Supplement: Generation of ordered protein assemblies using rigid three-body fusion

Ivan Vulovic¹,²,³, Qing Yao⁴†, Young-Jun Park²†, Alexis Courbet¹,², Andrew Norris⁵,⁶, Florian Busch⁵,⁶, Aniruddha Sahasrabuddhe⁵,⁶, Hannes Merten⁷, Danny D. Sahtoe¹,², George Ueda¹,², Jorge A. Fallas¹,², Sara J. Weaver⁴, Yang Hsia¹,², Robert A. Langan¹,², Andreas Plückthun⁷, Vicki H. Wysocki⁵,⁶, David Veesler², Grant J. Jensen⁴,⁸,⁹, David Baker¹,²,³,¹⁰

† These authors contributed equally to this work.

¹Institute for Protein Design, University of Washington, Seattle, WA, USA.
²Department of Biochemistry, University of Washington, Seattle, WA, USA
³Molecular Engineering and Sciences Institute, University of Washington, Seattle, WA, USA
⁴Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA USA
⁵Department of Chemistry and Biochemistry, The Ohio State University, Columbus, OH, USA
⁶Resource for Native Mass Spectrometry Guided Structural Biology, The Ohio State University, Columbus, OH, USA
⁷Department of Biochemistry, University of Zurich, Zurich, Switzerland
⁸Howard Hughes Medical Institute, Pasadena, CA, USA
⁹Department of Chemistry and Biochemistry, Brigham Young University, Provo, UT, USA
¹⁰Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA

Contents:

Methods
- Methods S1: Computational design
- Methods S2: Cryo-EM of coassembled DARPin and GFP
- Methods S3: Cryo-EM of coassembled DARPin and HSA
- Methods S4: Native Mass spectrometry
- Methods S5: Protein expression and purification
- Methods S6: Negative-stain electron microscopy
- Methods S7: Small angle x-ray scattering analysis

Figures
- Figure S1 - Size exclusion chromatography
- Figure S2 - Marginal designs
- Figure S3 - ankyrin/DARPin homo-dimer orientation selection criteria
- Figure S4 - D2-1.4H + GFP-DARPin alignment based grafting
- Figure S5 - Native mass spectrometry, round-1 designs
- Figure S6 - Native mass spectrometry, round-2 designs
- Figures S7, S8 - Negative stain EM and 2D class averages
- Figure S9 - Cryo-EM map and embedded model for scaffold-target complex of D2-21.8.GFP.v2
- Figure S10 - Cryo-EM micrographs and 2D class averages for scaffold-target complexes of D2-1.4H-GFP.v1 and D2-21.8.HSA-C9.v2
- Figure S11 - Processing scheme for D2-1.4H-GFP.v1 with GFP
- Figure S12 - Processing scheme for D2-21.8.HSA-C9.v2 with HSA
- Figure S13 - Fourier shell correlation for D2-21.8.HSA-C9.v2 with HSA

Tables
- Table S1 - Original design sequences
- Table S2 - Building block table for successful designs
- Table S3 - GFP and HSA-binding variant sequences
- Table S4 - Building blocks with crystal structures
- Table S5 - Native-MS expected and determined masses
- Table S6 - Cryo-EM data collection and refinement statistics

Text Files
- Text File S1 - RosettaScript design protocol
- Text File S2 - D2 symmetry definition
- Text File S3 - D3 symmetry definition

Supplement References
Methods S1: Computational design

A custom software library was built in C# with .NET Core (https://www.github.com/imvuw/trifuse, MIT License), which includes functionality for PDB parsing, alignment, symmetry/patterning, clash and contact checking, structure editing, and running the multi-domain fusion algorithm. The several parameters that control fusion were assigned based only on manual curation of outputs during testing in-silico and are likely not optimal for all scenarios. In particular, an 8-residue minimum overlap-length was selected because the idealized ankyrins used in this study have short helices compared to those of DHRs and helical bundles, but longer overlaps might be desirable with other starting components. Likewise, a lower angular-error tolerance might increase the success rate of tested designs, but it was kept at a moderately high 5 degrees, because the lack of crystal structures for so many designs introduced uncertainty about the initial model accuracy, so a tight angle tolerance would have been somewhat arbitrary.

As the method creates a much larger solution space than direct fusion, optimizations were necessary to keep runtimes reasonable while still exhaustively enumerating geometries. The most impactful optimization eliminates redundant alignments by greedily expanding the alignment windows of any valid 8-residue alignment until the R.M.S.D. threshold is exceeded or either secondary structure element ends - all shorter alignment windows contained within the expanded alignment need not be examined. The result is that fewer alignment combinations are considered than if every 8-residue window were examined and nearly identical outputs are largely avoided. The protocol could be run on all de-novo building-block combinations for a target geometry in less than 24 hours on a quad-core laptop.

The fusion output models were redesigned by Rosetta with a simple RosettaScripts protocol (Text File S1), involving only two Movers (operators that modify a design model): SetupForSymmetry and SymPackRotamersMover. These Movers respectively recreate the full symmetric assembly from the input single-chain asymmetric unit (output by the .Net library) and
redesign those residue side chains that were identified by output files in the Resfile format. After
the initial sidechain redesign pass, models deemed promising by a combination of total score
and manual inspection were subjected to one or more additional redesign passes with the same
protocol, but with user-generated Resfiles, to eliminate exposed hydrophobic residues, revert
residues to their original wildtype identity, or mutate Rosetta-designed glycines to alanines
within helices to improve helical propensity. The beta_nov16 score function was used
throughout. The introduction of non-native cysteines, prolines, or methionines was disallowed.

