Computational Design of Symmetric Protein Complexes with Implications for Vaccine and Biotherapeutic Development

George Ueda

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2018

Reading Committee:

David Baker, Chair

Phil Bradley

Kelly Lee

Program Authorized to Offer Degree:

Biochemistry

University of Washington

Abstract

Computational Design of Symmetric Protein Complexes with Implications for Vaccine and Biotherapeutic Development

George Ueda

Chair of the Supervisory Committee:

Professor David Baker

Biochemistry

Using a newly developed computational docking and scoring method combined with Rosetta two-sided interface design, we demonstrated accurate design of self-assembling oligomeric proteins that exhibit various degrees of symmetry. A set of designs were validated to match their respective models at the atomic-scale and we progressed to functionalize this class of proteins for targeted biological applications. With the unique ability to tailor new protein structures, we span a diversity of controllable geometric arrangements representing a molecular toolkit to probe biological systems at the subnanometer scale. Presented in this dissertation are examples of next-generation vaccine candidates that scaffold entire antigenic complexes, as well as potential biotherapeutics that activate signaling pathways through engagement and tunable clustering of cell surface receptors. These studies showcase the potential for computationally-generated molecules to trigger unique biological responses, providing novel insights, considerations, and future avenues for vaccine and biotherapeutic development across various disease spaces.

TABLE OF CONTENTS

List of Figures

List of Tables

ACKNOWLEDGEMENTS

In my first-year of graduate school, we were required to take a scientific grant-writing class. Although I didn't proceed to win any award, I obtained a piece of advice - that graduate school should be carried out not only in a professional sense, but also on a personal level provided it would be a different experience for each and every one of us. I would be remiss to not attest to this advice, and describe the experience as a tumultuous yet transformative journey.

Indeed, not all students enter from the same background, but from the first-year literature review classes to first scientific publications, all who make sufficient effort are guaranteed to grow scientifically. Project conception, experimental planning, technical troubleshooting, data interpretation, and peer-based discussion become well-learned and practiced key skills, among others of course I am forgetting to describe. With the large volume of work and ongoing mental engagement required, there is little driving force left than the pure passion for science. As a result however, a unique and exceptional environment manifests, for which I feel extremely fortunate and appreciative to have spent the past five years. This document serves to describe the work I took part in over the course of my graduate training in this invaluable environment, and encompasses the most scientifically productive and developmentally valuable years of my life.

It goes without saying that the experience was nothing without the peers, who have collectively included the most diverse and amazing set of individuals I have encountered in an academic environment or otherwise. None of the work presented in this document would have been possible without the guidance, help, and participation of countless individuals, many of whom I'd like to mention here. First I'd like to acknowledge William Sheffler for selflessly

helping me navigate the Rosetta software and quite frankly providing irreplaceable computational geneious, as well as Neil King and Jacob Bale who instigated many early discussions on project direction and provided exposure to collaboration within a large and growing scientific field. For other mentorship, guidance, troubleshooting, training, cooperative work, and scientific conversation, I'd like to specifically acknowledge Fabio Parmeggiani, Matthew Bick, Chris Bahl, Scott Boyken, Lauren Carter, Yang Hsia, Stephen Rettie, Brooke Nickerson, Bobby Langan, Una Natterman, Xinting Li, Jasmine Nguyen, Alex Kang, Gustav Oberdorfer, Alex Ford, Jason Gilmore, Anindya Roy, Michelle Matsunaga, Karla-Luise Herpoldt, Franziska Seeger, Stephanie Berger, Gaurav Bhardwaj, Marc Lajoie, TJ Brunette, David Younger, Cassie Brian, Danny Sahtoe, Alexis Courbet, Daniel Ellis, Hanlun Jiang, Peilong Lu, Ivan Vulovic, Christy Tinberg, Possu Huang, Tom Linsky, Lucas Nivon, and Jeremy Mills.

Science is also conducted (and thrives) based off of collaboration. I was lucky enough to be in one of the most well-connected academic labs in the world, and thereby had the opportunity to work with extremely talented individuals from all over the globe. This list comprises Aleksandar Antanasijevic at The Scripps Research Institute, whom has been an amazingly productive and impressive collaborator, Andrew Ward, Philip Brouwer, Rogier Sanders, Laurent Perez, Antonio Lanzavecchia, Young-Jun Park, David Veesler, Kritika Mohan, Kevin Jude, Christopher Garcia, Banumathi Sankaran, Peter Zwart, and Greg Hura.

Outside the sphere of science, many others have helped me in less tangible but equally valuable ways such as mental, emotional, or logistical support. Most directly, my mother Patricia Ueda, sister Emma Ueda, grandmother Dorothy Budinich, and grandfather Thomas Budinich

have been constant and reliable sources of support. Such personalities are not found ubiquitously, and as such I am exceptionally grateful. Additionally, many friends have persisted through graduate school and contributed to shaping my personality and scientific work, including Omar Farajallah, Alex Fu, Timothy Reilly-Crank, Matthew Loewen, Albert Chae, Cristian Guajardo, and Peter Lu.

During the course of graduate school, I also had several opportunities (requirements) to serve as a teaching assistant in undergraduate biochemistry courses and labs. While I found I was not the most keen on teaching large groups of students, I was also fortunate to have the opportunity to mentor students on individual projects in the laboratory. Of note, I'd like to specifically acknowledge Vanessa Nguyen whom provided one of the most engaging and rewarding mentorship dynamics imaginable. Her productivity and contributions were so great that she was included on our first publication, an occurrence that is generally rare and reserved for only exceptional undergraduates.

Returning to the first-year grant writing class, I did also obtain a second piece of advice - while you should choose your laboratory in line with your scientific passions, take consideration of your thesis advisor. I can say, over five years later, that David Baker was not even close to a wrong choice. Not only does he provide and somehow maintain an amazing academic environment, he is a gifted and driven scientist and possesses the astounding ability to think about the big picture and collectively guide technical efforts. In addition, he offers genuine mentorship, guidance, career advice, and conversation to the best of his abilities - a trait that should not be taken for granted with advisors.

Last and certainly not least, I'd like to express my deep gratitude for one extraordinary individual, Jorge Fallas. From my first quarter rotation in the Baker Laboratory (Winter, 2014), Jorge has fulfilled the role as my scientific mentor, everyday co-worker, and eventual best friend. Over the years, he has been an invariably intellectual individual providing means for scientific discussion (usually over a necessary coffee or meal), often regarding broader impacts of everyday work. I feel extremely fortunate to have met him those five years ago, and hope to continue collaborative work with him for years to come.

INTRODUCTION

Proteins are highly-evolved molecular machines that carry out a diverse range of exquisite reactions and functions critical to biology. Many proteins possess interfaces that form energetically-favorable interactions and guide self-assembly into elaborate complexes, an often essential process to function properly. Nature provides many examples of such structures, perhaps most impressive are viral capsids and other nanoscale architectures, but a multitude of macromolecular assemblies exist across a range of size scales[1,2]. Observation of these has motivated efforts to create comparable or even superior structures with hopes of achieving greater control over the biophysical determinants that govern protein folding and assembly. If this understanding could be employed in a robust way, the design capabilities would have implications in fields including biotherapeutic development, biomaterial design, and general interrogation of cell biology. While there has been success using nucleic acids as the building blocks to engineer supramolecular architectures[3], nature exhibits proteins as the molecules of choice for assemblies that possess much more diverse forms and functions. That being said, harnessing the complex physicochemical principles that govern protein folding and oligomerization has been an ongoing challenge, and strategic approaches have been required to improve our understanding as we continue to design new protein structures.

Nature provides a great design strategy with the degrees of symmetry exhibited in many natural large protein assemblies. The feature of symmetry, found in our expanding set of determined macromolecular structures[4], has historically been owed to the fact that symmetric assemblies require the fewest number of unique interface contacts between subunits[5]. There have been recent successes using the computational strategy of symmetric protein design to generate

assemblies of novel target architectures from naturally-occurring oligomeric components[6–8]. We sought to apply this design method but starting at the monomeric level and building up, thereby generating a set of new protein complexes across a range of symmetries with a high degree of configurational control at each step of the hierarchical design process. As a result, our designed proteins have been exceptionally amenable to functionalization across a range of size scales and configurations for applications in various biological spaces.

