*9*, *12*) in being able to generate pseudocyclic proteins around a target ligand, avoiding the need for large-scale ligand docking in a second step.

We used this pseudocycle CA-RFdiffusion approach to design protein structures around the JF494, JF596, and JF657 dyes. For each structure, we generated sequences using LigandMPNN(*13*). Each design is a single-domain protein of 115-160 aa, with a predicted cavity tailored to accommodate the desired JF dyes. We filtered the designs with Rosetta(*14*) for ligand binding and with AlphaFold2 (AF2)(*15*) for protein sequence-structure consistency. 4,000-6,000 top designs per dye were synthesized on oligonucleotide arrays (**see Methods** for statistics), displayed on the yeast surface, and screened by fluorescence-activated cell sorting (FACS). Next-generation sequencing of enriched pools revealed multiple high-affinity binders (**see Methods**).

The most enriched designed binders for each dye were expressed in *Escherichia coli* (*E. coli*), purified using affinity chromatography, and binding to each of the dyes were measured using fluorescent polarization (FP) assays (**see Methods**). We found that all designs bound their intended fluorophore targets with high affinity (**Fig. 1**); the estimated binding affinities of the best designs for each target (JF657-bp, JF596-bp, and JF494-bp) were 0.9 nM for JF657, 2.3 nM for JF596, and 19.3 nM for JF494. Using yeast surface display and flow cytometry, we found that each of these designs binds only the intended target dye, with negligible binding to the two non-targeted dyes (**Fig. 1**, **6th column**). JF657-bp and JF596-bp increased the fluorescence emission of their target dyes by nearly two fold compared to the dyes in buffer, likely by dye conformation restriction upon binding and decreased quenching by solvent (**Fig. 1**, **5th column**). As a result, the protein–dye complexes are exceptionally bright, going beyond the inherent brightness of these JF fluorophores. This property is advantageous for imaging applications, as it boosts signal-to-noise ratio (SNR) and allows the use of dye concentrations as low as 2 nM (**Extended Fig. 3**), which decreases the probability of unspecific dye staining and cell toxicity.

We purified each binder by size-exclusion chromatography, and obtained well-expressed, monomeric, soluble proteins (**see Methods**). Circular dichroism data confirmed predominantly α-helical secondary structures that match the design models and revealed high thermostability (**Extended Fig. 4**, **see Methods**). We determined the crystal structure of JF657-bp at 1.44 Å resolution (**Fig 2a**, **Extended Table S1**, **see Method**), and found that the crystal backbone closely matched the design model. AF3 predictions in the presence of JF657 generated ternary complexes also closely matching the design model **(Fig 2b)**. To examine the origins of the high specificity for JF657 over the other two dyes, we first superimposed each dye onto the ternary complex model **(Fig 2c)**. While JF657 makes extensive favorable interactions, there is a clear clash between JF596 and the JF657-bp backbone **(Fig 2d)**, and considerably fewer contacts between JF494 and JF657-bp **(Fig 2e)**. AF3 predictions of JF657-bp in complex with JF494 and JF596 yielded alternative binding modes with no polar contacts and fewer aliphatic contacts compared to the JF657–JF657-bp complex and their designed binders **(Extended Fig 5a-b**, **Extended Table S2**, **Fig 1)**, and with worse Rosetta binding energy and metric than the target design complex (**see Methods**). Our generative design approach thus produces binders with sufficiently high complementarity to the intended target which disfavor binding to the non-targeted dyes.

Cell imaging with organelle-specific dye binders

We sought to investigate the performance of JF657-bp, JF596-bp and JF494-bp as genetically encoded tags for fluorescence imaging. We began by fusing an N-terminal mitochondrial localization sequence (MitoTag from Mas70p) to JF657-bp, JF596-bp, and JF494-bp, along with eGFP (JF657-bp and JF596-bp) or mScarlet (JF494-bp) at the C-terminus to enable subcellular localization of the designed protein (**Fig 3a-c** and **Extended Fig 3**)(*16*, *17*). The Mito–binder–fluorescent protein constructs were individually transfected in HeLa cells. After incubation with 5 nM of the on-target dye in phenol-red-free media at 37 °C for 15 minutes, all three constructs exhibited co-localization of the dye and fluorescent protein signals at the mitochondrial surface **(Fig. 3a-c** and **Extended Fig 3)**. Thus, the three designed binders can effectively recruit the corresponding cell-permeable dye to specific subcellular locations at concentrations that produce little background staining of cells, a necessary pre-requisite for use in fluorescence imaging.

We next investigated the use of the dyes for multiplex multi-color imaging in cells. JF657-bp, JF596-bp, and JF494-bp were directed towards distinct subcellular compartments by fusion to Mas70p (targeting mitochondria), emerin (targeting the inner nuclear membrane), and 2xFYVE tag (targeting the early endosomal membrane), respectively(*16*, *18–20*). We transfected the three tagged designs simultaneously into HeLa cells, and incubated the cells with 5 nM JF657, 5nM JF494, and 20 nM JF596 in phenol-red-free media at 37 °C for 15 minutes. We observed three crisp non-overlapping fluorescent patterns corresponding to mitochondria (JF657 signal), the nuclear membrane (JF596 signal), and endosomes (JF494 signal) within the same cell **(Fig 3d)**. The three-color composite showed clear separation of signals with little bleed-through or cross-binding, underscoring the orthogonality and multiplexing capability of this system, which represents a substantial improvement over conventional fluorescent proteins, enabling more complex visualization of cellular architecture

with enhanced brightness in each channel. To assess the effect of cell fixation on JF657-bp, we added JF567 to cells transfected with MitoTag-JF657-bp-eGFP before and after cell fixation. Under both conditions, we observed a clear JF657 signal co-localized with eGFP (**extended Figure 6**); the ability to fix cells prior to staining is particularly advantageous for detection of protein interactions (see below).

Design of covalent binding

We sought to develop dye-specific covalent binders which could be useful for irreversible labeling and single-molecule imaging techniques. This approach would have the advantage over HaloTag-based systems(*5*) in that it requires no chemical modification of the dye, and the designs can be significantly smaller than HaloTag (297 residues). We identified positions on the JF657 dye susceptible to nucleophilic aromatic substitution (SNAr), and placed a cysteine in the JF657-binding protein with the thiolate side chain of the cysteine positioned adjacent to a susceptible aromatic carbon on the dye (**see Methods**). Upon binding, the cysteine can attack the dye and displace a leaving group, forming a covalent thioether bond between the dye and protein (**Fig. 4a** and **Extended Figure 7**).

We expressed and purified 47 such designs, incubated them with JF657, and investigated the formation of the protein-dye covalent adduct by high-resolution mass spectrometry (HR-MS, **Fig 4a**, **see Methods**). 12 out of 47 designs showed covalent binding signals. We characterized one of these designs (JF657-bp_cv) further and observed a mass shift corresponding to covalent modification of the design by JF657 dye ($[\text{Binder+JF657+H-Met}]^+$: observe = 18316.9 Da, expected = 18316.6 Da, error = 15.89 ppm). Sodium dodecyl sulfate polyacrylamide (SDS-PAGE) gel electrophoresis of denatured JF657-bp_cv–JF657 dye conjugates followed by staining with Coomassie Blue to visualize the protein revealed a clear JF657 derived fluorescent band at the same location as the protein staining (**see Methods, Extended Figure 7**). Covalently labeled JF657-bp_cv retained strong dye fluorescence and did not disassociate over extended time scales. Our designed system which effectively introduces a single step catalytic activity has the advantage over current HaloTag approaches in that the design has smaller size, is target-specific, and does not require chemical modification of the dye.

Design of a chemically induced dimerizer based on split JF657-bp

Chemically induced dimerization (CID) systems have been widely used to control protein and organelle localization, signal pathways, gene expression, and enzyme activity (*21*). Common CID systems, such as the rapamycin-inducible dimerization FKBP-FRB system, CATCHFIRE, gibberellin-based dimerizer, and ABA-based dimerizer (*21–24*), rely on non-fluorescent or relatively dim fluorogenic molecules for induction. Given the small size and nanomolar-affinity of our designed JF657-bp, we explored the potential of split variants for controlling protein-protein interactions with real-time fluorescence monitoring of the induced interactions in living cells **(Fig. 4c)**.

To evaluate the performance of the split system in cells, we utilized fluorescent colocalization analysis. We fused the split fragment JF657-bp_A to a MitoTag, and the red fluorescent protein, mScarelt(*25*), so that this fragment is localized to the mitochondrial outer membrane. The second fragment JF657-bp_B was fused to the green fluorescent protein, mNeonGreen(*26*), and expressed in the cytosol as a soluble, freely-diffusing protein. Upon induction of dimerization with 200 nM JF657 for 1 hour, the cytosolic fragment re-localized to the mitochondrial membrane and co-localized with the first fragment**.** The fluorescent signal of JF657 co-localized with the other two fluorescent protein signals as expected since the same molecule induces dimerization and provides a fluorescent indicator of the induced interaction (**Fig. 4d**). Titration of JF657 concentration indicated that 50 nM JF657 can induce dimerization, with higher dye concentrations yielding faster kinetics (**Extended Fig. 8**). Fluorescent live cell imaging showed a half-time of co-localization of $t_{1/2}$ = 23 ± 10 min (**Fig 4e,** and **Extended Movie 1**). The reconstituted dye binding can be visualized both in live cell imaging and post-fixation (**Fig 4d, e**). The split JF657 system provides a powerful new way to precisely visualize and control molecular interactions as the same molecule (the dye) both induces the interaction and serves as a fluorescent readout of that event.

Discussion

The de novo design of three new fluorophore-binding proteins highlights the power of computational protein design to generate new biological functions. By specifically designing proteins to bind synthetic dyes, we bridge the gap between the superior photophysical properties of small-molecule fluorophores and the genetic targetability of proteins. The three binder–dye pairs presented here provide an orthogonal toolkit of imaging tags, each providing an ultrabright fluorescent label in cells. As illustrated in **Fig. 3d**, researchers can now label multiple different proteins or structures in the same cell with three ultrabright JF dyes with minimal spectral overlap without need for large protein tags or secondary labeling steps. The potential advantages over current Halo-tag based systems extend beyond orthogonal binding, since binding is tight but reversible, dye molecules in solution can exchange with bound dye, reducing problems with photobleaching during long time imaging, and potentially enabling super-resolution techniques requiring fluorophore blinking in three different colors.