The input structure set consisted of 20 homo-dimer and 42 DHR spacer proteins already
verified within the lab, with 5 homo-dimers and 15 DHRs having been previously published with
solved crystal structures available in the Protein Data Bank 1,2,3, (Table S4). Two designed
crystal structures were unintentionally omitted from the input set (2L4HC2_4 and 3L6HC2_2
from Boyken et al). Two-helix dimers were removed from the scaffold set in the second round of
design, because better results were obtained from three-helix dimers. The third helix leads to a
larger hydrophobic core than exists in the two-helix dimers, which we expect leads to a higher
degree of order even in the monomeric form and might help to avoid aggregation and
misassembly. The other type of successfully incorporated dimer was based on ankyrins.
Although very similar, the minor binding orientation differences between the three ankyrin homo-
dimers was sufficient to make all three useful in finding distinct geometric solutions.

Methods S2: Cryo-EM of coassembled DARPin and GFP
Electron microscopy grids were prepared at 4°C at 100% humidity using vitrobot (FEI). In brief,
3 µl of purified sample at 1.0 mg/ml was applied to glow-discharged Quantifoil 200 mesh
R1.2/1.3 grid, and was manually blotted with a filter paper (Whatman No. 4) for approximately 3
seconds before plunging into liquid ethane. The grids were screened on a Talos Arctica 200 kV
with K3 direct electron detector for ice thickness and sample distribution. Micrographs of the
screened grid were collected on a Titan Krios microscope (FEI) operating at 300 kV with energy
filter (Gatan) and equipped with K2 Summit direct electron detector (Gatan), using data collection program SerialEM 4. A nominal magnification of 165,000x was used for data collection, corresponding to a pixel size of 0.834Å at specimen level, with the defocus ranging from -1.0 μm to -3.0 μm. Movies were recorded in super-resolution mode, with a total dose of 60 e/Å² and dose rate of 8.4 electron per pixel per second and fractionated into 40 frames. Movies were decompressed and gain-normalized using the program Clip in IMOD. Raw movies were corrected for beam-induced motion and binned by two using MotionCor2 5, and exposure-filtered in accordance with relevant radiation damage curves 6. The CTF estimation was performed with GCTF 7 on non-dose weighted micrographs. Micrographs with high CTF Figure of Merit scores and promising maximum resolution (better than 3.9 Å) were selected for further processing (total 1532 micrographs). Several rounds of autopicking using combinations of different references and manual picking were analyzed to determine optimal settings, and yielded similar results. These particles were subjected to iterative rounds of 2D classification, subset selection of high-quality classes, and re-extraction, yielding 138,348 particles from 1023 micrographs, all in RELION 3.0 8. The initial model was de novo generated and subsequent 3D heterogenous refinement was performed using cryoSPARC 9,10. Particles from the best quality 3D class were selected for further processing. The UCSF PyEM package 11 was used to convert the cryoSPARC coordinates into RELION. The resulting particles were analyzed by 3D refinement, Bayesian Particle Polishing and CTF Refinement in RELION with C1 or D2 symmetry and the raw map was sharpened using Phenix Autosharpen Map 12. All the reconstructions were analyzed using UCSF Chimera 13. The coordinate model was built by breaking the initial design model into domains and rigidly docking these individual protein structures into the EM map using Chimera. Once the orientation was identified, the model was then fit and adjusted manually in Chimera and Coot 14. The final Fourier shell correlation and local resolution was calculated with cryoSPARC and the local resolution was again computed in Resmap 15, though only the former was used in graphics. The core resolution was calculated
using the validation function in cryoSPARC. The cryo-EM maps have been deposited at the Electron Microscopy Data Bank with accession code EMD-23199.

**Methods S3: Cryo-EM of coassembled DARPin and HSA**

Co-complex of 21.8.HSA-C9.v2 with recombinant human albumin (Albumedix™ Veltis®) was purified by size-exclusion chromatography. 3 μL of 1 mg/ml of co-complex was loaded onto a freshly glow-discharged (30 s at 20 mA) 1.2/1.3 UltrAuFoil grid (300 mesh) prior to plunge freezing using a vitrobot Mark IV (ThermoFisher Scientific) using a blot force of 0 and 6 second blot time at 100% humidity and 25°C. Data were acquired using an FEI Titan Krios transmission electron microscope operated at 300 kV and equipped with a Gatan K2 Summit direct detector and Gatan Quantum GIF energy filter, operated in zero-loss mode with a slit width of 20 eV. Automated data collection was carried out using Leginon at a nominal magnification of 130,000x with a pixel size of 0.525 Å and stage tilt angles up to 30°. The dose rate was adjusted to 8 counts/pixel/s, and each movie was acquired in super-resolution mode fractionated in 50 frames of 200 ms. 1,011 and 1,866 micrographs were collected with 0° and 30° stage tilt, respectively. A defocus range comprised between -1.0 and -3.5 μm. Movie frame alignment, estimation of the microscope contrast-transfer function parameters, particle picking, and extraction were carried out using Warp. Particle images were extracted with a box size of 800 binned to 400 yielding a pixel size of 1.05 Å. Two rounds of reference-free 2D classification were performed using CryoSPARC to select well-defined particle images. The selected particles were subsequently subjected to ab initio 3D reconstructions and 3D refinement using CryoSPARC. Two rounds of 3D classification using RELION 3.0 were carried out with 50 iterations each (angular sampling 7.5° for 25 iterations and 1.8° with local search for 25 iterations) using ab initio generated models. Particle images were subjected to the Bayesian polishing procedure implemented in RELION 3.0 before performing another round of non-uniform refinement in cryoSPARC followed by per-particle defocus refinement and again non-
uniform refinement. To further improve the density of the human serum albumin (HSA), the particles were symmetry-expanded and subjected to focus 3D classification without refining angles and shifts using a soft mask encompassing the DARPin binding regions and HSA domain using a tau value of 60. Particles belonging to classes with the best resolved HSA density were selected and subjected to local refinement using cryoSPARC. Local resolution estimation, filtering, and sharpening were carried out using CryoSPARC. Reported resolutions are based on the gold-standard Fourier shell correlation (FSC) of 0.143 criterion and Fourier shell correlation curves were corrected for the effects of soft masking by high-resolution noise substitution. Detailed processing workflows are shown in Supplementary Figure 12. The cryo-EM maps have been deposited at the Electron Microscopy Data Bank with accession codes listed in table S6.