Many proteins self-assemble into cyclic homooligomers, one of the simplest symmetry groups, and play key roles in cell biology[1,9]. Naturally occurring cyclic oligomers have also been used as building blocks for protein-based materials that adopt cubic point-group symmetries (tetrahedral, octahedral, icosahedral) in a hierarchical manner[6–8]. With recent improvements in our understanding of the biophysical principles that govern protein folding and protein-protein interactions[10], we designed and validated a set of self-assembling *de novo* cyclic oligomers with targeted configurations and built exclusively from monomeric repeat proteins. During the initial docking stage, we employed a novel scoring method to rapidly evaluate symmetric protein configurations for energetically favorable interfacial contacts that recapitulated full atom Rosetta sequence design better than alternative low-resolution docking protocols. The method and results were published in 2017 in Nature Chemistry, entitled *Computational Design of Self-Assembling Cyclic Protein Homo-oligomers*[11]. The goal here was not only to expand our set of usable oligomeric scaffolds, but also to allow for user selection of configurations based on molecular and geometric properties entirely at the computational level. The specification of protein length, curvature, and oligomeric configuration brings a revolution in the level of control in protein assembly design, and allows for a much richer variety of downstream biomaterials.

To demonstrate the potential for designed cyclic oligomers to form higher-order assemblies, we leveraged a symmetric docking method[6,7] combined with our own developed scoring method[11] to generate a set of *de novo* two-component nanoparticles with idealized configurations for scaffolding functional domains through genetic fusion. These constructs provide new avenues for multivalent display of arbitrary proteins - in particular we screened for designed homotrimeric components with configurations ideal for fusion to several viral antigenic complexes to achieve a highly avid particle capable of triggering an enhanced immune response from activated B cells when compared to its monovalent format. Multivalent presentation of viral antigens is an acknowledged and active area of study for enhancing antigen-specific antibody titers during adaptive immune response[12], but computational methods have not been pursued to generate scaffolding platforms or particles especially capable of displaying complexes in their conformationally-relevant state[13,14]. Here, we validated a set of *de novo* two-component nanoparticles designed to scaffold viral fusion complexes human immunodeficiency virus (HIV) BG505 SOSIP[15], Influenza Hemagluttinin (HA)[16], and the respiratory syncitial virus (RSV) prefusion protein (DS-Cav1)[17]. The nanoparticles themselves were found to match the designed models by a combination of electron microscopy and small-angle X-ray scattering. Representative examples of antigen-displaying particles were also produced with genetically-fused constructs, and collectively comprise a panel differing antigen valencies and geometries capable of probing both intensity and epitope specificity of elicited antibody repertoires. The immunogenicity of these particles is under active investigation in rodent immunization, epitope mapping, non-human primate (NHP) trafficking, and viral protection studies. This work is currently being written up in a manuscript featuring the computational and

structure-based design and characterization of the antigen-displaying particles, while the molecules themselves are being investigated further in animal models as potential next-generation vaccines.

Oligomeric complexes are also prevalent in cytokine signaling, a critical part of regulation in cell biology and physiology. Many cytokines exist naturally as symmetric oligomers, capable of binding and cross-linking cell surface receptors to activate various cellular signals (agonism)[18] as opposed to inhibiting an interaction by binding alone (antagonism)[18,19]. Given the power of these determinants in physiology, many natural cytokines have been investigated as therapeutic agents, but most have severe shortcomings such as toxicity or lack of efficacy, which are contributed to by their stimulation of multiple downstream signaling responses and occasional dissimilar effects on differing cell types[20,21]. This is not necessarily surprising given that cytokines evolved to regulate highly complex and finely balanced homeostatic functions, and not to necessarily be repurposed as drugs. Nevertheless, cytokines have great potential as therapeutics if their actions could be controlled in a more rational way[22]. By analogy, tuning signaling pathways of G protein-coupled receptors (GPCRs) through small molecule intervention has proven to be a powerful therapeutic concept[23] but a similar strategy is not available for receptors that engage cytokines and growth factors in the form of higher-order complexes. Here, we present a general approach to modulate cytokine receptor signaling output on natural cells using a series of designed cytokine mimetics. Based on a repeat protein scaffold[24], we generated a synthetic cytokine mimetic capable of systematically controlling the type-I cytokine erythropoietin receptor (EpoR) complex configuration and corresponding biological phenotype in erythroids (red blood cells)[25]. We sampled a range of EpoR inter-dimer

arrangements with a series of designed dimers, corroborated by several design-EpoR complex crystal structures. The designed dimers exhibited a range of super, partial, and biased agonism, demonstrating a degree of signaling plasticity analogous to that seen with GPCRs. Using these ligands, we observe hematopoietic stage-selective and tunable stimulation of signaling pathways downstream from EpoR, and is the feature of a recent submission to Science entitled *Tuning Hematopoiesis through Topological Control of Cytokine Receptor Signaling*. This design strategy is suitable for diversifying signaling outputs in many other therapeutically impactful receptor systems such as the interleukin (IL)[26,27] and tumor-necrosis factor (TNF)[28] families for applications in autoimmune disease and cancer. This latter body of work is what will continue to guide my scientific work and career beyond graduation, in efforts to engineer new synthetic signaling molecules as biotherapeutic candidates for cancer and autoimmune disease.

CHAPTER 1. SYMMETRIC INTERFACE DESIGN

1.1 INTRODUCTION TO SYMMETRIC PROTEIN OLIGOMERS

Cyclic homo-oligomers, assembled from multiple identical protein subunits symmetrically arranged around a central axis, play key roles in many biological processes including catalysis, signaling, and allostery[1,29,30]. Despite their prevalence in natural systems, currently there is no systematic approach to design cyclic homo-oligomers starting from a monomeric protein structure. A number of prior design studies have relied on canonical structural motifs such as α-helical coiled coils[31], β-propeller motifs[32,33], unpaired β-strands[34] or metal binding sites[35]. Recently a C2 dimer mediated by an α-helical interface was reported but the design protocol required extensive iteration between computation and experiment[36]. In contrast, there has been considerable progress in designing proteins that fold into predetermined target structures ranging from idealized versions of natural folds[37–40] to topologies that appear not to have been explored during evolution[41,42]. Particularly interesting from an engineering perspective are *de novo* designed α-helical repeat proteins with a wide range of shapes which can be readily shortened or lengthened simply by changing the number of repeats in their sequence[42].

Here we present a general method for designing cyclic homo-oligomers *in silico* and use it to design interfaces onto recently developed repeat proteins (see **Appendix Methods** for scaffolds)[40,42,43] that direct their assembly into dimeric, trimeric, tetrameric and pentameric complexes. Structural characterization shows that many of the designs adopt the target oligomerization state and structure, demonstrating a basic understanding of the determinants for oligomerization. The capability of designing proteins with tunable shape, size, and symmetry enables rigid display of binding domains at arbitrary orientations and distances for a range of

biological applications.

1.2 Computational Cyclic Docking

The self-assembly of naturally occurring complexes is driven by chemical and shape complementarity. Protein-protein interfaces are generally comprised of a hydrophobic core that is buried upon binding and surrounded by a periphery of polar residues that prevent non-specific aggregation[44–47]. We developed a design strategy to generate such interfaces between protein monomers docked in a range of cyclic geometries. The strategy has two steps (**Figure 1.1**): first, low resolution docking to sample and rank symmetric arrangements of a given scaffold protein based on their designability (the likelihood of finding an amino acid sequence that can stabilize a given rigid body conformation), and second, full atom RosettaDesign[48] calculations to optimize the sequence at the protein-protein interfaces for high affinity binding. To explore the generality of the method, symmetries ranging from C2 through C6 were docked.

