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

Ivan Vulovic<sup>1,2,3</sup>, Qing Yao<sup>4†</sup>, Young-Jun Park<sup>2†</sup>, Alexis Courbet<sup>1,2</sup>, Andrew Norris<sup>5,6</sup>, Florian Busch<sup>5,6</sup>, Aniruddha Sahasrabuddhe<sup>5,6</sup>, Hannes Merten<sup>7</sup>, Danny D. Sahtoe<sup>1,2</sup>, George Ueda<sup>1,2</sup>, Jorge A. Fallas<sup>1,2</sup>, Sara J. Weaver<sup>4</sup>, Yang Hsia<sup>1,2</sup>, Robert A. Langan<sup>1,2</sup>, Andreas Plückthun<sup>7</sup>, Vicki H. Wysocki<sup>5,6</sup>, David Veesler<sup>2</sup>, Grant J. Jensen<sup>4,8,9</sup>, David Baker<sup>1,2,3,10</sup>

† These authors contributed equally to this work.

<sup>1</sup>Institute for Protein Design, University of Washington, Seattle, WA, USA.
<sup>2</sup>Department of Biochemistry, University of Washington, Seattle, WA, USA
<sup>3</sup>Molecular Engineering and Sciences Institute, University of Washington, Seattle, WA, USA
<sup>4</sup>Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA USA
<sup>5</sup>Department of Chemistry and Biochemistry, The Ohio State University, Columbus, OH, USA
<sup>6</sup>Resource for Native Mass Spectrometry Guided Structural Biology, The Ohio State University, Columbus, OH, USA
<sup>7</sup>Department of Biochemistry, University of Zurich, Zurich, Switzerland

<sup>8</sup>Howard Hughes Medical Institute, Pasadena, CA, USA

<sup>9</sup>Department of Chemistry and Biochemistry, Brigham Young University, Provo, UT, USA

<sup>10</sup>Howard 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 multidomain 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 <sup>1,2,3</sup>, (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<sup>4</sup>. 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<sup>-</sup>/Å<sup>2</sup> 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<sup>5</sup>, and exposurefiltered in accordance with relevant radiation damage curves <sup>6</sup>. The CTF estimation was performed with GCTF<sup>7</sup> 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<sup>8</sup>. The initial model was de novo generated and subsequent 3D heterogenous refinement was performed using cryoSPARC <sup>9,10</sup>. Particles from the best quality 3D class were selected for further processing. The UCSF PyEM package <sup>11</sup> 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 <sup>12</sup>. All the reconstructions were analyzed using UCSF Chimera<sup>13</sup>. 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<sup>14</sup>. The final Fourier shell correlation and local resolution was calculated with cryoSPARC and the local resolution was again computed in Resmap <sup>15</sup>, 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<sup>™</sup> Veltis<sup>®</sup>) 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 <sup>16</sup> at a nominal magnification of 130,000x with a pixel size of 0.525Å 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<sup>17</sup>. 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 nonuniform refinement in cryoSPARC followed by per-particle defocus refinement and again nonuniform 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 <sup>18</sup> using a Vanquish UHPLC coupled to a Q Exactive Ultra-High Mass Range (UHMR) mass spectrometer (Thermo Fisher Scientific) <sup>19,20</sup> modified to allow for surface-induced dissociation (SID) similar to that previously described <sup>21</sup>. With the exception of D3-19.14 (50  $\mu$ M), 1  $\mu$ L of 25  $\mu$ M protein in 25 mM Tris and 150 mM NaCI 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  $\mu$ L 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 <sup>22</sup>. 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 <sup>23</sup> 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 <sup>24</sup> or 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<sub>2</sub> or D<sub>3</sub>) was applied during reconstruction, except for designs D3-19.14 and D3-19.19, for which C<sub>1</sub> symmetry was applied (although the design model is D<sub>3</sub>).

## 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<sub>3</sub> 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 <sup>25</sup> 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 <sup>26,27</sup> for plotting.

## Figures:



**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.

Round-1 Design Example

Round-2 Design Example



**Figure S3**. Orientation criteria for the ankyrin/DARPin homo-dimer that were applied in the second design round as illustrated. Target-binding by DARPins 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 homooligomerization 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 S	erialEM movies collection
¥	
IMOD	Clip decompression and gain-normalization
¥	
Motior	nCor2 movement correction
¥	
GCTF:	Maximum CTF fit resolution (filtering)
¥	
1532 n	nicrographs
Ļ	Relion3 Autopicking (several rounds of manual and 2D class template picking)
138,34	l8 particles
Ļ	CryoSPARC ab initio 3d model generation (C1 symmetry) CryoSPARC heterogeneous refinement (D2 symmetry, three classes)
Selecte	ed highest quality class (52 % of total)
Ļ	UCSF pyEM coordinate conversion (CryoSPARC $ ightarrow$ RELION)
72140	particles
	Relion 3d refinement
¥	Relion CTF Refinement Relion Bayesian Particle Polishing
3D ma	p
Ļ	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.