Methods S4: Native mass spectrometry
Sample purity and oligomeric state were analyzed by online buffer exchange MS \(^{18}\) using a Vanquish UHPLC coupled to a Q Exactive Ultra-High Mass Range (UHMR) mass spectrometer (Thermo Fisher Scientific) \(^{19,20}\) modified to allow for surface-induced dissociation (SID) similar to that previously described \(^{21}\). With the exception of D3-19.14 (50 μM), 1 μL of 25 μM protein in 25 mM Tris and 150 mM NaCl were injected and online buffer exchanged into 200 mM ammonium acetate, pH 6.8 by a self-packed buffer exchange column (P6 polyacrylamide gel, Bio-Rad Laboratories) at a flow rate of 100 μL per min. A heated electrospray ionization (HESI) source with a spray voltage of 4 kV was used for ionization. Mass spectra were recorded for 1000 – 20000 m/z at 3125 resolution as defined at 400 m/z. The injection time was set to 200 ms. Voltages applied to the transfer optics were optimized to allow for ion transmission while minimizing unintentional ion activation, and a higher-energy collisional dissociation (HCD) of 5 V was applied. Mass spectra were deconvolved using UniDec V4.2.2 \(^{22}\). Deconvolution settings
included mass sampling every 10 Da, smooth charge states distributions, automatic peak width
tool, point smooth width of 1 or 10, and beta of 50 (artifact suppression).

**Methods S5: Protein Expression and purification**

DNA sequences encoding proteins with 6xHis tags were codon-optimized by Genscript and
cloned into pET28b+ or pET29b+ vector under the control of a T7 promoter. Plasmids were
transformed into BL21(DE3) *E. coli* and plated on LB agar plates. On different occasions, either
50 ml or 500 ml expression cultures were used. 50 ml expression cultures were directly
inoculated from plate colonies and grown for 24 hours in Studier’s autoinduction media with
shaking at 200 rpm. Alternatively, 5 ml starter cultures in TB were inoculated and grown for 9-12
hours before transfer to 500 ml autoinduction media for 16-18 hours. All growth media was
prepared with 100 μM kanamycin as a selection antibiotic. Expression cultures were spun down
for 10 minutes at 4,000 rcf, resuspended in 40 ml TBS (150 mM NaCl, 25 mM Tris) with Pierce
protease inhibitor (Product No. A32963), and lysed by sonication. Lysates were centrifuged at
25,000 rcf for 40 minutes to separate the insoluble fraction. The soluble fraction was purified by
affinity chromatography over Ni-NTA Agarose (Qiagen) gravity columns. Eluates were
concentrated and fractionated by SEC on a Superdex 200 Increase 10/300 GL.

**Methods S6: Negative-stain EM:**

PELCO 300 mesh Copper grids with Carbon film (Product 01843-F) were glow-discharged and
coated with 3μL of sample in TBS and then blotted immediately. 3 μL uranyl formate stain
(concentration either 0.75% or 2%) was applied and blotted immediately, twice, and then
allowed to dry. Two workflows were used for imaging. Approximately 50 micrographs per
construct were recorded either 1) on a Thermo Scientific Talos L120C transmission electron
microscope operating at 120kV with 4k x 4k Ceta CMOS camera at 57k magnification, followed
by contrast-transfer function (CTF) estimation, automatic reference-free particle picking,
classification, and ab-initio reconstruction in cisTEM of homogeneous subsets of particles \(^2\) on a FEI Tecnai T12 electron microscope using Leginon image collection software. In the latter case, parameters of the contrast transfer function (CTF) were estimated using CTFFIND4. Then, particles were picked in a reference-free manner using DoG Picker. Reference-free 2D classification was used to select homogeneous subsets of particles using CryoSPARC. The selected particles were subsequently subjected to ab initio 3D reconstructions and Homogenous 3D refinement using CryoSPARC. The known symmetry (D\(_2\) or D\(_3\)) was applied during reconstruction, except for designs D3-19.14 and D3-19.19, for which C\(_1\) symmetry was applied (although the design model is D\(_3\)).