To generate cyclic homooligomeric arrangements of n copies of a protein monomer, we center it at the origin and finely sample the degrees of freedom before generating a symmetric partner by a (360/n)° rotation around the Z-axis and translational slide into contact. The configurational space for cyclic docking is four dimensional: the usual six degrees of freedom required for orienting a rigid body, minus translations along and rotations around the symmetry axis of the oligomer (to which the structure is invariant). These four degrees of freedom can be reduced effectively to three by the requirement that the subunits must be roughly in contact. We realize this dimensionality reduction using a fast slide-into-contact algorithm along the X-axis allowing a small range of X offsets close to the contact value. For each of these, the axis of

symmetry is determined from the relative orientation of the two subunits, which can then be propagated to generate the full oligomer. Computing the slide distance along a given slide vector is accomplished using two two-dimensional arrays perpendicular to the slide direction into which the atoms along the leading face of each body are placed. Corresponding cells are checked, and the pair with the least separation provides an estimate of the slide distance. The bodies are placed according to this estimate, but may still have clashes. All contacting pairs of atoms across the bodies are checked using an octree-like data structure, and the bodies are backed off so as to relieve the largest clash found. This process is repeated until no clashes are found. In practice, only one or two iterations through the fast clash check are required in most cases, making the slide move rapid. The top ten non-redundant cyclic docked configurations were chosen as input for a Rosetta interface design protocol according to the scoring method described below.

1.3 RPX Score Method and Motif Database Construction

Existing methods for protein-protein docking fall into three general categories: (1) voxelized rigid representations with Fast Fourier Transform (FFT)-based docking[48–50], (2) docking based on patches of high-resolution local shape complementarity[51], and (3) Monte Carlo sampling with soft centroid models[52,53]. The first two categories are not ideal for the protein design problem because the precise shape and chemical detail of the docked surfaces are unavailable, as the interface residues are not known in advance and side chains are represented by one or two points in space where the interaction potential is evaluated as a function of the distance or distances between these points, and in some cases the associated angles[53–55]. The approach we take is most similar to (3), in which docked backbones are generated and then

scored using a low-resolution representation of the proteins (requiring only the backbone coordinates and secondary structure assignments) but with two notable improvements. First, we employ a six-dimensional implicit side chain scoring methodology, which better predicts the result of subsequent full atom design calculation than a traditional coarse-grained model, and second, we use an enumerative strategy to generate docked backbones, which samples more exhaustively the low-dimensional docking space than a Monte Carlo search.

This method, termed Residue Pair Transform (RPX), represents the interaction between two residues by the full six dimensional rigid body transformation between their respective backbone N, Cα, and C atoms. We employ a precompiled database of all favorable residue pair interactions (motifs) found in structures from the Protein Data Bank involving alanine, isoleucine, leucine, valine, and methionine, binning these data based on the rigid body transform between amino acids. The score of a given docked configuration is the sum, over each pair of residues across the interface (9Å cutoff), of the lowest Rosetta full atom energy found in the associated spatial transformation bin of the database. This approach predicts the interface energy resulting from full atom sequence design calculation better than the Rosetta centroid energy function (**Appendix Figure 1.1**). As the database is compiled offline, arbitrary data selection (different subsets of amino acid identities) and processing (alternative smoothing and scoring schemes) can be employed with no impact on runtime of the docking calculations.

The database of pairwise motifs used in this study was constructed from a culled set of deposited proteins from the RCSB Protein Data Bank as of August 2012, comprised of 23162 structures with a reported resolution of 2.5 Å or less. Because each independently crystallized protein in this set was thought to contain new structural information and afford high resolution

features, no sequence redundancy cut was used. Residue pair contact areas were filtered with a 1.7 $Å^2$ probe, at least 10 amino acid sequence separation, only highest occupancy state rotamers were used, and pairs without at least one partner within the first chain were ignored. Weights for each residue pair were introduced using the Rosetta Score12 score function[48,] and the Rosetta implementation of the DSSP secondary structure prediction server was used to assign a backbone type to each residue in a given pair. Residue pair energies greater than -0.001 energy units were discarded. Example usage file ("README") is included in **Supplementary Files** for the entire docking stage as well as the required motif database with precomputed scores ("xs_bb_ss_AILMV_sf5_std") and flags ("dock.flags").

1.4 Two-Sided Interface Design

An interface design protocol was implemented in RosettaScripts[56,57] (example designCx.xml provided in **Supplementary Files**) and is described briefly here. Each design trajectory took an input .pdb file containing a single subunit of each docked configuration with the cyclic axis aligned to the vector [0, 0, 1], as well as a Rosetta symmetry definition file (see **Appendix Methods**)[56]. The monomeric subunit was initially perturbed by a translation perpendicular to the axis of symmetry, as well as a random rotation in three-dimensional space. The applied perturbation was selected from a gaussian distribution bounded by user-defined distances and angles. An oligomer with the specified cyclic symmetry was then generated using the information stored in the symmetry definition file to cyclize the monomeric subunit. For the interface design protocol, designable positions were designated as residues that met the following criteria: beta carbon within 10 Å of at least one beta carbon from another subunit, at

least one atom within 5 Å of any other atom from the same subunit, a non-zero surface area accessible to solvent, identity of neither proline nor glycine. While keeping other subunit residues fixed, designable positions were optimized using the Rosetta packing algorithm with the default Talaris2013 score function[58] and the extended rotamer library available in Rosetta. Initially, a first packing round was executed with a modified score function using a full atom repulsive term weight of 0.05 (as opposed to a standard weight of 1.0) in order to sample more of the sequence space for the particular configuration. Once a sequence was converged upon, designable positions were allowed to minimize side chain torsion angles using the same reduced repulsive term weight. A subsequent round of packing and minimization was conducted, but with the repulsive term weight reset to 1.0 in order to to converge on a local minimum of standard Rosetta energy units. Individual design trajectories were filtered by the following criteria: difference between Rosetta energy of bound (oligomeric) and unbound (monomeric) states less than -20.0 Rosetta energy units, interface surface area greater than 700 $Å^2$, Rosetta shape complementarity greater than 0.65, and less than 45 mutations made from the respective native scaffold. Designs that passed these criteria were manually inspected and refined by single point reversions for mutations that were deemed non-contributory to stabilizing the bound state of the interface. The design with the best overall metrics for each unique docked configuration was then added to a set of 96 finalized proteins to be experimentally validated. The 11 dimers, 34 trimers, 19 tetramers, 17 pentamers and 15 hexamers are named according to the following nomenclature: the first 4 letters refer to the scaffold protein, the symmetry is denoted as Cn, and finally an integer is added to differentiate oligomers of identical symmetry and scaffold identity (**Appendix Table 1.1**).

1.5 Biophysical Characterization of Cyclic Oligomers

Synthetic genes encoding each of the 96 designs were synthesized and cloned into a vector with a T7 promoter system and either an N- or C-terminal $(His)_6$ tag, and the corresponding proteins expressed in *E. coli*. The proteins were purified by immobilized nickel-affinity chromatography (Ni^{2+} IMAC) and size-exclusion chromatography (SEC). 64 designs were soluble and amenable to purification (**Appendix Figure 1.3 and 1.4**). The oligomerization states for 44 designs that eluted from SEC with a single predominant species were determined by size-exclusion chromatography in tandem with multi-angle light scattering (SEC-MALS). For 21 of the designs, the molecular weights determined by light scattering agreed with the designed oligomerization state (**Appendix Table 1.2**).

To further assess the configuration of the designed proteins in solution, small-angle X-ray scattering (SAXS) measurements were performed on designs that had predominantly monodisperse traces in the SEC screen. A total of 26 designs (the 21 with consistent SEC-MALS data and 5 additional designs that had monodisperse SEC profiles) were characterized with this technique and the measured scattering profile was compared to that expected from the computational model[59,60]. Designs with a deviation of less than or equal to 3.1 using the χ measure[59] and a deviation of less than 11% between the computed and experimental radius of gyration were considered to be in the designed supramolecular arrangement (these thresholds were chosen based on the deviations between computed and measured values for designs with crystal structures consistent with the corresponding models; see below).

Of the 26 designs, 15 fulfill these criteria; 5 dimers, 6 trimers, 3 tetramers, and 1 pentamer. The docked configurations and designed interfaces of 13 of these are unique (three of the trimers have similar geometries with pairwise r.m.s.d. values between 1.9-2.5Å; the lowest pairwise r.ms.d. among the remaining designs is 5.3 Å with no similarity in designed interface). Computational models, *in silico* symmetric docking energy landscapes, SEC-MALS chromatograms and SAXS experimental and computed profiles for the 15 designs are summarized in **Figure 1.2** and **Appendix Figure 1.5**; data on the full set of designs is provided in **Appendix Tables 1.1-1.4**.