Our JF657 CID systems go beyond current approaches (such as the rapamycin-induced FKBP–FRB system) which enable small molecule control over protein interactions(*23*), but do not provide information on where in a cell the association is taking place. In our JF657-bp_A/JF657-bp_B system, the inducer is the JF657 fluorescent dye, such that dimerization is inherently coupled to a fluorescent signal, providing a direct readout of when and where the interaction occurs. Compared to previous systems(*23*) where fluorescent proteins were fused to CID systems for visualizing protein proximity, the fluorescent signal of our system directly come from the CID module, yielding a markedly simpler, smaller reporter that is brighter, more photostable, and effective at lower dye concentrations for both dimerization and imaging. The high affinity of the version of the system presented in **Fig. 4** allows direct control of protein interactions; for applications where detection of native protein-protein interactions (without perturbation by the dye) is desired, the CID pair can be reconstituted post cell fixation, allowing

visualization of protein pairs which are already in proximity, with reduced residual interaction between the split chains compared to current split protein reporters. In all the above scenarios, the superior photophysics properties of JF657 should enable super-resolution imaging of sites of interaction between two proteins, which is difficult with current split fluorescent protein systems.

A wide array of synthetic fluorophores have been described–including near-infrared dyes and those with special photophysical properties like photoactivation–and it should be readily possible to generate binders for many of these using the RFdiffusion-based approach described here, creating an even larger panel of orthogonal labels for multiplexed imaging beyond three colors. The ability to incorporate further functionality through catalytic sites which generate covalent interactions, and split CID/interaction reporter systems as described here should open up broad new vistas for cellular control and monitoring. Overall, our results provide a blueprint for marrying synthetic chemistry with protein engineering to advance the frontiers of biological imaging and beyond.

Figures

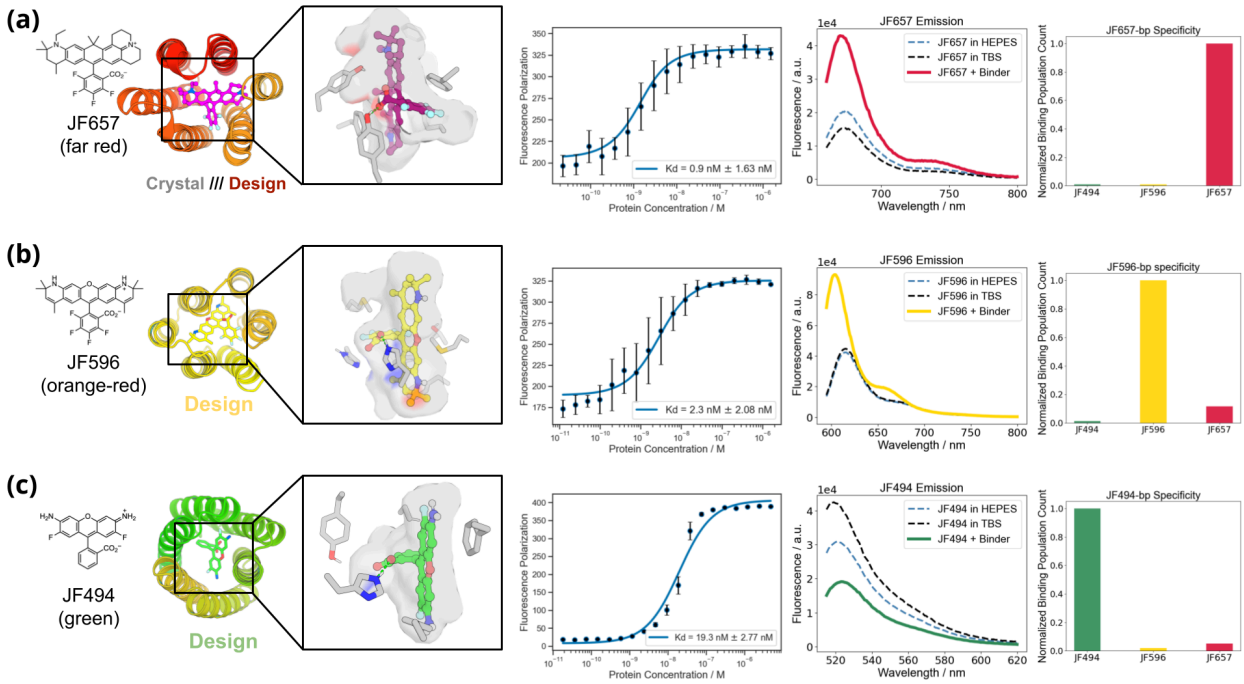


Figure 1: High-affinity and selectivity binders for Janelia Fluor® dyes.

Designs targeting JF657 **(a)**, JF596 **(b)** and JF494 **(c)**. For each dye-binder pair, from left to right are the (1) chemical structure of the target dye, (2) global and (3) zoom in views of the design models, (4) fluorescence polarization titrations with the corresponding Kd, (5) fluorescence spectra in the absence and presence of dye, and (6) dye-binding specificity measured by yeast surface display. For the specificity assay (6), binders bearing a C-terminal c-Myc tag were displayed on yeast, stained with fluorescein-labelled anti-c-Myc antibody to confirm expression, and incubated with 100 nM of each JF dye. Bar heights report the normalized number of yeast cells exhibiting both expression and ligand binding; taller bars indicate stronger binding. In the structure panels, the designs are displayed in cartoons, and the ligands in sticks. Nitrogen, oxygen, fluorine atoms are colored with blue, red, and teal, respectively. The binding pockets are shown in gray surface. Hydrogen bonds are shown in green dashes. Also see **Extended Fig. S5** and **Table. S2.**

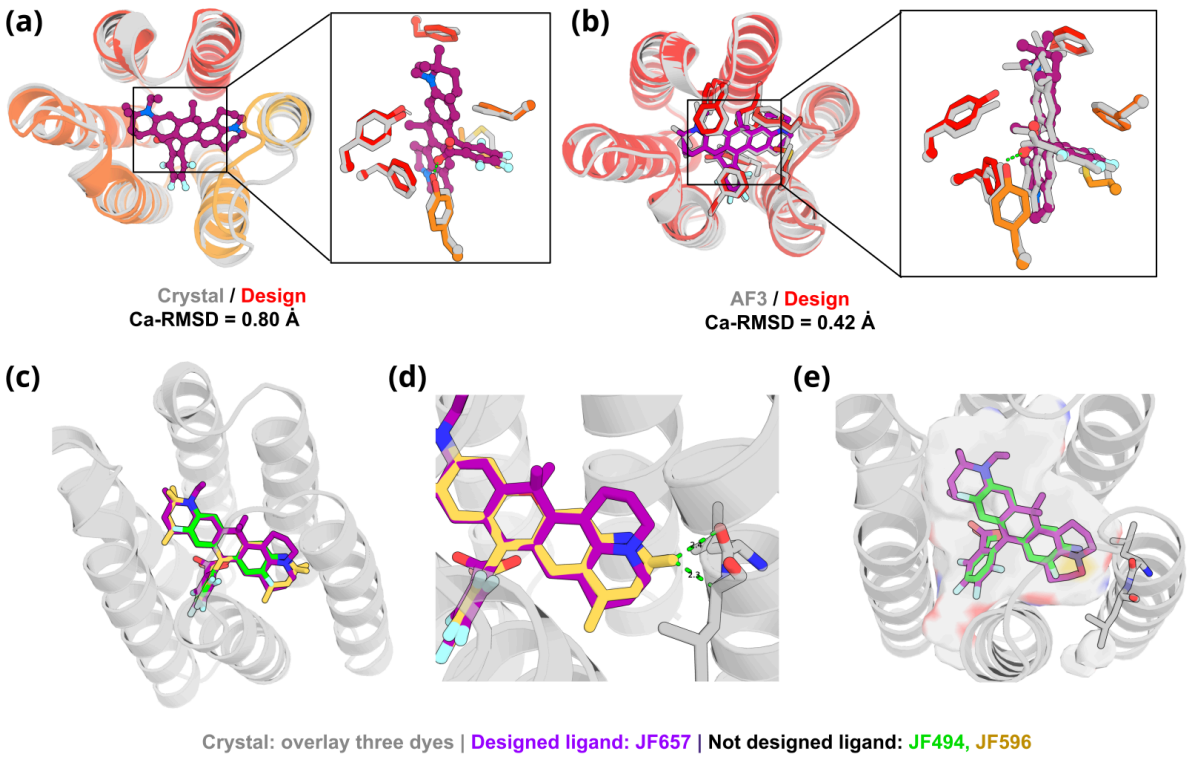
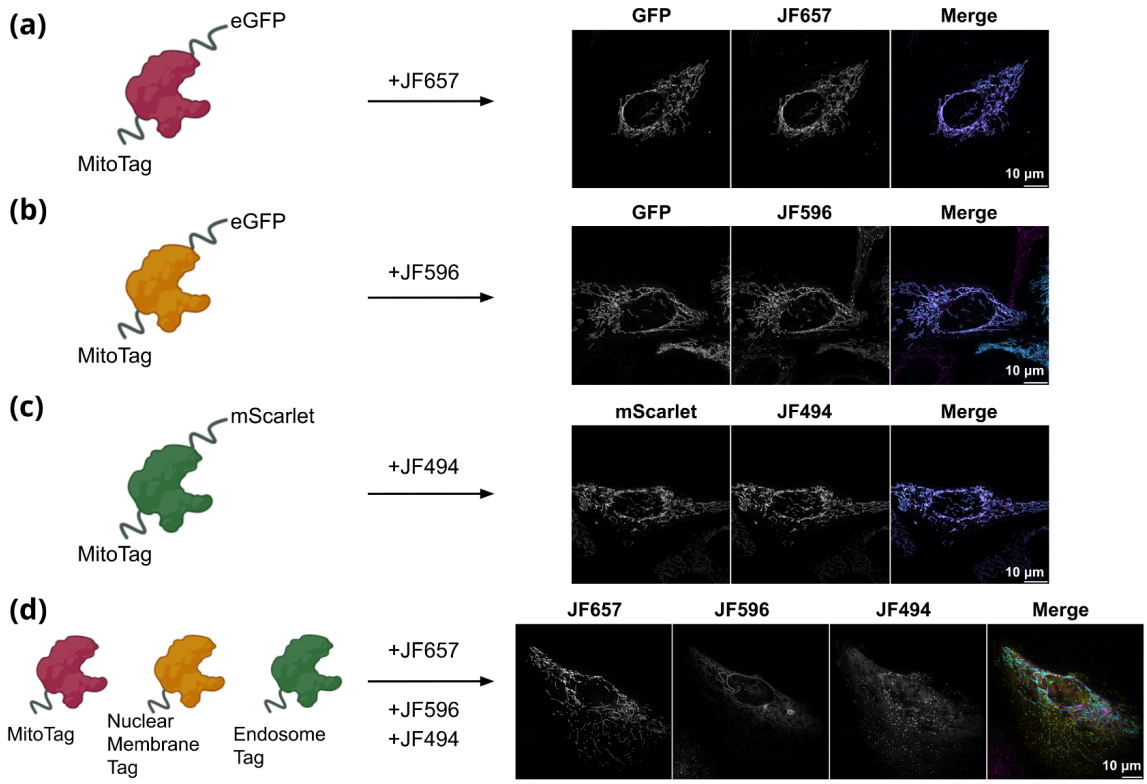


Figure 2: Structural characterization of JF657-bp.