**Methods S7: SAXS analysis:**

SAXS data were collected at the SIBYLS Beamline (Advanced Light Source in Berkeley, CA) via their Mail-In SAXS program. KNO\(_3\) was added to buffer solutions in the range of 2 to 5 mM to minimize radiation-damage induced aggregation. Samples were concentrated in Amicon Ultra 0.5ml centrifugal filters and flow-through was used as the background subtraction buffer. For each sample, the average scattering profile was computed, excluding data in the Guinier region for timepoints after radiation damage became observable. The Scatter 3.0 \(^{25}\) software was used for analysis; model and experimental Rg values were determined from their respective Guinier region data. Combined datasets (model-vs-experiment) were generated with the FOXS web server \(^{26,27}\) for plotting.
Figure S1: Size exclusion chromatography of post-IMAC eluate (not re-chromatography) shows a dominant species for constructs from the (A) first and (B) second design rounds.
**Figure S2:** Two designs showed promising results according to SAXS and native-MS, but appeared disordered by EM. It may be that the assemblies are sensitive to the low pH of uranyl formate stain or are simply more unstable. Consistent with lesser stability for design D3-1.5A2, incomplete assembly was observed when using offline buffer exchange for this design, indicating complex dissociation and/or unfolding as a result of extended protein storage in sub-optimal buffer (AmAC) that occurred between offline buffer exchange and native-MS measurement. In contrast, no complex dissociation was observed when using online buffer-exchange MS (which was used to generate all the data shown) as the time between buffer-exchange to AmAc and native-MS measurement is drastically reduced with this method. As online buffer exchange was a newer and improved protocol, it was not used with D2-21.22 or other round-2 designs.
Figure S3. Orientation criteria for the ankyrin/DARPin homo-dimer that were applied in the second design round as illustrated. Target-binding by DARPin usually occurs at the concave surface between loops and helices and it was thought that the flipped ankyrin/DARPin homo-dimer orientation in the round-2 designs (right) would generally orient binding-target copies away from one another.
Figure S4: Alignment-based construction of a hybrid DARPin scaffold “D2-1.4H.GFP.v1”. An alignment is performed during DARPin grafting to ensure that residues responsible for homo-oligomerization in the base construct (D2-1.4H) are preserved after hybridization with the DARPin. Binding residues (D2-1.4H homo-oligomerization and the DARPin-GFP interface) in the source constructs are darkened and the hybrid construct is colored according to whichever source construct the sequence was based on.
Figure S5: Raw spectra showing relative abundance vs m/z of the first-round designs. The charge state distributions are labeled with purple circles and the oligomeric states are noted as tetrameric (small 4, to the right of the charge state envelope) or hexameric (6). The peak labeled with an asterisk is an instrumental artifact.
Figure S6: Raw spectra showing relative abundance vs m/z of the second-round designs. The charge state distributions are labeled with purple circles and the oligomeric states are noted. Designed oligomers D2-21.22 and D3-19.24 appear to show low levels of self-association to form octamer and dodecamer, respectively. The peak labeled with an asterisk is an instrumental artifact.
Figure S7: Representative negative stain micrographs and 2D class averages for various designs. EMDB depositions are available for D2-1.1B (EMD-23534), D2-1.1D (EMD-23536), D2-1.4H (EMD-23535), and D3-1.5C (EMD-23533).
Figure S8: Representative negative stain micrographs and 2D class averages for various designs. EMDB depositions are available for D3-19.14 (EMD-23532) and D3-19.19 (EMD-23531).
Figure S9. Cryo-EM for GFP-related scaffold D2-21.8.GFP.v2 with docked models, resolved at approximately 6 Ångströms overall.
Figure S10. Representative micrographs and 2D class averages of scaffold and target complexes A) D2-4H.GFP.v1 with GFP and B) D2-21.8.HSA-C9.v2 with HSA.
1677 SerialEM movies collection
↓
IMOD Clip decompression and gain-normalization
↓
MotionCor2 movement correction
↓
GCTF: Maximum CTF fit resolution (filtering)
↓
1532 micrographs
↓
Relion3 Autopicking (several rounds of manual and 2D class template picking)

138,348 particles
↓
CryoSPARC ab initio 3d model generation (C1 symmetry)
CryoSPARC heterogeneous refinement (D2 symmetry, three classes)

Selected highest quality class (52% of total)
↓
UCSF pyEM coordinate conversion (CryoSPARC → RELION)

72140 particles
↓
Relion 3d refinement
↓
Relion CTF Refinement
Relion Bayesian Particle Polishing

3D map
↓
Phenix Autosharpen Map

Sharpened 3d map

**Figure S11.** Cryo-EM processing scheme for D2-4H.GFP.v1 with GFP complex.
Figure S12. Cryo-EM processing scheme for 21.8.HSA-C9.v2 with HSA complex.
Figure S13. Fourier shell correlation (FSC) curves calculated by cryoSPARC for the full complex (black) and local refinement (gray) of HSA with D2-21.8.HSA-C9.v2.
### Table S1: Design sequences

#### Round-1 Design Sequences

<table>
<thead>
<tr>
<th>Round-1 Design Sequences</th>
<th>D2-1.1B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MASDYLRLATEHNLATEAASLAELAASAVTLTVTDPSKTQEHTELASRLLEMEMSPLKAQELTREAIRKEGR NEESEKIKRSKSYKALKAKIALAIKEIKGVETAVRAAQEVAVLRAEAGNNVNLRAEVALAIKVAEIQGNVE EVAVKAQAVVSAALNAGDEVLKKVFAEQASRISKEAKQGNQEVKSASVSLSLIAAAASGDKLVKDLLESGADV NASSSDGKTPHLHAAENGHAKVLLLLEQGADPNAKDSGKTPHLAAENGHEEVVILLLAMGADPNTSDGRTPLDLARHGNEEVVKVEDHGWLEHHEHHHH</td>
</tr>
<tr>
<td></td>
<td>D2-1.1D</td>
</tr>
<tr>
<td></td>
<td>MASEKARIVENLEAALRNKAAEMKSAIKITRDNSSDEKATRYSLITKVLMSLEILITASLELAEKALREEG SDDSAEKVRKEAEIILSAKEAVARVMQEMQVMTIKRTGSSDLREVAELARVAKAAEKGNEVAVQARVAVEA AKQAGNDVLKVAEQALRIAKAEKQGNVEVAVKAVAEAAKQAGEDVLKKVAEASRISAEASKQGNKEVA SKALIVAQQAGSEKAVKAKIESGADVNASDSDGRTPLHLHAANENGHAEVALLIEKADVNAKDSGRTPLHLAAEN GHDEVVILLLIKGDVNASDSGRTPLHLHAANENGHADVNTSDLGRTPLDLAREHGNEEVKALE KQGWLEHHEHHHH</td>
</tr>
<tr>
<td></td>
<td>D2-1.4H</td>
</tr>
<tr>
<td></td>
<td>MGSEKARIVENLEAALRNAAAEMQSAIKIMMDDNSDEKLRLYRTLTVLRMSVELLRSLELAEKALREEG SDDSAEKVRKEAEIILSAKEAVARVMQEMQVMTIKRTGSSDLREVAELARVAKAAEKGNEVAVQARVAVEA AKQAGNDVLKVAEQALRIAKAEKQGNVEVAVKAVAEAAKQAGEDVLKKVAEASRISAEASKQGNKEVA SKALIVAQQAGSEKAVKAKIESGADVNASDSDGRTPLHLHAANENGHAEVALLIEKADVNAKDSGRTPLHLAAEN GHDEVVILLLIKGDVNASDSGRTPLHLHAANENGHADVNTSDLGRTPLDLAREHGNEEVKALE KQGWLEHHEHHHH</td>
</tr>
<tr>
<td></td>
<td>D3-1.5C</td>
</tr>
<tr>
<td></td>
<td>MGSEKARIVENLEAALRNAAAEMQSAIKIVADNASDEKALRYLRTKVLMSVELLRSLELAEKALREEG SDDSAEKVRKEAEIILKESETIEQITQEVADFAELANKATDEELRKEISKCARLELASRSTNDILKQIL EVAKLAFELASKATDEELIILKCCQAFFERASRTNDILKEIILEVAKAFETASKATDEEISKILLSILCAAL GNKDAVKSIAENGADVNASDSDGRTPLHLHAANENGHAEVALLIEKADVNAKDSGRTPLHLHAANENGHDEVVILL LLKGDADVNAKDSGRTPLHLHAANENGHADVNTSDLGRTPLDLAREHGNEEVKALEKQGWLEHHEHHHH</td>
</tr>
<tr>
<td></td>
<td>D3-1.5A2</td>
</tr>
<tr>
<td></td>
<td>MGSEKARIVENLEAALRNAAAEMQSAIKIVADNASDEKALRYLRTKVLMSVELLRSLELAEKALREEG SDDSAEKVRKEAEIILKESETIEQITQEVADFAELANKATDEELRKEISKCARLELASRSTNDILKQIL EVAKLAFELASKATDEELIILKCCQAFFERASRTNDILKEIILEVAKAFETASKATDEEISKILLSILCAAL GNKDAVKSIAENGADVNASDSDGRTPLHLHAANENGHAEVALLIEKADVNAKDSGRTPLHLHAANENGHDEVVILL LLKGDADVNAKDSGRTPLHLHAANENGHADVNTSDLGRTPLDLAREHGNEEVKALEKQGWLEHHEHHHH</td>
</tr>
</tbody>
</table>