Crystal structures that contain the designed interface were obtained for five of the designed proteins: two dimers, two trimers and one tetramer, and are compared to the design models in **Figure 1.3**. For each of the five cases the side chain rotamers of the hydrophobic residues are similar to those in the design model. The two dimers, ank3C2_1 and ank1C2_1, are both built from idealized ankyrin repeat proteins and are shown in **Figure 1.3a,b**. The ank3C2_1 design has a large hydrophobic patch (1100 $Å^2$) that is buried upon binding; all interface hydrophobic side chains are in the same rotameric state in the design model and the crystal structure with the exception of methionine 90 (**Figure 1.3a**, right panel). The backbone r.m.s.d. between the design model and the crystal structure is 1.0 Å. The agreement between the model and the structure of ank1C2_1 (**Figure 1.3b**) is even closer: both polar and hydrophobic side chain rotamers were correct and the r.m.s.d. to the model is only 0.9 Å.

The two trimeric designs with solved structures are 1na0C3_3 (**Figure 1.3c**) built from a consensus designed tpr protein[43], and HR00C3_2 (**Figure 1.3d**) built from a *de novo* designed repeat protein. 1na0C3_3 has a hydrophobic core that lies on the 3-fold axis formed by residues

in all subunits. The r.m.s.d. between the crystal structure and design model is 1.0 Å. HR00C3_2 contains a pore on the symmetry axis and is stabilized by three separate heterologous interfaces. This trimer was designed using the computational model of a designed repeat protein whose structure had not previously been confirmed by X-ray crystallography. Thus the crystal structure, which has a backbone r.m.s.d. to the model of 0.9 Å, validates the design of both the monomer and oligomer simultaneously. This ability to accurately design higher order structures based on design models of monomers will considerably streamline future computational design of nanomaterials using monomers with custom designed properties.

For the two dimers and the two trimers, the χ values between the measured SAXS scattering profiles and the profiles computed from either the corresponding design models or crystal structures are less than 3.1. In contrast, the experimental SAXS data for the designed tetramer, ank1C4_2 (**Figure 1.3e**), deviates considerably from that computed using the crystal structure (**Appendix Figure 1.6**). The ank1C4_2 crystal structure adopts a C2 symmetric tetrameric structure in which 2 pairs of chains accurately match the design model (r.m.s.d. of 1.1 Å), but exhibit clear overall distortion relative to the C4 symmetric design model (r.m.s.d. of 4.5 Å). There are two distinct interfaces present in the structure, one of which corresponds to the designed interface. The experimental SAXS profile is closer to the design model of the tetramer than the crystal structure, and hence it seems likely that the symmetry breaking in the crystal is due to lattice contacts.

A sixth structure was solved for design ank4C4, which shows a single symmetric peak by SEC and forms a tetrameric complex in solution as determined by MALS. The SAXS profile of this design does not match that computed from the design model (χ = 3.8), and the crystal

structure exhibits D2 symmetry rather than the target C4 symmetry. The SAXS profile computed from the D2 oligomer matches the measured scattering curve better than the target C4 model (χ = 1.2) indicating that the D2 state corresponds to the conformation of the design in solution (**Appendix Figure 1.7**).

1.6 Analysis of Rosetta Metrics in Characterized Homooligomer Designs

Progress in protein design will require study not only of the successes but also the failures. The results reported in this paper provide a valuable resource for understanding failure modes as the input scaffolds are all very stable designed proteins (in previous design studies, the often unknown stability of the starting native scaffolds and the robustness to amino acid substitutions were potentially confounding factors). We are able to distinguish distinct failure modes for the designs reported: 32 were not expressed solubly in *E. coli*, 24 adopted multiple oligomerization states, 4 were monomeric, 15 were monodisperse but had an oligomerization state different from that designed, and 6 occupied the designed oligomerization state but had unanticipated configurations based on SAXS data. Analysis of the properties of the design models revealed that designs with (1) a high total charge (greater than -50), (2) small (under 750 $Å^2$) interfaces, (3) poor shape complementarity (sc < 0.625), or (4) for which asymmetric pairwise docking calculations found much lower energy alternative arrangements than the two body interactions in the design model were generally unsuccessful. Furthermore, despite the success with HR00C3_2, designs based on monomers with crystal structures had higher success rates (19%) than those based on monomers validated only by SAXS (4%). The fraction of designs experimentally confirmed to be in the designed state increases from 15/96 in the overall

population to 14/45 restricting to models that satisfy the above criteria (low electrostatic repulsion, larger shape complementary interfaces, absence of much lower energy competing dimeric states, and crystallographically validated monomer structures). Evidently, we currently understand some but not all the factors determining the accuracy of the design calculations -- as this is clearly an important area for future investigation, we provide all of the experimental data for both unsuccessful and successful designs, the design sequences, and a variety of metrics computed from the models (**Appendix Figure 1.2**). The design success rate also clearly decreases with increasing oligomerization state -- indeed there were no successes with hexamers. Higher oligomerization states present several challenges: an increase in translational entropy loss (formation of 3 dimers from 6 subunits results in 3 independently translating bodies, whereas formation of a single hexamer results only in one), an increase in electrostatic repulsion, and a decrease in the difference in interface geometry between competing alternative oligomerization states (smaller reorientations are required to convert a pentamer to a hexamer than a dimer to a trimer). There are clear ways forward to address the second and third challenges: the total charge of the designs can be adjusted to be close to zero at pH 7.0 by suitable redesign of the surface (although some experimentation may be required to maintain solubility), and employing hydrogen bond networks[61] could provide the conformational specificity required to distinguish between higher order oligomerization states.

1.7 Stability and Extensibility

The repeat protein scaffolds used to construct the designed oligomers are very stable proteins, and thus guanidine denaturation can be used to probe the stability of the designed

interfaces independent of effects on the monomers. Four designed oligomers (one selected from each symmetry C2-C5) were purified in an initial round of IMAC and SEC, and subsequently run through SEC-MALS in TBS supplemented with 1M or 2M GuHCl. In both conditions, all four designs remained in their designed oligomeric state (as determined by MALS) without indications of smaller assembly formation (**Appendix Figure 1.8**).

To explore the modularity of the designs and the robustness of the designed interfaces, we extended two of the designed oligomers by appending two additional repeats to the original constructs. Extended versions of ank1C2_1 and HR04C4_1 were expressed and characterized as described above. SEC-MALS traces of the long constructs show the expected shifts to larger apparent sizes compared to the original constructs (**Figure 1.4**, third column), and the calculated molecular weights are close to those expected. Experimental SAXS profiles of the extended designs are in good agreement with the extended computational models (χ values are given in **Appendix Table 1.3**, see **Appendix Methods** for extension model generation) suggesting that the supramolecular arrangement of the subunits is maintained upon extending the scaffold protein. This ability to maintain oligomer geometry while extending the length of the monomers will be very useful for systematically varying the distance between binding moieties and for nanomaterial design.

1.8 Adaptation to Pseudosymmetric Interface Design

With the strategy developed to generate novel protein oligomers that possess cyclic symmetry, a natural goal was to extend the method towards heteromeric complexes. By restricting the subunits to be symmetrically related at the backbone level but subjected to

asymmetric interface design, pseudosymmetric complexes could be generated while keeping each subunit uniquely addressable. Designed proteins with these properties would be particularly useful as building blocks for hierarchical heteromeric protein complexes, resembling the architectures of many naturally-occurring viral capsids for example[62]. A set of pseudosymmetric heterodimers were designed using the previously described set of monomeric repeat proteins with a C2-specified docking search followed by a RosettaScripts[57,62] interface design protocol adapted from the cyclic oligomer design process but fully asymmetric interface design was allowed. Sequences for these final designs can be found in **Supplementary Table 1.1**.