The protein design model of the JF657-bp closely resembles the crystal structure **(a)**, and the protein and the ligand coordination match the AF3 prediction **(b)**. **(c)** Superposition of JF494 and JF596 on the JF657 model based on the shared rhodamine core. For JF596 **(d)** there are considerable clashes and for JF494 **(f)**, many interactions are lost, suggesting the same binding mode would not be feasible for not-designed targets. Nitrogen, oxygen, fluorine atoms are colored with blue, red, and teal, respectively. Proteins are shown in cartoons, ligands in sticks, and distances using green dashes. The binding pockets are shown in gray surface. Also see **Extended Fig. 5** and **Table S2**.



ma

Figure 3: Multiplexed imaging in live cells.

JF657-bp **(a)** JF596-bp **(b)**, and JF494-bp **(c)** were fused to MitoTag at their N terminus and eGFP or mScarlet at their C terminus and transfected into HeLa cells. The transfected cells were incubated with 5 nM of the respective dye and imaged live. Signals were visible in both fluorescent protein (left) and targeted JF dye (middle) channels, with clear co-localization of both colors (right) and little background signal. **(d)** Demonstration of multiplex imaging capabilities: JF657-bp, JF596-bp, and JF494-bp were fused with mitochondria, inner nuclear membrane, and endosomal localization tags, respectively, and transfected into HeLa cells. Fluorescence imaging of a single cell expressing all three dye binders post incubation with the three dyes showed JF657 (left), JF596 (middle-left), JF494 (middle-right), and 3-channel merge (right) specific labeling of the organelles to which their corresponding binder was targeted. A scale bar of 10 µm is shown in the bottom-right of each figure. Protein construct schematics on the left were generated with BioRender.

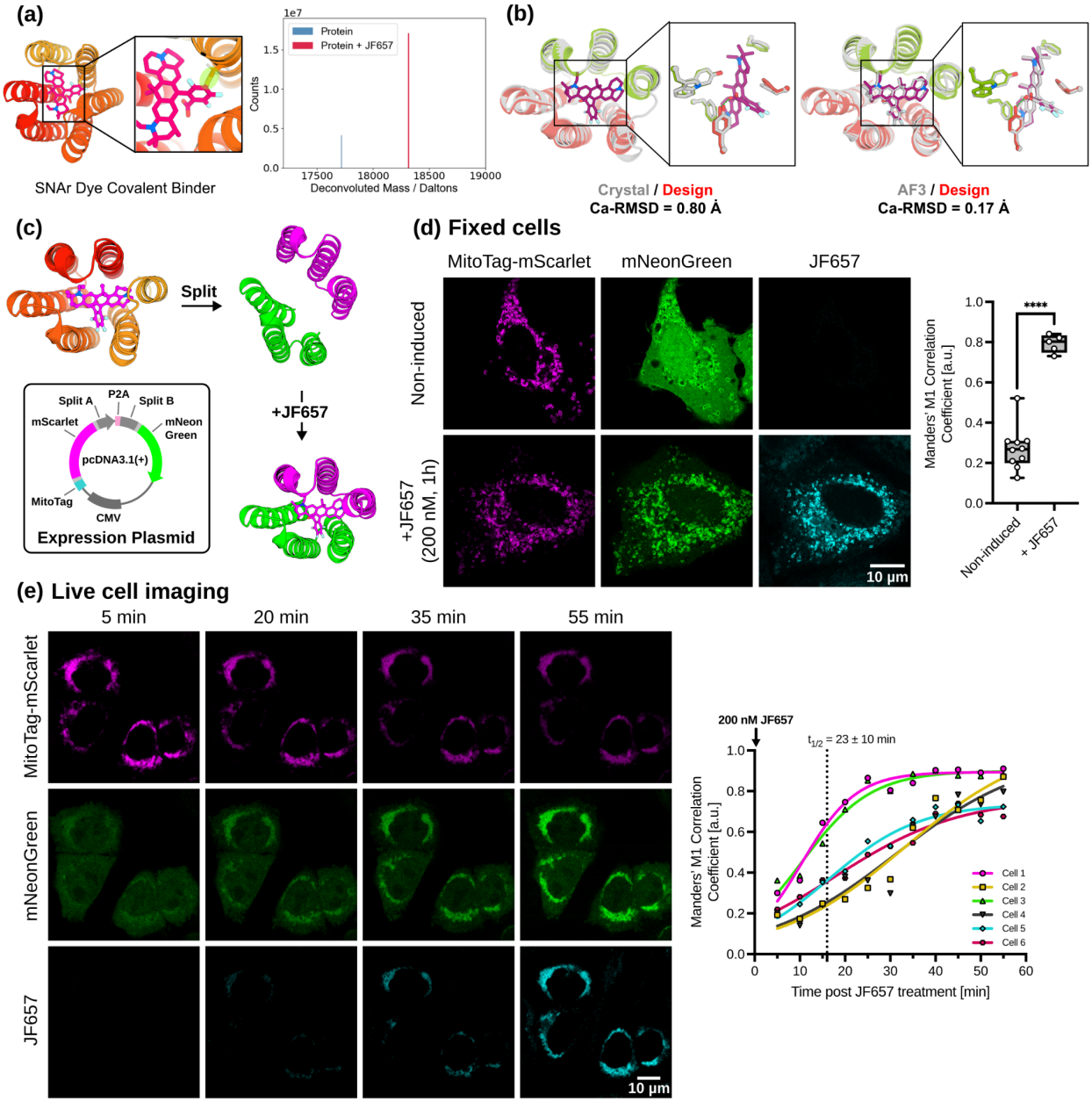
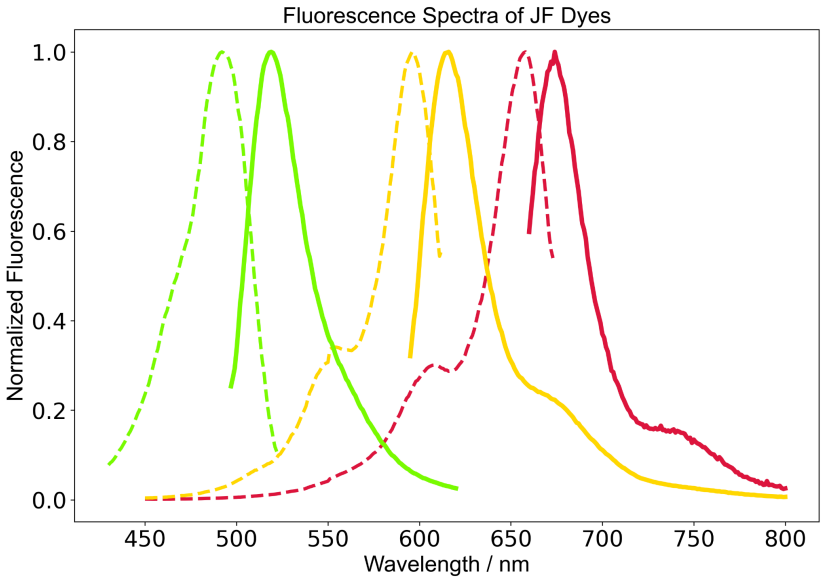


Figure 4: Design of covalent binder and chemically induced dimerization imaging systems.

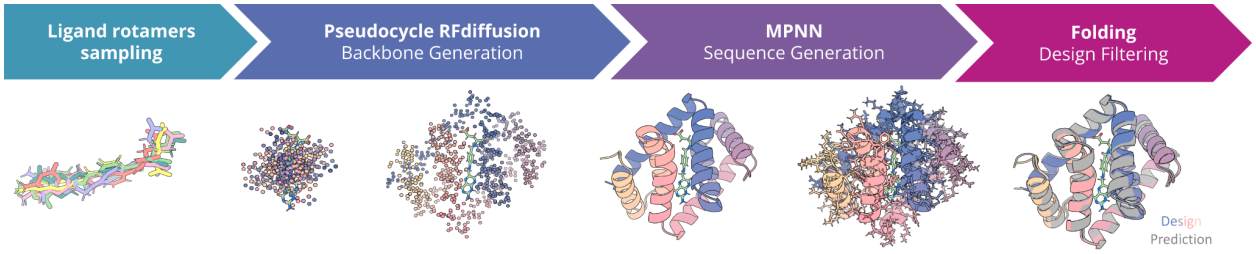
(a) JF657-bp_cv forms covalent bonds with JF657 through SNAr reaction. (left) A nucleophilic cysteine was designed into JF657-bp next to a susceptible aromatic carbon in the dye for SNAr reaction. (right) Mass spectrometry confirmed the formation of the binder-JF657 covalently linked species (expected mass 18316.6 Da, observed mass 18316.9 Da). **(b-e)** Design of a JF657-driven CID system. JF657-bp was split into two halves, such that association only occurs in the presence of JF657. **(b)** Crystal structure and AF3 prediction of the two halves with dye closely resemble the design model and crystal structure of JF657-bp. **(c)** Design strategy of a CID system using a split JF657-bp. The split system was evaluated in HeLa cells using the

shown expression plasmid with JF657-bp_A fused to mScarlet and MitoTag, and JF657-bp_B fused to mNeonGreen. Both constructs are separated by a ribosomal skipping site P2A. **(d)** HeLa cells expressing MitoTag-JF657-bp_A-mScarlet and freely diffusing cytoplasmic JF657-bp_B-mNeonGreen were imaged without (top row) or with (bottom row) incubation with 200 nM JF657 for 1 h at 37 ℃. In the absence of JF657, the mScarlet and mNeonGreen signals show weak co-localization, suggesting minimal interaction between JF657-bp_A and JF657-bp_B in the absence of JF657. In contrast, JF657-treated cells showed strong co-localization of mScarlet, mNeonGreen, and JF657 signals, indicating association of JF657-bp_A and JF657-bp_B interaction upon JF657 addition. **(e)** Live cell fluorescent imaging of HeLa cells expressing the same constructs described in **(d)**. Cells were treated with 200 nM JF657 and imaged every 5 minutes. Manders' M1 correlation coefficient was determined for individual cells at each time point and a logistic curve was fitted for each cell. The average half time with standard deviation for all analyzed cells is shown. Also see **Extended Fig. 7**, and **Extended Movie 1**.