#### Round-2 Design Sequences (D2 symmetry)

<table>
<thead>
<tr>
<th>Round-2 Design Sequences</th>
<th>D2-21.6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Round-2 Design Sequences (D3 symmetry)

<table>
<thead>
<tr>
<th>Design Sequence</th>
<th>Inserted Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>D2-21.8</td>
<td>MHHHHHGGSGSALEKIAKLAAARLSAELARARASAEMARLAIEAVSKERGSSELLKIVADLIVEAQEAVRRLTIESQQAIAKLIALDRIARAKAEASDEKMEEVAKEVQERAAEARARIEEEKLVLIILKIILRSIGDEEELKATKLAEEAIRVAREVDSLLERTIAEALKGDSRAAVKLAEEALAREAEERGDEEKVKAALIAAAAGDKDAVKDLIENGADVNGRDSGTPLHHAAENGHEEVIVALIAKADVNAKDSGDGTPLHHAAENGHEEVEVLLLLKGADVNADSGDTPHLHAAENGHEEVEVLLLAGADVNTSDSGDTRPDLAREHGEEEVKVEDHG</td>
</tr>
<tr>
<td>D2-21.22</td>
<td>MHHHHHGGSGSALEKIAVLNRLNRAAEMQSAIAKIAANNFSKKAEAYTRLTTKVLMSVLLRASLELAEKALREDSGDSAEKVRKAEAEILKEERLELKAARVAAAREAANATDEERKRIEIEELKAAREANRSTNEEIIKILEEARLRFLIAAEKGEAEVKLAEADGVDNGKDSGDKTPLHHAAENGHAKEVVLILLEQGADPNADKDSGTPHLHAAENGHEEVEVVLVILLLMHGADPNAKDSGDGTPLHHAAENGHEEVEVLLILLMGADPNTSDSGDTRPDLAREHGEEEVKVEDHG</td>
</tr>
<tr>
<td>D2-21.26</td>
<td>MHHHHHGGSGSALEKIAVLNRLNRAAEMQSAIAKIAANNFSKKAEAYTRLTTKVLMSVLLRASLELAEKALREDGDSAEKVRKAEAEILKEERLELKAARVAAAREAANATDEERKRIEIEELKAAREANRSTNEEIIKILEEARLRFLIAAEKGEAEVKLAEADGVDNGKDSGDKTPLHHAAENGHAKEVVLILLEQGADPNADKDSGTPHLHAAENGHEEVEVVLVILLLMHGADPNAKDSGDGTPLHHAAENGHEEVEVLLILLMGADPNTSDSGDTRPDLAREHGEEEVKVEDHG</td>
</tr>
<tr>
<td>D2-21.29</td>
<td>MHHHHHGGSGSALEKIAVLNRLNRAAEMQSAIAKIAANNFSKKAEAYTRLTTKVLMSVLLRASLELAEKALREDGDSAEKVRKAEAEILKEERLELKAARVAAAREAANATDEERKRIEIEELKAAREANRSTNEEIIKILEEARLRFLIAAEKGEAEVKLAEADGVDNGKDSGDKTPLHHAAENGHAKEVVLILLEQGADPNADKDSGTPHLHAAENGHEEVEVVLVILLLMHGADPNAKDSGDGTPLHHAAENGHEEVEVLLILLMGADPNTSDSGDTRPDLAREHGEEEVKVEDHG</td>
</tr>
<tr>
<td>D2-21.30</td>
<td>MHHHHHGGSGSALEKIAVLNRLNRAAEMQSAIAKIAANNFSKKAEAYTRLTTKVLMSVLLRASLELAEKALREDGDSAEKVRKAEAEILKEERLELKAARVAAAREAANATDEERKRIEIEELKAAREANRSTNEEIIKILEEARLRFLIAAEKGEAEVKLAEADGVDNGKDSGDKTPLHHAAENGHAKEVVLILLEQGADPNADKDSGTPHLHAAENGHEEVEVVLVILLLMHGADPNAKDSGDGTPLHHAAENGHEEVEVLLILLMGADPNTSDSGDTRPDLAREHGEEEVKVEDHG</td>
</tr>
</tbody>
</table>