## <u>Tables</u>

## Table S1: Design sequences

Round-1 Design Sequences

#### D2-1.1B

MASDYLRLATEHNKLATEAASLAAELAASAVTLTVTDPSKTAQEHTELASRLLEMMSQFLKAAQELTREAIRKEGR NEESEKTLRKSKSSYKALKALLKAIKAIEKGDVETAVRAAQEAVRLASEAGNNNVLRAVAEVALAIAKVAEEQGNV EVAVKAAQVAVSAALNAGDEDVLKKVAEQASRISKEAEKQGNQEVSKKALSVSLIAAAASGDKDLVKDLLESGADV NASSSDGKTPLHVAAENGHAKVVLLLLEQGADPNAKDSDGKTPLHLAAENGHAVVVALLLMHGADPNAKDSDGKTP LHLAAENGHEEVVILLLAMGADPNTSDSDGRTPLDLAREHGNEEVVKVLEDHGGWLEHHHHHH

#### D2-1.1D

MASEKARIAVENLEAALRLNKAAIEMAKSAIKITRDNSSDEKATRYSLLTAKVLVMSLELLTASLELAEKALREEG SDDSAEKVRKEAEEILSKAVEEAVRVMQEMVTIMKRTGSNDSLREVAELALRVAKAAEKAGNVEVAVQAARVAVEA AKQAGDNDVLRKVAEQALRIAKEAEKQGNVEVAVKAAKVAVEAAKQAGDEDVLKKVAEQASRIASEASKQGNKEVA SKALIVAAQAGSKEAVKKAIESGADVNASDSDGRTPLHHAAENGHAEVVALLIEKGADVNAKDSDGRTPLHHAAEN GHDEVVLILLLKGADVNAKDSDGRTPLHHAAENGHKRVVLVLILAGADVNTSDSDGRTPLDLAREHGNEEVVKALE KQGGWLEHHHHHH

#### D2-1.4H

MGSEKARIAVENLEAALRLNRAAAEMQKSAIKIMDDNSDDEKALRYLRLTTKVLRMSVELLRASLELAEKALREEG SDDSAEKVRKEAEEILKESTAILKLADAATKVADIKHDIKKAKEQQEQGNKEEAEKTLREATEKIKRVTEELEKIA KNSKTPEIALKAAEALVKLIKLLIEIAKLLQEQGNKEEAEKVLREATELIKRVTELLEKIAKNSDTPELALRAAEL LVRLIKLLIEIAKLLQEQGNKEEAEKVLREATELIKRVTELLEKIAKNSDTPELASRAAELLVRLIKLLQEIAKLL KEQGNKEEAEKVEREAKELLSRVLILAAEIGNKDIVKTALENGADVNASDSDGKTPLHLAAENGHKDVVELLLRQG ADPNAKDSDGKTPLHLAAENGHKVVVMLLLSQGADPNAKDSDGKTPLHLAAENGHEDVVLLLLLMGADPNTSDSDG RTPLDLAREHGNEDVVEALKAAGGWLEHHHHHH

#### D3-1.5C

MGSEKARIAVENLEAALRLNRAAAEMQKSAAKIVADNASDEKALRYLRLTTKVLRMSVELLRASLELAEKALREEG SDDSAEKVRKEAEEILKESTEILKEADQITEVADLAFELANKATDEELRKEISKCARLALELASRSTNDELIKQIL EVAKLAFELASKATDEELIKLILKCCQAAFERASRSTNDEEIKKILEVAKRAFETASKATDEEEIKSILLICAAAL GNKDAVKSAIENGADVNASDSDGRTPLHHAAENGNAEVVALLIEKGADVNAKDSDGRTPLHHAAENGHDEVVLILL LKGADVNAKDSDGRTPLHHAAENGHKRVVLVLILAGADVNTSDSDGRTPLDLAREHGNEEVVKALEKQ<u>GGWLEHHH</u> HHH