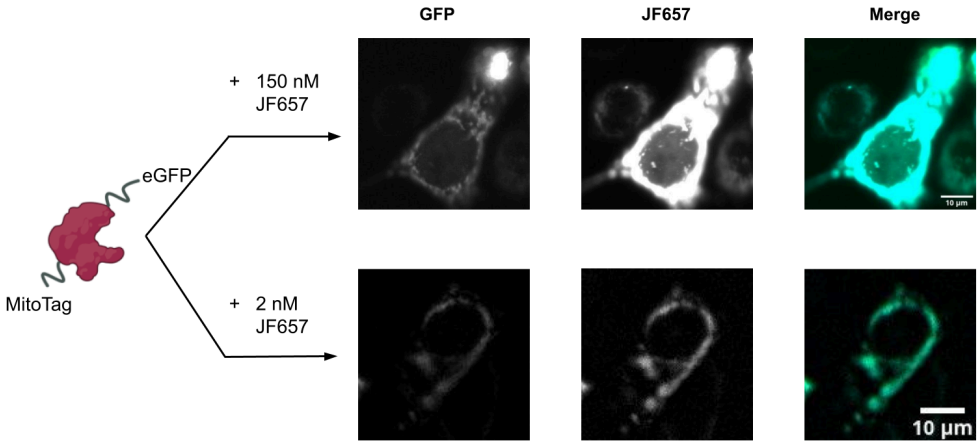
Extended Figure 1. Normalized fluorescence spectra of the three dyes.

Fluorescent spectra measurement of JF657 (red), JF596 (yellow-orange), and JF494 (green). Solid lines are the emission spectra, dotted lines are the excitation spectra. The selected fluorophores are spectrally distinct, with minimal overlap between their excitation and emission profiles, ensuring low signal crosstalk between fluorescent channels (**see Methods**).



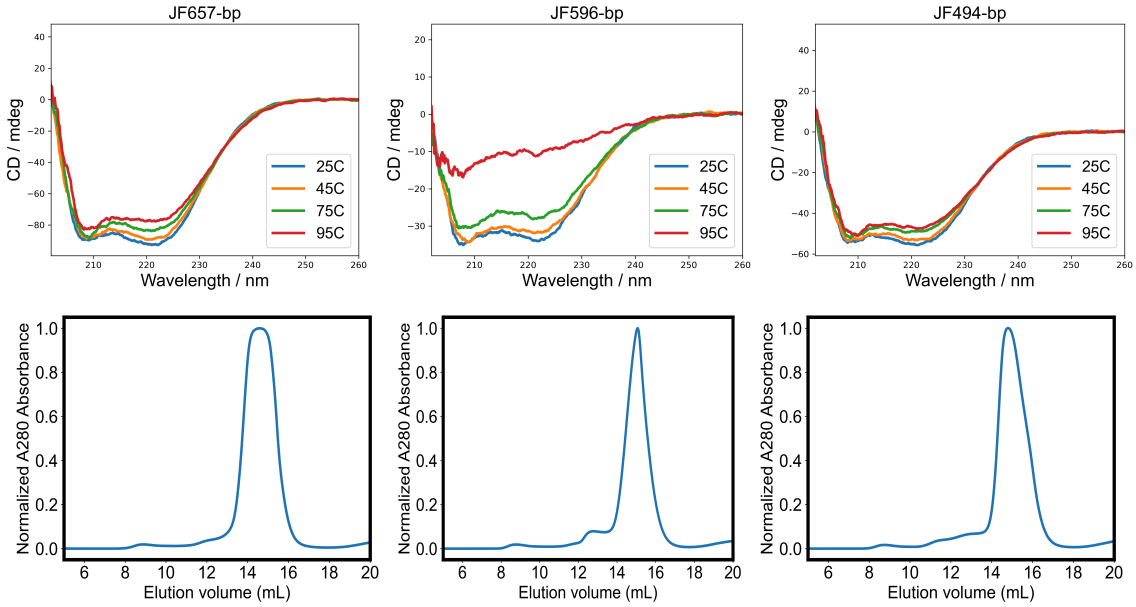
Extended Figure 2. Computational Design pipeline.

We generated the ligand rotamers using CREST(*27*). The protein backbones were generated with the pseudocycle CA-Diffusion(*9*, *10*). LigandMPNN was then performed on these generated backbones. Using Rosetta(*14*) and AlphaFold2(*15*), we performed the protein folding and ligand binding design winner selection.



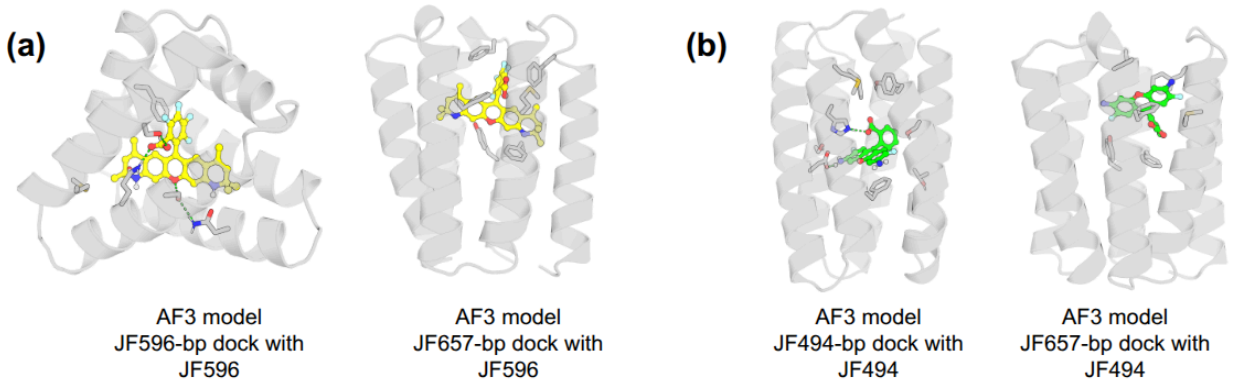
Extended Figure 3. Incubation at low concentration of dye enables adequate SNR subcellular organelle imaging.

A construct of MitoTag-JF657-bp-eGFP was transfected to HeLa cells. Specific labeling was achieved at both high (150 nM, as comparison, HaloTag experiments commonly require using of 25 nm to 500 nM dyes for live cell imaging) and low (2 nM) JF657 concentrations The protein construct schematic on the left was generated with BioRender.



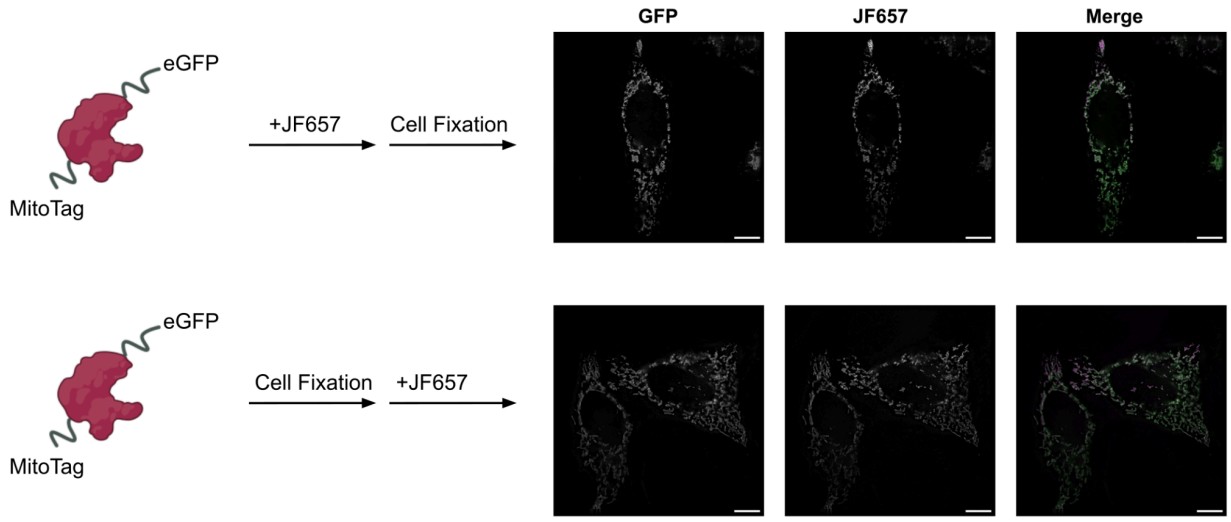
Extended Figure 4. Biochemical characterization of three binders.

Top row, left to right, JF657-bp, JF596-bp, JF494-bp were incubated at different temperatures and scanned for their secondary structure characteristics under circular dichroism. Bottom row, left to right JF657-bp, JF596-bp, JF494-bp were purified using size exclusion chromatography yielding monomeric protein profiles (**see Methods**).



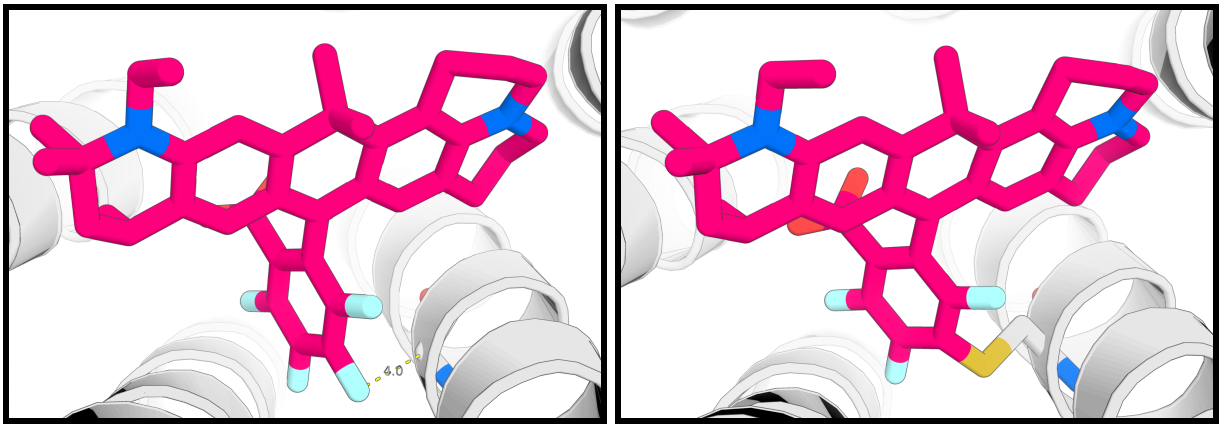
Extended Figure 5. Co-folding with targeted and non-targeted ligands for specificity analysis.

Using AF3(*28*), we predicted the co-folding complex model between the targeted and not-targeted ligand-protein pairs. Visually, a significantly smaller number of polar and aliphatic contacts is visible between not-targeted complexes compared to designed complexes. Energy-based analysis also suggested worse binding energy, less contact surface between not-designed pairs compared to designed pairs (See **Extended Table S2** for binding energy analysis). Collectively, these data supported our hypothesis that through optimizing shape complementary, we can optimize different pseudocyclic proteins towards their own designed targets, and achieve highly specific binders towards highly similar ligands.