1.9 Biophysical Characterization of Pseudosymmetric Heterodimers

Synthetic genes encoding each of the designed two-component heterodimers were ordered and constructed in the pET28-NESG vector and cloned into BL21 (DE3) *E. coli* competent cells. A second ribosome-binding site was inserted between the open-reading frames of each of the oligomer components ("TAAAGAAGGAGATATCAT"), such that the two proteins would be co-expressed in the same cell. Only one component was specified to possess a terminal $(His)_6$ tag for expression, purification (Ni^{2+} IMAC), and co-elution screening by SDS-PAGE. The $(His)_6$-tagged component was intentionally extended by +2 internal repeat modules to clarify co-elution by SDS-PAGE. Designs that presented two polypeptide species by SDS-PAGE were further purified by SEC, molecular weights of assemblies and components were measured by SEC-MALS and liquid-chromatography mass spectrometry (LC-MS), and supramolecular configurations were assessed by SAXS.

Of the 60 experimentally-tested designs, 38 were found to associate into multimers (co-elution by SDS-PAGE), and 10 formed monodisperse species in solution with the predicted polypeptide masses as confirmed by SEC-MALS and LC-MS, and experimental SAXS profiles that demonstrated agreement between the scattering profiles computed from the design model (**Appendix Figure 1.12**). A crystal structure of one designed interface heterodimer, ank3C2_het1 was obtained by the Garcia Laboratory at Stanford University using the Stanford Synchrotron Radiation Lightsource, and refined to a resolution of 2.5Å. When compared to the computational model, the backbones and side chain rotamers of the hydrophobic residues are similar - the heavy atom r.m.s.d. between structure and model was 0.6Å (**Appendix Figure 1.13**).

1.10 Discussion on Two-Sided Symmetric Interface Design

These results show that oligomeric protein complexes with cyclic symmetry can be generated from repeat protein building blocks by computationally designing geometrically complementary, low-energy interfaces. A key advance is the new fast method for assessing designability that provides a reasonable estimate of the energy obtained after a full atom combinatorial sequence design calculation with roughly six orders of magnitude less computational cost. This allows exhaustive evaluation of the possible cyclically docked configurations of a monomer, which would not be possible with a combinatorial, all-atom sequence design calculation. The broad applicability of the computational pipeline developed here is highlighted by the number of successful designs and symmetries (C2-C5). **Appendix Figure 1.11** provides an overview of all of the experimentally validated dimers, trimers,

tetramers and pentamer - the broad range of structures and the variety of interface geometries and architectures far exceeds that reported in any previous study (beta propeller designs are shown for comparison[33,41]). The combination of RPX search for designable interfaces followed by Rosetta all atom design calculations can clearly generate a wide range of new interfaces involving three to five alpha helices; the ability of the approach to design new beta sheet and loop containing interfaces is an area for future investigation.

The design pipeline presented here can be combined with the modularity of computationally designed repeat proteins to control the three-dimensional arrangement of the protomers at multiple length scales. While the designed interfaces control the nanoscale three-dimensional arrangement, extensions of the repeat proteins allow for the placement of functional motifs with sub-nanometer resolution in each of the interacting proteins. Designed proteins can remain folded under strongly denaturing conditions[41], and the design process provides unparalleled control over their geometry and amino acid composition allowing for reactive chemical moieties, such as thiols or aromatic rings, to be reserved to engineer function in downstream applications. An immediate use for these designed oligomers is to probe how the geometry and valency of tethered signaling molecules affects the clustering of receptors and cellular response. The relationship between ligand valency, spatial orientation, and signaling outcome is not well understood, and designed homoligomerization with systematically tunable lengths should be very well suited for investigating this and other basic biochemical questions.

Figures

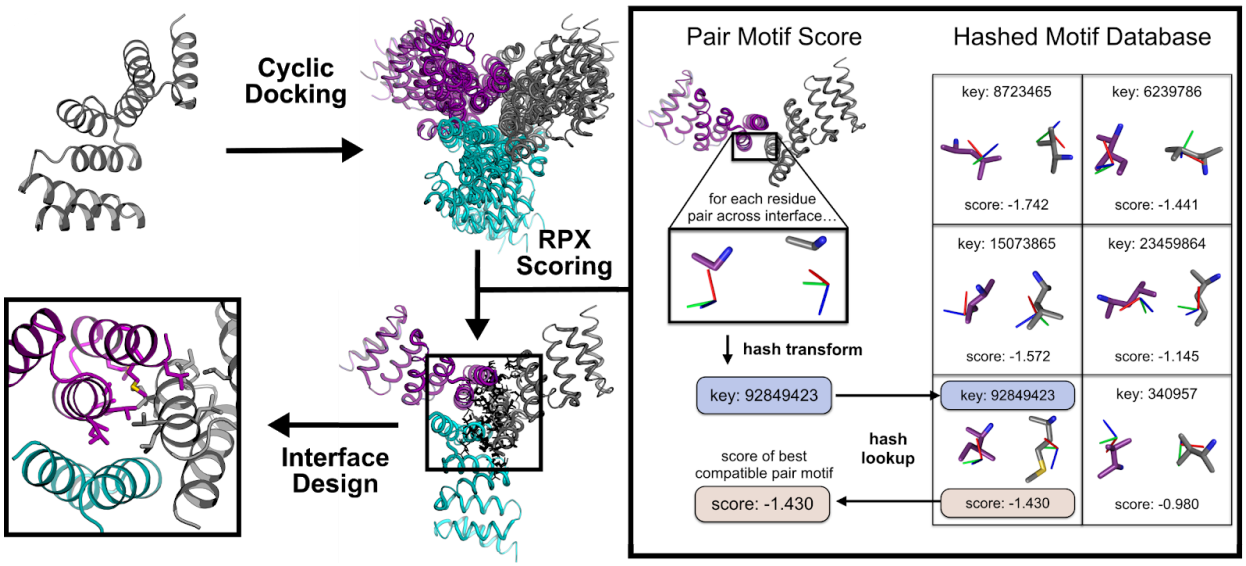


Figure 1.1. Computational design protocol for cyclic oligomers. Left, starting with a monomeric protein we exhaustively sample cyclic docked configurations, score them using the RPX method and generate sequences to drive the complex formation using a full atom RosettaDesign[48] calculation. Right, schematic representation of the RPX model scoring procedure.