### Round-3 Design Sequences (D3 symmetry)

<table>
<thead>
<tr>
<th>Design Sequence</th>
<th>Inserted Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>D3-19.14</td>
<td>MHHHHHGGSGSALEKIAKLAAARLSAELARARASAEMARLAIEAVSKERGSSELLKIVADLIVEAQEAVRRLTIESQQAIAKLIALDRIARAKAEASDEKMEEVAKEVQERAAEARARIEEEKLVLIILKIILRSIGDEEELKATKLAEEAIRVAREVDSLLERTIAEALKGDSRAAVKLAEEALAREAEERGDEEKVKAALIAAAAGDKDAVKDLIENGADVNGRDSGTPLHHAAENGHEEVIVALIAKADVNAKDSGDGTPLHHAAENGHEEVEVLLLLKGADVNADSGDTPHLHAAENGHEEVEVLLLAGADVNTSDSGDTRPDLAREHGEEEVKVEDHG</td>
</tr>
<tr>
<td>D3-19.19</td>
<td>MHHHHHGGSGSALEKIAKLAAARLSAELARARASAEMARLAIEAVSKERGSSELLKIVADLIVEAQEAVRRLTIESQQAIAKLIALDRIARAKAEASDEKMEEVAKEVQERAAEARARIEEEKLVLIILKIILRSIGDEEELKATKLAEEAIRVAREVDSLLERTIAEALKGDSRAAVKLAEEALAREAEERGDEEKVKAALIAAAAGDKDAVKDLIENGADVNGRDSGTPLHHAAENGHEEVIVALIAKADVNAKDSGDGTPLHHAAENGHEEVEVLLLLKGADVNADSGDTPHLHAAENGHEEVEVLLLAGADVNTSDSGDTRPDLAREHGEEEVKVEDHG</td>
</tr>
<tr>
<td>Sequence</td>
<td>SEC, nMS, and SAXS Data Agreement</td>
</tr>
<tr>
<td>----------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>DVNAKDSDGRTPLHAAENGHDEVVLILLKADVNAKDSDGRTPLHAAENGHVVLILLAGADVNTSDSGRTTPLDHARENGNEKVKALQE</td>
<td>List of sequences whose SEC, nMS, and SAXS data were in agreement with the corresponding design model. Underlined linker and (His)$_6$-tag denote a region added after computational design and not modeled during SAXS analysis.</td>
</tr>
</tbody>
</table>

D3-19.20
MHHHHHHGSSEKARIAVENLEAALRNRAAAEMQKSAIKIALDNSDEKAIRYALTTKVLKMSVEILLSLELAEKALREEGSDLASEKVRKEAEELKESTILILAADALTALLLLQKVRKVEKEIKSNKDEEAVETAARLAIELRVAKREELELAKLGFLKKAELAIKIAARKLALEAVALEAGYDVNAKDSDGRTPLHAAENGALEVVLALLNGADVNAKDSGRTPLHAAENGNKRVVLILAGADVNTSDDGTPLDLARENGNEEVKALERR

D3-19.24
MHHHHHHGSSEKARIAVENLEAALRNRAAAEMQFLAIKIMLLNSSDEKAARFLRLETTKVLKMSVEILLSLELAEKALREEGSDLASEKVRKEAEELKESTILILAADALTALLLLQKVRKVEKEIKSNKDEEAVETAARLAIELRVAKREELELAKLGFLKKAELAIKIAARKLALEAVALEAGYDVNAKDSDGKTPLHAAENGHEDVVLILLLMGADPNTSDDGRTPLDLAREHGNEDVVKALKAAG
Table S2: Building block verification level for successful designs

<table>
<thead>
<tr>
<th>Design ID</th>
<th>N-terminal oligomer</th>
<th>Spacer</th>
<th>C-terminal oligomer</th>
</tr>
</thead>
<tbody>
<tr>
<td>D2-1.1B</td>
<td>rop4 (*)</td>
<td>DHR62 (*)</td>
<td>ank3C21</td>
</tr>
<tr>
<td>D2-1.1D</td>
<td>rop20 (*)</td>
<td>DHR62 (*)</td>
<td>ank1C2G3</td>
</tr>
<tr>
<td>D2-1.4H</td>
<td>rop20 (*)</td>
<td>DHR68 (*)</td>
<td>ank3C22 (*)</td>
</tr>
<tr>
<td>D3-1.5C</td>
<td>rop20 (*)</td>
<td>DHR15 (*)</td>
<td>ank1C2G3</td>
</tr>
<tr>
<td>D3-1.5A2 (marginal)</td>
<td>rop20 (*)</td>
<td>DHR15 (*)</td>
<td>ank1C2G3</td>
</tr>
<tr>
<td>D2-21.8</td>
<td>rop20 (*)</td>
<td>DHR15 (*)</td>
<td>ank3C21</td>
</tr>
<tr>
<td>D2-21.22 (marginal)</td>
<td>rop20 (*)</td>
<td>DHR57 (*)</td>
<td>ank1C2G3</td>
</tr>
<tr>
<td>D2-21.29 and D2-21.30</td>
<td>rop20 (*)</td>
<td>DHR82 (*)</td>
<td>ank3C21</td>
</tr>
<tr>
<td>D2-21.26</td>
<td>rop20 (*)</td>
<td>DHR71</td>
<td>ank1C2G3</td>
</tr>
<tr>
<td>D3-19.19</td>
<td>rop20 (*)</td>
<td>DHR82 (*)</td>
<td>ank1C2G3</td>
</tr>
<tr>
<td>D3-19.14</td>
<td>rop20 (*)</td>
<td>DHR76</td>
<td>ank1C2G3</td>
</tr>
<tr>
<td>D3-19.20</td>
<td>rop20 (*)</td>
<td>DHR82 (*)</td>
<td>ank1C2G3</td>
</tr>
<tr>
<td>D3-19.24</td>
<td>rop20 (*)</td>
<td>DHR82 (*)</td>
<td>ank3C22 (*)</td>
</tr>
</tbody>
</table>