#### D3-1.5A2

MGSEKARIAVENLEAALRLNRAAAEMQKSAAKIVADNASDEKALRYLRLTTKVLRMSVELLRASLELAEKALREEG SDDSAEKVRKEAEEILKESTEILKLADQITEVADLAFELANKATDEELRKEIVKCAKLALELASRSTNDELKKQIL EVAKLAFELASKATDEELIKEILKCCQLAFELASRSTNDELIKQILEVAKLAFELASKATDEELIKEILKCCQLAF ELASRSTNDELIKLILEVAKAAFERASKATDEEEIKEILKVCQEAFEEASRSTNDEEIASILLVAAALLGNKDAVK DAIENGADVNASDSDGRTPLHHAAENGNAEVVALLIEKGADVNAKDSDGRTPLHHAAENGHDEVVLILLLKGADVN AKDSDGRTPLHHAAENGHKRVVLVLILAGADVNTSDSDGRTPLDLAREHGNEEVVKALEKQGGWLEHHHHHH

Round-2 Design Sequences (D2 symmetry)

D2-21.6

MHHHHHHGSGSALEKIAKLIIEAARLSAELARRAARASAEMARLAIEAVSKERGSSELLKIVADLIVEAQEAVVRL IIESQQIAAKLAEDLIRAAKEAASDESKMEEVAKEVQERAERAARDIEEKLLKVLIELIKKLARSIGDEERLKATK LAEEAIRVAREVGDSLLERIALEAALKGDSRAAKAVLKAAELAREAAERGDEEKVKAAALIAAAAAGDKDAVKDLI ENGADVNGRDSDGRTPLHHAAENGNAEVVALLIAKGADVNAKDSDGRTPLHHAAENGHEEVVLILLLKGADVNAKD SDGRTPLHHAAENGHKRVVLVLILAGADVNTSDSDGRTPLDLAREHGNEEVVKALEKQ

#### D2-21.8

MHHHHHHGSGSEKARIAVENLEAALRLNRAAAEMQKSAIKIMRDNSSDEKAFRYLLLTTKVLKMSVELLRASLELA EKALREEGSDDSAEKVRKEAEEILKESTEILKRAELETLKAAVRVAAEAAARNATDEEERKRIEEELKKAEERANR STNEEEIKKILEEALARFLIILAEKGAKEAVKLALEAGADVNGKDSDGKTPLHLAAENGHAKVVLLLLEQGADPNA KDSDGKTPLHLAAENGHAVVVALLLMHGADPNAKDSDGKTPLHLAAENGHEEVVILLLAMGADPNTSDSDGRTPLD LAREHGNEEVVKVLEDHG

#### D2-21.22

MHHHHHHGSGSEKARIAVENLEAALRLNRAAAEMQLSAAKIAADNFSDKKAAEYTRLTTKVLEMSVELLRASLELA EKALREEGSDDSAEKVRKEAEEILKESKEILEAAEALTRIAHLARKAAESTDPEEALKIAKEAIEIALKTVKENPS ELALQAVLAAVILASAVAKRVTDPDKALKIAKLVIELALEAVKEDPSTDALRAVLEAVRLAEEVARRVTDPIKALK IAALVIQLAAEAAKEDPSEEAQRALKLAAELAAEALERGADVNYHDEDGRTPLHHAAEAGADEAVLILLLKGADVN AKDSDGRTPLHHAAENGNKRVVLVLILAGADVNTSDSDGRTPLDHAREHGNEEVVKALEKQ

#### D2-21.26

MHHHHHHGSGSEKARIAVENLEAALRLNRAAAEMQASAAKIVADNTSDEKAYRYLELTTKVLLMSVELLRASLELA EKALREEGSDDSAEKVRKEAEEILKESTLALEAAEELTKAAKAALRAREASERGDEEEFRKAAEEALEAAKRVVER AKKAGIPELVAAAAAVALAIAELAAKNGDKEVFKKAAESALEVAKRLVEVASKEGDPELVLEAAKVALRVAELARK NGDKEVLKKAALSAAEVALRLAEVAKKEGDPDLVREAAEILADALEKGLDVNIHDEDGRTPLHHAAELGADEAVLI LLLAGADVNAKDSDGRTPLHHAAENGHKRVVLVLILAGADVNTSDSDGRTPLDHAREHGNEEVVKALEKQ