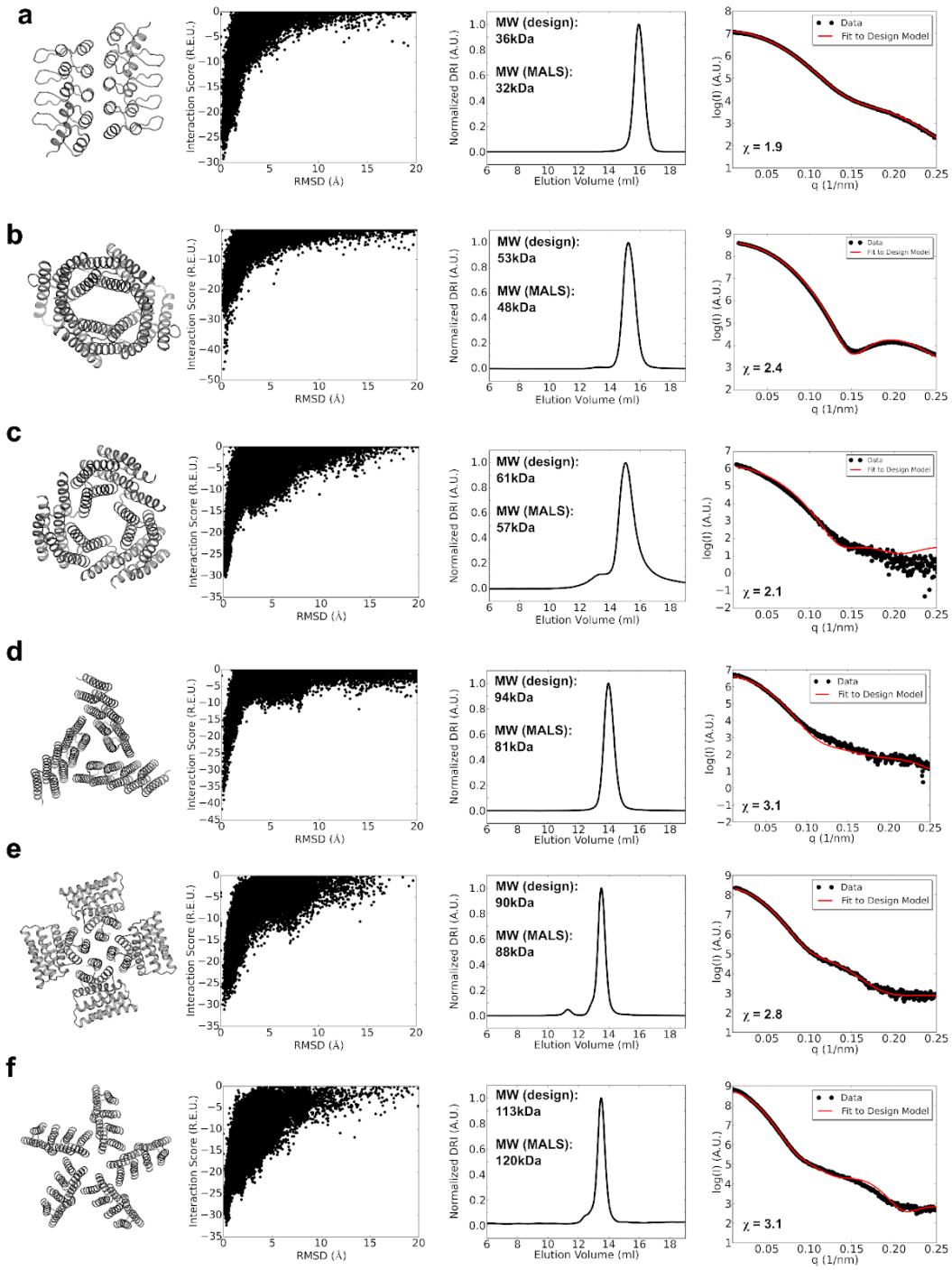


Figure 1.2. Assessment of the solution conformation of selected cyclic oligomers. From left to right: computational model, symmetric docking energy landscape, SEC chromatogram used for molecular weight determination, and SAXS scattering profiles experimentally measured (black dots) and computed from the model (red line). "MW (design)" refers to the molecular weight of the oligomer design and "MW (MALS)" refers to the experimentally determined molecular weight. **a,** ank3C2_1. **b,** HR79C2. **c,** HR08C3 **d,** HR00C3_2. **e,** HR04C4_1. **f,** HR10C5_2. Data for the nine other successful designs are provided in **Appendix Figure 1.5**.

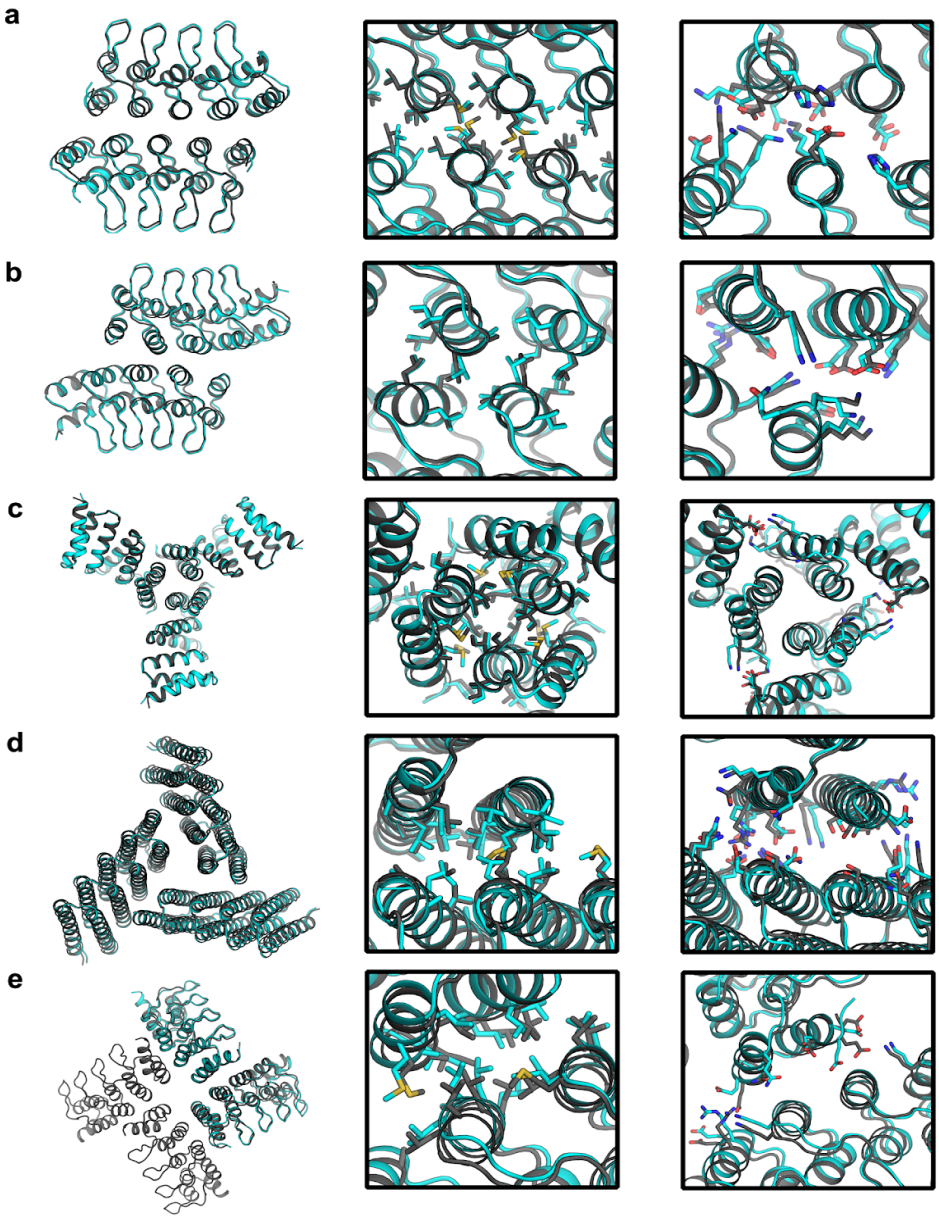


Figure 1.3. Comparison between experimentally determined crystal structures and corresponding design models. Crystal structures are shown in cyan and models in gray. Left column, full model and crystal structure superposition; Right column, superposition showing hydrophobic side chains at the designed interface. **a,** ank3C2_1 (r.ms.d. to model 1 Å) **b,** ank1C2_1 (r.ms.d. to model 0.9 Å) **c,** 1na0C3_3 (r.ms.d. to model 1 Å) **d,** HR00C3_2 (r.ms.d. to model 0.9 Å) **e,** ank1C4_2 pair of chains (r.ms.d. to model 1.1 Å).

