De novo design of modular protein oligomers to investigate cell signaling

Natasha Ilyana Edman

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2022

Reading Committee:

David Baker, Chair

Phil Bradley

Ning Zheng

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

Abstract

De novo design of modular protein oligomers to investigate cell signaling

Natasha Ilyana Edman

Chair of the Supervisory Committee:

David Baker

Department of Biochemistry

Clustering of receptors is a critical step in activation of signaling, but this phenomenon is difficult to investigate with native soluble ligands that bind only 1-2 copies of a receptor. I designed a series of modular protein scaffolds with 4-, 6-, and 8-fold symmetry, offering up to 8 sites for display of a receptor-binding domain. Scaffolds were extended by adding repeat units to vary the spacing of the bound receptors. To target receptors, a de novo miniprotein binder to the fibroblast growth factor receptor (FGFR) was attached to scaffolds through genetic fusion. Treatment of cells with the designed scaffolds resulted in colocalization and decreased membrane diffusion of receptors, increased Erk phosphorylation, and intracellular calcium release indicating activation of FGF signaling. These designed protein scaffolds can be applied as a universal tool to a variety of systems to dissect the role of clustering in cell surface receptor-mediated signaling pathways.

This thesis contains work from the following publications:

Hsia Y, Mout R, Sheffler W, Edman NI, Vulovic I, Park Y-J, et al. Design of multi-scale protein complexes by hierarchical building block fusion. *Nat Commun* 2021;12:2294.

From Sahtoe DD, Praetorius F, Courbet A, Hsia Y, Wicky BIM, Edman NI, et al. Reconfigurable asymmetric protein assemblies through implicit negative design. *Science* 2022;375(6578): eabj7662. Reprinted with permission from AAAS.

A version of portions of this thesis are currently in preparation as a journal article entitled:

De novo design of modular protein oligomers to investigate cell signaling
Natasha Edman, Rachel Redler, Thomas Schlichthaerle, Ashish Phal, Seong An, Ali Etemadi, Florian Praetorius, Zhe Li, Neville Bethel, Longxing Cao, Brian Coventry, Lauren Carter, Marcos Miranda, Gira Bhabha, Damian Ekiert, Hannele Ruohola-Baker, Joseph Schlessinger, and David Baker

Acknowledgments

First and foremost, I thank my PhD adviser David Baker. David has believed in me from the beginning and has always been my #1 supporter and cheerleader. My PhD would have been a lot less fun without all the David-endorsed lab shenanigans, including the 2019 lab retreat, the 2020 joke abstract competition, covering David's desk in wrapping paper for his birthday, and dressing up as a banana for my Halloween lab meeting. As I continue my career in research, I will strive to emulate David's optimism, intellectual humility, and ability to attract the brightest minds in the world to work together on the most audacious scientific problems.

Within David's lab, I have had several outstanding mentors; chief among them are Thomas Schlichthärle, Robby Divine, Florian Praetorius, Rubul Mout, and Xinru Wang. In addition, I thank my numerous colleagues and coauthors, including Ali Etemadi, Joseph Harman, Mohamad Abedi, Neville Bethel, Andrew Favor, Ryan Kibler, Stacey Gerben, Brian Coventry, Alexis Courbet, Marc Éxposit-Goy, Buwei Huang, Erin Yang, Yang Hsia, Josh Lubner, and Danny Sahtoe. If you threw a dart at the Baker lab timeline board posted outside MolES 420, no matter which face you hit, I would have a story of how that person helped me, learned from me, or simply made me laugh.

The interdisciplinary nature of my project means that I have several outstanding collaborators outside the IPD. This includes the brilliant cryo-EM specialist Rachel Redler of Gira Bhabha and Damian Ekiert's lab, as well as Ashish Phal in Hannele Ruohola-Baker's lab, Seong An in Joseph "Yossi" Schlessinger's lab, and Sanjay Srivatsan, fellow MSTP student. I am honored to count these scientists among my co-authors on the forthcoming manuscript. I thank the staff of the Institute for Protein Design for their work, particularly Lauren Carter, Kandise VanWormer, Michelle DeWitt, Asim Bera, Andrew Borst, Xinting Li, and Marcos Miranda.

I have been immensely fortunate in my thesis committee consisting of Phil Bradley, Conrad Liles, Julie Mathieu, and Ning Zheng, as well as my MSTP advisor Mary-Claire King. Three years is far too short a time to learn from this wise, enthusiastic, and supportive group of scientists. Outside of the program, I have received excellent scientific training from my longtime mentors David Low and Matthew Sweet, who continue to encourage me while still puzzling over my decision to pursue an MD or a PhD (respectively).

Finally, I thank my family and friends who have stood by me throughout my training. A special thanks to my brother Lukas, who answered my most elementary coding questions without judgment; my mother Ursula for troubleshooting the AKTA with me on slow weekends when my labmates weren't around; and Austin, who produced my defense and leveraged his degree in Communication to help me write many important emails.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Receptor tyrosine kinases

Receptor tyrosine kinases (RTKs) play critical roles in development, response to injury, and various pathologies including cancers.[1,2] RTKs are single-pass transmembrane proteins with an extracellular ligand-binding domain and an intracellular domain containing a kinase and phosphorylation sites. Canonically, upon ligand binding, the receptor dimerizes and undergoes a conformational change and the intracellular domains cross-phosphorylate, allowing binding of various effector proteins to transduce a signal to the cell. Cell responses to RTK signaling include proliferation or cell cycle control, differentiation, migration, and metabolic regulation.[2]

RTK activation has layers of complexity beyond the single ligand-receptor interaction. For example, clustering of RTKs can amplify the signaling effect of ligand binding and even activate signaling in absence of the canonical ligand.[3–7] This appears to result from formation of higher-order assemblies containing multiple receptor and effector proteins that exhibit avid binding and slowed dissociation.[7]

Due to RTKs' numerous roles in physiology and disease, there is an investigative and clinical need for methods to generate multivalent ligands that can induce formation of receptor clusters. This requires designing a construct with 2 functions: multimerization and receptor binding. I took a modular approach to design, in which each function is designed separately and then connected by genetic fusion into a larger construct. I first designed novel oligomers and then attached receptor-binding domains.

1.2 Oligomers

Multivalent receptor-binding constructs to date have included protein oligomers, nanoparticles, unbounded arrays such as hydrogels, and DNA nanostructures.[3,4,8] Our group has previously designed *de novo* protein oligomers with up to 8-fold symmetry, as well as unbounded assemblies and nanoparticles, all of which are highly thermostable and can be abundantly expressed in *Escherichia coli*. This has enabled us to display multiple copies of receptor-binding domains to agonize a pathway, in some cases with greater potency than the native ligand.[9–12] These technologies typically rely on either the native ligand for the receptor or an antibody isolated for the receptor of interest.

A key advantage of oligomers, as opposed to nanoparticles and unbounded assemblies, is the precise control over the number of receptors engaged. This allowed to tease out valency-related differences in activation of the Tie2 signaling pathway, using a series of designed protein oligomers.[9,13,14] In this study, only the oligomers with 6-fold or greater valency activated Tie2 signaling. I reasoned that receptor clustering would similarly require higher-valency oligomers. Though many dimers and trimers have been successfully designed, there were few examples of higher valency from our group: two C5 (cyclic symmetry of 5), two C6, and one C8. The asymmetric units of these proteins were all 2-helix bundles, which limits the potential space of receptor topology to be explored; additionally, their small diameter of ~40 angstroms (Å) may sterically hinder clustering of the full number of receptors in such a compact space. I therefore set out to design a new series of oligomeric proteins of C4-C8 valency with varied geometry and greater distance between subunits.

1.3 Receptor-binding domains

A challenge of native ligands is their promiscuity for multiple receptors and differential effects in different cell types and tissues.[15] At the same time, native ligands and antibodies have limited ability to vary geometry of a receptor pair, which is known to have significant effect on downstream signaling. We have shown previously that *de novo* designed ligands to a receptor can elicit novel biological effects not seen with the native ligand. In this study, a series of dimeric ligands was designed to the erythropoietin receptor, holding the receptor pairs in a variety of orientations and inducing biased agonism of the downstream signaling pathway.[16] This demonstrated that *de novo* designed ligands are a promising tool for exploring signal transduction beyond what is possible with native ligands.

In parallel with the creation of the dimeric erythropoietin receptor binder, design methods were developed to generate monomeric, high-affinity miniprotein binders or "minibinders". Designed minibinders to date include targets such as the fibroblast growth factor receptor (FGFR), PD-1, SARS-CoV-2 spike protein, and influenza hemagglutinin, with nanomolar to picomolar affinity.[10,17–19] In addition to their strong affinity, minibinders carry the same thermostability and ease-of-production advantages of *de novo* proteins. We have shown that binders can be attached to multivalent scaffolds to improve affinity/avidity to a target (in the case of SARS-CoV-2) or to agonize a receptor (PD-1).[10,20] Using minibinders avoids problems such as the promiscuity of native ligands for multiple receptors: for example, we recently designed a binder with high affinity and selectivity for the C isoforms of the fibroblast growth factor receptor.[19](Sangwon Lee, personal communication, 2022-02-13)

I hypothesized that display of a monomeric binder on a multivalent scaffold would allow us to engage multiple targets at once and induce clustering of the target receptor on the cell membrane (**Figure 1.1**).

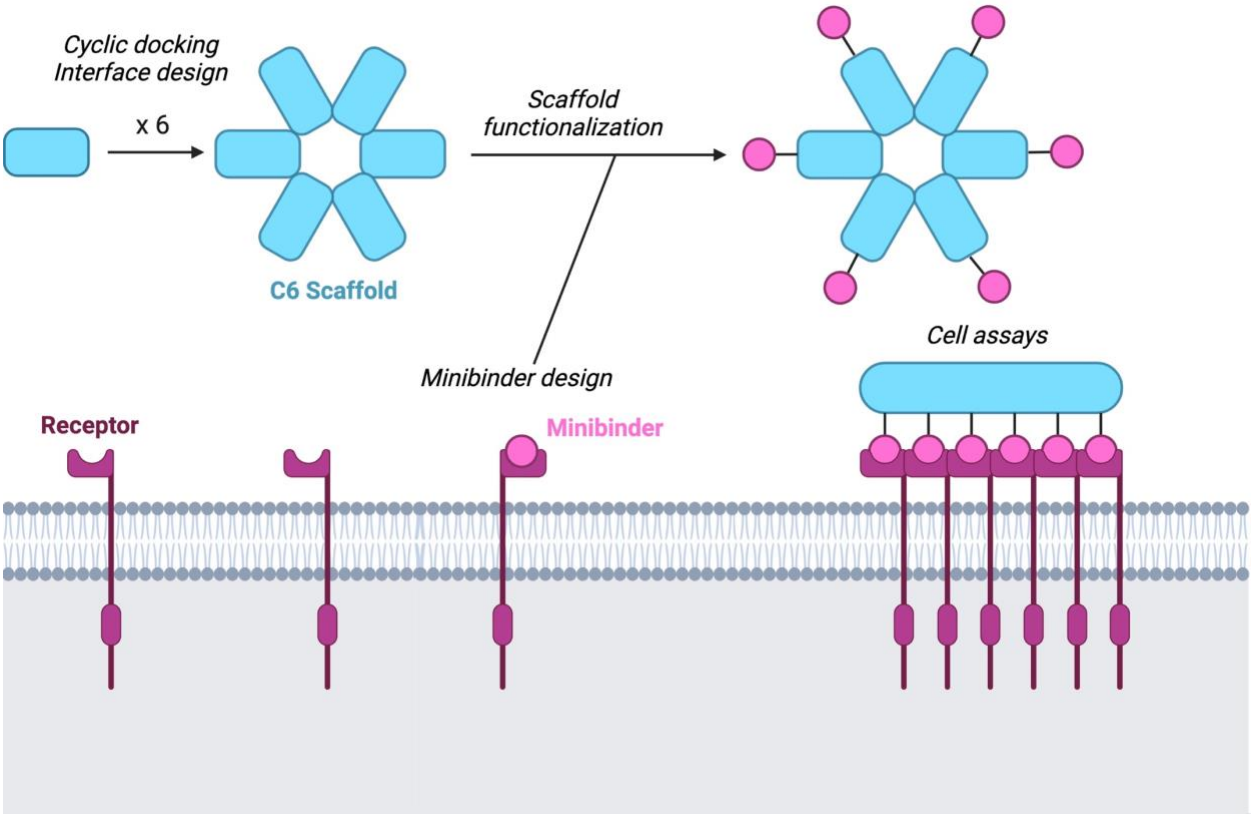

Figure 1.1. Design approach. First, a multimeric protein scaffold is generated through cyclic docking and interface design. The scaffold is capable of presenting multiple copies of a receptor-binding domain (minibinder) by genetic fusion of the coding sequences. Binding multiple receptors in close physical proximity is expected to cluster receptors and stimulate activation of the signaling pathway.

Chapter 2: Oligomer design

2.1 Docking

I set out to create new C4-C8 proteins by docking repeat proteins into the desired symmetry and designing an oligomeric interface. A set of 19 proteins with high-resolution crystal structures and/or small-angle X-ray (SAXS) validation was used as scaffolds for the asymmetric unit (Table 1). These designed helical repeat proteins (DHRs) contain 4 repeat units consisting of paired helices, permitting the addition or subtraction of repeat units to change the length of the protein.[21,22] A subset of these DHRs had previously been designed into C2-C5 oligomers using a docking protocol that first arranges asymmetric units into the desired symmetry and then optimizes sequences at the new protein-protein interface for binding between subunits.[13] These authors reported success for 4/11 dimers, 6/34 trimers, 6/19 tetramers, 1/17 pentamers, and 0/15 hexamers, with success defined as SAXS data consistent with the design model. We expanded the scaffold set to include newer designed repeat proteins and applied the method to C7 and C8 symmetry for the first time.

Docking and design of oligomer interfaces was performed as previously described.[13] Briefly, this protocol consists of generating low-resolution docks and scoring by the residue-pair transform (RPX) metric, which evaluates designability of an interface based on pairs of amino acids across the junction that can stabilize the conformation. After this initial fast scoring metric, RosettaDesign calculations were performed to optimize the protein-protein interface. Designs were filtered by $\Delta\Delta G$ (between -35 and -70), solvent accessible surface area (SASA > 1400), shape complementarity (sc > 0.65), and fewer than 2 unsatisfied hydrogen bonds. This resulted in 150

outputs, 109 of which were selected for structural characterization: 15 tetramers, 16 pentamers, 24 hexamers, 24 heptamers, and 30 octamers.

Designs were ordered as synthetic genes in plasmids with a C-terminal 6xHis tag and expressed in *Escherichia coli*, followed by purification by immobilized nickel affinity chromatography (IMAC). Of the 60 designs that expressed with a soluble product on IMAC, 28 had single monodisperse peaks on size exclusion chromatography (SEC), indicating relative homogeneity of the size of the particle. SEC has sufficient resolution to detect an assembling design compared with a monomeric unit, which typically elutes 3-4mL later than a C4+ assembly. The resolution is not high enough to determine the precise oligomeric state, e.g. hexamer vs. heptamer, and it cannot detect whether multiple oligomeric states exist, which is a known issue in designing oligomers with valency of 4+.[13,14] For this reason, designs were further characterized by SAXS[23,24], which provides low-resolution information about particle shape, and SEC with multi-angle light scattering (SEC-MALS), which can estimate particle mass by comparing against a set of proteins with known mass. Ten of the 28 designs had a single oligomeric state confirmed by both methods (**Figure 2.1, Figure 2.2B-D**). Five of the successful oligomers were tetramers, four were hexamers, and one was an octamer. Of note, the starting scaffold PDB ID 5CWN[21] generated 5 successful designs: C4-71, C4-717, C6-714, C6-71, and C8-71.

Two designs had off-target symmetry: the design C6-714 was designed as a C7 but behaved as a C6 by SEC-MALS; similarly, C6-79 was designed as a C8 but behaved as a C6. For the designs with an off-target symmetry, SAXS experimental data were plotted against theoretical traces for both the on-target and off-target symmetries. For C6-714, there was little difference in the theoretical traces for C6 and C7 (**Figure 2.1C**), while for C6-79, the C6 theoretical trace closely

matched the experimental data in stark contrast to the C8 theoretical trace (**Figure 2.1G**). None of the ordered pentamers had a single oligomeric state on characterization.

To further analyze the structures, the 10 designs were screened by negative-stain electron microscopy and cryo-electron microscopy (cryo-EM), and 6 were selected for 3D map reconstruction. Two designs, C4-131 and C6-71, had severe preferred orientation on cryo-EM grids and were not pursued further. Final resolution for C8-71 and C6-79 was 3-5Å, compared with 5-7Å for C4-71 and C4-81. All maps were in good agreement of the design model, except for C6-79, which formed a hexamer instead of the designed octamer.

The designs C4-81 and C6-4 both contain disulfide bridges within the asymmetric unit, originating in the base building blocks 5CWQ (a.k.a. DHR81) and 5CWB (a.k.a. DHR4), respectively. The presence of cysteines can make future modification difficult if one plans to incorporate unpaired cysteines for maleimide labeling, for example. Both 5CWQ and 5CWB have validated cognate designs (DHR80 and DHR9, respectively) in which the cysteine residues were changed to alanine, serine, or other residues as appropriate.[21] I made the same sequence changes in C4-81 and C6-4, named the designs C4-80 and C6-9, and ordered genes for experimental characterization. The disulfide-free designs eluted at a similar volume as the original design on SEC, indicating that they maintain the original oligomeric state and facilitating use of these proteins for future materials design.

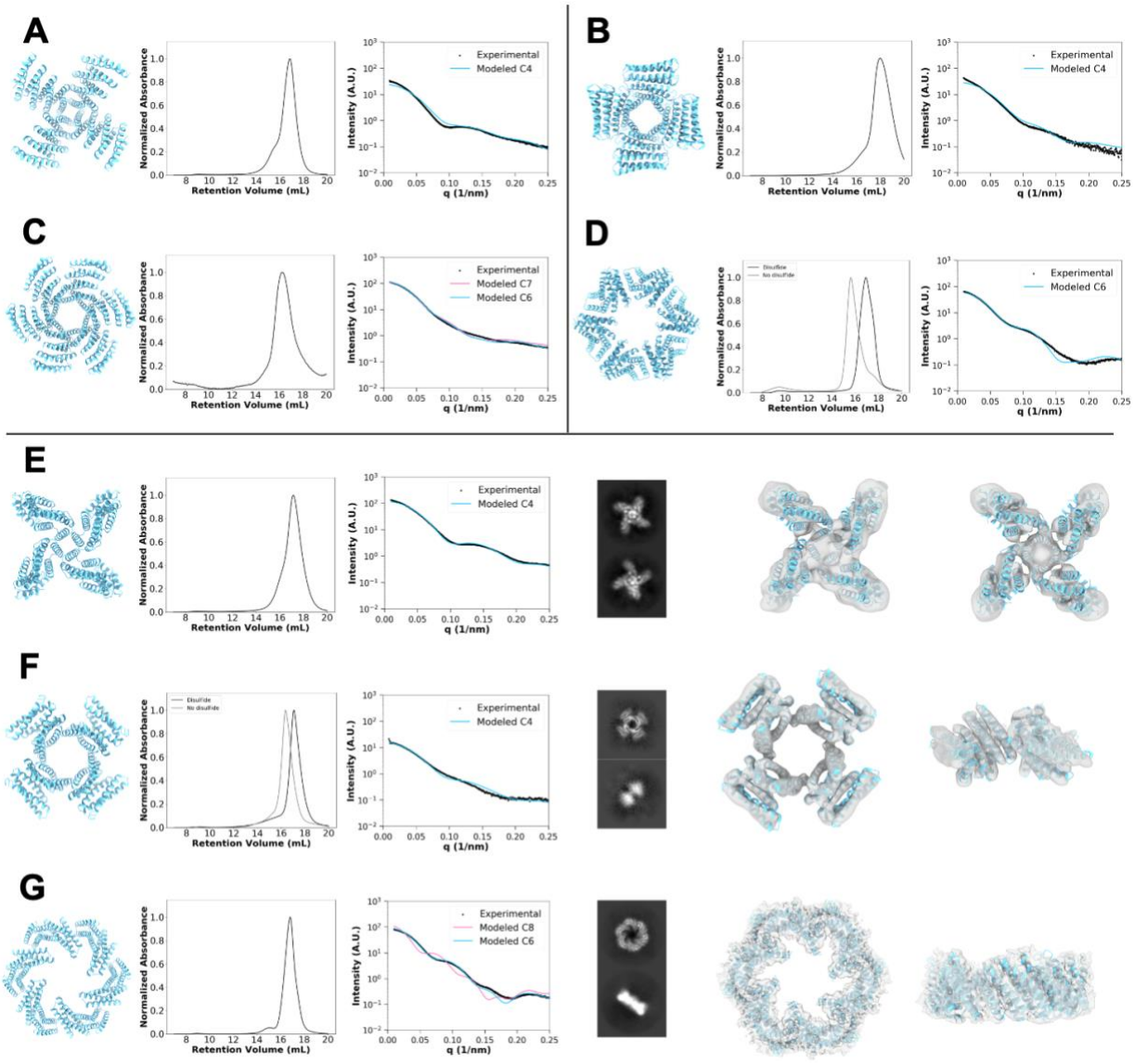

Figure 2.1. Characterization of oligomers designed by cyclic docking. Design protocol adapted from Fallas et al. 2017. The asymmetric unit is docked and scored by RPX, then an interface is designed using Rosetta, following which designs are filtered by several metrics. Designs are characterized by SEC and SAXS, then screened by negative-stain electron microscopy for particle formation on grids to select designs for cryo-EM. The designs C4-81 and C6-79 were selected for characterization and map-building by cryo-EM. (**A**) C4-717, (B) C4-18, (**C**) C6-714, (**D**) C6-4 (black) and disulfide-free variant C6-9 (grey), (**E**) C4-131, (**F**) C4-81 (black) and disulfide-free variant C4-80 (grey), (**G**) C6-79.