#### D2-21.29

MHHHHHHGSGSEKARIAVENLEAALRLNRAAAEMQKSAIKIADDNRSDEKALRYALLTTKVLEMSVELLRASLELA EKALREEGSDDSAEKVRKEAEEILEKSSRILAEAFVITARLATELARLLQEKAKKTGDAKELREAKRALKEAAEYV EKALKINKDDDEARELLERIEEELKKVEKLLEEILIKAAARGDKDLVKLALKAGADVNASDSDGKTPLHKAAENGH AKVVLLLLEQGADPNAKDSDGKTPLHLAAENGHAVVVALLLMHGADPNAKDSDGKTPLHLAAENGHEEVVILLLAM GADPNTSDSDGRTPLDLAREHGNEEVVKVLEDHG

#### D2-21.30

MHHHHHHGSGSEKARIAVENLEAALRLNRAAAEMQKSAIKIAEDNSSDEKAIRYTLLTTRVLEMSFELLRASLELA EKALREEGSDDSAEKVRKEAEEILEESLRILAEAFVRTARFLKELAERLQERAKKTGDPELLAEAYEALREAVEFV KKAEKINPDNERAKKTLEELKEELRKVEELLKELLIRAAERGDKDTVRRALEAGADVNAKDSDGKTPLHLAAENGH AKVVLLLLEQGADPNAKDSDGKTPLHLAAENGHAVVVALLLMHGADPNAKDSDGKTPLHLAAENGHEEVVILLLAM GADPNTSDSDGRTPLDLAREHGNEEVVKVLEDHG

Round-2 Design Sequences (D3 symmetry)

#### D3-19.14

MHHHHHHGSSEKARIAVENLEAALRLNRAAAEMQKSAIKIARDNRSDKKALLYLLLATYVLEMSLELLRASLELAE KALREEGSDDSAEKVRKEAEEILKESKEIFLRAALETAKAAAEYVEEAAREAERRGNPELRDAAKALRKYLEEANE EAAKQGNAEKILRVALAALLIAAAALGDKDLVKDLIEMGADVNGHDLDGRTPLHLAAAAGHAEVVALLIEKGADVN AKDSDGRTPLHHAAENGHDEVVLILLLKGADVNAKDSDGRTPLHHAAENGHKRVVLVLILAGADVNTSDSDGRTPL DLARENGNEEVVKVLEKA

#### D3-19.19

MHHHHHHGSSEKARIAVENLEAALRLNRAAAEMQKSAIKIVDDNSSDVRAIEYLALTSAVLAESLLLLASLELAE KALREEGSDDSAEKVRKEAEEILEESARIAAEAAEESLRAAEEAIELARKTGDSDALRAAAEALKAARAAVRAAIA ANPDDDKAEEIAKRLEEALNRVLHEAAERGDQDAVKLVIEAGGDVNARDSDGRTPLHHAAENGHAEVVALLIRKGA  $\label{eq:constraint} DVNAKDSDGRTPLHHAAENGHDEVVLILLLKGADVNAKDSDGRTPLHHAAENGHKRVVLVLILAGADVNTSDSDGRTPLDHARENGNEKVVKALQEQ$ 

### D3-19.20

MHHHHHHGSSEKARIAVENLEAALRLNRAAAEMQKSAIKIALDNSSDEKAIRYARLTTKVLKMSVELLRASLELAE KALREEGSDDSAEKVRKEAEEILKESTLILEAADLATALLDLLQKVRKVEKEIKSNKDNEEAVETAARLAIELARV AKRLEELAKKLGDGFLKKLAEKAIKIAARALEVALEAGYDVNAKDSDGRTVLHHAAENGALEVVLLALLNGADVNA KDSDGRTPLHHAAENGNKRVVLVLILAGADVNTSDSDGRTPLDLARENGNEEVVKALERR

## D3-19.24

MHHHHHHGSSEKARIAVENLEAALRLNRAAAEMQFLAIKIMLLNSSDEKAARFLRLTTKVLKMSVELLRASLELAE KALREEGSDDSAEKVRKEAEEILKESTEILEATEEATKLLELLEEARKVEEAIKSNPDNDEAVETAKRIAEEARKV ALKLFEYASKLGIPLLAKAAAEALAVALKAGADPNAKDSDGKTPLHHAAEAGVKLAVMLLLSHGADPNAKDSDGKT PLHLAAENGHEDVVLLLLLMGADPNTSDSDGRTPLDLAREHGNEDVVKALKAAG

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.