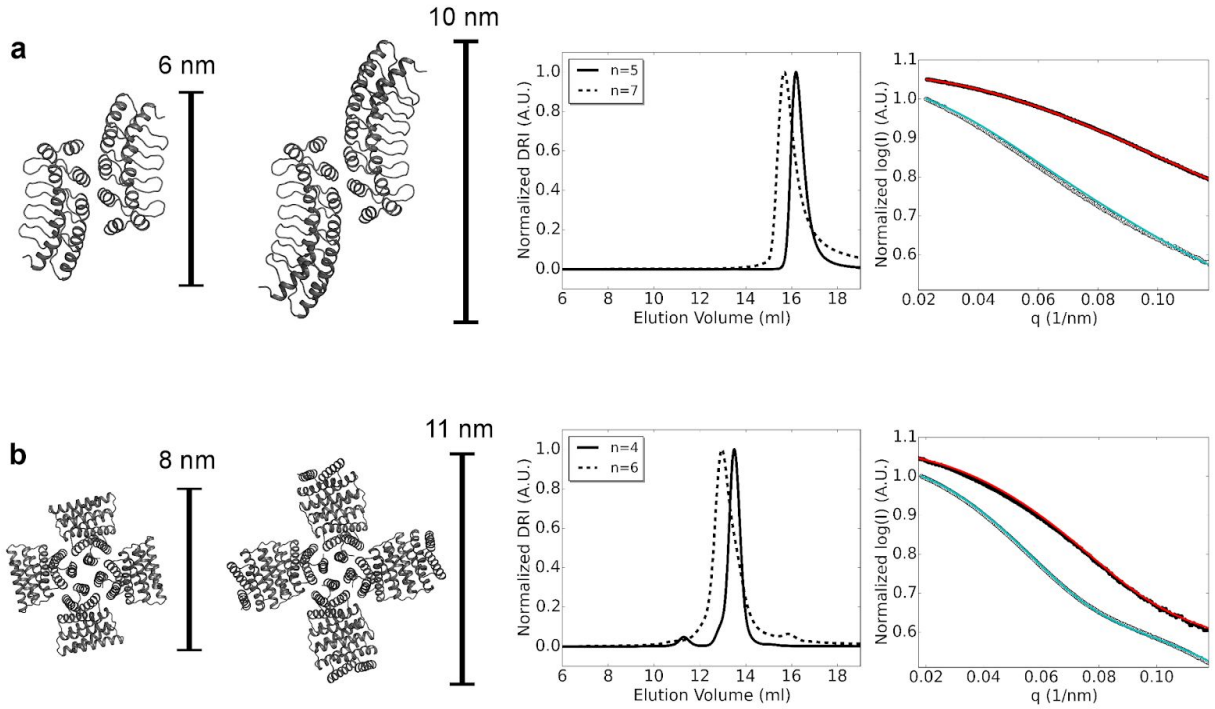


Figure 1.4. Designed oligomer subunit extensions. Top panel from left to right: computational model of the design (n repeats), model of the extension (n+2 repeats), SEC-MALS chromatogram (original design: solid line; extended design: dotted line), SAXS profiles (original design: experimental data in black circles and computed profile in red; extended design: experimental data open circles and computed profile in cyan). **a,** ank1C2_1. **b,** HR04C4_1.

Tables

Table 1.1. Summary of experimental results for the designed cyclic homooligomers.

Symmetry	Designs	Soluble Expression	Target Molecular Weight	Structural Validation
C2	11	11/11	7/11	5/11
C3	34	20/34	6/34	6/34
C4	19	13/19	6/19	3/19
C5	17	9/17	1/17	1/17
C6	15	11/15	1/15	0/15
total	96 (100 %)	64 (67 %)	21 (22 %)	15 (16 %)

Chapter 2. Targeted Design of Protein-Based Vaccines

2.1 Introduction to Virus-Like Particles as Vaccine Candidates

Some of the earliest vaccines, and several still today, are classified as killed or inactivated viruses in which native-like viruses are defunctionalized through heat or chemical denaturation methods[63]. Although the resulting products contain linear epitopes that might resemble those found from natural viruses, they are not necessarily delivered in their native conformations. Efforts to develop conformationally relevant forms of fusion proteins that fold independently from other viral machinery have been pursued to induce antibody repertoires capable of recognizing key structural features of antigen complexes[17,64,65]. An ongoing challenge has been to develop platforms capable of presenting idealized or engineered immunogens in a strategic pattern to the immune system, such that a more broad and potent neutralizing response can be triggered by virtue of B cell receptor (BCR) crosslinking[12]. Recombinant protein vaccines largely aim to accomplish this by displaying target immunogens in a symmetric array whilst not presenting a significant likelihood of adverse effect[14,66]. Recent reports have suggested the added feasibility of using non-viral scaffolding nanoparticles, such as self-assembling ferritin, to act as platforms for antigens such as influenza HA[67,68]. This particular example resulted in greater protection against virus challenge in animals than with the corresponding inactivated virus, and may represent a general strategy to confer advantageous biophysical characteristics compared to conventional viral vectors such as increased stability, avidity, and ease of production.

Although recombinant vaccine strategies have been attempted in the past, cases where target antigens have been successfully fused to an underlying scaffold and yielded well-behaved

multivalent carriers are rather serendipitous - often times there are issues expressing the fusion constructs or they are found to be prone to non-specific aggregation. Among several factors, one of the primary causes for fusion incompatibility lies in the arbitrary positioning of the polypeptide's free termini available for fusion. This observation prompted investigation into whether recent methods for symmetric protein oligomer design could be adapted to develop a new generation of geometrically-targeted complexes built in a customized antigen-specific manner. We set out to generate a set of targeted homotrimers with the capability of scaffolding viral fusion domains with shared C3-symmetry (**Figure 2.1a,b**). These trimers were then intended to be used as components for higher order two-component complexes[6,7], capable of presenting 4-20 copies of the antigenic protein complex on the same particle (**Figure 2.1c,d**). The human immunodeficiency virus spike protein (BG505 SOSIP)[15], influenza hemagglutinin (H1 HA)[16], and an engineered respiratory syncytial virus F glycoprotein (DS-Cav1)[17] complexes were selected as viral antigen targets for multivalent display, with intentions of introducing a sufficient degree of platform versatility and applicability across different disease spaces from a shared set of underlying designs.

2.2 Targeted Docking and Design Against Trimeric Antigens

A selected set of designed repeat proteins (shared from chapter 1 in this document) were used as scaffolds for symmetric trimer design and combined with a targeted computational search against viral fusion proteins: human immunodeficiency virus BG505 SOSIP, influenza HA, and respiratory syncitial virus DS-Cav1 (PDB IDs 5VJ6, 5W5S, 5TNP). These fusion proteins were selected due to their C3 symmetry (making them ideal candidates for symmetric scaffolding), availability of high-resolution structural data, and variety in termini geometries to

select for a diverse set of oligomeric scaffolds. In order to maximize the likelihood of yielding well-behaved trimers, a C3-specified docking search using the repeat proteins was first performed and output configurations were assessed by the previously described RPX scoring method for interface design[11]. Up to the top-scoring 100 docks for each scaffold were then aligned against each of the three antigen complexes along the shared axis of symmetry and sampled translationally along and rotationally about the axis in 1 Å and 1° increments respectively. For each sample, the distance between the C-terminus of each antigen structure and N-terminus of the docked trimer was measured until a minimum, non-clashing distance was determined (**Appendix Figure 2.1**). Eight unique repeat protein scaffold monomers were found to generate C3-symmetric trimeric configurations with adequately small distance solutions between termini of the scaffold and target antigen. Non-redundant alignment solutions with <15 Å between the termini of any one antigen were selected for full Rosetta C3-symmetric interface design and computational filtering using an input .pdb file containing a single subunit with the cyclic axis aligned to the vector [0, 0, 1] as well as a cyclic symmetry definition file as described in previously published methods (see section 1.4)[11]. Following design selection and manual refinement, a set of 23 designed trimers met our computational screening criteria and exhibited a variety of geometric configurations. Sequences for the designed trimers selected for characterization can be found in **Appendix Table 2.1**.

2.3 Biophysical Characterization of Antigen-Targeted Trimers

Synthetic genes encoding each of the designed trimers were purchased and constructed in the pET21-NESG vector which includes a standard T7 promoter system and C-terminal $(His)_6$ tag, and cloned into BL21 (DE3) *E. coli* competent cells and expressed by galactose-mediated

autoinduction overnight. Proteins were isolated by Ni^{2+} immobilized metal affinity chromatography (Ni^{2+} IMAC) from lysate obtained from microfluidized cell pellets and further purified using size-exclusion chromatography (SEC; failed design SEC chromatograms presented in **Appendix Figure 2.3**). 21 designs were found to express in the soluble fraction and 11 were determined to form the intended oligomeric state as assessed by size-exclusion chromatography in tandem with multi-angle light scattering (SEC-MALS; examples in **Figure 2.2** top panel, second row; metrics in **Appendix Table 2.3** top panel, left; SEC-MALS chromatograms acquired for other designs presented in **Appendix Figure 2.2**). Small-angle X-ray scattering (SAXS) further confirmed 4 of these designs to occupy a supramolecular configuration close to the designed model (**Figure 2.2**. top panel, third row; metrics in **Appendix Table 2.3** top panel, right), and were carried forward into crystallization screens (Index, PEG/Ion Hampton Research, and Morpheus Molecular Dimensions). Design 1na0C3_2 was found to crystallize in 1M LiCl, 100mM citrate, 20% w/v PEG 6000, pH 4, and was frozen using 25% glycerol as cryoprotectant. Design 3ltjC3_1 crystallized in 1mM DL-glutamic acid monohydrate, 100mM DL-alanine, 100mM glycine, 100mM DL-lysine monohydrochloride, 100mM DL-serine; 100mM Tris (base), 100mM BICINE, 20% v/v ethylene glycol, 10 % w/v PEG 8000, pH8.5. Diffraction data for each of these designs were collected at the Advanced Light Source (Beamline 8.2.1) at Lawrence Berkeley National Laboratory in Berkeley, California and crystal structures were solved for the two designed trimers. The structures for 1na0C3_2 and 3ltjC3_1v2 were refined to 2.6 Å and 2.3 Å resolution demonstrating a full heavy-chain R.M.S.D between the designed models and structures of less than 1.4 Å and 0.8 Å respectively. Crystallographic metrics and structure-to-model comparisons can be found in **Appendix Table 2.4** and **Appendix**

Figure 2.4 respectively.