2.2 Repeat extension

By using repeat proteins as the asymmetric unit, the length or diameter of a protein can be increased by inserting additional repeat units.[13,16,21] Three designs derived from 5CWN were selected for repeat extension: C4-71, C6-71, and C8-71. We employed a secondary structure-based helical fusion approach, aligning the original repeat protein to the asymmetric unit, fusing, then redesigning residues around the fusion site to match the original oligomer where appropriate (**Figure 2.2A**). An additional 2 or 4 repeat units were added, to create a 6-repeat variant and an 8-repeat variant of each design. Designs were characterized to ensure oligomeric state by SEC-MALS, SAXS, and cryo-EM (**Figure 2.2B-D**). Single-particle cryo-EM reconstructed 3D maps of the extended oligomers agree well with the design models, with secondary structure clearly visible. This confirmed that oligomers can be extended in diameter by repeat extension without disrupting oligomeric state.

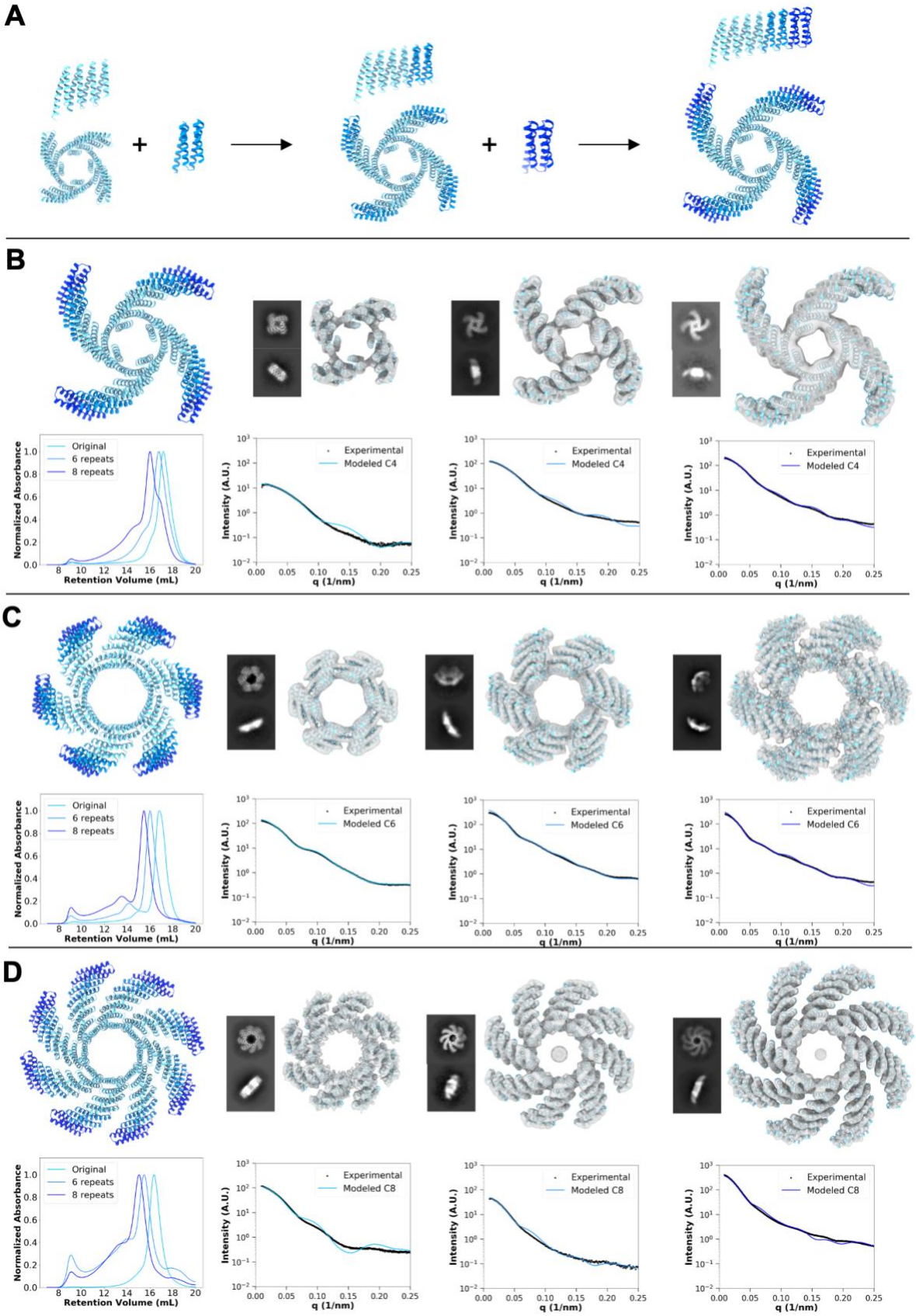
A
B
C
D
Original
6 repeats
8 repeats
Normalized Absorbance
Retention Volume (mL)
Experimental
Modeled C4
Modeled C6
Modeled C8
Intensity (A.U.)
q (1/nm)

Figure 2.2. Oligomers with repeat extensions. (**A**) Extension schematic. Repeat units are added by aligning the repeat protein to the oligomer asymmetric unit, creating a fusion junction, and redesigning residues where appropriate. 6- and 8-repeat versions were generated. Designs were validated by SEC, SAXS, and cryo-EM. (**B**) C4-71, (**C**) C6-71, (**D**) C8-71.

2.3 Cryo-electron microscopy backbone models

The high resolution of the C6-79 and C8-71 maps enabled us to build backbone models of these designs (**Figures 2.3-4**). Alignments of backbone models to original design models were performed using MM-align[25] to calculate root mean square deviation (RMSD) in Å of the crystal structure from the design model. The C6-79 cryo-EM model aligns well with the docked C6 prediction with an RMSD of 2.85 Å, and the C6 state bears a geometric resemblance to the C8 design model. Unlike the C8 design model, the cryo-EM model contains an undesigned salt bridge between Glu38 and Arg34, together with Thr200 on the neighboring chain. Arg34 is also participating in a cation-pi interaction with Trp201 on the neighboring chain at a distance of 3.1Å. These favorable interactions, which were not present in the C8 design, may explain why C6-79 assembles into a hexamer rather than the designed octamer.

The cryo-EM model of C8-71 is also in good agreement with the design model with an RMSD of 1.79 Å, with the exception of the final 10 C-terminal residues, which show backbone deviation from the design model. This is possibly to better accommodate Trp152 on the neighboring chain and results in Leu198 being more buried than in the design model, likely also a favorable change. Mutation of these residues to hydrophilic residues (W152E and/or L198D) disrupted oligomer formation as shown by broadening of the SEC trace, further supporting that these hydrophobic residues are buried rather than surface exposed as in the design model.

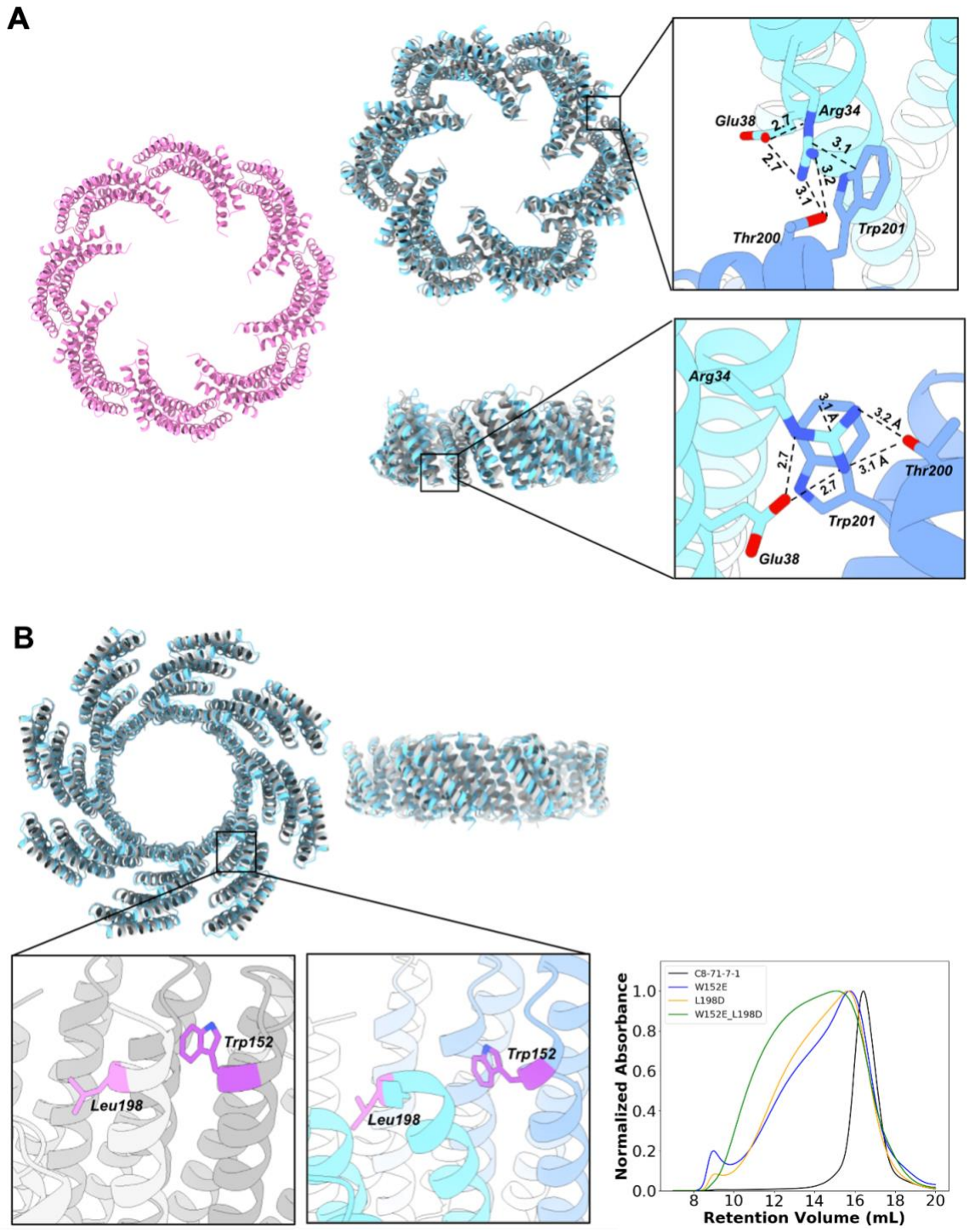

Figure 2.3. Cryo-EM backbone models of C6-79 and C8-71. (**A**) C6-79 model. An undesigned salt bridge is present in the cryo-EM model but not in the original design. The C6 shape bears a geometric resemblance to the original C8 design model. (**B**) C8-71 model. Residues Trp152 and Leu198 deviate from the design model, bending toward the core of the protein. Mutation of either or both of these residues to a hydrophilic residue disrupts oligomer formation as demonstrated by SEC.

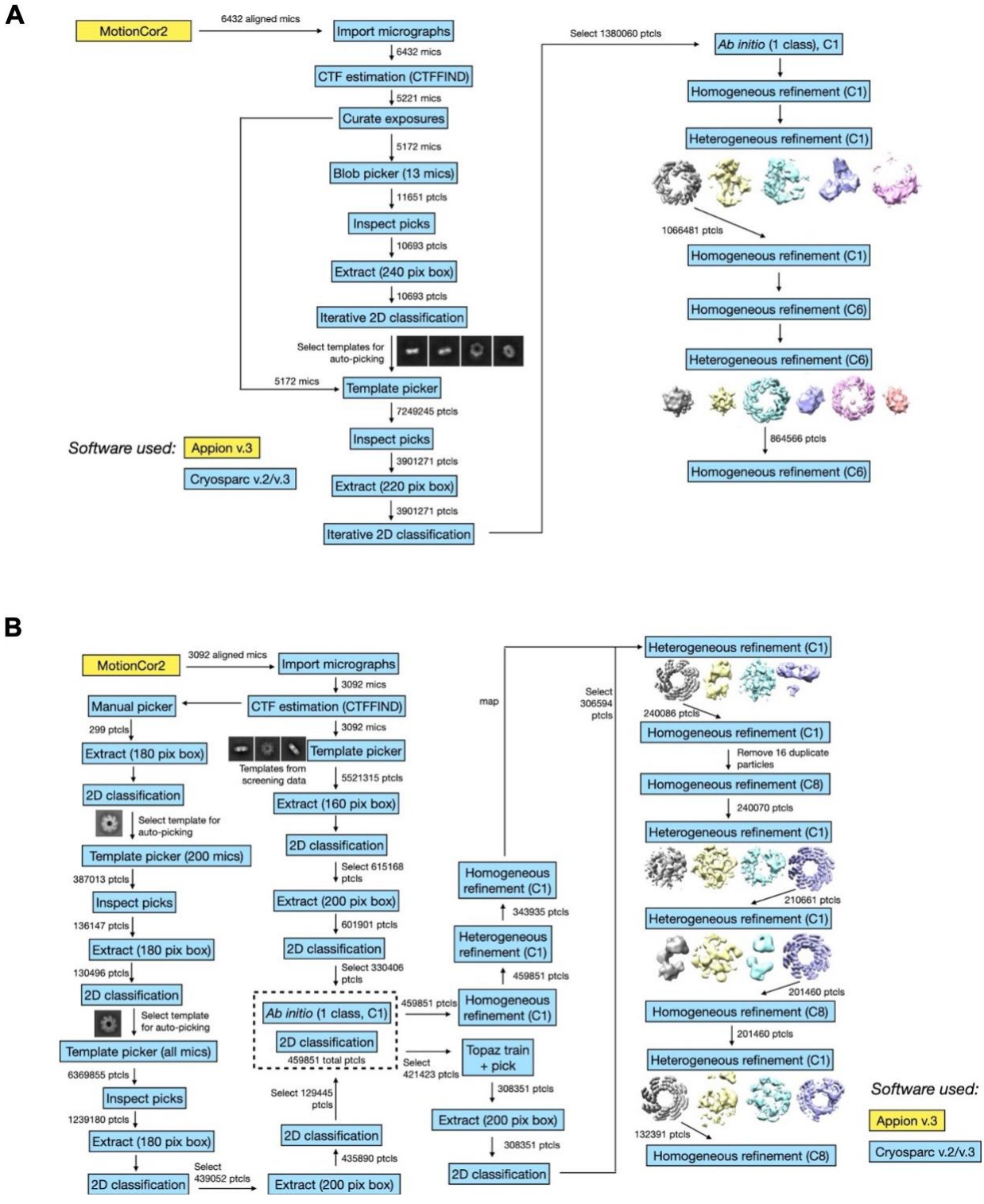

Figure 2.4. Cryo-EM data processing workflow for (A) C6-79 and (B) C8-71.

2.4 Oligomer design using unvalidated building blocks

Using previously experimentally validated proteins as the asymmetric unit increases the likelihood of success but decreases the potential geometric diversity of the oligomer. We next attempted to design new oligomers with the same docking protocol, but now using unvalidated repeat protein designs[Yang W et al. in preparation] or 3-helix bundles.[19] A total of 30 proteins of C2-C8 valency were ordered; of these, 6 had the correct oligomeric state, all dimers (**Figure 2.5**). A crystal structure was obtained for the designs C2-01018 and C2-02802, showing good agreement with the design model.

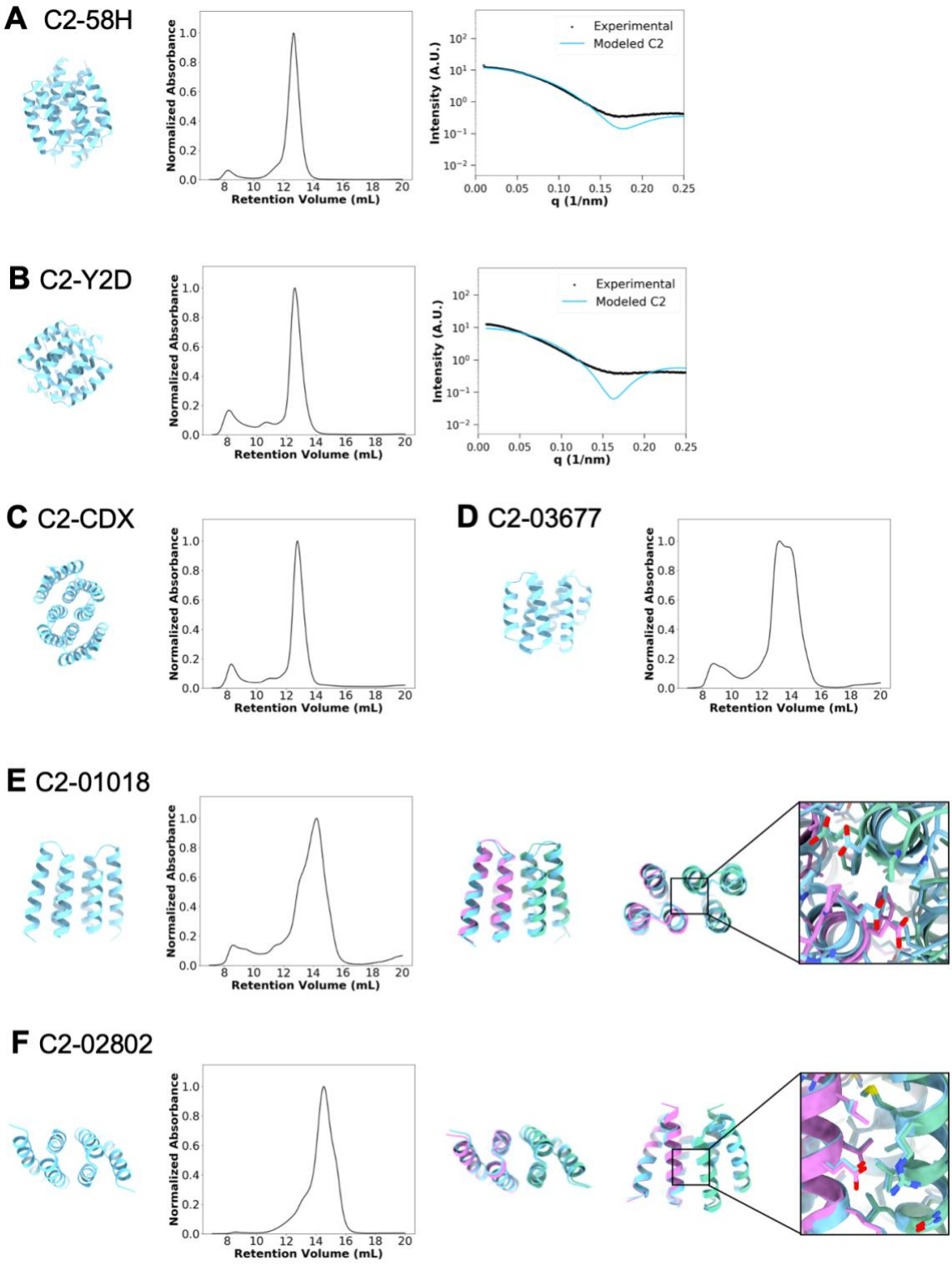

Figure 2.5. Dimers designed by cyclic docking using unvalidated building blocks. Designs were validated by SEC +/- SAXS. (**A**) C2-58H, (**B**) C2-Y2D, (**C**) C2-CDX, (**D**) C2-01018, (**E**) C2-03677, (**F**) C2-02802. For C2-01018 and C2-02802, a crystal structure (pink and green) was obtained and shows good agreement with the design model (blue).