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

Table S2: Building block verification level for successful designs

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

D2-1.1D.GFP.v1: Aggregated

SEKARIAVENLEAALRLNKAAIEMAKSAIKITRDNSSDEKATRYSLLTAKVLVMSLELLTASLELAEKALREEGSD DSAEKVRKEAEEILSKAVEEAVRVMQEMVTIMKRTGSNDSLREVAELALRVAKAAEKAGNVEVAVQAARVAVEAAK QAGDNDVLRKVAEQALRIAKEAEKQGNVEVAVKAAKVAVEAAKQAGDEDVLKKVAEQASRIASEASKQGNKEVASK ALIVAAQAGSKEAVKKAIESGADVNASDSDGRTPLHHAAENGHAEVVALLIEKGADVNAKDSNGHTPLLHAARNGH DEVVLILLLKGADVNAKDDVGVTPLHLAAQRGHKRVVLVLILAGADVNTADLWGQTPLHLAATAGHLEVVKALLKQ GADVNARDNIGHTPLHLAAWAGHLEIVEVLLKYGADVNAQDKFGKTPFDLAIDNGNEDIAEVLQKA

D2-1.4H.GFP.v1: 4.8 Å overall resolution SEKARIAVENLEAALRLNRAAAEMQKSAIKIMDDNSDDEKALRYLRLTTKVLRMSVELLRASLELAEKA LREEGSDDSAEKVRKEAEEILKESTAILKLADAATKVADIKHDIKKAKEQQEQGNKEEAEKTLREATEK IKRVTEELEKIAKNSKTPEIALKAAEALVKLIKLLIEIAKLLQEQGNKEEAEKVLREATELIKRVTELL EKIAKNSDTPELALRAAELLVRLIKLLIEIAKLLQEQGNKEEAEKVLREATELIKRVTELLEKIAKNSD TPELASRAAELLVRLIKLLQEIAKLLKEQGNKEEAEKVEREAKELLSRVLILAARIGNKDIVKTALENG ADVNASDDVGVTPLHLAAQRGHKDVVELLLRQGADPNAKDLWGQTPLHLAATAGHKVVVMLLLSQGADP NAKDNIGHTPLHLAAWAGHEDVVLLLLLMGADPNTSDKFGKTPFDLAIDNGNEDVVEALKAAGG

D2-1.5C.GFP.v1: Disordered SEKARIAVENLEAALRLNRAAAEMQKSAAKIVADNASDEKALRYLRLTTKVLRMSVELLRASLELAEKALREEGSD DSAEKVRKEAEEILKESTEILKEADQITEVADLAFELANKATDEELRKEISKCARLALELASRSTNDELIKQILEV AKLAFELASKATDEELIKLILKCCQAAFERASRSTNDEEIKKILEVAKRAFETASKATDEEEIKSILLICAAALGN KDAVKSAIENGADVNASDSDGRTPLHHAAENGNAEVVALLIEKGADVNAKDSDGHTPLLHAARNGHDEVVLILLLK GADVNAKDDVGVTPLHLAAQRGHKRVVLVLILAGADVNTADLWGQTPLHLAATAGHEEVVKALIKQGADVNARDNI GHTPLHLAAWAGHLEIVEVLLKYGADVNAQDKFGKTPFDLAIDNGNEDIAEVLQKA

D2-21.8.GFP.v1: Slight aggregation MHHHHHGSGSEKARIAVENLEAALRLNRAAAEMQKSAIKIMRDNSSDEKAFRYLLLTTKVLKMSVELLRASLELA EKALREEGSDDSAEKVRKEAEEILKESTEILKRAELETLKAAVRVAAEAAARNATDEEERKRIEEELKKAEERANR STNEEEQKKILEEALGRFLIILARKGAKEAVKLALEAGADVNAADDVGVTPLHLAAQRGHAKVVLLLLEYGADPNA ADLWGQTPLHLAATAGHAVVVALLLMHGADPNARDNIGHTPLHLAAWAGHEEVVILLLAMGADPNAQDKFGKTPLD LARDNGNEEVVKVLEDHAA

D2-21.8.GFP.v2: 6-7 Å overall resolution with target and minor preferred orientation

MHHHHHHGSGSEKARIAVENLEAALRLNRAAAEMQKSAIKIMRDNSSDEKAFRYLLLTTKVLKMSVELL RASLELAEKALREEGSDDSAEKVRKEAEEILKESTEILKRAELETLKAAVRVAAEAAARNATDEEERKR IEEELKKAEERANRSTNEEEQKKILEEALGRFLIILARKGAKEAVKLALEAGADVNAADDVGVTPLHLA AQRGHAKIVLLLLEYGADPNAADLWGQTPLHLAATAGHAVIVALLLMHGADPNARDNIGHTPLHLAAWA GHEEIVILLLAMGADPNAQDKFGKTPLDLARDNGNEEVVKVLEDHAA