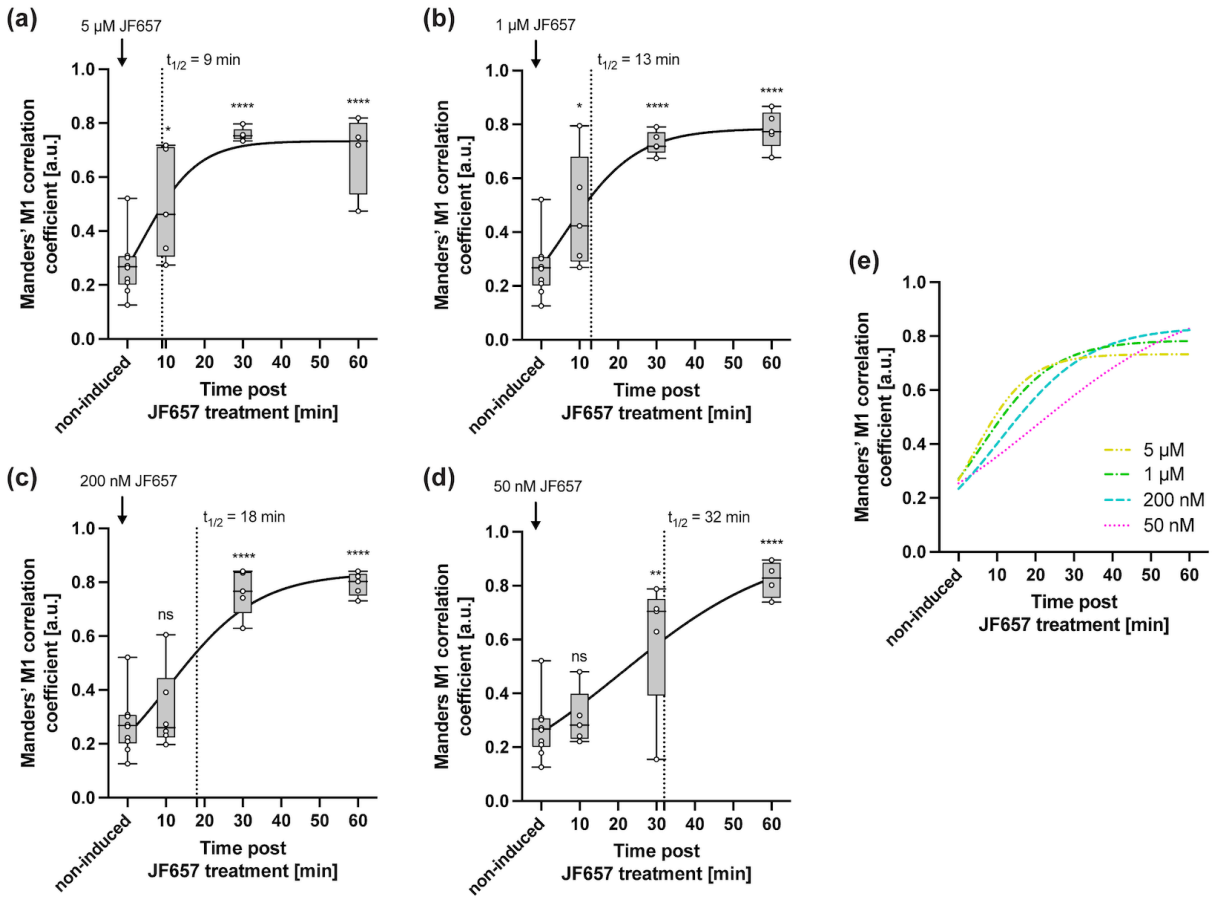
Extended Figure 6: Fluorescence imaging of fixed cells expressing MitoTag-JF657-bp-eGFP.

HeLa cells were transfected with the mitoTag-JF657-bp-eGFP. Cells were either first incubated with 5 nM JF657 for 15 minutes at 37 °C and washed 3 times with phenol-red-free media prior to fixed fixation **(Top panel)** or first fixed, then incubated with 5 nM JF657 for 15 minutes at 37 °C and washed 3 times with phenol-red-free media **(Bottom panel)**. A 10 µm white scale bar is displayed at the bottom right corner on every image.



Extended Figure 7. The covalent dye binding protein.

Top: coomassie blue gel image (**left**), and fluorescence gel image (**right**). The protein ladder and the designs were loaded to the gel for imaging. A few potential SNAr design hits show up in both coomassie blue and fluorescence gel images. The best design, JF657-bp_cv (highlighted in green box), shows strongest co-localization of expected designed protein staining (upper band) and far red signal, suggesting successful SNAr reaction happening. Column 1 and 10 were loaded with BioRad Precision Plus Protein™ Dual Xtra Prestained Protein Standards. **Bottom:** Design model of JF657-bp (**left**) and JF657-bp_cv (**right**) with catalytic cysteine installed.



Extended Figure 8. Kinetics study of JF657 concentration for CID design in cell. HeLa cells were transfected with the CID construct (for details see **Fig. 4c**), and treated with different concentrations of JF657: 5 µM **(a)**, 1 µM **(b)**, 200 nM **(c)**, and 50 nM **(d)**. Cells were fixed at specific time points, fluorescent images were acquired and Manders' M1 correlation coefficients between MitoTag-JF657-bp_A-mScarlet and JF657-bp_B-mNeonGreen were determined. Each data point represents one cell. For each time point the data is represented as a box and whisker plot with all data points shown. A logistic growth curve was fitted and the half-time $t_{1/2}$ was determined. In **(e)** the fitted curves for all dye concentrations are shown.

Extended Table S1. Crystal structure data collection and refinement statistics.

	JF657-bp (PDB Code: 9PVL)
Data Collection	
Space group	*P* 2_1
Cell parameters	
a,b,c (Å)	33.36, 123.99, 56.64
α, β, γ (°)	90, 92.28, 90
Resolution (Å)	33.33 - 1.44 (1.51 - 1.44)
R_{merge}	0.076 (0.897)
R_{pim}	0.035 (0.430)
$I/\sigma(I)$	10.9 (1.9)
$CC_{1/2}$	0.998 (0.625)
Completeness (%)	100 (100)
Redundancy	6.6 (6.1)
Refinement	
Resolution (Å)	33.33 - 1.44 (1.47 - 1.44)
No. reflections	83269
R_{work} / R_{free}	0.2112 (0.3099)/ 0.2447 (0.2970)
No. atoms	
Protein	4432
Solvent	392
Ramachandran Favored/allowed/ Outlier (%)	99.81/019/0.00
R.m.s. deviations	
Bond lengths (Å)	0.008
Bond angles (˚)	1.027
$B_{factors}$ (Å^2)	
Protein	23
Solvent	29

Extended Table S2. Rosetta analysis of bindings between designed and not-designed binding complexes.

Binder/Dye	**Binder**	**ddG**/REU	**Hydrogen bond number**	**contact_molecular_surface**
JF494	JF494-bp	-29.798	2	297.216
JF494	JF657-bp	-21.890	0	275.778
JF596	JF596-bp	-37.846	3	460.086
JF596	JF657-bp	-36.571	0	410.588

Author contribution

Long Tran: project design and execution, manuscript preparation; Shajesh Sharma#: library preparation; Steffen Klein#: pipeline design and cellular image for CID; David Jurgens#: computational scripts collaboration; Justin Decarreau#: cellular imaging and discussion, manuscript preparation; Bingxu Liu: discussion and project preparation; Yujia Wang: design pipeline design; Asim Bera: crystal curation; Xinting Li: HRMS data curation; Alex Kang: xtal curation; Dionne K Vafeados: library preparation; Nicole Roullier: library preparation; Wei Chen: experimental preparation; Gyu Rie Lee: design pipeline and experimental preparation; Julia Mahamid: CID concept and imaging supervision; Luke Lavis: project design; Linna An: project design, computational analysis, and manuscript preparation; David Baker: project design and manuscript preparation.

Acknowledgement

Crystallographic diffraction data was collected at either the Advanced Photon Source 24-ID-E or the CBMS/NSLS2. This research used resources 17-ID-1/17-ID-2 of the National Synchrotron Light Source II, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Brookhaven National Laboratory under Contract No. DE-SC0012704. The Center for BioMolecular Structure (CBMS) is primarily supported by the National Institutes of Health, National Institute of General Medical Sciences (NIGMS) through a Center Core P30 Grant (P30GM133893), and by the DOE Office of Biological and Environmental Research (KP1605010). This research was performed on APS beam time award(s) (DOI: https://doi.org/10.46936/APS-189693/60013820) from the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science user facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. We thank the Advanced Light Microscopy Facility (ALMF) and the cryo-electron microscopy platform at the EMBL, and Nikon and Zeiss for support. S.S., S.K., D.J., J.D. are co-second authors; they contributed equally, and everyone agrees that their authorship orders can be exchanged to benefit their own career development. The authors thank David Agard, Helen Eisenach, Shingo Honda, Thomas Schlichthaerle, Avi Swartz, Jonas Wilhelm, Xinru Wang, Yulai Liu, Yang Bo, Hojun Choi, and Green Ahn for helpful discussions and comments during the preparation of this manuscript. We acknowledge the use of ChatGPT, a language model developed by OpenAI, for minor suggestions with respect to the text.

Funding

The research is supported by Bill and Melinda Gates Foundation INV-043758 (GR019486: BMGF CORE LABS - 63-9201 - 2021, to J.D., A.K.B., W.C., L.A.); Gift from Microsoft (GF117374: Microsoft Protein Prediction Research, S.S.); Howard Hughes Medical Institute (GR020267: C19 HHMI INITIATIVE - 66-0656 - 2021, to A.K., E.J., N.R.); Merck (GR024322: MERCK COLLABORATION - 66-8270 - 2021, to L.T., D.V.); PG225041: Breakthrough Fund Therapeutic Machines (to B.X.L., Y.J.W.); The Audacious Project at the Institute for Protein Design (PG117878: Audacious Hub, PG117866: Audacious Discretionary Sub, to D.B.); European Commission through the ARISE programme (Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement number 945405, to S.K.); Chan Zuckerberg Initiative (Grant Number 2023-332009, to J.M and D.B.); Cancer Grand Challenges team MATCHMAKERS funded by Cancer Research UK CGCATF-2023/100008 (B.L., D.B.); the Damon Runyon Cancer Research Foundation (DRG-2507-23 to W.C.)

Conflict of Interest

L.T., S.K., J.M., L.A. abd D.B. filled a provisional patent for this work. Other authors claim no conflict of interest.