2.5 Helical fusion

An attractive feature of helical proteins is the ability to rigidly fuse two proteins through shared alpha helical secondary structure, a process known as helical fusion. An existing oligomer can be modified in its geometry (through attaching extensible DHR "arms") or functionality (e.g. adding heterodimers or receptor-binding domains). As we have shown in this study and previously, the success rate for C4-C8 designs generated through docking protocols is relatively low due to the need to design a new oligomeric interface. Therefore, helical fusion represents an efficient method to diversify function and geometry of an existing oligomer.

We developed a method known as HelixFuse to generate a library of fusions between DHRs and oligomeric proteins. DHRs and oligomers were combinatorially fused together by overlapping the terminal helix residues in both directions ("AB": c-terminus of HB to n-terminus of DHR, "BA": n-terminus of HB to c-terminus of DHR).[22] On the oligomer end, up to 4 residues were allowed to be deleted to maximize the sampling space of the fusion while maintaining the structural integrity of the oligomeric interface. On the DHR end, deletions up to a single repeat were allowed. After the C-beta atoms are superimposed, an RMSD check across 9 residues was performed to ensure that the fusion results in a continuous helix. If no residues in the fused structure clash (Rosetta centroid energy < 10), sequence design was carried out at all positions within 8Å of the junction. This first step of the fusion sampling is wrapped into the Rosetta MergePDBMover[22]. After sequence design around the junction region[26,27], fusions were then evaluated based on the number of helices interacting across the interface (at least 3), buried surface (sasa > 800) across the junction, and shape complementarity (sc > 0.6) to identify designs likely to be rigid across the junction point. In total, the building block library generated in silico by HelixFuse using DHR arms consisted of 107 fusions to a C5 helical bundle, 5H2LD-10.[28]

As a proof of concept, 22 designs were selected for experimental characterization and genes encoding the designs were synthesized and subsequently expressed in E. coli. Of these designs, 12 were soluble, 7 were monodisperse by SEC, and 2 had SEC-MALS mass estimates at the predicted range as well as matching SAXS profiles (**Figures 2.6-7**). Cryo-electron microscopy of C5_HF-3921 followed by 3D reconstruction showed that the positions of the helical arms are close to the design model. By negative-stain electron microscopy (EM), C5_HF-2101 was class averaged and the top-down view clearly resembles that of the designed model and its predicted projection map (**Figure 2.8**).

I also applied the method to the non-helical bundle oligomer tpr1C4_pm3, a designed homotetramer[13]. As described above, I fused DHRs to the C-terminal helix of tpr1C4_pm3. A total of 16 tpr1C4_pm3 fusions were tested, 14 found to be soluble, and 10 displayed monodisperse peaks by SEC. The best behaving designs were analyzed by electron microscopy. C4_nat_HF-7900 was found to form monodisperse particles by cryo-EM, with the 3D reconstruction modeled to 3.7Å resolution (**Figure 2.6B**). The model of the cryo-EM reconstruction of C4_nat_HF-7900 shows very good matching near the oligomeric hub of the protein where side chains are clearly resolved and as expected. However, the structure deviates from the design model at the most distal portions of the structure. This is likely due to the inherent flexibility of the unsupported terminal helices of the DHRs[21,22,29] and lever arm effects which increase with increasing distance from the fusion site.

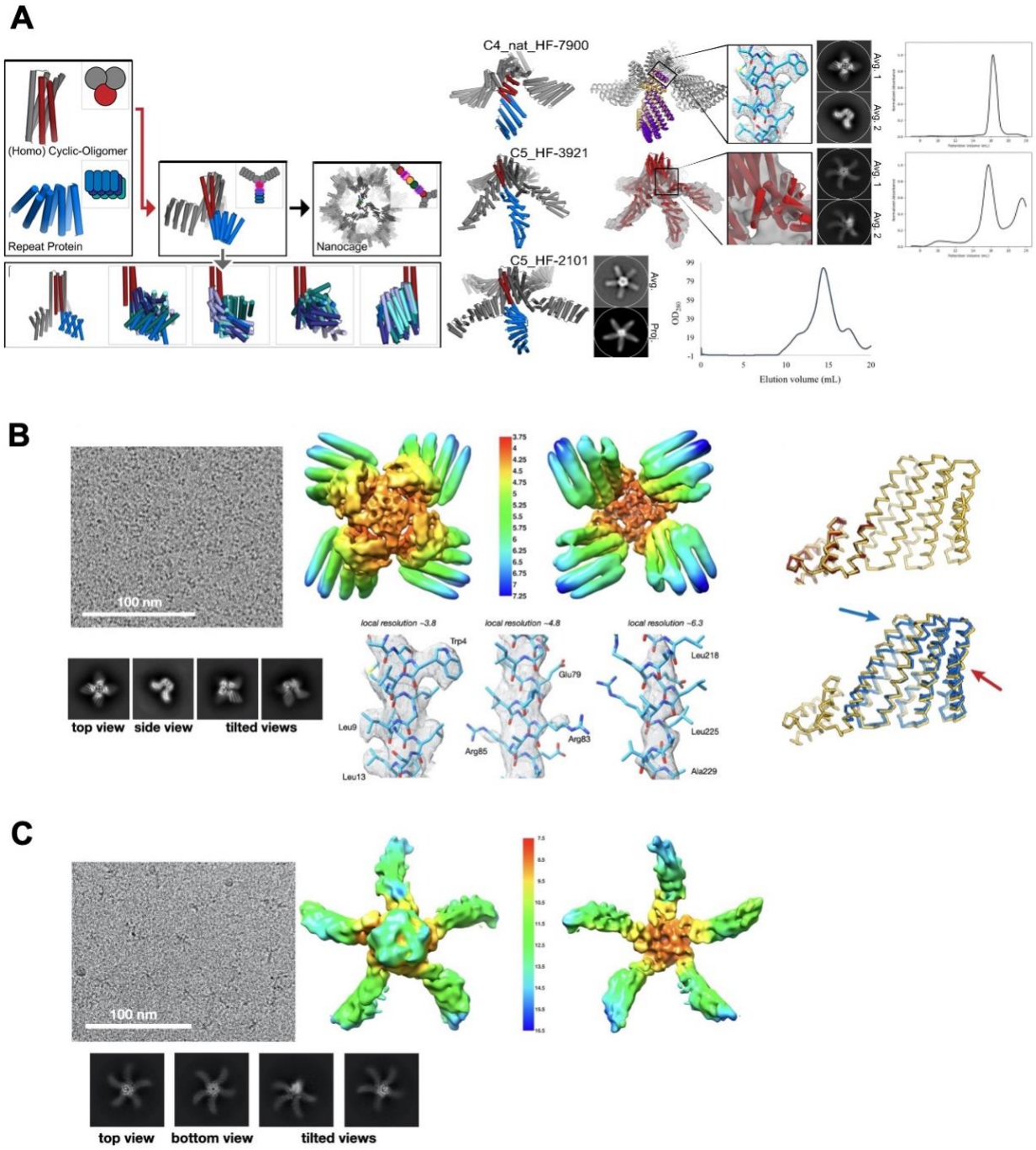

Figure 2.6. Rigid helical fusion to oligomers using HelixFuse. Adapted from Hsia, Mout et al. 2021. (**A**) Homo-oligomers (red) are fused to *de novo* helical repeat proteins (shades of blue) to create a wide range of building blocks. Symmetric units shown in grey. SEC and EM data shown for 3 designs, with a 3.8Å resolution map of C4_nat_HF-7900. All showed a good overall match to negative-stain EM or cryo-EM 2D class averages (top) from one direction, using a predicted projection map (bottom). (**B**) Cryo-EM data and reconstruction for C4_nat_HF-7900. A total of 3752 images were acquired from a single sample. Representative motion-corrected micrograph, representative 2D class averages, locally-filtered cryo-EM map colored by local resolution, fit of

cryo-EM structure (sticks) to density (mesh) in areas of high, intermediate, and low local resolution, and alignment of tpr1C4_pm3 (red) or DHR79 (blue) to C4_nat_HF-7900's cryo-EM model (yellow). (**C**) Cryo-EM data for C5_HF-3921. A total of 6744 images were acquired from a single sample. Representative motion-corrected micrograph, representative 2D class averages, and cryo-EM map colored by local resolution.

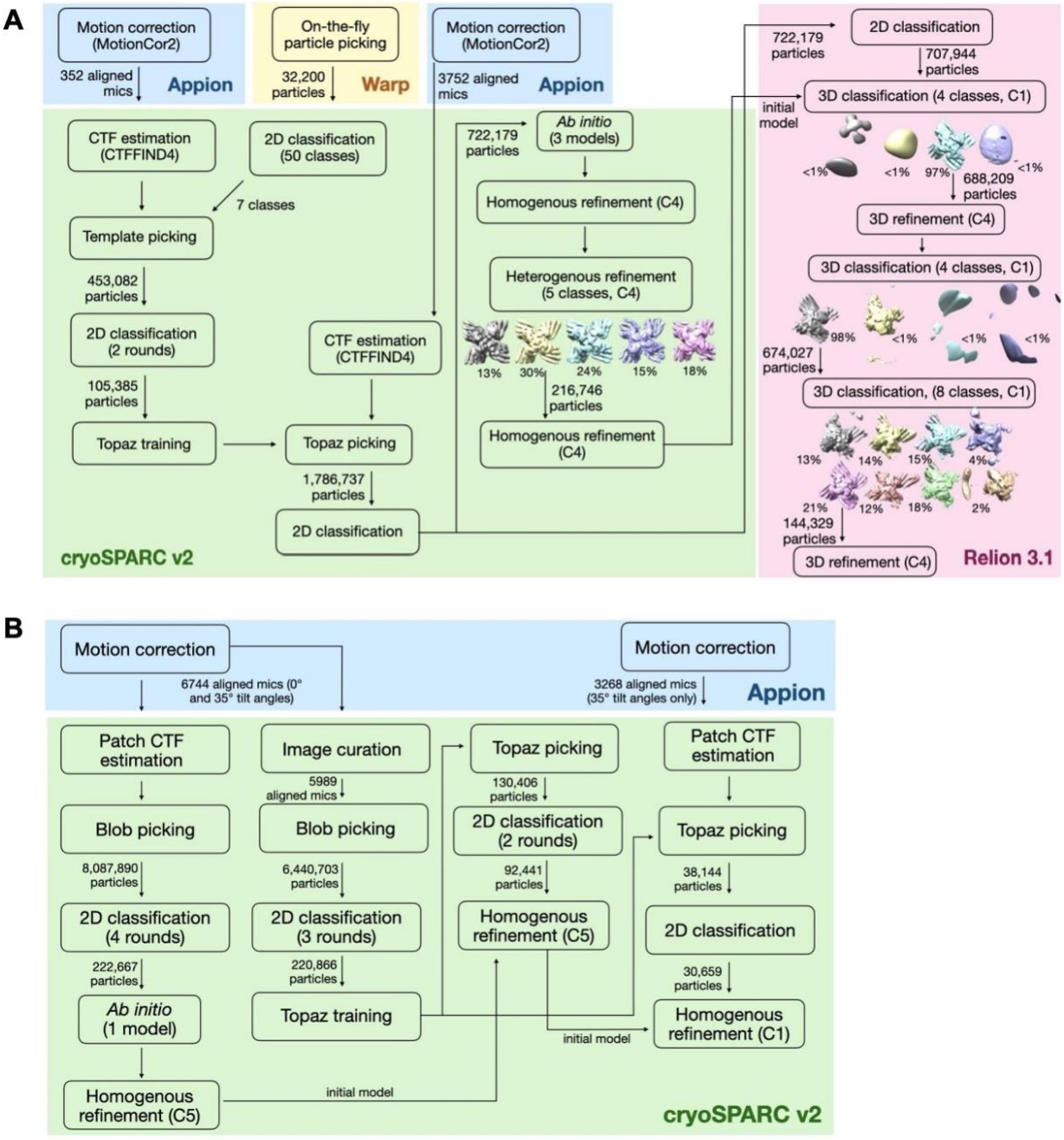

Figure 2.7. Cryo-EM data processing workflow for (A) C4_nat_HF-7900 and (B) C5_HF-3921. Adapted from Hsia, Mout et al. 2021.

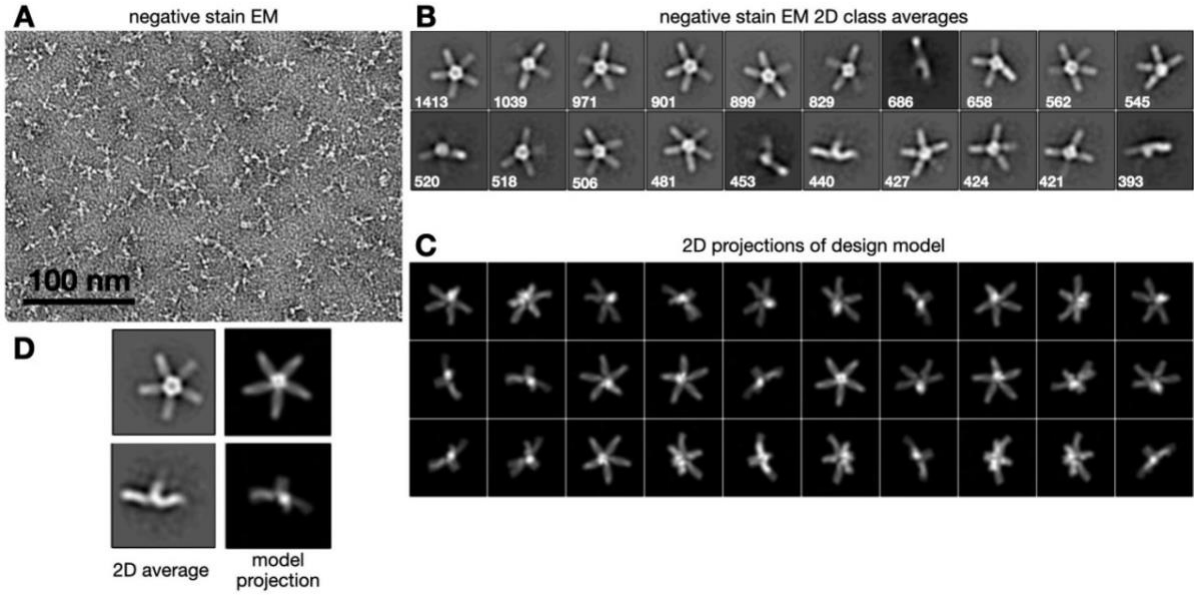

Figure 2.8. Negative stain EM data for C5_HF-2101. Adapted from Hsia, Mout et al. 2021. (**A**) Representative micrograph. (**B**) Most populated 2D class averages; numbers on each class image indicate the number of particles in that class. (**C**) 2D projections of a 20 Å-filtered volume generated from the atomic coordinates of the design model. (**D**) Selected 2D class averages shown alongside matching model projections. The micrograph in panel (**A**) is representative of 24 images acquired from a single sample. Particles were extracted from all micrographs for 2D classification, with the number of particles comprising each class indicated in panel (**B**).

2.6 Two-component systems: flexible

Two-component systems offer an additional layer of control and complexity that can be useful for various materials. Separating functional domains into two associating components allows for "mix-and-match" of a previously designed component with a new component, without the need to redesign the entire structure, as was elegantly demonstrated by Divine et al. in their description of antibody nanocages.[12] I sought to add this functionality to my designed oligomers and apply these "smart" systems to contemporary problems in molecular biology. Importantly, I chose to use only *de novo* designed proteins in my constructs for their specific control over protein function.

Protein domains may be connected either flexibly or rigidly. In flexible fusion, one domain is attached to the oligomer with a short glycine-serine linker between the two protein sequences. This method does not require any computational design and gives the functional domain some freedom of movement in relation to the oligomeric scaffold. I designed flexible linker fusions of the de novo heterodimer LHD101[30] to oligomers C4-71, C4-81, C6-71, C6-79, and C8-71, including the extension series. The LHD101B component was attached to the "outermost" terminus of each design (N or C terminus where appropriate) with a single glycine-serine sequence (GS). I purified the constructs from E. coli by IMAC and SEC. The oligomeric construct was then assembled with its cognate LHD101A at a 1:1.5 molar ratio of oligomer to free heterodimer, and assemblies were analyzed by SEC. Addition of the A component resulted in a leftward shift in the chromatogram peak, corresponding to assembly and increased size of the complex (**Figure 2.9**).