D2-21.8.GFP.v3: Preferred orientation MHHHHHHGSGSEKARIAVENLEAALRLNRAAAEMQKSAIKIMRDNSSDEKAFRYLLLTTKVLKMSVELLRASLELA EKALREEGSDDSAEKVRKEAEEILKESTEILKRAELETLKAAVRVAAEAAARNATDEEERKRIEEELKKAEERANR STNEEEQKKILEEALGRFLIILARKGAKEAVKLALEAGADVNAADDVGVTPLHLAAQRGHAKVVLLLLEQGADPNA ADLWGQTPLHLAATAGHAVVVALLLMHGADPNARDNIGHTPLHLAAWAGHEEVVILLLAMGADPNAQDKFGKTPLD LARDNGNEEVVKVLEDHAA

D2-21.29.GFP.v1: Severe aggregation MHHHHHHGSGSEKARIAVENLEAALRLNRAAAEMQKSAIKIADDNRSDEKALRYALLTTKVLEMSVELLRASLELA EKALREEGSDDSAEKVRKEAEEILEKSSRILAEAFVITARLATELARLLQEKAKKTGDAKELREAKRALKEAAEYV EKALKINKDDDEARELLERIEEELKKVEDLLGKILLEAARAGDKDLVKLALKAGADVNAADDVGVTPLHLAAQRGH AKVVLLLLEYGADPNAADLWGQTPLHLAATAGHAVVVALLLMHGADPNARDNIGHTPLHLAAWAGHEEVVILLLAM GADPNAQDKFGKTPLDLARDNGNEEVVKVLEDHAA

D2-21.29.GFP.v2: Preferred orientation and aggregated MHHHHHGSGSEKARIAVENLEAALRLNRAAAEMQKSAIKIADDNRSDEKALRYALLTTKVLEMSVELLRASLELA EKALREEGSDDSAEKVRKEAEEILEKSSRILAEAFVITARLATELARLLQEKAKKTGDAKELREAKRALKEAAEYV EKALKINKDDDEARELLERIEEELKKVEKLLGEILLEAARAGDKDLVKLALKAGADVNAADDVGVTPLHLAAQRGH AKIVLLLLEYGADPNAADLWGQTPLHLAATAGHAVIVALLLMHGADPNARDNIGHTPLHLAAWAGHEEIVILLLAM GADPNAQDKFGKTPLDLARDNGNEEIVKVLEDHAA

D2-19.20.GFP.v1: Aggregated MHHHHHGSGSEKARIAVENLEAALRLNRAAAEMQKSAIKIALDNSSDEKAIRYARLTTKVLKMSVELLRASLELA EKALREEGSDDSAEKVRKEAEEILKESTLILEAADLATALLDLLQKVRKVEKEIKSNKDNEEAVETAARLAIELAR VAKRLEELAKKLGDGFLKKLAEKAIKIAARALEVALEAGYDVNAKDSDGATVLHHAARNGALEVVLLALLNGADVN AADDVGVTPLHLAAQRGNKRVVLVLILAGADVNAADLWGQTPLHLAATAGHLEVVKALLKRGADVNARDNIGHTPL HLAAWAGHLEIVEVLLKYGADVNAQDKFGKTPFDLAIDNGNEDIAEVLQKA

D2-21.8.HSA-C9.v2: 4.5 Å overall resolution with target, 4.0 Å HSA and DARPin with local refinement MHHHHHSEKARIAVENLEAALRLNRAAAEMQKSAIKIMRDNSSDEKAFRYLLLTTKVLKMSVELLRASLELAEKA LREEGSDDSAEKVRKEAEEILKESTEILKRAELETLKAAVRVAAEAAARNATDEEERKRIEEELKKAEERANRSTN EEEIKKILEEALARFLLEAAWKGAKEAVKLALEAGADVNAADYFGHTPLHLAARNGHAKVVLLLLEQGADPNADDF AGSTPLHLAARAGHAVVVALLLMHGADPNAVDSNGFTPLHLAAQKGHEEVVILLLAMGADPNAQDKFGKTPFDLAI DNGNEEVVKVLEDHG

Original DARPin C9 (anti-HSA) - FLAG tag underlined MRGSHHHHHHGSDLGKKLLEAAWWGQDDEVRILMANGADVNAADYFGHTPLHLAARNGHLEIVEVLLKTGADVNAD DFAGSTPLHLAARAGHLEIVEVLLKAGADVNAVDSNGFTPLHLAAQKGHLEIVEVLLKHGADVNAQDKFGKTPFDL AIDNGNEDIAEVLQKAAKLN<u>DYKDDDDK</u>

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

# DHR spacers:

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).