Reference

1. M. Hirano, R. Ando, S. Shimozono, M. Sugiyama, N. Takeda, H. Kurokawa, R. Deguchi, K. Endo, K. Haga, R. Takai-Todaka, S. Inaura, Y. Matsumura, H. Hama, Y. Okada, T. Fujiwara, T. Morimoto, K. Katayama, A. Miyawaki, A highly photostable and bright green fluorescent protein. *Nat. Biotechnol.* **40**, 1132–1142 (2022).

2. J. B. Grimm, A. N. Tkachuk, R. Patel, S. T. Hennigan, A. Gutu, P. Dong, V. Gandin, A. M. Osowski, K. L. Holland, Z. J. Liu, T. A. Brown, L. D. Lavis, Optimized Red-Absorbing Dyes for Imaging and Sensing. *J. Am. Chem. Soc.* **145**, 23000–23013 (2023).

3. J. B. Grimm, L. D. Lavis, Synthesis of rhodamines from fluoresceins using Pd-catalyzed C-N cross-coupling. *Org. Lett.* **13**, 6354–6357 (2011).

4. K. Yan, Z. Hu, P. Yu, Z. He, Y. Chen, J. Chen, H. Sun, S. Wang, F. Zhang, Ultra-photostable small-molecule dyes facilitate near-infrared biophotonics. *Nat. Commun.* **15**, 2593 (2024).

5. G. V. Los, L. P. Encell, M. G. McDougall, D. D. Hartzell, N. Karassina, C. Zimprich, M. G. Wood, R. Learish, R. F. Ohana, M. Urh, D. Simpson, J. Mendez, K. Zimmerman, P. Otto, G. Vidugiris, J. Zhu, A. Darzins, D. H. Klaubert, R. F. Bulleit, K. V. Wood, HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis. *ACS Chem. Biol.* **3**, 373–382 (2008).

6. A. Keppler, S. Gendreizig, T. Gronemeyer, H. Pick, H. Vogel, K. Johnsson, A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat. Biotechnol.* **21**, 86–89 (2003).

7. M. Holtmannspötter, E. Wienbeuker, T. Dellmann, I. Watrinet, A. J. Garcia-Sáez, K. Johnsson, R. Kurre, J. Piehler, Reversible Live-Cell Labeling with Retro-engineered HaloTags Enables Long-Term High- and Super-Resolution Imaging. *Angew. Chem. Int. Ed.* **62**, e202219050 (2023).

8. E. A. Specht, E. Braselmann, A. E. Palmer, A Critical and Comparative Review of Fluorescent Tools for Live-Cell Imaging. *Annu. Rev. Physiol.* **79**, 93–117 (2017).

9. L. An, M. Said, L. Tran, S. Majumder, I. Goreshnik, G. R. Lee, D. Juergens, J. Dauparas, I. Anishchenko, B. Coventry, A. K. Bera, A. Kang, P. M. Levine, V. Alvarez, A. Pillai, C. Norn, D. Feldman, D. Zorine, D. R. Hicks, X. Li, M. G. Sanchez, D. K. Vafeados, P. J. Salveson, A. A. Vorobieva, D. Baker, Binding and sensing diverse small molecules using shape-complementary pseudocycles. *Science* **385**, 276–282 (2024).

10. A. Lauko, S. J. Pellock, K. H. Sumida, I. Anishchenko, D. Juergens, W. Ahern, J. Jeung, A. F. Shida, A. Hunt, I. Kalvet, C. Norn, I. R. Humphreys, C. Jamieson, R. Krishna, Y. Kipnis, A. Kang, E. Brackenbrough, A. K. Bera, B. Sankaran, K. N. Houk, D. Baker, Computational design of serine hydrolases. *Science* **388** (2025).

11. L. Lu, X. Gou, S. K. Tan, S. I. Mann, H. Yang, X. Zhong, D. Gazgalis, J. Valdiviezo, H. Jo, Y. Wu, M. E. Diolaiti, A. Ashworth, N. F. Polizzi, W. F. DeGrado, De novo design of drug-binding proteins with predictable binding energy and specificity. *Science* **384**, 106–112 (2024).

12. L. An, D. R. Hicks, D. Zorine, J. Dauparas, B. I. M. Wicky, L. F. Milles, A. Courbet, A. K. Bera, H. Nguyen, A. Kang, L. Carter, D. Baker, Hallucination of closed repeat proteins containing central pockets. *Nat. Struct. Mol. Biol.* **30**, 1755–1760 (2023).

13. J. Dauparas, G. R. Lee, R. Pecoraro, L. An, I. Anishchenko, C. Glasscock, D. Baker, Atomic context-conditioned protein sequence design using LigandMPNN. *Nat. Methods* **22**, 717–723 (2025).

14. A. Leaver-Fay, M. Tyka, S. M. Lewis, O. F. Lange, J. Thompson, R. Jacak, K. W. Kaufman, P. D. Renfrew, C. A. Smith, W. Sheffler, I. W. Davis, S. Cooper, A. Treuille, D. J. Mandell, F. Richter, Y.-E. A. Ban, S. J. Fleishman, J. E. Corn, D. E. Kim, S. Lyskov, M. Berrondo, S. Mentzer, Z. Popović, J. J. Havranek, J. Karanicolas, R. Das, J. Meiler, T. Kortemme, J. J. Gray, B. Kuhlman, D. Baker, P. Bradley, "Rosetta3" in *Methods in Enzymology* (Elsevier, 2011; https://linkinghub.elsevier.com/retrieve/pii/B9780123812704000196)vol. 487, pp. 545–574.

15. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

16. D. G. Millar, G. C. Shore, The signal anchor sequence of mitochondrial Mas70p contains an oligomerization domain. *J. Biol. Chem.* **268**, 18403–18406 (1993).

17. B. P. Cormack, R. H. Valdivia, S. Falkow, FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173**, 33–38 (1996).

18. H. Stenmark, R. Aasland, B. H. Toh, A. D'Arrigo, Endosomal localization of the autoantigen EEA1 is mediated by a zinc-binding FYVE finger. *J. Biol. Chem.* **271**, 24048–24054 (1996).

19. S. Manilal, N. thi Man, C. A. Sewry, G. E. Morris, The Emery-Dreifuss Muscular Dystrophy Protein, Emerin, is a Nuclear Membrane Protein. *Hum. Mol. Genet.* **5**, 801–808 (1996).

20. A. Nagano, R. Koga, M. Ogawa, Y. Kurano, J. Kawada, R. Okada, Y. K. Hayashi, T. Tsukahara, K. Arahata, Emerin deficiency at the nuclear membrane in patients with Emery-Dreif uss muscular dystrophy. *Nat. Genet.* **12**, 254–259 (1996).

21. B. Z. Stanton, E. J. Chory, G. R. Crabtree, Chemically induced proximity in biology and medicine. *Science* **359**, eaao5902 (2018).

22. S. Bottone, O. Joliot, Z. V. Cakil, L. El Hajji, L.-M. Rakotoarison, G. Boncompain, F. Perez, A. Gautier, A fluorogenic chemically induced dimerization technology for controlling, imaging and sensing protein proximity. *Nat. Methods* **20**, 1553–1562 (2023).

23. D. M. Spencer, T. J. Wandless, S. L. Schreiber, G. R. Crabtree, Controlling Signal Transduction with Synthetic Ligands. *Science* **262**, 1019–1024 (1993).

24. F.-S. Liang, W. Q. Ho, G. R. Crabtree, Engineering the ABA Plant Stress Pathway for Regulation of Induced Proximity. *Sci. Signal.* **4**, rs2–rs2 (2011).

25. D. S. Bindels, L. Haarbosch, L. Van Weeren, M. Postma, K. E. Wiese, M. Mastop, S. Aumonier, G. Gotthard, A. Royant, M. A. Hink, T. W. J. Gadella, mScarlet: a bright monomeric red fluorescent protein for cellular imaging. *Nat. Methods* **14**, 53–56 (2017).

26. N. C. Shaner, G. G. Lambert, A. Chammas, Y. Ni, P. J. Cranfill, M. A. Baird, B. R. Sell, J. R. Allen, R. N. Day, M. Israelsson, M. W. Davidson, J. Wang, A bright monomeric green fluorescent protein derived from Branchiostoma lanceolatum. *Nat. Methods* **10**, 407–409 (2013).

27. P. Pracht, F. Bohle, S. Grimme, Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Phys. Chem. Chem. Phys. PCCP* **22**, 7169–7192 (2020).

28. J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard, J. Bambrick, S. W. Bodenstein, D. A. Evans, C.-C. Hung, M. O'Neill, D. Reiman, K. Tunyasuvunakool, Z. Wu, A. Žemgulytė, E. Arvaniti, C. Beattie, O. Bertolli, A. Bridgland, A. Cherepanov, M. Congreve, A. I. Cowen-Rivers, A. Cowie, M. Figurnov, F. B. Fuchs, H. Gladman, R. Jain, Y. A. Khan, C. M. R. Low, K. Perlin, A. Potapenko, P. Savy, S. Singh, A. Stecula, A. Thillaisundaram, C. Tong, S. Yakneen, E. D. Zhong, M. Zielinski, A. Žídek, V. Bapst, P. Kohli, M. Jaderberg, D. Hassabis, J. M. Jumper, Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).

29. C. Zhang, M. Shine, A. M. Pyle, Y. Zhang, US-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes. *Nat. Methods* **19**, 1109–1115 (2022).

30. D. M. Hoover, DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res.* **30**, e43 (2002).

31. D. Baker, L. F. Milles, B. I. M. Wicky, COMPOSITIONS AND METHODS FOR SEMI-AUTOMATED PROTEIN PRODUCTION.

32. M. Glögl, A. Krishnakumar, R. J. Ragotte, I. Goreshnik, B. Coventry, A. K. Bera, A. Kang, E. Joyce, G. Ahn, B. Huang, W. Yang, W. Chen, M. G. Sanchez, B. Koepnick, D. Baker, Target-conditioned diffusion generates potent TNFR superfamily antagonists and agonists. *Science* **386**, 1154–1161 (2024).

33. A guided tour into subcellular colocalization analysis in light microscopy - BOLTE - 2006 - Journal of Microscopy - Wiley Online Library. https://onlinelibrary.wiley.com/doi/10.1111/j.1365-2818.2006.01706.x.

34. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

35. S. Bolte, F. P. Cordelières, A guided tour into subcellular colocalization analysis in light microscopy. *J. Microsc.* **224**, 213–232 (2006).

36. W. Kabsch, *XDS*. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).

37. A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).

38. P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, *PHENIX* : a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).

39. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).

40. C. J. Williams, J. J. Headd, N. W. Moriarty, M. G. Prisant, L. L. Videau, L. N. Deis, V. Verma, D. A. Keedy, B. J. Hintze, V. B. Chen, S. Jain, S. M. Lewis, W. B. Arendall, J. Snoeyink, P. D. Adams, S. C. Lovell, J. S. Richardson, D. C. Richardson, MolProbity: More and better reference data for improved all-atom structure validation: PROTEIN SCIENCE.ORG. *Protein Sci.* **27**, 293–315 (2018).