Constituent building blocks for successful and marginal designs are shown. Those building blocks that have been SAXS-verified, but not crystallized, are starred (*); most successful designs are composed with two of three components having only SAXS verification. Two successful designs were created without any building block crystal verification whatsoever.
Table S3: DARPin-binding variant sequences and cryo-EM result

<table>
<thead>
<tr>
<th>DARPin</th>
<th>Binding Variant</th>
<th>Resolution</th>
<th>EM Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>D2-1.1D.GFP.v1</td>
<td>Aggregated SEKARIAVENLEAALRLNAAEMQSAIKIMRDNSSDEKAFLRYLLLTTLKMLKSMVLLRASLELEA KRALREESGDSSEAKVRKAEAEILEEKSEILKRAELETILKAAVRAAAEAAARNAITEERKRIEEELKKAEERAN RSTNEEQKKILEKALGFRIL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D2-1.4H.GFP.v1</td>
<td>4.8 Å overall resolution SEKARIAVENLEAALRLNAAEMQSAIKIMRDNSSDEKAFLRYLLLTTLKMLKSMVLLRASLELEA KRALREESGDSSEAKVRKAEAEILEEKSEILKRAELETILKAAVRAAAEAAARNAITEERKRIEEELKKAEERAN RSTNEEQKKILEKALGFRIL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D2-1.5C.GFP.v1</td>
<td>Disordered SEKARIAVENLEAALRLNAAEMQSAIKIMRDNSSDEKAFLRYLLLTTLKMLKSMVLLRASLELEA KRALREESGDSSEAKVRKAEAEILEEKSEILKRAELETILKAAVRAAAEAAARNAITEERKRIEEELKKAEERAN RSTNEEQKKILEKALGFRIL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D2-21.8.GFP.v1</td>
<td>Slight aggregation MHHHHHGGSEKARIAVENLEAALRLNAAEMQSAIKIMRDNSSDEKAFLRYLLLTTLKMLKSMVLLRASLELEA KRALREESGDSSEAKVRKAEAEILEEKSEILKRAELETILKAAVRAAAEAAARNAITEERKRIEEELKKAEERAN RSTNEEQKKILEKALGFRIL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D2-21.8.GFP.v2</td>
<td>6-7 Å overall resolution with target and minor preferred orientation MHHHHHGGSEKARIAVENLEAALRLNAAEMQSAIKIMRDNSSDEKAFLRYLLLTTLKMLKSMVLLRASLELEA KRALREESGDSSEAKVRKAEAEILEEKSEILKRAELETILKAAVRAAAEAAARNAITEERKRIEEELKKAEERAN RSTNEEQKKILEKALGFRIL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D2-21.8.GFP.v3</td>
<td>Preferred orientation MHHHHHGGSEKARIAVENLEAALRLNAAEMQSAIKIMRDNSSDEKAFLRYLLLTTLKMLKSMVLLRASLELEA KRALREESGDSSEAKVRKAEAEILEEKSEILKRAELETILKAAVRAAAEAAARNAITEERKRIEEELKKAEERAN RSTNEEQKKILEKALGFRIL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D2-21.29.GFP.v1</td>
<td>Severe aggregation MHHHHHGGSEKARIAVENLEAALRLNAAEMQSAIKIMRDNSSDEKAFLRYLLLTTLKMLKSMVLLRASLELEA KRALREESGDSSEAKVRKAEAEILEEKSEILKRAELETILKAAVRAAAEAAARNAITEERKRIEEELKKAEERAN RSTNEEQKKILEKALGFRIL</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

GFP- and HSA-binding variants tested are listed, along with the corresponding resolution achieved or the observed failure mode, where applicable. Design IDs are comprised of the underlying scaffold ID plus the suffix v[#], to separate variants of the same scaffold, which were produced by shifting the grafted residues up or down by one or more ankyrin repeats.
Table S4: Listing of X-ray verified building blocks

<table>
<thead>
<tr>
<th>DHR spacers:</th>
</tr>
</thead>
<tbody>
<tr>
<td>5CWB (DHR4), 5CWD (DHR7), 5CWF (DHR8), 5CWG (DHR10), 5CWH (DHR14), 5CWI (DHR18), 5CWJ (DHR49), 5CWK (DHR53), 5CWL (DHR54), 5CWM (DHR64), 5CWN (DHR71), 5CWO (DHR76), 5CWP (DHR79) and 5CWQ (DHR81).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C2 homo-dimers:</th>
</tr>
</thead>
<tbody>
<tr>
<td>5KBA (Ank1C2), 5HRY (Ank3C2_1), 5J73 (2L4HC2_9), 5J0K (2L4HC2_23), 5J10 (2L4HC2_24)</td>
</tr>
</tbody>
</table>