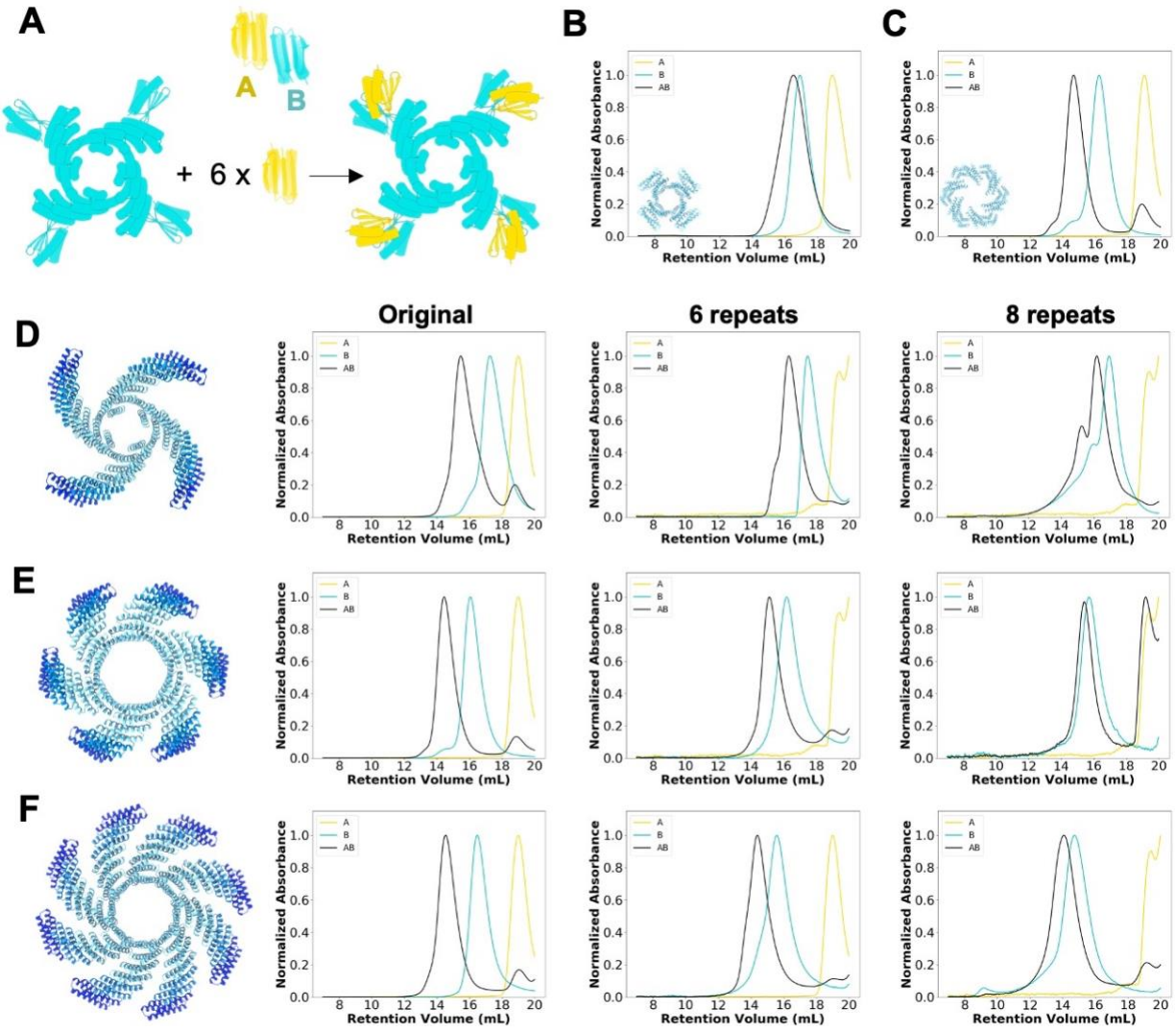

Figure 2.9. Flexible linker fusion of heterodimers to oligomers. (**A**) Assembly schematic. Heterodimer chain "B" is linked to the oligomer (shown in aqua) via a glycine-serine sequence. Mixing with chain "A" (yellow) results in spontaneous assembly of the 2-component structure (black). Assembly is verified by SEC of the oligomer without cognate heterodimer (labeled B), with cognate heterodimer (labeled AB), or free cognate heterodimer alone (labeled A). SEC traces depict fusions of LHD101 to (**B**) C4-71, (**C**) C4-81, (**D**) C6-71, (**E**) C6-79, and (**F**) C8-71. For **D-F**, columns left-to-right correspond to the original 4-repeat version, the 6-repeat version, and the 8-repeat version.

2.7 Two-component systems: rigid

Rigid fusion offers further control over topology of the complex. Functional domains can be rigidly fused to an oligomeric scaffold either through HelixFuse or through overlapping shared

repeat protein domains. Overlapping by repeat protein domain does not involve formation of a new fusion junction, which may theoretically increase success rate beyond the 10-15% seen with HelixFuse. I used a series of heterodimer-DHR fusions and previously published C4-symmetric homo-oligomers that already contained a terminal DHR. Applying the same align and splice approach as for the repeat extension, I generated overlap designs of the heterodimer LHD274B-DHR53 fusion to a previously designed C4-DHR fusion, 137_tj10C4_G1_lego_6, which contains DHR53 repeats at its "outermost" N terminus. I purified the designs and assembled with their cognate heterodimer as described above (**Figure 2.10**).

The C4 oligomeric hub was stable and soluble in isolation and readily forms the target complex when mixed with its binding partners, as confirmed by SEC chromatography and negative stain EM class averaging (**Figures 2.10-11**). In the absence of its binding partner, we observed an additional concentration-dependent peak on SEC, indicating formation of a higher-order complex. This is likely a dimer of C4 hubs, since the C4 hub contains the redesigned protomer LHD274B, which is known to weakly homodimerize. Addition of the binding partner drives reconfiguration of this higher order assembly into the on-target octameric (A4B4) complex. Though the homodimeric interaction is weak in comparison to the heterodimer, I chose not to pursue this design further out of concerns that it would not tolerate further perturbations in the form of added minibinders. [Modified from Sahtoe DD, Praetorius F, Courbet A, Hsia Y, Wicky BIM, Edman NI, et al. Reconfigurable asymmetric protein assemblies through implicit negative design. *Science* 2022;375(6578): eabj7662. Reprinted with permission from AAAS.]

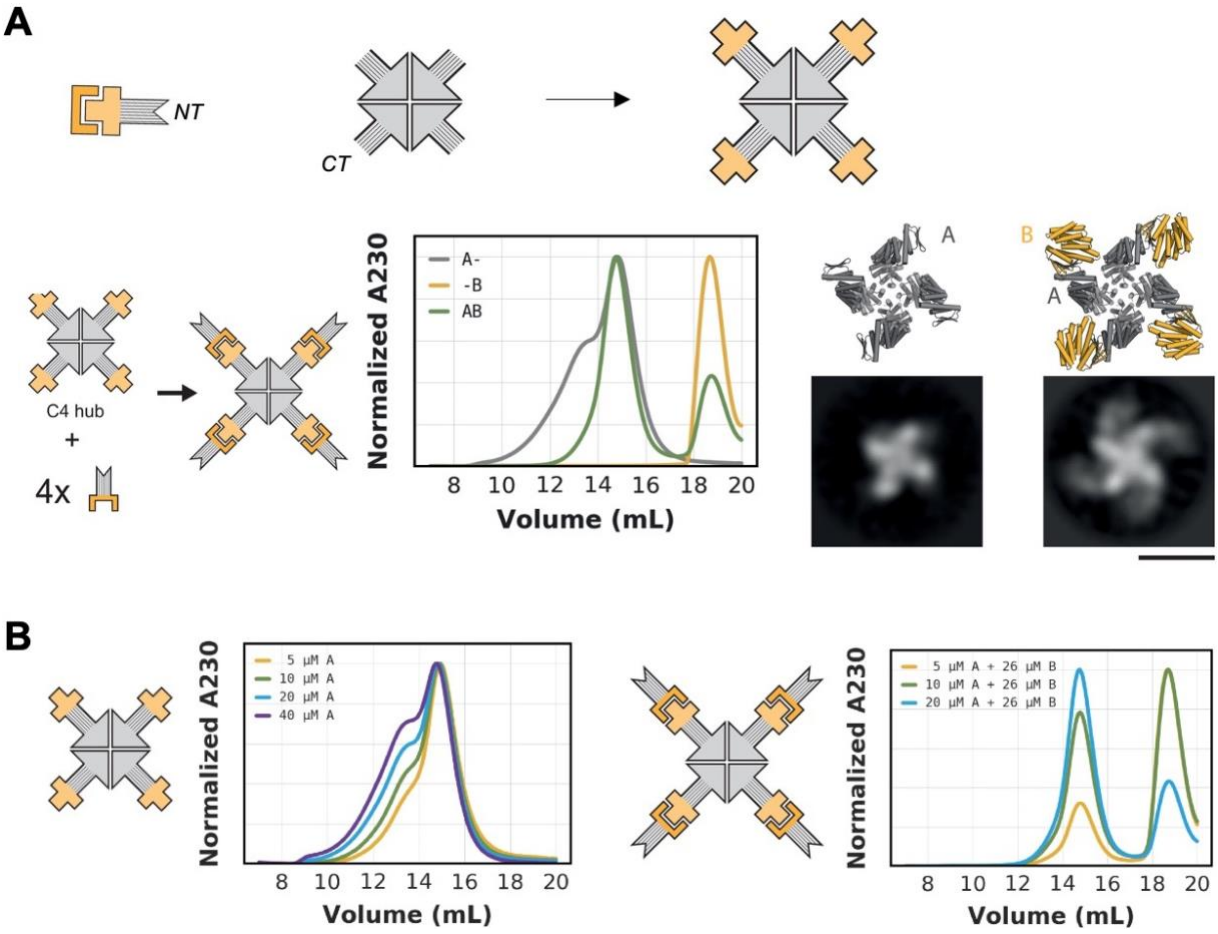

Figure 2.10. Rigid helical fusion of heterodimers to oligomers. (**A**) Schematic depictions showing matching repeat protein regions aligned to generate a 2-component rigid fusion heterodimer-oligomer. The C4-symmetric hub presenting four copies of LHD274B binds its cognate binding partner (274A53). SEC trace shows assembly after mixing of the components; representative negative stain EM class averages are shown on the right. (**B**) SEC traces of the C4-symmetric hub at different concentrations without binding partner and with a constant concentration of binding partner. Concentrations are given per monomer (5 μM corresponds to 1.25 μM tetramer). From Sahtoe DD, Praetorius F, Courbet A, Hsia Y, Wicky BIM, Edman NI, et al. Reconfigurable asymmetric protein assemblies through implicit negative design. *Science* 2022;375(6578): eabj7662. Reprinted with permission from AAAS.

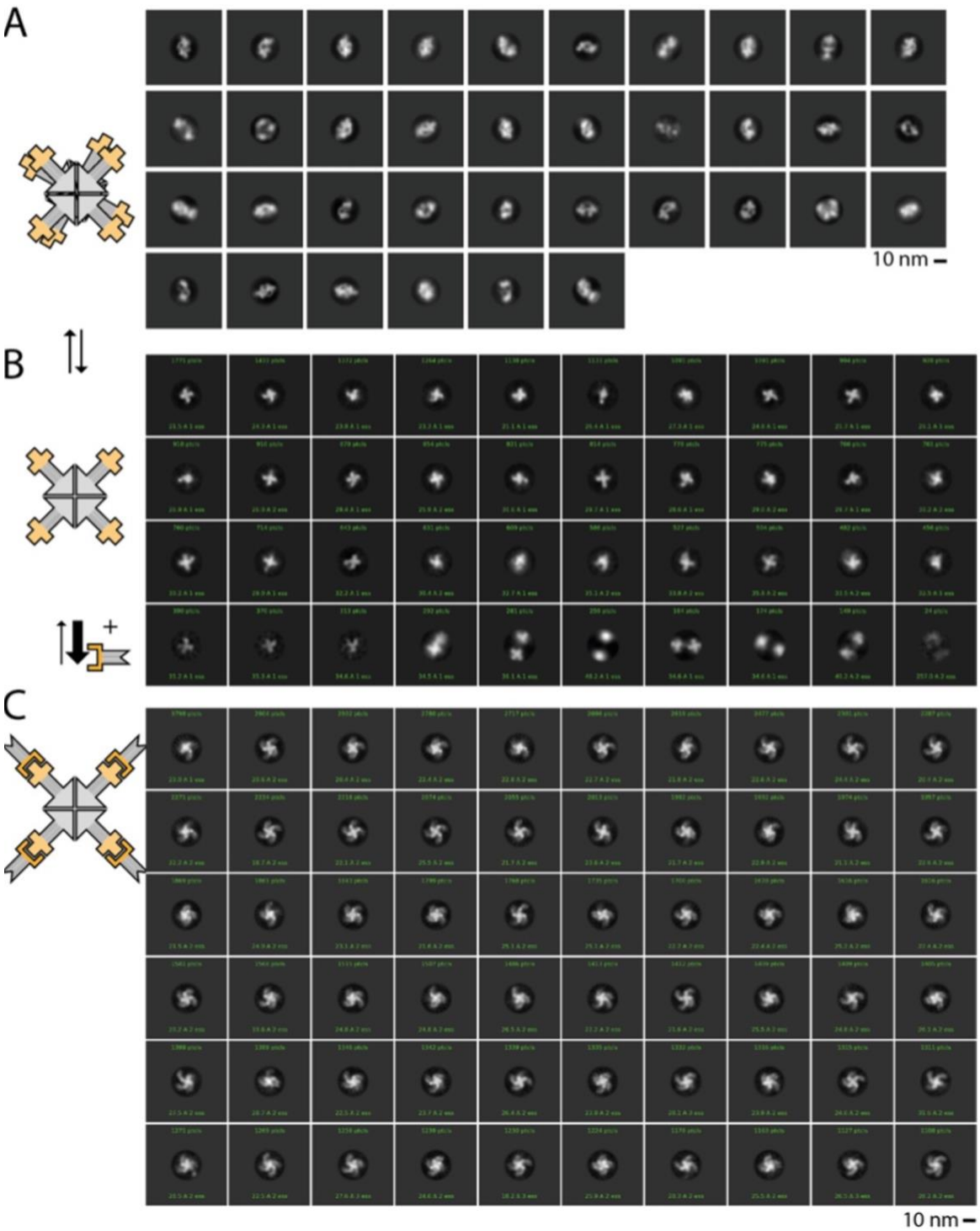

Figure 2.11. Negative stain EM class averages of C4 symmetric hetero-oligomer. Schematic representations (left; (dark grey: C4 hub, gold: binding partner) and negative stain EM class averages (right) of the C4-symmetric hub without (**A**) and with (**C**) binding partner. In absence of the binding partner, the C4 hub exists in equilibrium between a higher order complex (**A**) and the designed C4 complex (**B**). From Sahtoe DD, Praetorius F, Courbet A, Hsia Y, Wicky BIM, Edman NI, et al. Reconfigurable asymmetric protein assemblies through implicit negative design. *Science* 2022;375(6578): eabj7662. Reprinted with permission from AAAS.

Chapter 3: Targeting the fibroblast growth factor receptor

3.1 De novo minibinders

Miniprotein binders or "minibinders" are small *de novo* proteins designed to bind their target with a high degree of affinity. Briefly, design of minibinders consists of 3 phases, as previously described.[19] Initially, a set of 10,000+ designs is created by docking of amino acid motifs combined with stable backbone scaffolds against the target protein and interface design. Genes encoding these proteins are ordered as DNA segments with flanking regions homologous to the pETCON plasmid and the Aga2 yeast surface protein, then co-transformed with plasmid into yeast. Transformed cells are sorted in the presence of a fluorescently labeled target protein, and collected cells are eventually sequenced to identify designs that enriched in successive sorts. Following this phase, a site saturation mutagenesis (SSM) library is generated, in which each amino acid position in the protein is substituted with all 20 amino acids. These designs are again sorted from yeast display, and sequencing results identify specific mutations that enriched, indicating an increased affinity for the target. In the final stage, a combinatorial library is generated, with specific mutations indicated from the SSM library to have increased affinity. Following this, individual designs are expressed in *E. coli* and tested by biolayer interferometry (BLI) to assess affinity to the target.

Minibinders have been successfully designed against a number of targets, with potent antagonistic activity.[10,17–19] This includes a minibinder "mb7" designed against the fibroblast growth factor receptor 2 (FGFR2); mb7 binds FGFRs 1-4 with varied affinities but is selective for the C isoform, with little to no binding against the B isoform. mb7 is a potent antagonist of signaling through FGFR C isoforms, in part due to blocking the interaction with the partner Klotho

proteins.(Sangwon Lee, personal communication, 2022-02-13) I hypothesized that we could link the growth factor receptor minibinders to oligomeric scaffolds and induce clustering and activation of the target receptor, thus converting a monomeric antagonist into a multimeric agonist.

3.2 Minibinder-oligomer fusions

To create a multivalent receptor-binding construct, mb7 was genetically fused at the amino and carboxyl termini of oligomers with a short glycine-serine linker, with linker length selected based on the success of the trimeric spike protein minibinder designs.[20] Four oligomers were selected for fusion: C4-71, C6-71, C6-79, and C8-71. Depending on the fusion terminus and the geometry of the oligomer, the binding domains are displayed at different spacings on adjacent subunits: for example, C6-79C_mb7 displays the minibinders 54Å apart, while C6-79N_mb7 displays the binders 18Å apart. The fusions eluted at the same volume as the oligomers determined by SEC, with the exception of C6-71C_mb7, which eluted significantly earlier than the base design (**Figure 3.1**). Characterization by negative stain EM 2D class averages showed that C6-71C_mb7 particles were self-associating into dihedral structures, presumably via the hydrophobic interface of the minibinder domain being presented in a favorable conformation for this interaction. For all other fusion constructs characterized by negative stain EM, the designs retained the oligomeric state of the original structure.

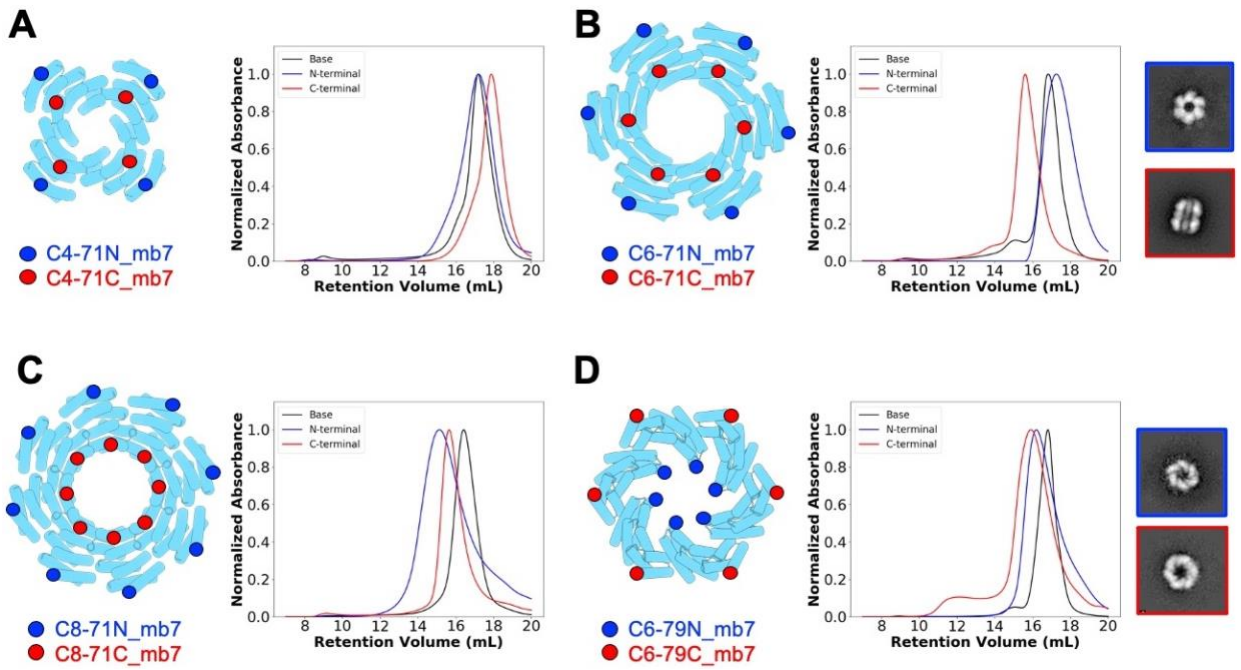

Figure 3.1. Flexible linker fusion of the FGFR minibinder mb7 to oligomers. mb7 is attached either at the amino terminus (N, shown in blue) or the carboxy terminus (C, shown in red) of the oligomer. Constructs are labeled according to the terminus of the oligomer with mb7 attached. Fusions are analyzed by SEC to compare elution profile to the building block without oligomer. Fusions to C6-71 and C6-79 were further characterized by electron microscopy to examine differences in elution time; C6-71C_mb7 forms a dihedral with 2 copies self-associating via their mb7 domains. (**A**) C4-71, (**B**) C6-71, (**C**) C8-71, (**D**) C6-79.

3.3 Signal transduction in target cells

FGFR signaling has several downstream effectors, including phosphorylation of extracellular signal-regulated kinase 1 and 2 (ERK1/2) and activity of phospholipase C-gamma (PLC-ɣ) leading to intracellular calcium release.[31–33] ERK phosphorylation (pERK) was selected as an initial screen for FGFR activity in response to the mb7 oligomers. Designs were screened in a Chinese hamster ovary (CHO) cell line expressing human FGFR1c, and in human umbilical vein endothelial cells (HUVECs) as a more clinically relevant primary cell line. Most of the designed oligomers induced pERK in the CHO cell line, with the exception of C8-71N_mb7; bands were

also faint for C6-71C_mb7, C6-71N_mb7, and C6-79N_mb7 (**Figure 3.2A**). This trend was similar in HUVECs (**Figure 3.2B**). The design C6-79C_mb7 had the strongest bands in both HUVECs and CHO cells, so this design and its cognate C6-79N_mb7 were chosen for more detailed dose-response and functional analysis. C6-79C_mb7 showed increased pERK at a binding-site concentration as low as 1nM, which compares favorably with the native ligand FGF2 (**Figure 3.2C**). In contrast, C6-79N_mb7 showed an increase in pERK only after 1 hour of treatment (**Figure 3.2D**). Phosphorylation of ERK was stronger for C6-79C_mb7 at 15min than C6-79N_mb7 at 1 hour (**Figure 3.2E**).