Material and Methods

Computational design pipeline

We generate the ligand rotamers using CREST(*27*). Protein backbones are generated with the pseudocycle CA-Diffusion(*9*, *10*). We then filter these backbones by measuring surface area around the ligand, keeping only those with satisfactory contact area.
To design sequences, the LigandMPNN-FastRelax cycle was then performed on these filtered backbones(*9*). Using Rosetta(*14*) and AlphaFold2(*15*), we evaluated the protein folding and ligand binding design for final selection, following the previously published selection criteria(9). Generally, the major filters used are: plddt >= 0.85, Cα-RMSD <= 1.5, high contact_molecular_surface, low ddG, and low holes_around_lig.
In the end, we experimentally tested 6032, 5800, and 4843 variants of binder designs to JF657, JF596, and JF494, respectively. For the JF657-bp library, from 6032 designs, we got 334 binders enriched by yeast surface display screening, at 5 uM or tighter binding affinity. Based on this observation, we shifted our strategy to focusing on the best binders based on the enrichment.

The scripts will be available at github repo. (the repo will turn public after manuscript acceptance).

Computational SNAr covalent dye binder design pipeline

The design of JF657 covalent binder started from experimentally validated JF657-bp. These JF657-bp were first filtered to select for structures containing at least one residue with a 'Cα' atom to the para-fluorine from the JF657 at distance below 4.0 Å (see **Extended Figure 7**). The residues satisfying this geometric criterion were then mutated to cysteine as the designated nucleophile for an attack on the para-fluorine-bearing carbon. We then used Rosetta(*14*) to further filter for designs with optimal distance and orientation between the thiol group to para-fluorine through first run `FastRlex` protocol, then filter based on the measured distance metric.

Computational Chemically induced dimerization design pipeline

The design of JF657-induced dimerization started from experimentally validated JF657-bp and split into two fragments (denoted "A" and "B"). In total, we created 132 split variants. These designs were filtered for high confident protein-small molecule complexes formation, as reflected by RoseTTAFold-All-Atom(19) protein-ligand-pae score (<= 13), AF2 multimer(13) pLDDT score (>= 92) and US-align (*29*) TMscore between design and AF2 prediction (>= 0.92).

Computational Rosetta-based analysis of binders

To analyze the specificity between binders and their corresponding dyes, we performed Rosetta energy(*14*) analysis between the binder and the designed and not-designed dyes.
Using AF3(*28*), we generated the complexes between the JF657-bp bound to JF494 and JF596. The predicted protein scaffolds (JF657-bp) are highly similar to the crystal structure, indicating

the AF3 prediction results can be used for downstream analysis with high fidelity. Similarly, we generated the complexes between JF494-bp bound to JF494, and JF596-bp and JF596.The same Rosetta protocol was used for complex scoring. The complex was first minimized, then scored for binding ddG, counted for hydrogen bond number, and scored for other metrics commonly used for ligand-protein interactions(*9*).
Based on the results (**Extended Table 2**), both ligands JF494 and JF596 exhibited worse ddG, hydrogen bond number, and contact molecular surface to JF657-bp than their designed binder, supporting our hypothesis that through optimizing shape complementary, we can differentiate different binders towards highly similar ligand variants.

Yeast surface display library preparation

All experiments were performed following previously published protocols(*9*). All library oligos were obtained from Twist Bioscience (South San Francisco, CA). Roche (Basel, Switzerland) supplied the KAPA HiFi HotStart Uracil+ReadyMix (KK2802) and KAPA HiFi HotStart Kit (KK2101). Gene fragments were sourced from Twist Bioscience (South San Francisco, CA) or IDT (Coralville, IA). EvaGreen Dye, 20 × in water (#31000), was purchased from Biotium (Fremont, CA). The USER enzyme (NEB#M5508) and NEBNext End Repair Module (E6050L) came from New England BioLabs (Ipswich, MA). DNA purification kits were purchased from QIAgen (Hilden, Germany). Unless otherwise specified, all chemicals and consumables were purchased from Thermo Fisher (Bothell, WA). Anti-cMyc-R, Phycoerythrin (anti-cMyc-PE) was obtained from Cell Signaling Technologies (Danvers, MA).

Designed protein sequences were back-translated with DNAworks(*30*), ensuring that common restriction enzyme cut sites were excluded. Each nucleotide sequence was split into two segments—a 5′ part and a 3′ part—sharing complementary overlap in the center; this was automated with an in-house Python script. To minimize qPCR bias from oligo-length variation, designs were binned into groups of ~1,000 by length. Primers matching the pETCON3 yeast-surface-display vector were appended to the 5′ end of the 5′ segment and the 3′ end of the 3′ segment, while inner primers containing AA or TT were added to the opposite ends to streamline library PCR. Depending on design length, 250- or 300-nt oligo pools were ordered from Twist Bioscience.

Oligo pools were resuspended in water to 100 ng/μL and stored at −20 °C, then diluted to 2.5 ng/μL for qPCR. Individual 5′ and 3′ segments were amplified with custom primers using KAPA HiFi HotStart Uracil+ReadyMix, with DNA purification between qPCR steps. USER enzyme and the NEBNext End Repair Module removed inner primers per the manufacturers' protocols. After purification, the digested 5′ and 3′ fragments were assembled by the same qPCR protocol, and assemblies were verified by gel electrophoresis. Desired DNA bands were excised and purified for further amplification. Approximately 4–6 μg of the pooled oligos were used to transform electrocompetent yeast. The oligo pool and linearized pETCON3 vector were mixed at a 3:1 ratio, and yeast were transformed by electro-transfer following the previously published protocol(*9*).

Yeast surface display experiment for enrichment of the potential binders

The established yeast-surface-display protocol(*9*) was followed with only minor adjustments. PBSF (PBS containing 0.1 % w/v BSA) served as the buffer for every yeast-sorting step. Strain EBY1 was used throughout all surface-display experiments. Library fluorescence-activated cell sorting (FACS) was conducted on a Sony SH800S Cell Sorter fitted with a 100 µm chip and 488 nm/561nm/638nm lasers.

Each yeast library first underwent an expression sort to enrich binders showing adequate surface display. Binder-expressing cells were stained 1 : 50 (v/v) with anti-cMyc-PE for 30 min at 4 °C; cells exhibiting elevated PE signal were collected for downstream binding assays. These cells were regrown in C-Trp-Ura medium with 2 % glucose (CTUG) and then induced overnight at 30 °C in SGCAA medium.

The induced yeast were stained 1 : 35 (v/v) with anti-cMyc-PE for 45 min at 4 °C, washed twice in PBSF, and incubated with the dye ligand for 60–120 min at 4 °C. After two further PBSF washes, cells were sorted. Unstained induced yeast served as a negative control; cells positive in both dye fluorescent channel and PE channel above this control were collected and expanded in CTUG. Following 1–2 days of growth, cultures were re-induced in SGCAA medium and sorted again as described. Typically, two to three rounds were required to achieve clear enrichment, and when binding populations exceeded 2 %, titrations were performed to pinpoint the strongest binders.

Samples from the expression step and every sorting round were reserved for MiSeq analysis. Yeast-sorting data were processed and visualized using FlowJo v10 (FlowJo, Inc.).

Expression and purification of binders

The workflow largely followed an earlier publication(*9*), with all reagents and consumables obtained from Thermo Fisher unless noted otherwise. Selected designs were back-translated with DNAWorks, and eblocks (IDT or Twist Bioscience) were inserted into a pET29b-based vector bearing a C-terminal, SNAC-cleavable His-tag by Golden Gate assembly (New England Biolabs, Ipswich, MA) before transformation into *E. coli* BL21.

Small-scale expression and purification of verified binders

Designs exhibiting promising activity were subjected to rapid, small-scale expression and purification to enable timely validation, following previous protocol(*31*, *32*). 1 µL Golden Gate subcloning reactions were assembled in 96-well PCR plates, transformed into chemically competent BL21(DE3) cells, and, after a 1 h outgrowth, distributed into four 96-deep-well plates, each containing 1 mL autoclaved TB auto-induction medium supplemented with kanamycin, 2 mM $MgSO_4$, and 1 × 5052, for a final culture volume of ~4 mL per construct. After 20–24 h incubation, cells were harvested, lysed, and the clarified lysates were loaded directly onto 100 µL Ni-NTA agarose beds in 96-well fritted plates pre-equilibrated with Tris wash buffer. Following thorough washing, bound proteins were eluted in 200 µL Tris buffer containing 500 mM imidazole.

Eluates were sterilized through a 0.22 µm 96-well filter plate (Agilent 203940-100) before size-exclusion chromatography (SEC). Samples were analysed on an ÄKTA FPLC equipped with an autosampler compatible with 96-well source plates. Monomeric binders were resolved on a Superdex 75 Increase 5/150 GL column (Cytiva 29148722).

For SNAr reaction designs, the purified binders were incubated with the dyes, followed by HR-MS analysis.

Large-scale expression and purification of verified binders

For designs with validated binding, catalyzing activities, we perform large-scale expression and purification to prepare more material.

Cultures (50 mL) were grown in auto-induction Terrific Broth within 250 mL baffled Erlenmeyer flasks—6 h at 37 °C followed by 24 h at 18 °C, shaken at ~200 g in New Brunswick Innova 44 shakers. Cells were harvested, resuspended in 25 mL lysis buffer (25 mM Tris, 100 mM NaCl, pH 8, a protease-inhibitor tablet) and sonicated for 4.5 min (10 s on/10 s off, 65 % amplitude). After a 30-min spin at 14,000 g, the supernatant was incubated for 1 h with 1 mL Ni-NTA resin (Qiagen) in a BIO-RAD Econo-Pac gravity column, rotating gently. Resin washes comprised 20 CV of low-salt buffer (50 mM Tris, 100 mM NaCl, 10 mM imidazole, pH 8) and 20 CV of high-salt buffer (50 mM Tris, 1000 mM NaCl, 10 mM imidazole, pH 8). Proteins were eluted with 2 CV of elution buffer (20 mM Tris, 100 mM NaCl, 500 mM imidazole, pH 8) and polished on a Superdex 75 Increase 10/300 GL column (ÄKTA system) in isocratic TBS (20 mM Tris, 100 mM NaCl, pH 8).