## C2 homo-dimers:

5KBA (Ank1C2), 5HRY (Ank3C2\_1), 5J73 (2L4HC2\_9), 5J0K (2L4HC2\_23), 5J10 (2L4HC2\_24)

The building blocks used in this study that have been solved by X-ray crystallography are listed. Although solved by crystallography, DHR5 was not not included in the set because homooligomerization was detected in the original study.

# Table S5: Native-MS verification

Design ID	Oligomeric state	Oligomer mass (expected, kDa)	Oligomer mass (native-MS, kDa)	Error (%)	Intensity (%)
D2-1.1B	4	154.8	155.0	0.13	100
D2-1.1D	4	167.2	167.3	0.09	100
D2-1.4H	4	215.8	216.1	0.14	100
D3-1.5A2	6	296.4	296.6	0.07	100
D3-1.5C	6	250.1	250.3	0.06	100
D2-21.8	4	140.2	140.6	0.30	100
D2-21.22	4	156.9	157.4	0.34	100
	8 (artificial 4-mer dimerization)	313.8	315.0	0.39	10
D2-21.26	4	160.0	160.7	0.37	100
D2-21.29	4	147.2	147.7	0.35	100
D3-19.14	6	209.1	209.9	0.40	100
D3-19.19	6	208.4	209.2	0.39	100
D3-19.20	6	188.8	189.3	0.53	100
D3-19.24	6	182.5	183.4	0.49	100
	12 (artificial 6-mer dimerization)	365.0	367.3	0.64	10

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

Data Collection	DARPin scaffold (D2-1.4H.GFP.v1) & GFP	DARPin scaffold (D2-21.8.HSA-C9.v2) & HSA	DARPin scaffold (D2-21.8.HSA-C9.v2) & HSA (local refinement)
EMDB ID	EMD-23199	EMD-23537	EMD-23538
Particle count	138,348	487,905	164,745
Magnification	165,000x	130,000x	
Pixel size (Å)	0.834	0.525	
Defocus range (µm)	-1 to -3	-1 to -3.5	
Voltage (kV)	300	300	
Electron dose (e <sup>-</sup> /Å <sup>2</sup> )	60	76	
Refinement			
Whole-map resolution (Å) with 0.143 cutoff	4.78	4.53	4.0
Map sharpening B factor (Å <sup>2</sup> )	-315	-250	-82

Cryo-EM data collection and refinement statistics for scaffold-target complexes.

## Text File S1: Example design script

```
<ROSETTASCRIPTS>
        <SCOREFXNS>
                <ScoreFunction name="sfx hard symm" weights="beta.wts"</pre>
symmetric="1" >
                </ScoreFunction>
        </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 non
native"></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.

# Text File S2: Example symmetry definition (D<sub>2</sub>)

```
symmetry name d2
subunits 4
number of interfaces 3
E = 4*VRT0001 + 2*(VRT0001:VRT0002) + 2*(VRT0001:VRT0003) +
2*(VRT0001:VRT0004)
anchor residue COM
virtual_transforms start
start -1,0,0 0,1,0 0,0,0
rot Rz angle 180.0
rot Rx angle 180.0
rot Rz angle 180.0
virtual transforms stop
connect virtual JUMP1 VRT0001 VRT0002
connect virtual JUMP2 VRT0002 VRT0003
connect virtual JUMP3 VRT0003 VRT0004
set dof BASEJUMP x(50) angle x(0:360) angle y(0:360) angle z(0:360)
set dof JUMP2 z(50) angle z(0:90.0)
```

A symmetry definition file for D<sub>2</sub> symmetry, for redesign with Rosetta and RosettaScripts

## Text File S3: Example symmetry definition (D<sub>3</sub>)