To assess FGFR activity through PLC-ɣ (which operates through a different signal transduction pathway from pERK), the C6-79 designs were screened by a calcium release assay in CHO cells expressing human FGFR1c and deficient in heparan. Heparan sulfate (frequently substituted with heparin in experiments) is a cofactor for FGFs that appears to enhance FGF-FGFR binding and signaling through oligomerization of FGF1 molecules.[34–37] For this reason, heparin must be added together with FGF1 in these cells to activate FGFR1c. Cells treated with C6-79C_mb7 showed increased fluorescence comparable to that of treatment with FGF1 + heparin, while cells treated with C6-79N_mb7 showed minimal increase in fluorescence (**Figure 3.3**). Notably, unlike the native ligand FGF1, C6-79C_mb7 induced calcium release without the addition of exogenous heparin. Together these results indicated that C6-79C_mb7 leads to activation of two distinct downstream effectors of FGFR signaling.

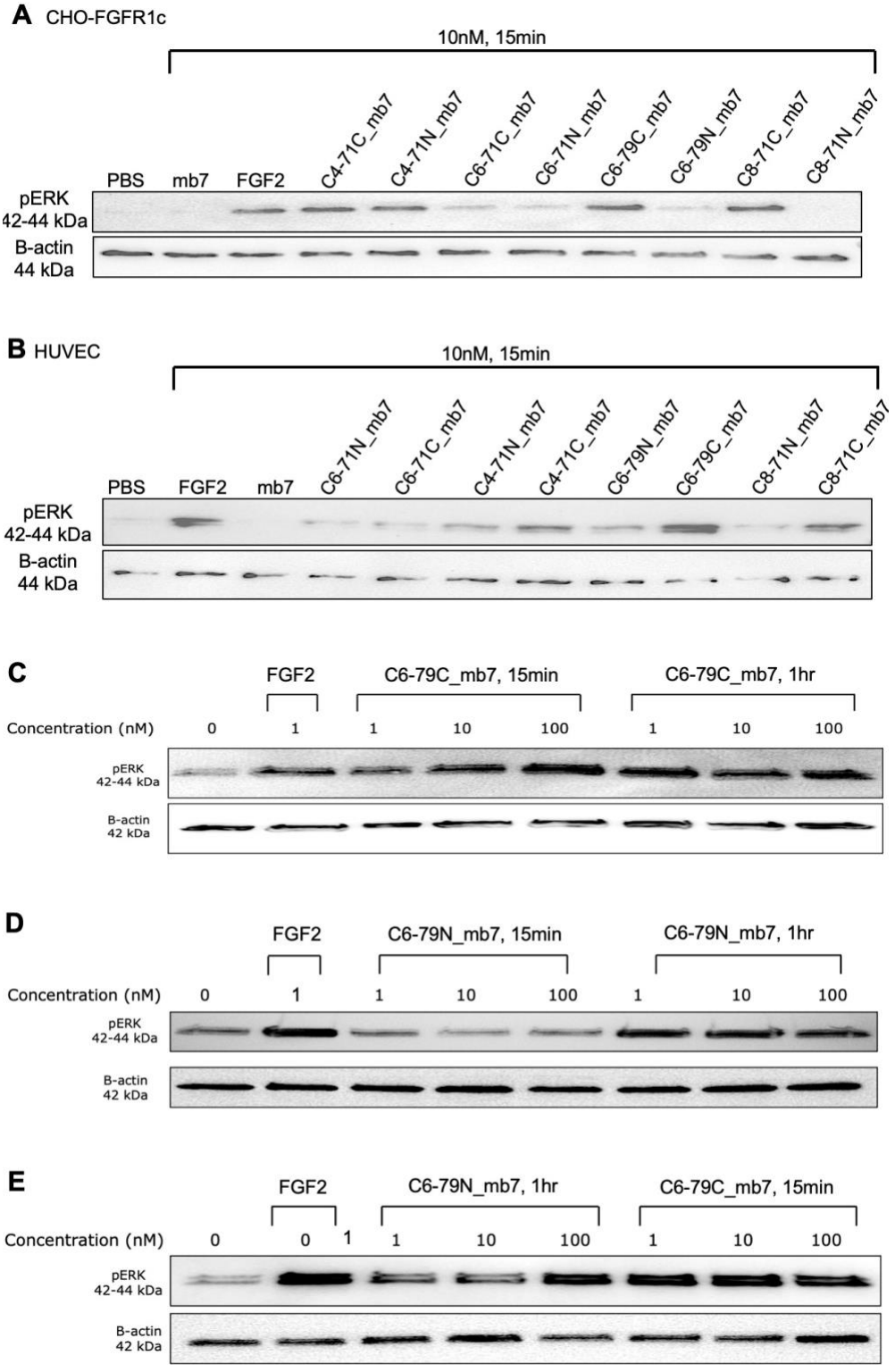

Figure 3.2. ERK phosphorylation in response to mb7 scaffolds. (**A**) Screening oligomer fusions in CHO-FGFR1c cells. (**B**) Screening oligomer fusions in HUVECs. (**C**) Dose and time dependence of C6-79C_mb7 treatment. (**D**) Dose and time dependence of C6-79N_mb7 treatment. (**E**) Comparing C6-79N_mb7 at 1 hour to C6-79C_mb7 at 15 minutes.

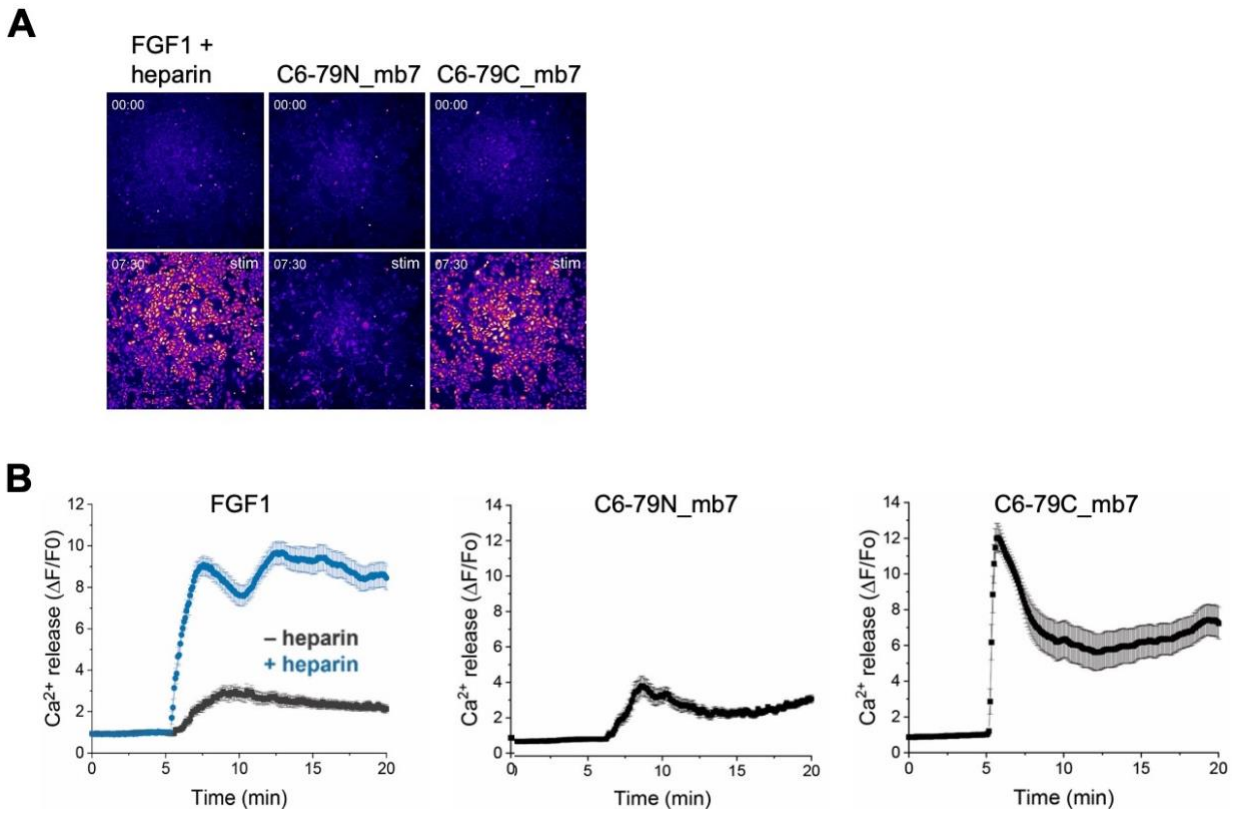

Figure 3.3. Intracellular calcium release in response to C6-79 mb7 scaffolds. (A) Calcium release assay using fluorescent indicator. Heparan-deficient CHO cells expressing human FGFR1c were treated with 10nM FGF + heparin or 100nM scaffold for 2.5min. C6-79C_mb7 shows a robust increase in fluorescence corresponding to increased calcium release. (**B**) Quantification of calcium release.

3.4 Receptor diffusion and distribution

The designed scaffolds are intended to activate FGFR signaling by clustering multiple copies of the receptor and inducing cross-phosphorylation. To confirm that the FGFR activity is due to clustering, we analyzed receptor movement via total internal reflection fluorescence (TIRF) single-molecule imaging using a fluorescent Halo-ligand in heparan-deficient CHO cells expressing Halo-tagged human FGFR1c[38]. Cells treated with C6-79C_mb7 showed decreased movement of FGFR1c (**Figure 3.4A**). The calculated diffusion coefficient was lowest for cells

treated with C6-79C_mb7, followed by those treated with FGF1 + heparin; cells treated with C6-79N_mb7 had a similar diffusion coefficient to untreated cells (**Figure 3.4B**).

Fluorescent labeling also permitted us to gather information on colocalization of multiple receptors by calculating fluorescence intensity of spots on the surface of the cell. Stepwise increases in intensity of the analyzed and detected spots indicate higher numbers of FGFR1c receptors colocalized. Cells treated with C6-79C_mb7 or FGF1 + heparin showed stepwise increases up to 2000AU from a baseline peak near 500AU, suggesting the formation of clusters containing multiple FGFR1c molecules (**Figure 3.4C**). This supports C6-79C_mb7 activating FGFR signaling through clustering of multiple receptor copies, again without requiring heparan as a cofactor.

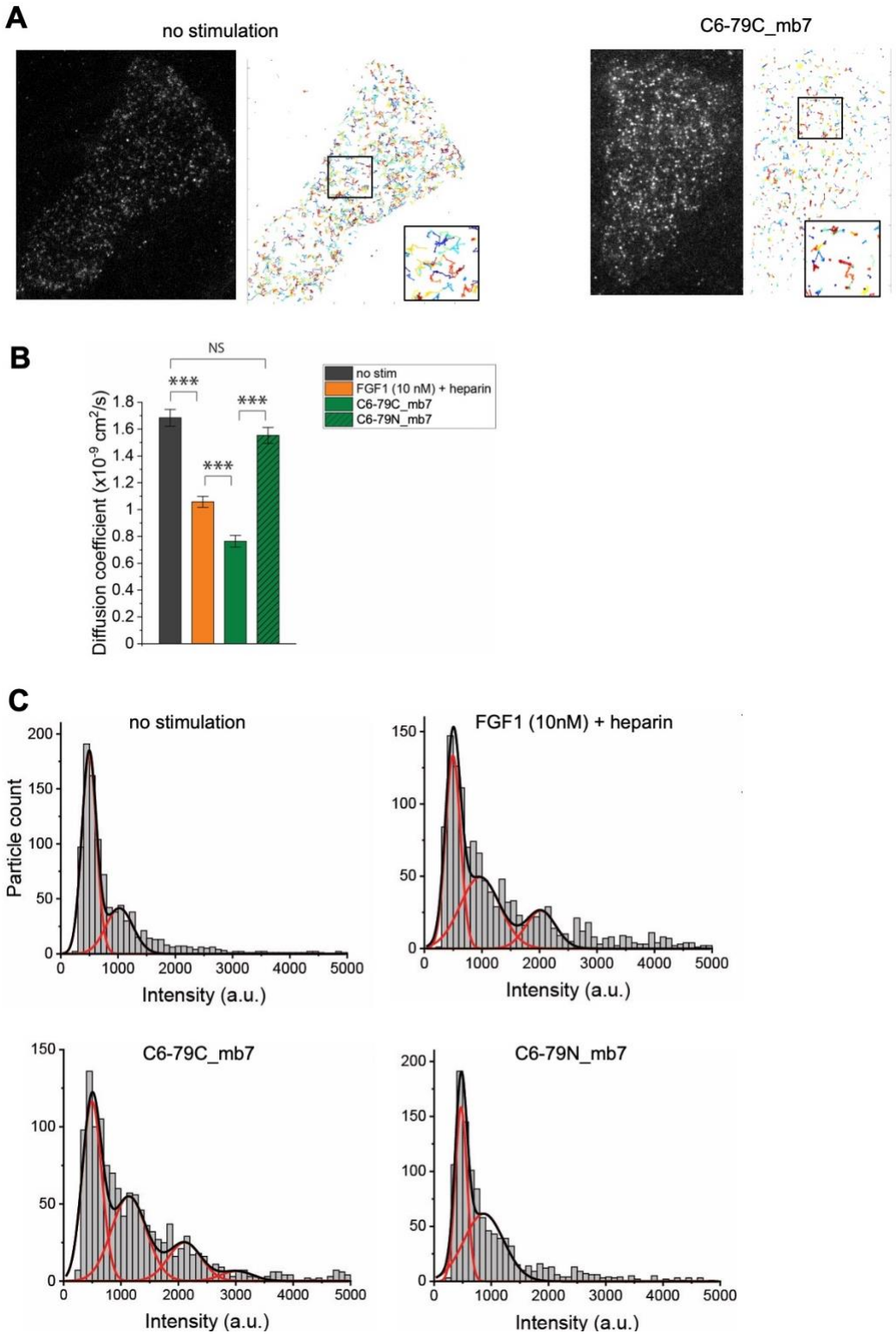

Figure 3.4. Receptor diffusion and distribution in response to C6-79 mb7 scaffolds. Treatment with 100nM C6-79 mb7 scaffolds. (**A**) Single-molecule imaging of HaloTag-FGFR1c in heparan-deficient CHO cells expressing human FGFR1c using TIRF microscopy. (**B**) Calculated diffusion coefficient of FGFR1C. (**C**) Intensity distribution of HaloTag-FGFR1c detected spots.

Chapter 4: Discussion

In this work, I present a generalizable and modular platform for studying receptor clustering using entirely *de novo* protein scaffolds. *De novo* oligomers including tetramers, hexamers, and an octamer are described, with cryo-EM structures containing side-chain resolution; larger versions are generated by adding repeat units; and minibinders are attached by short flexible linkers without disturbing target binding. Expression in E. coli streamlines production without the need for chemical conjugation or assembly steps. Targeting receptors with a *de novo* minibinder scaffold permits exploration of clustering biology without relying on the interactions of the native ligand, for example avoiding the need for heparin as a cofactor in FGF signaling. This platform offers broad applicability for studying the role of receptor clustering in various systems.

The designed scaffolds offer several novel advantages for the study of receptor clustering. Using a cyclic oligomer as a scaffold means that a precise number of receptors can be engaged and colocalized on the membrane. This has been limited by the availability of C4+ oligomers, which are challenging to design due to the decreased angle between subunits leading to flexibility with multiple oligomeric states.[13,14] In fact, the octamer presented here will represent only the 102nd deposition of a C8 structure in the PDB to date. While the oligomers represent a significant expansion in the number of *de novo* designed oligomers available, the ~10% success rate in design highlights the need for further work on this challenging problem.

Using the designed protein mb7 allowed us to explore novel biology beyond that offered by natural ligands for FGFR. Multimerizing mb7 on an oligomeric scaffold activated FGFR signaling pathways, converting an antagonist into an agonist simply by clustering the receptor. Notably, this occurred in absence of the coreceptor heparan sulfate, which is normally required for

signaling through FGF1. In the case of fibroblast growth factors, the requirement for heparin to stabilize FGF has been a major limitation in clinical development of FGFs, which have shown promise in various wound healing applications.[15] A ligand that can stimulate FGFR signaling without co-administration of heparin would thus have immediate therapeutic relevance in regenerative medicine. This shows the potential of designed ligands to go beyond limitations of naturally occurring ligands for the receptor.

Activation using the designed scaffolds may be specific to the C isoforms of FGFRs, which show predominantly mesenchymal expression and for which mb7 shows high selectivity; mb7 also blocks interaction of FGFR with Klotho proteins to selectively antagonize endocrine FGF signaling.[39](Sangwon Lee, personal communication, 2022-02-13) Klotho proteins act as molecular "zip codes" to facilitate interaction of endocrine FGFs with FGFRs.[40,41] It is possible that C6-79C_mb7 may agonize FGFR signaling preferentially in mesenchymal cells while blocking activity of endocrine FGFs, which would enhance targeting to specific tissues/cell types or bias downstream signaling.

FGFRs are not the only receptor tyrosine kinases that require a cofactor or coreceptor such as heparan or Klotho proteins. Examining the roles of these coreceptors is particularly difficult as the receptor ligand may have weak to no interaction with the coreceptor. In this case, designed ligands offer promise as they can be targeted to the coreceptor specifically without binding the main receptor. This opens up the possibility to engage both receptor and coreceptor with the same scaffold, by attaching designed ligands targeting each component individually. Such multivalent scaffolds will permit more detailed analysis of the composition of receptor clusters and their downstream implications.

Chapter 5: Materials and Methods

5.1 Scaffold selection and cyclic docking

Subunit scaffolds consisted of a set of 18 monomeric designed repeat proteins with high-resolution crystal structures or SAXS data (Brunette et al. 2015 and Brunette et al. 2020). PDB IDs for building blocks are provided in **Table 1**. Docking was performed as previously described [Fallas et al. 2017]. Briefly, the protocol aligns subunits along the desired symmetry axis and scores these using a residue pair-motif database derived from PDB structures. This resulted in 829 outputs among all 5 symmetries attempted (C4, C5, C6, C7, and C8). Outputs were then fed into FastDesign to generate an oligomeric interface. These design outputs were filtered by ΔΔG (between -35 and -70), solvent accessible surface area (SASA > 1400), shape complementarity (sc > 0.65), and fewer than 2 unsatisfied hydrogen bonds. This resulted in 150 outputs. Docking for alternative symmetries was performed as above without the interface design step.

5.2 Expression and purification

Sequences of the designed proteins were reverse translated with optimization for Escherichia coli expression, with a C-terminal glycine-serine linker followed by a 6x histidine tag. Sequences were ordered as synthetic genes from Integrated DNA Technologies within the pET29b+ vector between NdeI and XhoI cloning sites. This vector contains a kanamycin resistance marker and a T7 promoter. Plasmids were transformed into E. coli BL21 (DL3) competent cells and plated on LB with kanamycin 50mg/L. Transformants were inoculated into 50mL of autoinduction expression media (for 1L: 12g tryptone, 24g yeast extract, 20 mL 50×M, 20 mL 50x5052, 2 mL 1M MgSO4, 200 µL Studier Trace metals, 100 µg kanamycin, q.s. to 1L with

filtered water) in a 250mL flask. Expression cultures were grown for 20 hours at 37˚C with 200rpm shaking. Cells were pelleted by centrifugation at 4000xg and resuspended in a lysis buffer consisting of 25mM Tris pH 8, 300mM NaCl, and 20mM imidazole with added protease inhibitor and DNase. Cells were lysed by sonication at 85% amplitude with 8 x 15 second pulses. Lysate was separated into soluble and insoluble fractions by centrifugation at 18,000xg. Immobilized metal affinity chromatography (IMAC) was used to purify designed protein. Nickel-nitrilotriacetic acid (Ni-NTA) resin was initially equilibrated with 5 column volumes (CV) lysis buffer. Supernatant was poured over the columns, followed by 20CV wash buffer (25mM Tris pH 8, 400mM NaCl, 30mM imidazole). Protein was eluted using 5CV elution buffer (25mM Tris pH 8, 300mM NaCl, 500mM imidazole). Eluate was purified by size exclusion chromatography (SEC) on an AKTA PURE FPLC system using an S6 column with Tris-buffered saline (TBS; 25mM Tris pH 8, 150mM NaCl) at a speed of 0.75mL/min. Fractions corresponding to the peak trace were collected and combined for further analysis.