For proteins used in fluorescence polarization and crystallography experiments, His-tags were removed via on-bead SNAC cleavage. After binding, columns were rinsed twice with lysis buffer to eliminate imidazole, equilibrated with 10 CV SNAC buffer (100 mM CHES, 100 mM acetone oxime, 100 mM NaCl, 500 mM guanidine chloride, pH 8.5), and then incubated overnight at room temperature with 20 CV fresh cleavage solution (2 mM $NiCl_2$ in SNAC buffer). Resin was washed with 10 CV weak-elution buffer (20 mM Tris, 100 mM NaCl, 10 mM imidazole, pH 8); both the cleavage and wash fractions were collected and assessed by SDS-PAGE for tag-free protein. These proteins were pooled, concentrated, subjected to the same SEC protocol as above, and verified by mass spectrometry.

Fluorescent polarization assay for binder identification

The protein concentrations were measured using UV-Vis.
All FP studies and analyses were performed following a previously established protocol(*9*) on a Biotek Synergy Neo2 Reader equipped with Dual PMT optical filter cube (part number: 1035108).
The proteins were serial diluted and 0.8 nM of dye ligands were added to the diluted protein. The mixtures were transferred to a Corning 96 Well Half Area Flat Bottom Non-Binding Surface Black plate (Corning REF 3686). After 5 min incubation at 25 ℃, the parallel and perpendicular

light were read with scale to the highest-intensity well. Two independent FP were performed as replicates. FP signal was calculated based on below equation:

$$FP = I_{parallel} - I_{perpendicular}$$

The data were fit with the standard binding isotherm model.

Circular dichroism spectra

Circular dichroism (CD) spectra for the chosen designs were acquired on a Jasco J-1500 spectrometer. Each sample (0.5 mg/mL) was prepared in TBS pH 8 and placed in a 1 mm-path-length cuvette. Raw CD data were converted to molar mean residue ellipticity by dividing the signal by $N \times C \times L \times 10$, where N is the residue count, C is the protein concentration, and L is the optical path length (0.1 cm). A set of key binders was examined (**Extended Fig. 4**); as anticipated, all displayed fold-specific CD profiles, and most retained their folded state throughout melting–refolding experiments.

High-resolution Mass spectrometry

Protein samples after size exclusion chromatography purification were normalized to 20 uM (about 0.3 mg/mL), and subsequently incubated with 400 uM of JF657 overnight at 37 °C in TBS (20 mM Tris, 100 mM NaCl, pH 8). Prior to mass spectrometry measurement, the incubated samples were buffer exchanged to TBS using Zeba Spin Desalting Columns (Thermo Fisher Scientific) to remove excess JF657. HR-MS is obtained under positive mode on an Agilent G6230B ToF equipped with an AdvanceBio RP-Desalting column, and subsequently deconvoluted by way of Bioconfirm using a total entropy algorithm.

Cellular imaging

For live-cell imaging, HeLa cells (ATCC CCL-2) were cultured at 37 °C with 5% CO_2 in flasks with Dulbecco's modified Eagle medium (DMEM) (Gibco) supplemented with 10% fetal bovine serum (FBS) (Hyclone) and 1% penicillin–streptomycin (PenStrep) (Gibco). For transient transfection, cells were seeded at 1.2×10^4 cells per well in a 18-well glass-bottom plates (Ibidi USA, μ-slide, cat# 81817), and allowed to adhere and grow overnight. 24 hours after plating, each well was transfected with 100 ng of the plasmid encoding the binder construct using JetPrime, following the manufacturer's protocol. Exactly one day later, culture medium was exchanged for phenol-red-free medium supplemented with the indicated concentration of JF dye (5 nM JF657, 20 nM JF596, 5 nM JF494), and cells were incubated at 37 °C for 15 min to permit labelling. Wells were then washed three times with fresh phenol-red-free medium; the final wash was left in place for imaging. Fluorescence images were acquired immediately thereafter under live-cell conditions using standard filter sets matched to the respective dye.

Co-localization images were automatically sampled using an IN Cell Analyzer 2500HS (Molecular Devices) and a 60× 0.95 NA CFI Plan Apo objective (Nikon). Cells were illuminated with a seven-color Solid State Illuminator (SSI) for fluorescence excitation. Fluorescence signals were acquired sequentially using the following filters: Green (excitation 473/28 nm, emission 511/23 nm), yellow/orange (excitation nm, emission nm), and far-red (excitation 631/28 nm,

emission 684/24 nm). Imaging was controlled using the IN Cell Analyzer 2500 HS software version 7.4 and light collected on sCMOS camera without binning.

Three-color, multiplexed images were acquired with a commercial OMX-SR system (Leica). Toptica diode lasers with excitations of 488 nm, 568 nm, and 640 nm were used. Emission was collected on three separate PCO.edge sCMOS cameras using an Olympus 60× 1.420-NA PlanApochromat oil immersion lens; 1024 × 1024 images (pixel size, 6.5 μm) were captured with no binning. Acquisition was controlled with AcquireSR Acquisition control software. Images were deconvolved in SoftWoRx 7.0.0 (GE Healthcare) using the enhanced ratio method and 200 nm noise filtering, followed by background correction using the rolling-ball algorithm in ImageJ/Fiji (sigma= 35).

Fluorescence spectra scanning

Fluorescence spectra scanning was performed on a Biotek Synergy Neo2 Reader (Agilent). To measure the fluorescence spectra of bound and unbound dyes, we incubated freshly purified JF dyes binders with their respective JF dyes, with an excess molar ratio of the binders to the dyes. After incubation at room temperature for 5 minutes, the mixture was transferred to a Corning 96 Well Half Area Flat Bottom Non-Binding Surface Black plate (Corning REF 3686). Subsequently, equal volume and concentration of JF dyes, in different buffers, including protein buffer TBS and 10 mM HEPES pH 7.3, were transferred to the plate. For fluorescence emission spectral scanning, samples containing JF657 were excited at 650 nm, and emission was recorded from 660 to 800 nm in 1-nm increments. Samples containing JF596 were excited at 575 nm, with emission collected from 595 to 800 nm. Samples containing JF494 were excited at 505 nm, and emission was scanned from 515 to 620 nm, all using a 1-nm step size.

SNAr protein identification through fluorescent SDS-PAGE protein gel electrophoresis

To screen for JF657-bp_cv, we first expressed designed proteins in an in-house high throughput expression and selection procedure(*31*). These proteins were then purified on High-Performance Liquid Chromatography (HPLC) as protocol described above. Proteins at the expected peaks were chosen to be incubated with 5 uM JF657 at 37 °C in 20 mM Tris, 100 mM NaCl pH 8 for 12 hours. After incubation, an equal volume of BIORAD 2x Laemmli Sample Buffer was added to the incubated protein - JF657 mixture, then the mixture was heated at 95 °C for 5 minutes, and loaded on SDS-PAGE protein gel. Gel electrophoresis performed at 180 V for 30 minutes. The gel was then imaged under LI-COR Odyssey M imager to localize the far red fluorescent signal. Afterwards, the gel was stained with coomassie blue, and finally analyzed for protein bands using BIORAD ChemiDoc XRS+.

Fluorescence microscopy co-localization evaluation for CID

Fluorescent co-localization analysis was used to evaluate the top 8 CID designs. We began by building plasmids encoding for the mitochondrial localization tag MitoTag (N-terminal peptide from the Mitochondrial Import Receptor Subunit TOM70: ‘EMKSFITRNKTAILATVAATGTAIGAYYYYNQLQQ’), followed by JF657-bp_A, mScarlet, then ribosomal skipping P2A, followed by JF657-bp_B, fused to mNeonGreen **(Fig 4c)**. We

envisioned the JF657-bp_A would be effectively localized at the mitochondria, whereas the JF657-bp_B expressed freely in the cytosol. Constructed plasmids were transiently transfected into HeLa Kyoto cells using Lipofectamine™ 3000 Transfection Reagent (Thermo Fisher Scientific). Cells were then incubated overnight at 37 °C, 5% CO2 prior to imaging the following day. The experimental group was incubated with 200 nM JF657 for 1 h at 37 °C. Cells were fixed with 4% paraformaldehyde (PFA) in PBS and wide-field fluorescence images were acquired on a Nikon Ti-Eclipse microscope using a CFI P-Apo 40x/0.95NA air objective. Co-localization analysis between MitoTag-JF657-bp_A-mScarlet and JF657-bp_B-mNeonGreen was performed using the JaCoP plugin in Fiji (*33*, *34*).

Fluorescence live cell microscopy for CID

Fluorescent live cell imaging was used to evaluate kinetics of dimerization for the top CID design based on the fluorescence co-localization screening utilizing the same plasmid design as described above. HeLa Kyoto cells were transfected with TransIT HeLa (Mirus). Cells were then incubated overnight at 37 °C, 5% CO_2 prior to imaging the following day. Cells were treated with 200 nM JF657 and confocal fluorescence images were acquired every 5 minutes post treatment on a Zeiss LSM880 AiryScan microscope using a Plan-Apochromat 40x/1.4NA oil DIC objective. Co-localization analysis between MitoTag-JF657-bp_A-mScarlet and JF657-bp_B-mNeonGreen was performed using the JaCoP plugin(*35*) in Fiji. For each cell, a logistic growth curve (Y=YM*Y0/((YM-Y0)*exp(-k*x) +Y0)) was fitted and the half time ($t_{1/2}$=(1/k)*ln((Y0+YM)/Y0)) was determined with GraphPad Prism.

Cell Fixation

Fixation was performed by incubating cells with a 4% v/v paraformaldehyde (PFA) solution in PBS for 10 minutes at room temperature. Following fixation, cells were washed three times with PBS to remove residual PFA.

Crystallography conditions

Crystallization experiments were conducted using the sitting drop vapor diffusion method. Initial crystallization trials were set up by Mosquito LCP from SPT Labtech. Formulated total 200 nL drops using the 96-well plate format at 20˚C. Plates were imaged using UVEX microscopes and UVEX PS-256 from JAN Scientific.
Diffraction quality crystals formed in 0.2 M ammonium sulfate, 0.1 M Tris pH 7.5 and 20 % (w/v) PEG 5000 MME for in 11 % (w/v) PEG 3350, 0.1 M HEPES pH 7.5 for JF657-bp (**Extended Table S1**).

Diffraction data was collected at the National Synchrotron Light Source II on beamline either 17-ID-1 or 17-ID-2. X-ray intensities and data reduction were evaluated and integrated using XDS(*36*) and merged/scaled using Pointless/Aimless in the CCP4 program suite(62). Structure determination and refinement starting phases were obtained by molecular replacement using Phaser(*37*) using the designed model for the structures. Following molecular replacement, the models were improved using phenix.autobuild(*38*); with rebuild-in-place to false, and using simulated annealing. Structures were refined in Phenix(*38*). Model building was performed using

COOT(*39*). The final model was evaluated using MolProbity(*40*). Data collection and refinement statistics are recorded in the statistics table below. Data deposition, atomic coordinates, and structure factors reported in this paper have been deposited in the Protein Data Bank (PDB), http://www.rcsb.org/ with accession code 9PVL.