The building blocks used in this study that have been solved by X-ray crystallography are listed. Although solved by crystallography, DHR5 was not included in the set because homooligomerization was detected in the original study.
<table>
<thead>
<tr>
<th>Design ID</th>
<th>Oligomeric state</th>
<th>Oligomer mass (expected, kDa)</th>
<th>Oligomer mass (native-MS, kDa)</th>
<th>Error (%)</th>
<th>Intensity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D2-1.1B</td>
<td>4</td>
<td>154.8</td>
<td>155.0</td>
<td>0.13</td>
<td>100</td>
</tr>
<tr>
<td>D2-1.1D</td>
<td>4</td>
<td>167.2</td>
<td>167.3</td>
<td>0.09</td>
<td>100</td>
</tr>
<tr>
<td>D2-1.4H</td>
<td>4</td>
<td>215.8</td>
<td>216.1</td>
<td>0.14</td>
<td>100</td>
</tr>
<tr>
<td>D3-1.5A2</td>
<td>6</td>
<td>296.4</td>
<td>296.6</td>
<td>0.07</td>
<td>100</td>
</tr>
<tr>
<td>D3-1.5C</td>
<td>6</td>
<td>250.1</td>
<td>250.3</td>
<td>0.06</td>
<td>100</td>
</tr>
<tr>
<td>D2-21.8</td>
<td>4</td>
<td>140.2</td>
<td>140.6</td>
<td>0.30</td>
<td>100</td>
</tr>
<tr>
<td>D2-21.22</td>
<td>4</td>
<td>156.9</td>
<td>157.4</td>
<td>0.34</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>8 (artificial 4-mer dimerization)</td>
<td>313.8</td>
<td>315.0</td>
<td>0.39</td>
<td>10</td>
</tr>
<tr>
<td>D2-21.26</td>
<td>4</td>
<td>160.0</td>
<td>160.7</td>
<td>0.37</td>
<td>100</td>
</tr>
<tr>
<td>D2-21.29</td>
<td>4</td>
<td>147.2</td>
<td>147.7</td>
<td>0.35</td>
<td>100</td>
</tr>
<tr>
<td>D3-19.14</td>
<td>6</td>
<td>209.1</td>
<td>209.9</td>
<td>0.40</td>
<td>100</td>
</tr>
<tr>
<td>D3-19.19</td>
<td>6</td>
<td>208.4</td>
<td>209.2</td>
<td>0.39</td>
<td>100</td>
</tr>
<tr>
<td>D3-19.20</td>
<td>6</td>
<td>188.8</td>
<td>189.3</td>
<td>0.53</td>
<td>100</td>
</tr>
<tr>
<td>D3-19.24</td>
<td>6</td>
<td>182.5</td>
<td>183.4</td>
<td>0.49</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>12 (artificial 6-mer dimerization)</td>
<td>365.0</td>
<td>367.3</td>
<td>0.64</td>
<td>10</td>
</tr>
</tbody>
</table>

Expected oligomer masses versus those determined by native-MS. Differences between expected and measured values are within the limits of method accuracy and can be explained by a combination of adducts, oligomer size, signal quality, mass resolution and data processing settings. Artificial dimerization between oligomers can occur dependent on concentration and droplet size during the electrospray process. This was notably observed for designs D2-21.22 and D3-19.24.
Table S6: Cryo-EM refinement statistics

<table>
<thead>
<tr>
<th>Data Collection</th>
<th>DARPin scaffold (D2-1.4H.GFP.v1) &amp; GFP</th>
<th>DARPin scaffold (D2-21.8.HSA-C9.v2) &amp; HSA</th>
<th>DARPin scaffold (D2-21.8.HSA-C9.v2) &amp; HSA (local refinement)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMDB ID</td>
<td>EMD-23199</td>
<td>EMD-23537</td>
<td>EMD-23538</td>
</tr>
<tr>
<td>Particle count</td>
<td>138,348</td>
<td>487,905</td>
<td>164,745</td>
</tr>
<tr>
<td>Magnification</td>
<td>165,000x</td>
<td>130,000x</td>
<td></td>
</tr>
<tr>
<td>Pixel size (Å)</td>
<td>0.834</td>
<td>0.525</td>
<td></td>
</tr>
<tr>
<td>Defocus range (μm)</td>
<td>-1 to -3</td>
<td>-1 to -3.5</td>
<td></td>
</tr>
<tr>
<td>Voltage (kV)</td>
<td>300</td>
<td>300</td>
<td></td>
</tr>
<tr>
<td>Electron dose (e/Å²)</td>
<td>60</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whole-map resolution (Å) with 0.143 cutoff</td>
<td>4.78</td>
<td>4.53</td>
<td>4.0</td>
</tr>
<tr>
<td>Map sharpening B factor (Å²)</td>
<td>-315</td>
<td>-250</td>
<td>-82</td>
</tr>
</tbody>
</table>

Cryo-EM data collection and refinement statistics for scaffold-target complexes.
Text File S1: Example design script

```xml
<ROSETTASCRIPTS>
  <SCOREFXNS>
    <ScoreFunction name="sfx_hard_symm" weights="beta.wts" symmetric="1" />
  </SCOREFXNS>
  <TASKOPERATIONS>
    <InitializeFromCommandline name="init" />
    <RestrictIdentities name="nomutate_VIRTUAL" identities="XXX" prevent_repacking="1" />
    <DisallowIfNonnative name="disallow_nonnative" disallow_aas="CPM" />
    <ReadResfile name="resfile_designable" filename="%%resfile%%" />
  </TASKOPERATIONS>
  <MOVERS>
    <SetupForSymmetry name="symmetry_setup" definition="%%symdef%%"></SetupForSymmetry>
    <SymPackRotamersMover name="design_rotamers_resfile" scorefxn="sfx_hard_symm"
      task_operations="init,nomutate_VIRTUAL,resfile_designable,disallow_nonnative"></SymPackRotamersMover>
  </MOVERS>
  <PROTOCOLS>
    <Add mover_name="symmetry_setup" />
    <Add mover_name="design_rotamers_resfile" />
  </PROTOCOLS>
</ROSETTASCRIPTS>
```

bash$: <rosetta_scripts_path> -ignore_zero_occupancy false -database <rosetta_database_path> -linmem Ig 10 -lazy Ig true -parser:protocol <rosettascripts_xml_path> -s <pdb_path> -native <pdb_path> -nstruct 1 -parser:script_vars symdef=<D2_or_D3_symmetry_definition>
resfile=<resfile> -ex1 -ex2 -unmute all -out:pdb_gz true -out:path:all ./ -beta -overwrite -scorefile <scorefile_name>.sc

A simple design script and command-line example applies symmetry and designs sidechains with whatever score-function is in beta at the time of use (beta_nov16 during this work). Symmetry definition files are provided in Supplementary Text Files S2 and S3.
A symmetry definition file for $D_2$ symmetry, for redesign with Rosetta and RosettaScripts
1. A symmetry definition file for D₃ symmetry, for redesign with Rosetta and RosettaScripts.
References