5.3 Size exclusion chromatography with multi-angle light scattering (SEC-MALS)

Samples were run in TBS (50mM Tris-HCL, 150mM NaCl pH 8.0) at 1mL/min over a Superose 6 10/300 column using an Agilent 1260 HPLC. The HPLC is in line with a Heleos multi-angle static light scattering and Optilab T-rEX detector (Wyatt Technology Co.). Using ASTRA (Wyatt Technology Co.), a weighted average molecular weight (Mw) and number average molar mass (Mn) were calculated to determine monodispersity-by-polydispersity index (PDI), with PDI = Mw/Mn.

5.4 Small-angle X-ray scattering

SEC-purified samples were prepared for small-angle X-ray scattering (SAXS) by concentrating (if needed) with a 10K molecular weight cut-off spin concentrator followed by filtration with a 0.22µm spin filter. Samples were sent in low (1mg/mL) and high (3-5mg/mL) concentrations in TBS, with flow-through from concentrators used as blanks for later buffer subtraction during data analysis. Scattering data were collected at the SIBYLS High Throughput SAXS Advanced Light Source in Berkeley, California and analyzed with FrameSlice and ScÅtter software package.[23]. Experimental data were compared to design model predictions using the FOXS server (https://modbase.compbio.ucsf.edu/foxs/).[24] For samples with clear deviation from the design model and indication of off-target symmetry by SEC-MALS and electron microscopy, data were additionally compared to a theoretical design model of the off-target symmetry.

5.5 Negative stain EM grid preparation, data collection, and data processing

Proteins were diluted to 20 µg/ml with TBS, then immediately applied to freshly glow-discharged Formvar/carbon 400 mesh copper grids (Ted Pella catalog #01754-F). After incubation for 45s, excess protein solution was removed by blotting from the side with filter paper, then grids were inverted onto two successive drops of sample buffer followed by three to five successive drops of 2% uranyl formate, with excess solution removed by blotting after each application. The final stain applied was incubated for 15s before blotting. Air-dried grids were imaged using a FEI Talos L120C TEM equipped with a 4K × 4K Gatan OneView camera, at a nominal magnification of 73,000x and pixel size of 2.0 Å. Micrographs were imported to Relion 3.1[42] and/or cryoSPARC v2[43] and, after picking using automated protocols in each program, particles were subjected to 2D

classification. Design model projections were generated using EMAN2[44] and Relion, and projections were aligned with experimental 2D class averages using Sparx[45].

5.6 Cryo-EM grid preparation and data collection

All grids were plunge-frozen into liquid ethane using a Vitrobot Mark IV with a chamber maintained at 100% humidity and 22C. Prior to plunge-freezing, 3.5 μL of each design at 0.1 - 1.0 mg/ml was applied to freshly glow-discharged grids of the following types: QUANTIFOIL® R 1.2/1.3 on Cu 400 mesh grids (C6-79), QUANTIFOIL® R 1.2/1.3 on Cu 400 mesh grids + graphene oxide (C8-71; Electron Microscopy Sciences cat. #GOQ400R1213Cu), and/or QUANTIFOIL® R 2/2 on Cu 300 mesh grids + 2 nm C (C4-71 and extensions, C4-81, C8-71 and extensions, C6-71 and extensions, and C4-131). All grids were first screened at NYU on a Talos Arctica microscope operated at 200 kV with a Gatan K3 camera. Larger datasets were acquired for C4-71, C4-81, C6-79, and C8-71 on a Titan Krios microscope operated at 300 kV with a Gatan K3 camera and BioQuantum energy filter ("Krios 6" at the New York Structural Biology Center). In both imaging setups, data acquisition was controlled via Leginon[46] and pre-processing (including motion correction and binning, as applicable) was performed with MotionCor2 as integrated in Appion[47,48]. Further data collection parameters are shown in **Tables II-IV**.

5.7 Processing of 200 kV cryo-EM screening datasets

These datasets include C4-71 extensions, C6-71 and extensions, C8-71 extensions, and C4_nat_HF-7900. Aligned, dose-weighted micrographs and STAR files for particles picked "on the fly" with Warp4 were imported to cryoSPARC v.3 for CTF estimation[49], particle picking, 2D classification, and 3D classification/refinement. 2D classification of particles imported from Warp

was used to identify suitable starting classes for template-based auto-picking. In cases where Warp picking did not yield meaningful templates (or omitted certain particle views), unrepresented particle views were located using manual and/or blob picking and classified in 2D to generate additional templates for auto-picking. Initial maps were generated from 2D-curated particles by ab initio reconstruction in C1 followed by iterative rounds of heterogeneous and homogeneous refinement. Local resolution ranges were estimated using cryoSPARC and the full range of values is reported in Tables III and VII. Global resolution (using independent half-maps from refinement and FSC = 0.143 threshold) and sphericity were estimated using the 3DFSC server (https://3dfsc.salk.edu/)[50]. The FSC mask automatically tightened during the final round of homogenous refinement in cryoSPARC (volume file output with suffix "mask_fsc_auto.mrc") was used for resolution and sphericity calculations. Additional processing details for each dataset are shown in **Tables III and VII**.

5.8 Processing of 300 kV cryo-EM datasets

These datasets include C4-71, C4-81, C6-79, C8-71, and C5_HF-3921. Detailed processing workflows are shown in **Figures 2.4 and 2.7**. For all 300 kV datasets, aligned and dose-weighted micrographs were imported to cryoSPARC v.2/v.3 for CTF estimation, particle picking, 2D classification, and initial 3D curation and refinement. Templates for C6-79 auto-picking were generated using cryoSPARC's "blob picker". For C4-81, C4-71, and C8-71, 2D averages generated from cryo-EM pre-screening data were used for initial template-based auto-picking. Template-based auto-picking of the C8-71 ring was dominated by side views; to retain top views during auto-picking and curation, these views were picked separately using a single auto-picking template generated from 2D classification of manually-picked top views. For C1-41-8-1 and C8-

71, curated particles from template picking were used as a training set for Topaz[51] picking within cryoSPARC. The final sets of curated particles from cryoSPARC were imported to Relion v.3[42] for further 2D/3D classification and 3D refinement, which improved map quality for C4-71 and C4-81. Final 3D refinements were performed with the highest expected symmetry imposed, as well as in C1 and with lower-order symmetries imposed. For all datasets, imposing the highest designed circular symmetry improved map quality without introducing substantial artifacts. Global resolution (using independent half-maps from refinement and FSC = 0.143 threshold) and sphericity were estimated using the 3DFSC server. The FSC mask automatically tightened during the final round of homogenous refinement in cryoSPARC was used for resolution and sphericity calculations for C6-79 and C8-71. For C4-71 and C4-81 maps refined in Relion, soft masks were generated using relion_mask_create for resolution and sphericity calculations.

5.9 C6-79 and C8-71 model building and refinement

De novo designed model coordinates for the C8-71 octamer were first docked into the cryo-EM map as a single rigid body using UCSF Chimera[52]. Initial fitting for C6-79 was performed in Chimera with six copies of the designed monomer manually placed into the map, fit as six individual rigid bodies, and merged into a single set of coordinates for the hexamer. In PHENIX v.1.16[53], docked coordinates were stripped of hydrogens using phenix.pdbtools and refined in real space using iterative rounds of phenix.real_space_refine and manual model adjustment in COOT[54,55]. For C8-71, a single instance of simulated annealing was performed at the beginning of automated refinement in PHENIX; non-crystallographic symmetry, secondary structure, Ramachandran, and rotamer restraints were enabled throughout. Additional density is present in the C8-71 cryo-EM maps near W113, at the interface between subunits of the octamer. In addition

to the C8 map used for model refinement, this density is also visible at comparable thresholds in C1 and C4 maps refined from the same particles. As no obvious candidate molecule could be identified for this density, it was left unmodelled.

For C6-79, rigid-body refinement and a single instance of simulated annealing were used in early rounds of automated real-space refinement. Non-crystallographic symmetry, secondary structure, Ramachandran, and rotamer restraints were enabled throughout refinement. Additionally, the final round of phenix.real_space_refine included ADP refinement and reference model restraints (using the starting model as a reference to restrain residues 46-53 to manually-adjusted positions and strictly match rotamers). Additional model statistics are shown in **Table V**.

5.10 Low-endotoxin protein production

Genes were expressed as described above. Cultures were resuspended and lysed in a phospho-buffered saline (PBS)-based lysis buffer with added protease inhibitor and DNase. Cells were sonicated and pelleted as described above. Supernatant was filtered through a 0.45µm filter prior to loading onto IMAC columns. IMAC columns were pre-washed with PBS + 1% Triton X-100 + 0.75% CHAPS to remove any residual endotoxin and equilibrated with PBS + 5 mM imidazole. Supernatant was poured onto the column and followed by washing with 5CV PBS + 30 mM imidazole. To remove endotoxin, 4 wash steps were performed using 5CV PBS + 1% Triton X-100 + 0.75% CHAPS, with 30min 37C incubations on the first and third wash. This was followed by 2 washes with 10CV PBS, then elution with 5CV PBS + 400 mM imidazole. SEC was performed as described above on a dedicated AKTA PURE FPLC with lines, loops, and fraction dispenser pre-washed using 500mM NaOH + 0.75% CHAPS. Endotoxin levels were measured with the LAL endotoxin testing system (Charles River Laboratories).

5.11 Cell culture

Human umbilical vein endothelial cells (HUVECs) were obtained from Lonza, Germany (#C2519AS). Cells were grown in EGM2 media (20% fetal bovine serum, 1% penicillin-streptomycin, 1% Glutamax [Gibco #35050061], 1% endothelial cell growth factor, 1mM sodium pyruvate, 7.5mM HEPES, 0.08mg/mL heparin, 0.01% amphotericin B, a mixture of 1 x RPMI 1640 +/- glucose for a final concentration of 5.6mM glucose; filtered through 0.45um) on 0.1% gelatin-coated 35mm cell culture dishes. Cells were cryopreserved at passage 4 for later thawing and use in Western blots.

CHO cells were a gift from the Schlessinger Lab (Yale University). Parental heparan-deficient Chinese hamster ovary (CHO) cells (pgsD-677 cells; ATCC, #CRL-2244,) or pgsD-677 cells stably transfected with human FGFR1c were maintained in F-12K medium (ATCC, #30-2004) containing 10% (vol/vol) fetal bovine serum (FBS; Gibco, #16140071) with or without 10 μg/mL puromycin (InvivoGen, #ant-pr-1) for selection in 100-mm treated dishes (Corning, #430167) at 37°C and 5% CO_2. Cells were split every 2 d at a 1:4 ratio by dissociation with TrypLE Express Enzyme (Gibco, #12605010).

5.12 Treatment and protein isolation for Western blot

Cells were passaged at 80,000 cells/well onto a 12-well plate and grown to 80% confluence. Media was aspirated and 500uL of new media was added. Cells were treated with designed scaffold at 1, 10, or 100nM and FGF at 1 or 10nM for 15 minutes or 1 hour. Concentration is reported as the concentration of the mb7 domain, not the oligomeric particle; therefore, 100nM of C4-71N_mb7 corresponds to 100nM of mb7 and 25nM of C4 oligomers.

Following treatment, the media was aspirated and cells were rinsed with PBS. Cells were lysed in tubes with 130 uL of lysis buffer, consisting of 20mM Tris-HCl (pH=8), 150mM NaCl, 15% w/v glycerol, 1% w/v Triton, 3% w/v SDS, 25mM beta-glycerophosphate, 50mM NaF, 10mM sodium pyrophosphate, 0.5% orthovanadate, 1% PMSF, 25U Benzonase Nuclease (EMD Chemicals, Gibbstown, NJ), protease inhibitor tablet (Thermo Scientific), and phosphatase inhibitor cocktail 2 (#P5726) (all chemicals from Sigma-Aldrich unless otherwise specified). Cell lysate was collected into a fresh tube; 43.33uL of 4x Laemmli sample buffer with 10% beta-mercaptoethanol was added and cells were heated at 95C for 10 minutes. Samples were either used immediately for Western blot analysis or stored at -80C.

5.13 Western blotting

If frozen, protein samples were thawed and heated at 95C for 10 minutes. A 10% SDS-PAGE gel was loaded with 30uL of protein per sample well and run for 30 minutes at 250V. Proteins were then transferred onto a nitrocellulose membrane for 12 minutes, using the semi-dry turbo transfer Western blot apparatus (Bio-Rad); and the membrane was blocked in 5% bovine serum albumin for 1 hour. The membrane was then probed with antibodies for pERK1/2 (1:5000 dilution; Cell Signaling) and beta-actin (1:10,000 dilution; Cell Signaling), with incubation on a rocker at 4C overnight. Membranes were washed 3 times with 1x TBST at 10 minute intervals, then incubated with the respective HRP-conjugated secondary antibody (1:10,000 dilution in 5% milk; Bio-Rad) at room temperature for 1 hour. Membranes were blocked in 5% milk at room temperature for 1 hour and washed 3 times with 1x TBST at 10 minute intervals. Membranes were developed using the Chemiluminescence developer and imaged with the Bio-Rad ChemiDoc Imager.

5.14 Calcium release assay

pgsD-677-FGFR1c cells were plated to 50% confluence on uncoated 35mm glass-bottom dishes (MatTek Corporation, #P35G-1.5-14-C) in phenol-red free DMEM (Gibco, #21063029) supplemented with 4.5 g/L glucose and 10% (vol/vol) FBS (FBS; Gibco, #16140071) and 10 μg/mL puromycin (InvivoGen, #ant-pr-1). The following day, cells were starved for 2-3 hours in serum-free media. The Calbryte 520 AM calcium indicator (AAT Bioquest, Product number 20651) was dissolved in DMSO and added to serum-free media to a final concentration of 5.7μM. After incubation with the indicator for 30 min at 37°C and 5% CO_2, cells were washed 3x with serum free media and imaged with a 10x objective microscope. The imaging protocol consisted of filter cube with excitation/emission filters for GFP fluorescence, with an exposure time of 100-200ms, acquisition rate of 1 frame/5sec, and 20min recording duration (5min baseline recording + 15min ligand treatment time). In the movies, Calbryte fluorescence was measured by subtracting the mean fluorescence of a local background region from that of the cell using ImageJ (NIH).

5.15 TIRF microscopy

For single-molecule imaging experiments, pgsD-677 cells were plated on 35-mm glass-bottom dishes (MatTek Corporation,#P35G-1.5-14-C) to 75% confluence in phenol-red free DMEM (Gibco, #21063029) supplemented with 4.5 g/L glucose and 10% (vol/vol) FBS (FBS; Gibco, #16140071) and transfected with 0.25 μg HaloTag-FGFR1c plasmid the next day using Lipofectamine 3000 reagent (Invitrogen, #L3000001), according to the manufacturer's instructions. The following day, cells were starved for 2-3 hours in serum-free media, labeled

with 0.25 μM cell-impermeant Alexa488 HaloTag ligand (Promega, #G1001) for 15 min at 37°C and 5% CO_2, and then washed 3x with phenol-red free media. After labeling, cells were immediately imaged at 37°C and 5% CO2 in a cage incubator (OkoLab) housing a Nikon Eclipse Ti2 microscope (Nikon) equipped with a motorized Ti-LA-HTIRF module with a 15-mW LU-N4 488 laser, using a CFI Plan Apochromat Lambda 100x/1.45 Oil TIRF objective and a Prime95B cMOS camera (110-nm pixel size; Teledyne Photometrics). Images were acquired using a 100-ms exposure time at 10 Hz with the laser power set at 100%. The penetration depth of the evanescent field was ~118 nm.

5.16 Single-particle tracking

Particles were localized and tracked using the MATLAB software GaussStorm.[56,57] Briefly, particles were automatically detected by application of a bandpass filter to remove noise, followed by convolution with a Gaussian kernel, and then the selection of above-threshold pixels. Particles were then fitted with elliptical two-dimensional Gaussian functions, which yielded their intensities expressed as the volume under the curve, as well as their positions with subpixel accuracy. Particles were tracked frame to frame using a tracking algorithm with a tracking window of 7 pixels between consecutive frames. The distribution of the displacements of single particles was used to calculate mean diffusion coefficient in a field of view encompassing an entire cell.

References

1. Karpov OA et al. Receptor tyrosine kinase structure and function in health and disease. *AIMS Biophysics* 2015;2:476-502.
2. Lemmon MA, Schlessinger J. Cell signaling by receptor tyrosine kinases. *Cell* 2010;141:1117-1134.
3. Westerfield JM, Barrera FN. Membrane receptor activation mechanisms and transmembrane peptide tools to elucidate them. *Journal of Biological Chemistry* 2020;295:1792-1814.
4. Zhang K, Gao H, Deng R, Li J. Emerging Applications of Nanotechnology for Controlling Cell-Surface Receptor Clustering. *Angewandte Chemie - International Edition* 2019;58:4790-4799.
5. Wu H. Higher-order assemblies in a new paradigm of signal transduction. *Cell* 2013;153:287-292.
6. Garcia-Parajo MF, Cambi A, Torreno-Pina JA, Thompson N, Jacobson K. Nanoclustering as a dominant feature of plasma membrane organization. *Journal of Cell Science* 2014;127:4995-5005.
7. Mayer BJ, Yu J. Protein Clusters in Phosphotyrosine Signal Transduction. *Journal of Molecular Biology* 2018;430:4547-4556.
8. Nuñez-Prado N et al. The coming of age of engineered multivalent antibodies. *Drug Discovery Today* 2015;20:588-594.
9. Zhao YT et al. F-domain valency determines outcome of signaling through the angiopoietin pathway. *EMBO Reports* 2021;22:e53471.
10. Bryan CM et al. Computational design of a synthetic PD-1 agonist. *Proceedings of the National Academy of Sciences* 2021;118:e2102164118.
11. Ben-Sasson AJ et al. Design of biologically active binary protein 2D materials. *Nature* January 2021.
12. Divine R et al. Designed proteins assemble antibodies into modular nanocages. *Science* 2021;372:eabd9994.
13. Fallas JA et al. Computational design of self-assembling cyclic protein homo-oligomers. *Nature Chemistry* 2017;9:353-360.
14. Xu C et al. Computational design of transmembrane pores. *Nature* 2020;585:129-134.
15. Beenken A, Mohammadi M. The FGF family: Biology, pathophysiology and therapy. *Nature Reviews Drug Discovery* 2009;8:235-253.
16. Mohan K et al. Topological control of cytokine receptor signaling induces differential effects in hematopoiesis. *Science* 2019;364:eaav7532.
17. Chevalier A et al. Massively parallel de novo protein design for targeted therapeutics. *Nature* 2017;550:74-79.
18. Cao L et al. De novo design of picomolar SARS-CoV-2 miniprotein inhibitors. *Science* 2020;370:426-431.
19. Cao L et al. Robust de novo design of protein binding proteins from target structural information alone. *bioRxiv* September 2021:2021.09.04.459002.
20. Hunt AC et al. Multivalent designed proteins protect against SARS-CoV-2 variants of concern. *bioRxiv* 2021.
21. Brunette T et al. Exploring the repeat protein universe through computational protein