```
symmetry name d3
subunits 6
number of interfaces 4
E = 6*VRT0001 + 6*(VRT0001:VRT0002) + 3*(VRT0001:VRT0004) +
3*(VRT0001:VRT0005) + 3*(VRT0001:VRT0006)
anchor residue COM
virtual transforms start
start -1,0,0 0,1,0 0,0,0
rot Rz angle 120.0
rot Rz angle 120.0
rot Rx angle 180.0
rot Rz angle 120.0
rot Rz angle 120.0
virtual transforms stop
connect virtual JUMP1 VRT0001 VRT0002
connect virtual JUMP2 VRT0002 VRT0003
connect virtual JUMP3 VRT0003 VRT0004
connect virtual JUMP4 VRT0004 VRT0005
connect virtual JUMP5 VRT0005 VRT0006
set dof BASEJUMP x(50) angle x(0:360) angle y(0:360) angle z(0:360)
set dof JUMP3 z(50) angle z(0:60.0)
```

1. A symmetry definition file for D<sub>3</sub> symmetry, for redesign with Rosetta and RosettaScripts.

# References

- 1. JA Fallas, G Ueda, W Sheffler, et al. Computational design of self-assembling cyclic protein homooligomers. *Nat Chem.* 9, 353-360. (2017)
- 2. TJ Brunette, F Parmeggiani, PS Huang, et al. Exploring the repeat protein universe through computational protein design. *Nature*. 528, 580-584. (2015)
- 3. SE Boyken, Z Chen, B Groves, et al. De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity *Science*. 352, 680-687. (2016)
- 4. DN Mastronarde. Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol.* 152, 36-51.
- 5. S Zheng, E Palovcak, JP Armache. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* 14, 331–332 (2017)
- 6. T Grant, N Grigorieff. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *Elife*. 4, e06980. (2015)
- 7. K Zhang. Gctf: Real-time CTF determination and correction. J Struct Biol. 193, 1-12. (2016)
- 8. J Zivanov, T Nakane, BO Forsberg, et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife*. 7, e42166. (2018)
- 9. A Punjani, JL Rubinstein, JL Fleet, et al. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods*. 14, 290-296. (2017)
- 10. A Punjani, H Zhang, DJ Fleet. Non-uniform refinement: adaptive regularization improves singleparticle cryo-EM reconstruction. *Nat Methods*. 17, 1214-1221 (2020)
- 11. D Asarnow, E Palovcak, Y Cheng. UCSF pyem v0.5. Zenodo (2019). Available at https://www.doi.org/10.5281/zenodo.3576630
- 12. TC Terwilliger, O V Sobolev, P V Afonine, et al. Automated map sharpening by maximization of detail and connectivity. *Acta Crystallogr D Struct Biol* 74, 545–559 (2018)
- 13. EF Pettersen, TD Goddard, CC Huang, et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem.* 25, 1605-1612. (2004)
- 14. P Emsley, K Cowtan. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 60, 2126-2132. (2004)
- 15. A Kucukelbir, FJ Sigworth, HD Tagare. Quantifying the local resolution of cryo-EM density maps. *Nat Methods.* 11, 63-65. (2014)
- 16. C Suloway, J Pulokas, D Fellmann, et al. Automated molecular microscopy: the new Leginon system. *J Struct Biol.* 151, 41-60. (2005)
- 17. D Tegunov, P Cramer. Real-time cryo-electron microscopy data preprocessing with Warp. *Nat Methods*. 16, 1146-1152. (2019)
- 18. ZL VanAernum, F Busch, BJ Jones, et al. Rapid online buffer exchange for screening of proteins, protein complexes and cell lysates by native mass spectrometry. *Nat Protoc.* 15, 1132-1157. (2020)
- 19. M van de Waterbeemd, KL Fort, D Boll, et al. High-fidelity mass analysis unveils heterogeneity in intact ribosomal particles. *Nat Methods*. 14, 283-286. (2017)
- KL Fort, M van de Waterbeemd, D Boll, et al. Expanding the structural analysis capabilities on an Orbitrap-based mass spectrometer for large macromolecular complexes. *Analyst.* 143, 100-105. (2017)
- ZL VanAernum, JD Gilbert, ME Belov, et al. Surface-Induced Dissociation of Noncovalent Protein Complexes in an Extended Mass Range Orbitrap Mass Spectrometer. *Anal Chem.* 91, 3611-3618. (2019)
- 22. MT Marty, AJ Baldwin, EG Marklund, et al. Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. *Anal Chem.* 87, 4370-4376. (2015)
- 23. FW Studier. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif.* 41, 207-234. (2005)
- 24. T Grant, A Rohou, N Grigorieff. cisTEM, user-friendly software for single-particle image processing. *Elife*. 7, e35383. (2018)
- 25. RP Rambo, JA Tainer, Accurate assessment of mass, models and resolution by small-angle scattering. *Nature*. 496, 477-481 (2013)

- 26. D Schneidman-Duhovny, M Hammel, JA Tainer, Sali A. Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophys J*. 105, 962-974. (2013)
- 27. D Schneidman-Duhovny, M Hammel, JA Tainer, A Sali. FoXS, FoXSDock and MultiFoXS: Singlestate and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Res.* 44, W424-W429. (2016)