design. *Nature* 2015;528:580-584.
22. Brunette T, Bick MJ, Hansen JM, Chow CM, Kollman JM, Baker D. Modular repeat protein sculpting using rigid helical junctions. *Proceedings of the National Academy of Sciences* 2020;117:8870-8875.
23. Dyer KN et al. *High-Throughput SAXS for the Characterization of Biomolecules in Solution: A Practical Approach*. Vol 1091.; 2014.
24. Schneidman-Duhovny D, Hammel M, Sali A. FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Research* 2010;38:W540-W544.
25. Mukherjee S, Zhang Y. MM-align: a quick algorithm for aligning multiple-chain protein complex structures using iterative dynamic programming. *Nucleic Acids Research* 2009;37:e83-e83.
26. Leaver-Fay A et al. Rosetta3: An Object-Oriented Software Suite for the Simulation and Design of Macromolecules. *Methods in Enzymology* 2011;487:545-574.
27. Coventry B, Baker D. Protein sequence optimization with a pairwise decomposable penalty for buried unsatisfied hydrogen bonds. *PLOS Computational Biology* 2021;17:e1008061.
28. Hsia Y et al. Design of a hyperstable 60-subunit protein icosahedron. *Nature* 2016;535:136-139.
29. Geiger-Schuller K, Sforza K, Yuhas M, Parmeggiani F, Baker D, Barrick D. Extreme stability in de novo-designed repeat arrays is determined by unusually stable short-range interactions. *Proceedings of the National Academy of Sciences of the United States of America* 2018;115:7539-7544.
30. Sahtoe DD et al. Reconfigurable asymmetric protein assemblies through implicit negative design. *Science* 2022;375:eabj7662.
31. Turner N, Grose R. Fibroblast growth factor signalling: From development to cancer. *Nature Reviews Cancer* 2010;10:116-129.
32. Ornitz DM, Itoh N. The Fibroblast Growth Factor signaling pathway. *WIREs Developmental Biology* 2015;4:215-266.
33. Ferguson HR, Smith MP, Francavilla C. Fibroblast Growth Factor Receptors (FGFRs) and Noncanonical Partners in Cancer Signaling. *Cells* 2021;10:1-35.
34. Brown A, Robinson CJ, Gallagher JT, Blundell TL. Cooperative Heparin-Mediated Oligomerization of Fibroblast Growth Factor-1 (FGF1) Precedes Recruitment of FGFR2 to Ternary Complexes. *Biophysical Journal* 2013;104:1720-1730.
35. Presta M, Dell'Era P, Mitola S, Moroni E, Ronca R, Rusnati M. Fibroblast growth factor/fibroblast growth factor receptor system in angiogenesis. *Cytokine & Growth Factor Reviews* 2005;16:159-178.
36. Plotnikov AN, Hubbard SR, Schlessinger J, Mohammadi M. Crystal structures of two FGF-FGFR complexes reveal the determinants of ligand-receptor specificity. *Cell* 2000;101:413-424.
37. Nunes QM, Li Y, Sun C, Kinnunen TK, Fernig DG. Fibroblast growth factors as tissue repair and regeneration therapeutics. *PeerJ* 2016;2016:1-31.
38. Los G V. et al. HaloTag: A novel protein labeling technology for cell imaging and protein analysis. *ACS Chemical Biology* 2008;3:373-382.
39. Gong SG. Isoforms of Receptors of Fibroblast Growth Factors. *Journal of Cellular Physiology* 2014;229:1887-1895.
40. Lee S et al. Structures of β-klotho reveal a 'zip code'-like mechanism for endocrine FGF

signalling. *Nature* 2018;553:501-505.
41. Suzuki Y et al. FGF23 contains two distinct high-affinity binding sites enabling bivalent interactions with α-Klotho. *Proceedings of the National Academy of Sciences* 2020;117:31800-31807.
42. Zivanov J et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* 2018;7.
43. Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods* 2017;14:290-296.
44. Bell JM, Chen M, Durmaz T, Fluty AC, Ludtke SJ. New software tools in EMAN2 inspired by EMDatabank map challenge. *Journal of Structural Biology* 2018;204:283-290.
45. Hohn M et al. SPARX, a new environment for Cryo-EM image processing. *Journal of Structural Biology* 2007;157:47-55.
46. Suloway C et al. Automated molecular microscopy: the new Leginon system. *Journal of Structural Biology* 2005;151:41-60.
47. Lander GC et al. Appion: an integrated, database-driven pipeline to facilitate EM image processing. *Journal of Structural Biology* 2009;166:95-102.
48. Zheng SQ, Palovcak E, Armache J-P, Verba KA, Cheng Y, Agard DA. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature Methods* 2017;14:331-332.
49. Rohou A, Grigorieff N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology* 2015;192:216-221.
50. Tan YZ et al. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nature Methods* 2017;14:793-796.
51. Bepler T et al. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nature Methods* 2019;16:1153-1160.
52. Pettersen EF et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 2004;25:1605-1612.
53. Adams PD et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *International Tables for Crystallography* 2012:539-547.
54. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 2010;66:486-501.
55. Echols N et al. Graphical tools for macromolecular crystallography in PHENIX. *J Appl Crystallogr* 2012;45:581-586.
56. Holden SJ et al. Defining the Limits of Single-Molecule FRET Resolution in TIRF Microscopy. *Biophysical Journal* 2010;99:3102-3111.
57. Fontana M, Fijen C, Lemay SG, Mathwig K, Hohlbein J. High-throughput, non-equilibrium studies of single biomolecules using glass-made nanofluidic devices. *Lab on a Chip* 2018;19:79-86.

Tables

Table I. Designed monomeric repeat proteins used as building blocks.

Name	PDB ID	Reference
DHR4	5CWB	Brunette et al. 2015
DHR5	5CWC	Brunette et al. 2015
DHR7	5CWD	Brunette et al. 2015
DHR8	5CWF	Brunette et al. 2015
DHR10	5CWG	Brunette et al. 2015
DHR14	5CWH	Brunette et al. 2015
DHR18	5CWI	Brunette et al. 2015
DHR49	5CWJ	Brunette et al. 2015
DHR53	5CWK	Brunette et al. 2015
DHR54	5CWL	Brunette et al. 2015
DHR64	5CWM	Brunette et al. 2015
DHR71	5CWN	Brunette et al. 2015
DHR76	5CWO	Brunette et al. 2015
DHR79	5CWP	Brunette et al. 2015
DHR80		Brunette et al. 2015
DHR81	5CWQ	Brunette et al. 2015
TJ116	6W2R	Brunette et al. 2020
TJ120	6W2V	Brunette et al. 2020
TJ121	6W2W	Brunette et al. 2020
TJ131	6W2Q	Brunette et al. 2020

Table II. Structural characterization of docked protein oligomers, stratified by symmetry.

	Ordered	Soluble expression	Monodisperse by SEC	Single, correct oligomeric state	Single, off-target oligomeric state
C4	15	12	6	5	0
C5	16	9	3	0	0
C6	24	12	8	2	0
C7	24	16	5	0	1 (C6)
C8	30	11	6	1	1 (C6)
Total	109	60	28	8	2

Table III. Cryo-EM data acquisition and processing for 300kV datasets.

	C4-71	C4-81	C6-79	C8-71
Image acquisition				
Sample name (Grid ID)	RR0231	RR0233, RR0243	RR0204	RR0201, RR0202
Microscope	Krios	Krios	Krios	Krios
Voltage (kV)	300	300	300	300
Exposure navigation	Image shift	Image shift	Image shift	Image shift
Nominal magnification	105000X	105000X	81000X	81000X
Detector	Gatan K3	Gatan K3	Gatan K3	Gatan K3
Pixel size (Å/pix)	0.4124 (binned 2X)	0.4124 (binned 2X)	1.069	1.069
No. of frames	40	40	50	50
Dose rate (e-/Å/s)	29.23	29.23	24.52	24.52
Exposure per frame (e-/Å^2)	1.169	1.169	1.226	1.226
Total dose (e-/Å^2)	46.77	46.77	61.3	61.3
Exposure time (s)	1.6	1.6	2.5	2.5
Defocus range (μm)	0.3 - 1.3	0.8 - 2.5	0.8 - 2.5	0.8 - 2.5
No. of micrographs	5221	18412	6434	3092
Data processing				
Software used	Appion v.3, Cryosparc v.2/v.3, RELION v.3	Appion v.3, Cryosparc v.2/v.3, RELION v.3	Appion v.3, Cryosparc v.2/v.3	Appion v.3, Cryosparc v.2/v.3
No. of extracted particles	3355984	1878998	3901271	308351
No. of final particles	107483	259010	864566	132391
Box size (pixels)	212	256	220	200
Symmetry Imposed	C4	C4	C6	C8
Global map resolution (Å) (FSC 0.143 unmasked / masked):	5.1 / 4.0	5.4 / 4.9	4.80 / 4.0	4.3 / 3.6
Map resolution range (Å)	4.9 - 9.4	5.0 - 10.1	3.5 - 8.0	3.2 - 7.5

Map sharpening B factor (Å^2)	N/A	N/A	-313.2	-223.3
Sphericity from 3DFSC (unmasked/masked):	0.75 / 0.76	0.86 / 0.88	0.92 / 0.95	0.82 / 0.98

Table IV. Cryo-EM data acquisition and processing for 200kV datasets.

Image acquisition	**C4-71_6x**	**C4-71_8x**	**C6-71**	**C6-71_6x**	**C6-71_8x**	**C8-71_6x**	**C8-71_8x**
Sample name (Grid ID)	RR0301	RR0302	RR0303	RR0304	RR0305	RR0306	RR0328
Microscope	Arctica	Arctica	Arctica	Arctica	Arctica	Arctica	Arctica
Voltage (kV)	200	200	200	200	200	200	200
Exposure navigation	Stage position	Stage position	Stage position	Stage position	Stage position	Stage position	Stage position
Nominal magnification	36,000X	36,000X	36,000X	36,000X	36,000X	36,000X	36,000X
Detector	Gatan K3	Gatan K3	Gatan K3	Gatan K3	Gatan K3	Gatan K3	Gatan K3
Pixel size (Å/pix)	0.548 (binned 2X)	0.548 (binned 2X)	0.548 (binned 2X)	0.548 (binned 2X)	0.548 (binned 2X)	0.548 (binned 2X)	0.548 (binned 2X)
No. of frames	40	40	40	40	40	40	40
Dose rate (e-/Å/s)	20.37	20.37	20.37	20.37	20.28	21.89	21.89
Exposure per frame (e-/Å^2)	1.426	1.426	1.426	1.426	1.419	1.314	1.314
Total dose (e-/Å^2)	57.04	57.04	57.04	57.04	56.77	52.54	52.54
Exposure time (s)	2.8	2.8	2.8	2.8	2.8	2.4	2.4
Defocus range (μm)	2.1 - 2.9	2.1 - 2.9	2.1 - 2.9	2.1 - 2.9	1.1 - 2.9	2.0 - 3.0	2.0 - 3.0
No. of micrographs	121	127	81	87	223	263	96
Data processing							
Software used	Appion v.3, Cryospar c v.3	Appion v.3, Cryospar c v.3	Appion v.3, Cryospar c v.3	Appion v.3, Cryospar c v.3	Appion v.3, Cryospar c v.3	Appion v.3, Cryospar c v.3	Appion v.3, Cryospar c v.3
No. of extracted particles	9276	7265	16793	15332	93046	28236	62332

No. of final particles	4919	2694	4593	7460	23855	9092	3584
Box size (pixels)	200	220	200	220	300	256	256
Symmetry Imposed	C4	C4	C6	C6	C6	C8	C8
Global map resolution (Å, FSC 0.143 unmasked / masked):	9.1 / 6.6	13.4 / 7.8	9.1 / 7.6	8.3 / 6.7	6.7 / 4.1	7.4 / 4.7	7.8 / 5.0
Map resolution range (Å)	4.4 - 13.9	5.8 - 14.7	5.0 - 15.0	4.1 - 14.0	2.5 - 12.5	2.9 - 14.3	3.8 - 14.7
Map sharpening B factor ($Å^2$)	N/A	N/A	N/A	N/A	-166.2	N/A	N/A
Sphericity from 3DFSC (unmasked/masked):	0.59 / 0.88	0.95 / 0.85	0.66 / 0.82	0.50 / 0.93	0.78 / 0.95	0.72 / 0.78	0.45 / 0.87

Table V. Model statistics for C6-79 and C8-71 cryo-EM structures.

	C6-79	**C8-71**
Map CC (mask)	0.814	0.824
Map CC (volume)	0.808	0.813
Map CC (peaks)	0.709	0.696
R.m.s. deviations (bonds)	0.005	0.006
R.m.s. deviations (angles)	0.728	0.711
Ramachandran plot values (%)		
outliers	0.00	0.00
allowed	1.38	2.35
favored	98.62	97.65
Rotamer outliers (%)	0.00	0.00
C-beta deviations (%)	0.00	0.00
CaBLAM outliers (%)	0.93	1.03
Overall score (Molprobity[16])	1.75	1.87
Clashscore	17.80	20.00

Table VI. Crystal structure data collection and refinement statistics for C2-02802 and C2-01018.

	C2-02802	**C2-01018**
Data Collection		
Space group	R 3 :H	P 32 2 1
Cell parameters		
a,b,c (Å)	56.429, 56.429, 96.624	69.31, 69.31, 49.41
α, β, γ (°)	90, 90, 120	90, 90, 120
Resolution (Å)	43.61 - 1.49 (1.54 - 1.49)	38.15 - 3.10 (3.21 - 3.10)
R_{merge} (%)	0.067 (1.185)	0.276 (0.689)
R_{pim} (%)	0.032 (0.605)	0.137 (0.344)
$I/\sigma(I)$	11.88 (0.87)	14.75 (2.89)
$CC_{1/2}$	0.997 (0.469)	0.979 (0.877)
Completeness (%)	99.71 (99.14)	98.49 (97.68)
Redundancy	5.1 (4.8)	4.9 (5.2)
Refinement		
Resolution (Å)	43.61 - 1.49 (1.543 - 1.49)	
No. reflections	18703 (1843)	2615 (253)
R_{work} / R_{free} (%)	19.15 (32.41) /21.17 (33.05)	38.15 - 3.101 (3.21 - 3.10)
No. atoms		
Protein	984	954
Water	46	0
Ramachandran Favored/allowed Outlier (%)	100.00/0.00/0.00	99.10/0.90/0.00
R.m.s. deviations		
Bond lengths (Å)	0.002	0.004
Bond angles (°)	0.37	0.58
$B_{factors}$ (Å^2)		
Protein	43.14	109.12
Water	61.94	n/a

Table VII. Cryo-EM data collection parameters for C4_nat_HF-7900 and C5_HF-3921.

	C4_nat_HF-7900 (6XSS, EMD-22305)	**C5_HF-3921** (EMD-22306)
Microscope	Talos Arctica	Titan Krios
Electron energy	200 kV	300 kV
Pixel size	0.859 Å	1.083 Å
Total electron dose	57.05 $e^{-}/Å^2$	68.61 $e^{-}/Å^2$
Number of frames in each movie	40	50
Exposure time	2800 ms	2500 ms
Defocus range	-0.2 - -4.2 µm	-1.9 - -5.0 µm
Tilt angle(s)	0 °	0, 35 °
Number of images acquired	3,752	6,744
Number of particles used in final map	144,329	30,659
Final map resolution (FSC = 0.143)	3.70	8.06
B-factor for map sharpening	-180 $Å^2$	-500 $Å^2$
Sphericity of 3DFSC	0.895	0.786

Table VIII. Model statistics for C4_nat_HF-7900 cryo-EM structure.

Map CC (mask)	0.78
Map CC (volume)	0.77
Map CC (peaks)	0.63
rmsd (bonds)	0.003 Å
rmsd (angles)	0.605°
Ramachandran plot values	
outliers	0.00%
allowed	2.52%
favored	97.48%
Rotamer outliers	0.00%
C-beta deviations	0.00%
Overall score (Molprobity[16])	2.04
PDB ID	6XSS

Table IX. Sequences of designed oligomers and fusions. Sequences include N-terminal Met +/- Gly or MEKKI expression tag (DNA sequence: atggagaaaaaaatc) and C-terminal or N-terminal 6xHis tags with Gly-Ser linker and/or TEV cleavage site (underlined).

Name	Sequence
C4-18	MGSIEKLCKKAESEAREARSKAEELRQRHPDSQAARDAQKLASQAEEAVKLACELAQEH PNAWIARACIRAASEAAEAASKAAELAQRHPDSKAARDAIKLASQAAEAVKLACELAQEH PNADIAELCILAAWAAARAASLAAELAQRHPDLWAANLAIRLASQAAEAVKLACELAQEH PNAEIARECIWLAWEAALLAALAAEEAQRHPNDIRAMLLFIEAIRKAEEVKKRCERGSLEH HHHHH
C4-717	MGPEEILERAKESLERAREASERGDEEEFRKAAEKALELAKRLVEQAKKEGDPELVLEAA KVALRVAELAAKNGDKEVFKKAAESALEVAKRLVEVASKEGDPELVLEAAKVALRVAELA AKNGDKEVFKKAARSALWVAFILVKVALKEGDPELVEEAAKVAIRVFELAWEQGDEDVLR LALLTMIVVLILLILVLLKKGGWGSLEHHHHHH
C4-71	MGPEEILERARESLERAREASERGDEEEFRKAAEKALELAKRLVEQAKKEGDPWMVMW AALVALWVALLALRNGDKEVFKKAAESALEVAKRLVEVASKEGDPEMVLLAAWVALFVA WLAWLFGDKEVFKKAAESALEVAKRLVEVASKEGDPELVEEAAKVAEEVEKLAEKQGDE EVREKAWETWMEVWLLWLEVRLRKGGGGSLEHHHHHH
C4-80	MGELERESREAEKRLKEARLFAWAARLLGDLKLLAKALIEEARAVQELARVASERGNRD EAWDAFEKALEVFEEAVKVSEEAREQGDDEVLALALIAIALAVLALAEVASSLGISELAELA WKMAEWVLEEARKVSEEAREQGDDEVLALALIAIALAVLALAEVASSRGNKEEAERAYE DARRVEEEARKVKESAEEQGDSEVKRLAEEAEQLAREARRHVQETRGGWLEHGSLEH HHHHH
C4-81	MGELERESREAEKRLKEARLFAWAARLLGDLKLLAKALIEEARAVQELARVACERGNRD EAWDAFEKALEVFEEAVKVSEEAREQGDDEVLALALIAIALAVLALAEVACCLGISELAEL AWKMAEWVLEEARKVSEEAREQGDDEVLALALIAIALAVLALAEVACCRGNKEEAERAY EDARRVEEEARKVKESAEEQGDSEVKRLAEEAEQLAREARRHVQECRGGWLEHGSLE HHHHHH
C4-131	MGLKELLKRAEELAKSPDPEDLKEAVRLAEEVVRERPGSEAAKKALEIIQEAAEKLKKSP DPEAIIAAARALLKIAATTGDNEAAKQAIEAASKAAQLAEQRGDDELVCEALALLIAAQVLL LKQQGVPMLEVAIHVAETILQILQRLKRKGASEEVRKECLKRILREIAEALQRSGVPEEEIA LIMLLIILLLMML GSLEHHHHHH
C6-4	MGDECEKKAREVALRVLVLWAKGTSEDEIAEEVAREISEVIRTLKESGSSYEVICECVARI VAFIVEVLVLMGTSEDEIAEIVARVISEVIRTLKESGSSYEVICKCVAFIVAEIVEALKRAGTS EDEIAEIVARVISEVIRTLKESGSSEDIIWECIMLIMIFIAEALLRSGTSEDEIREILRRVRSEV ERTLKESGSGSLEHHHHHH
C6-9	MGDEAEKKAREVALRVLVLWAKGTSEDEIAEEVAREISEVIRTLKESGSSYEVIAEIVARIV AFIVEVLVLMGTSEDEIAEIVARVISEVIRTLKESGSSYEVIAKIVAFIVAEIVEALKRAGTSE DEIAEIVARVISEVIRTLKESGSSEDIIWEIIMLIMIFIAEALLRSGTSEDEIREILRRVRSEVER TLKESGSGSLEHHHHHH
C6-714	MGPEEILERAKESLERAREASERGDEEEFRKAAEKALELAKRLVEQAKKEGDPELVLEAA KVALRVAELAAKNGDKEVFKKAAESALEVAKRLVEVASKEGDPELVLEAAKVALEVARLA AENGDKEVFKKAAESALEVAAKLVWVAMKEGDPRMVINALMVALWVLLLAFLQGDEEVF ERARTLFELVRNFIEALEMREGGGGSLEHHHHHH
C6-71	MGPEEILERAKESLERAKEAFERGDEEEFRKAAEKALELAKRLVEQAKKEGDPELVFEAA RVALWVAWLAAWFGDKEVFKKAAESALEVAKRLVEVAKEEGDPELVLKAAFVALLVAIM AVILGDKEVFKKAAESALEVAKRLVEIAAREGDPELVEEAAKVAELVRELAKLMGDEEVY EKARETAREVRLFLLFVRIWEGGGGSLEHHHHHH

C6-79	MGSSDEEEARELEERAREAAKRAIEAAKRTGDPRVRELAEELVKLAIWAAVEVWLDPSS SDVNEALKLIVEAIEAAVRALEAAERTGDPEVRELARELVRLAVEAAEEVQRNPSSSDVN EALKLIVIAIEAAVRALEAAERTGDPEVRELARELVRLAVEAAEEVQRNPSSEEVNEALRKI IKLILFAVMVLELAEEIGDPTWREMARRAVREAVELAEEVQRDPSGWLGHGSLEHHHHH H
C8-71	MGPEEILERAKESLERAREASERGDEEEFRKAAEKALELAKRLVEQAKKEGDPELVLEAA RVALWVAELAAKNGDKEVFKKAAESALEVAKRLVEVASKEGDPDLVAWAALVALWVAFL AFLNGDKEVFKKAAESALEVAKALMEVAMKVGAPWLVELAIAVARAVWLLAELFGDEEV RRRAEAFEIILRIAAIAVKAWLGGGGSLEHHHHHH
C4-71-6x	MGPEEILERAKESLERAREASERGDEEEFRKAAEKALELAKRLVEQAKKEGDPELVLEAA KVALRVAELAAKNGDKEVFKKAAESALEVAKRLVEVASKEGDPELVLEAARVALEVARLA AENGDKEVFKKAAESALEVAKRLVEVASKEGDPWMVMWAALVALWVALLALRNGDKEV FKKAAESALEVAKRLVEVASKEGDPEMVLLAAWVALFVAWLAWLFGDKEVFKKAAESAL EVAKRLVEVASKEGDPELVEEAAKVAEEVEKLAEKQGDEEVREKAWETWMEVWLLWLE VRLRKGGGGSLEHHHHHH
C4-71-8x	MGPEEILERAKESLERAREASERGDEEEFRKAAEKALELAKRLVEQAKKEGDPELVLEAA KVALRVAELAAKNGDKEVFKKAAESALEVAKRLVEVASKEGDPELVLEAAKVALRVAELA AKNGDKEVFKKAAESALEVAKRLVEVASKEGDPELVLEAAKVALRVAELAAKNGDKEVF KKAAESALEVAKRLVEVASKEGDPELVLEAARVALEVARLAAENGDKEVFKKAAESALEV AKRLVEVASKEGDPWMVMWAALVALWVALLALRNGDKEVFKKAAESALEVAKRLVEVA SKEGDPEMVLLAAWVALFVAWLAWLFGDKEVFKKAAESALEVAKRLVEVASKEGDPELV EEAAKVAEEVEKLAEKQGDEEVREKAWETWMEVWLLWLEVRLRKGGGGSLEHHHHHH
C6-71-6x	MGPEEILERAKESLERAREASERGDEEEFRKAAEKALELAKRLVEQAKKEGDPELVLEAA KVALRVAELAAKNGDKEVFKKAAESALEVAKRLVEVASKEGDPELVLEAAKVALEVAKLA FENGDKEVFKKAAESALEVAKRLVEVASKEGDPELVFEAARVALWVAWLAAWFGDKEV FKKAAESALEVAKRLVEVAKEEGDPELVLKAAFVALLVAIMAVILGDKEVFKKAAESALEV AKRLVEIAAREGDPELVEEAAKVAELVRELAKLMGDEEVYEKARETAREVRLFLLFVRIW EGGGGSLEHHHHHH
C6-71-8x	MGPEEILERAKESLERAREASERGDEEEFRKAAEKALELAKRLVEQAKKEGDPELVLEAA KVALRVAELAAKNGDKEVFKKAAESALEVAKRLVEVASKEGDPELVLEAAKVALRVAELA AKNGDKEVFKKAAESALEVAKRLVEVASKEGDPELVLEAAKVALRVAELAAKNGDKEVF KKAAESALEVAKRLVEVASKEGDPELVLEAAKVALEVAKLAFENGDKEVFKKAAESALEV AKRLVEVASKEGDPELVFEAARVALWVAWLAAWFGDKEVFKKAAESALEVAKRLVEVAK EEGDPELVLKAAFVALLVAIMAVILGDKEVFKKAAESALEVAKRLVEIAAREGDPELVEEA AKVAELVRELAKLMGDEEVYEKARETAREVRLFLLFVRIWEGGGGSLEHHHHHH
C8-71-6x	MGPEEILERAKESLERAREASERGDEEEFRKAAEKALELAKRLVEQAKKEGDPELVLEAA KVALRVAELAAKNGDKEVFKKAAESALEVAKRLVEVASKEGDPELVLEAAKVALEVARLA AENGDKEVFKKAAESALEVAKRLVEVASKEGDPELVLEAARVALWVAELAAKNGDKEVF KKAAESALEVAKRLVEVASKEGDPDLVAWAALVALWVAFLAFLNGDKEVFKKAAESALE VAKALMEVAMKVGAPWLVELAIAVARAVWLLAELFGDEEVRRRAEAFEIILRIAAIAVKAW LGGGGSLEHHHHHH
C8-71-8x	MGPEEILERAKESLERAREASERGDEEEFRKAAEKALELAKRLVEQAKKEGDPELVLEAA KVALRVAELAAKNGDKEVFKKAAESALEVAKRLVEVASKEGDPELVLEAAKVALRVAELA AKNGDKEVFKKAAESALEVAKRLVEVASKEGDPELVLEAAKVALRVAELAAKNGDKEVF KKAAESALEVAKRLVEVASKEGDPELVLEAAKVALEVARLAAENGDKEVFKKAAESALEV AKRLVEVASKEGDPELVLEAARVALWVAELAAKNGDKEVFKKAAESALEVAKRLVEVAS KEGDPDLVAWAALVALWVAFLAFLNGDKEVFKKAAESALEVAKALMEVAMKVGAPWLV ELAIAVARAVWLLAELFGDEEVRRRAEAFEIILRIAAIAVKAWLGGGGSLEHHHHHH
C2-58H	MGDEELLRELLKLLLKLLEQMGDEEARRVVEELREELEKKGDPRALVLAFALVILVFLLRIL RELGDEELVRRVEELWEELLKEGDPQAMMEVFKLVQELQERRGSLEHHHHHH

C2-Y2D	MGGEELLQEVARVLLKLAQELGDPDVERVVRELLERLERKGDPRIVIRILLLLVALLLLWIA RELGDPEVVRELEELLKRLIKKGDPRLFAEILRIVLELEEEVGGSLEHHHHHH
C2-CDX	MGPKKQLMKLLFKVLEALFRGDEETLRELAREAVELAERLLKLGDPELLFLALAIAIIVAWA VGDEELLKRLAQIIKELLKRAEELGDPDLRRLIEELVEFVERLGSLEHHHHHH
C2-01018	MSEESEKLLKEAKEEVRRLFEEGRPEDAARVAFEFLQRLLDLGDPDAVKELLQFLRELLG SLEHHHHHH
C2-3677	MDRKEELLERMAELLIELLEQIGDPRLKEILKLLLELMRMEGNEEVVERVEEVLKRLGSLE HHHHHH
C2-02802	MDLEERFERLYEKAKKLAEERGDERARRMIELLRQLFETVGDPRILELLELLLQLLEGGSL EHHHHHH
C5_HF-3921	MGHHHHHHGSGSENLYFQGGSSDLQEVADRIVEQLKREGRSPEEARKEARRLIEEIKQS AGGDSELIEVAVRIVKELEEQGRSPSEAAKEAVELIERIRRAAGGDSELIEVAVRIVKELEE QGRSPSEAAKEAVELIERIRRAAGGDSELIEVAVRIVKELEEQGRSPSEAAKEAVELIERIR RAAGGDSELIEVAVRIVKELEEQGRSPSEAAKEAVELIERIRRAAGGDSELIEVAVRIVKFL EEAGMSPSEAAKVAVELIERIRRAAGGDSELIEKAVRIVRRLERRGLSPAEAAKIAVAIIAA EVLSREAEKIREETEEVKKEIEESKKRPQSESAKNLILIMQLLINQIRLLALQIQMLRLQLEL
C5_HF-2101	MGHHHHHHGSGSENLYFQGGSSEKEKVEELAQRIREQLPDTELAREAQELADEARKSD DSEALKVVYLALRIVQQLPDTELAREALELAKEAVKSTDSEALKVVYLALRIVQQLPDTEL AREALELAKEAVKSTDSEALKVVYLALRIVQQLPDTELAREALELAKEAVKSTDSEALKVV YLALRIVQQLPDTELAREALELAKEAVKSTDSEALKVVYLALRIVQQLPDTELAREALELAK EAVKSTDSEALKVVYLALRIVQLLPDTDLARKALELAKEAVKMDDQEVLKVVYKALQIVAD KPNTEEADEALRDARLKLEAARLRREMEKIREETEEVKKEIEESKKRPQSESAKNLILIMQ LLINQIRLLALQIRMLDLQL KL
C4_nat_HF-7900	MASSWVMLGLLLSLLNRLSLAAEAYKKAIELDPNDALAWLLLGSVLLLLGREEEAEEAAR KAIELKPEMDSARRLEGIIELIRRAREAAERAQEAAERTGDPRVRELARELKRLAQEAAEE VRRDPDSKDVNEALKLIVEAIEAAVRALEAAERTGDPEVRELARELVRLAVEAAEEVQRN PSSSDVNEALKLIVEAIDAAVRALEAAEKTGDPEVRELARELVRLAVEAAEEVQRNPSSE EVNEALKDIVKAIQEAVESLREAEESGDPEKREKARERVREAVERAEEVQRDPSSGGSW GLEHHHHHH
C4-71-101B	MGRQEKVLKSIEETVRKMGVTMETHRSGNEVKVVIKGLHESQQEQLLEDVLRTAEKQGV RVRIRFKGDTVTIVVREGSGPEEILERARESLERAREASERGDEEEFRKAAEKALELAKR LVEQAKKEGDPWMVMWAALVALWVALLALRNGDKEVFKKAAESALEVAKRLVEVASKE GDPEMVLLAAWVALFVAWLAWLFGDKEVFKKAAESALEVAKRLVEVASKEGDPELVEEA AKVAEEVEKLAEKQGDEEVREKAWETWMEVWLLWLEVRLRKGGGGSLEHHHHHH
C4-81-101B	MGELERESREAEKRLKEARLFAWAARLLGDLKLLAKALIEEARAVQELARVACERGNRD EAWDAFEKALEVFEEAVKVSEEAREQGDDEVLALALIAIALAVLALAEVACCLGISELAEL AWKMAEWVLEEARKVSEEAREQGDDEVLALALIAIALAVLALAEVACCRGNKEEAERAY EDARRVEEEARKVKESAEEQGDSEVKRLAEEAEQLAREARRHVQECRGGWLEHGSGR QEKVLKSIEETVRKMGVTMETHRSGNEVKVVIKGLHESQQEQLLEDVLRTAEKQGVRVR IRFKGDTVTIVVREGSLEHHHHHH
C6-71-101B	MGRQEKVLKSIEETVRKMGVTMETHRSGNEVKVVIKGLHESQQEQLLEDVLRTAEKQGV RVRIRFKGDTVTIVVREGSPEEILERAKESLERAKEAFERGDEEEFRKAAEKALELAKRLV EQAKKEGDPELVFEAARVALWVAWLAAWFGDKEVFKKAAESALEVAKRLVEVAKEEGD PELVLKAAFVALLVAIMAVILGDKEVFKKAAESALEVAKRLVEIAAREGDPELVEEAAKVAE LVRELAKLMGDEEVYEKARETAREVRLFLLFVRIWEGGGGSLEHHHHHH
C6-79-101B	MGSSDEEEARELEERAREAAKRAIEAAKRTGDPRVRELAEELVKLAIWAAVEVWLDPSS SDVNEALKLIVEAIEAAVRALEAAERTGDPEVRELARELVRLAVEAAEEVQRNPSSSDVN EALKLIVIAIEAAVRALEAAERTGDPEVRELARELVRLAVEAAEEVQRNPSSEEVNEALRKI

	IKLILFAVMVLELAEEIGDPTWREMARRAVREAVELAEEVQRDPSGWLGHGSGRQEKVL KSIEETVRKMGVTMETHRSGNEVKVVIKGLHESQQEQLLEDVLRTAEKQGVRVRIRFKG DTVTIVVREGSLEHHHHHH
C8-71-101B	MGRQEKVLKSIEETVRKMGVTMETHRSGNEVKVVIKGLHESQQEQLLEDVLRTAEKQGV RVRIRFKGDTVTIVVREGSGPEEILERAKESLERAREASERGDEEEFRKAAEKALELAKRL VEQAKKEGDPELVLEAARVALWVAELAAKNGDKEVFKKAAESALEVAKRLVEVASKEGD PDLVAWAALVALWVAFLAFLNGDKEVFKKAAESALEVAKALMEVAMKVGAPWLVELAIA VARAVWLLAELFGDEEVRRRAEAFEIILRIAAIAVKAWLGGGGSLEHHHHHH
C4Hub_274A53	MGNTHFIVVHGSEDAAQLAEELVREISKKEGTEVRFEKKDGLLSIEVKNLSEERQREIQKA LQLVQDVANAERVVRERPGSNLAKKALEIILRAAEELAKLDLKASLKAAVRAAEKVVREQ PGSELAKKALEIILRAAEELAKLPDPEALHEAVRAAEHVVRSQPGSEAAKEALRIIQEAAEL LKESPDPTAIIRAARALLKIARTTGDEEAAKEAIEAAKKAADLARERGDDELVCEALALLVA AQVELLKQQGTSAVEIAKIVARVISEVIRTLKEKGSSYEVICECVARIVAEIVEALKRSGTSA AIIALIVALVISEVIRTLKESGSSFEVILECVIRIVLEIIEALKRSGTSEQDVMLIVMAVLLVVLA TLQLSGSLEHHHHHH
C4-71C_mb7	MGPEEILERARESLERAREASERGDEEEFRKAAEKALELAKRLVEQAKKEGDPWMVMW AALVALWVALLALRNGDKEVFKKAAESALEVAKRLVEVASKEGDPEMVLLAAWVALFVA WLAWLFGDKEVFKKAAESALEVAKRLVEVASKEGDPELVEEAAKVAEEVEKLAEKQGDE EVREKAWETWMEVWLLWLEVRLRKGGGGSDRRKEMDKVYRTAYKRITSTPDKEKRKE VVKEATEQLRRIAKDEEEKKKAAYMISFLKTLGSLEHHHHHH
C6-71C_mb7	MGPEEILERAKESLERAKEAFERGDEEEFRKAAEKALELAKRLVEQAKKEGDPELVFEAA RVALWVAWLAAWFGDKEVFKKAAESALEVAKRLVEVAKEEGDPELVLKAAFVALLVAIM AVILGDKEVFKKAAESALEVAKRLVEIAAREGDPELVEEAAKVAELVRELAKLMGDEEVY EKARETAREVRLFLLFVRIWEGGGGSDRRKEMDKVYRTAYKRITSTPDKEKRKEVVKEA TEQLRRIAKDEEEKKKAAYMISFLKTLGSLEHHHHHH
C6-79C_mb7	MGSSDEEEARELEERAREAAKRAIEAAKRTGDPRVRELAEELVKLAIWAAVEVWLDPSS SDVNEALKLIVEAIEAAVRALEAAERTGDPEVRELARELVRLAVEAAEEVQRNPSSSDVN EALKLIVIAIEAAVRALEAAERTGDPEVRELARELVRLAVEAAEEVQRNPSSEEVNEALRKI IKLILFAVMVLELAEEIGDPTWREMARRAVREAVELAEEVQRDPSGWLGHGSDRRKEMD KVYRTAYKRITSTPDKEKRKEVVKEATEQLRRIAKDEEEKKKAAYMISFLKTLGSLEHHHH HH
C8-71C_mb7	MEKKIGPEEILERAKESLERAREASERGDEEEFRKAAEKALELAKRLVEQAKKEGDPELV LEAARVALWVAELAAKNGDKEVFKKAAESALEVAKRLVEVASKEGDPDLVAWAALVALW VAFLAFLNGDKEVFKKAAESALEVAKALMEVAMKVGAPWLVELAIAVARAVWLLAELFGD EEVRRRAEAFEIILRIAAIAVKAWLGGGGSDRRKEMDKVYRTAYKRITSTPDKEKRKEVVK EATEQLRRIAKDEEEKKKAAYMISFLKTLGSLEHHHHHH
C4-71N_mb7	MGDRRKEMDKVYRTAYKRITSTPDKEKRKEVVKEATEQLRRIAKDEEEKKKAAYMISFLK TLGSPEEILERARESLERAREASERGDEEEFRKAAEKALELAKRLVEQAKKEGDPWMVM WAALVALWVALLALRNGDKEVFKKAAESALEVAKRLVEVASKEGDPEMVLLAAWVALFV AWLAWLFGDKEVFKKAAESALEVAKRLVEVASKEGDPELVEEAAKVAEEVEKLAEKQGD EEVREKAWETWMEVWLLWLEVRLRKGGGGSLEHHHHHH
C6-71N_mb7	MGDRRKEMDKVYRTAYKRITSTPDKEKRKEVVKEATEQLRRIAKDEEEKKKAAYMISFLK TLGSPEEILERAKESLERAKEAFERGDEEEFRKAAEKALELAKRLVEQAKKEGDPELVFE AARVALWVAWLAAWFGDKEVFKKAAESALEVAKRLVEVAKEEGDPELVLKAAFVALLVAI MAVILGDKEVFKKAAESALEVAKRLVEIAAREGDPELVEEAAKVAELVRELAKLMGDEEV YEKARETAREVRLFLLFVRIWEGGGGSLEHHHHHH
C6-79N_mb7	MGDRRKEMDKVYRTAYKRITSTPDKEKRKEVVKEATEQLRRIAKDEEEKKKAAYMISFLK TLGSSSDEEEARELEERAREAAKRAIEAAKRTGDPRVRELAEELVKLAIWAAVEVWLDPS SSDVNEALKLIVEAIEAAVRALEAAERTGDPEVRELARELVRLAVEAAEEVQRNPSSSDV NEALKLIVIAIEAAVRALEAAERTGDPEVRELARELVRLAVEAAEEVQRNPSSEEVNEALR

	KIIKLILFAVMVLELAEEIGDPTWREMARRAVREAVELAEEVQRDPSGWLGHGSLEHHHHHH
C8-71N_mb7	MEKKIGDRRKEMDKVYRTAYKRITSTPDKEKRKEVVKEATEQLRRIAKDEEEKKKAAYMI SFLKTLGSPEEILERAKESLERAREASERGDEEEFRKAAEKALELAKRLVEQAKKEGDPE LVLEAARVALWVAELAAKNGDKEVFKKAAESALEVAKRLVEVASKEGDPDLVAWAALVA LWVAFLAFLNGDKEVFKKAAESALEVAKALMEVAMKVGAPWLVELAIAVARAVWLLAELF GDEEVRRRAEAFEIILRIAAIAVKAWLGGGGSLEHHHHHH