2.4 Antigenicity of Trimeric Fusions

Provided the degree of structural validation with the trimer designs, one construct was selected for assessment of immediate antigen-scaffolding capabilities. The designed trimer, 1na0C3_2, was selected as first test-cases for fusion to DS-Cav1 engineered to maintain its prefusion conformation[17]. The fused antigen-trimer construct was found to secrete out of HEK293F cells, and was assessed for its ability to present DS-Cav1 in its conformationally-relevant state. A conformationally-specific antibody Palivizumab[69], also an approved prophylactic for RSV, was used in a dilution-series ELISA assay demonstrating a substantial response for cell supernatant with secreted DS-Cav1-1na0C3_2 compared to DS-Cav1 alone (**Appendix Figure 2.5**). The sequence for this construct can be found in **Appendix Table 2.1.**

2.5 *De Novo* Nanoparticle Generation with Designed Trimers

At least one of the four SAXS-validated trimers (**Figure 2**, top panel) were included as a component for generation of a new set of two-component protein nanoparticles tailored for antigen display across tetrahedral, octahedral, and icosahedral symmetries (architectures) analogous to those previously described. By pairing the trimers with other designed symmetric oligomers[11], components were first docked using the tcdock program[6,7] into two-component architectures. To enrich the variety of oligomeric states in the input oligomers, three additional naturally-occurring homopentamers (PDB IDs 2JFB, 2OBX, 2B98) were included, specifically with the I53 architecture in mind (sequences in **Appendix Table 2.1**). In order to pre-select

nanoparticle configurations with a high number of energetically-favorable side chain contacts for Rosetta interface design, docks were scored and ranked using the RPX method (section 1.3) as opposed to prior methods involving pure interface residue contact count. High-scoring and non-redundant nanoparticle configurations were selected for Rosetta interface design with an added caveat that the included trimer face outward for antigen display. The design protocol took a single-chain input .pdb of each component, and a symmetry definition file containing information for a specified cubic point group symmetry using a Rosetta symmetry definition file[56]. The oligomers were then aligned to the corresponding symmetry axes of the architecture using the Rosetta SymDofMover taking into account the slide translations and rotations retrieved from the .dok file output from tcdock. Each configuration was then subjected to a symmetric interface design protocol similar to previously described (see section 1.4). Designs were selected based on the difference in Rosetta energy units between unbound and bound states of the two subunits, Rosetta shape complementarity (sc) of 0.6 or higher, buried solvent-accessible surface area (sasa) of 700 Å or higher, buried unsatisfied polar groups, and number of total mutations from the input oligomer. Designs that passed these criteria were manually inspected and refined by single point reversions for mutations that were deemed non-contributory to stabilizing the bound state of the interface. The sequence with the best overall metrics for each unique docked configuration was then added to a set of 52 finalized two-component nanoparticle designs to be experimentally characterized. The 11 tetrahedra, 20 octahedra, 21 icosahedra are named according to the following nomenclature: the first letter refers to the symmetry of the nanoparticle, the oligomeric state of the first component (A) and n-fold symmetry axis to which it was aligned is denoted by the first integer, while the second component (B) is denoted by the

second integer. Following are the letters "dn" reflecting the "*de novo"* designed input oligomers used for this study, and an integer denoting the n-th configuration ranked by RPX score from the docking stage (for example, "I53_dn5" indicates an icosahedral nanoparticle comprised of a pentameric and trimeric component, occupying the 5th highest RPX-scoring nanoparticle configuration). Sequences for the designed nanoparticles can be found in **Appendix Table 2.2**, also noting each of the oligomeric components used to comprise the design.

2.6 Biophysical Characterization of *De Novo* Nanoparticles

Synthetic genes encoding each of the designed two-component nanoparticles were purchased (Gen9) and constructed in the pET28-NESG vector that includes a standard T7 promoter system and C-terminal $(His)_6$ tag and cloned into BL21 (DE3) *E. coli* competent cells. A second ribosome-binding site was included between the open-reading frames of each of the oligomer components ("TAAAGAAGGAGATATCAT"), such that the two proteins would be co-expressed in the same cell. The 52 designs were isolated from lysate (Ni^{2+} IMAC) and 23 co-eluted by SDS-PAGE, suggesting at least a degree of spontaneous co-assembly provided only second (B) components were $(His)_6$-tagged. These designs were further purified by SEC (chromatograms for failed designs presented in **Appendix Figure 2.6**, and molecular weights of assemblies were measured by MALS. 4 designs, T33_dn2, T33_dn5, T33_dn10, and I53_dn5, ran as monodisperse particles of the predicted molecular weight (**Figure 2.2**, bottom panel middle row, metrics in **Appendix Table 2.3**, bottom panel left) and further investigated by solution SAXS experiments. Solution-based scattering curves were produced indicating adoption of the predicted supramolecular configuration as compared to the especially-distinctive computed scattering curves derived from the models[59] (**Figure 2.2**, bottom panel bottom row;

metrics in **Appendix Table 2.3**, bottom panel right).

2.7 Structural Characterization of *De Novo* Nanoparticles and Antigen-Fused Counterparts by Electron Microscopy

Particle field images were collected for each of the 4 designed nanoparticles by negative-stain electron microscopy (NS-EM) as described in previous methods[70,71]. Approximately 1000 particles were manually picked from the micrographs of each design, and two-dimensionally classified using the Iterative MSA/MRA algorithm (see **Appendix Methods** for methods). Three-dimensional classifications and refinements were continued in Relion/2.1[72], indicating agreement between the designed models and the experimental electron density (**Figure 2.3**).

At this stage, antigen fusions to each of the four designed nanoparticles were investigated for fully-valent particle assembly. Synthetic genes for antigen-fused trimeric components including BG505 SOSIP fused to T33_dn2A, T33_dn10A, and I53_dn5B (various truncations and linker lengths) while influenza HA fused to I53_dn5B (single minimal linker) were transfected, and individually expressed and secreted out of HEK293F mammalian cells (sequences can be found in **Appendix Table 2.2)**. The proteins were then purified using a combination of immuno-affinity and gel filtration chromatography. Partner assembly components were produced recombinantly in *E. coli* as described with previous oligomers. Several reactions containing 5-10 μg of the purified fusion component and equimolar amount of the partner assembly component were prepared and incubated for 24 hours while optimizing buffer and temperature conditions. Assembled particles were then purified using a gel filtration Sephacryl S500 HR column and assessed by electron microscopy similarly to described above.

Three-dimensional classifications for four nanoparticle-antigen fusion complexes: T33_dn2-SOSIP, T33_dn10-SOSIP, I53_dn5-SOSIP, and I53_dn5-HA were obtained indicating a variety of multivalent particles displaying each of the antigen complexes with fairly confined electron density suggesting a low degree of complex flexibility (**Figure 2.4**).

2.8 Discussion on Hierarchical Protein-Based Vaccine Design

A strong activation of B cell receptor signaling is a hallmark of successful immune responses. Nonetheless, the molecular mechanism of this complex phenomenon is not fully understood. Historically, live-attenuated or inactivated viruses and engineered virus-like particles (VLPs) have been able to confer protective immunity without pathogenicity but the empirically-driven development process of these therapeutic agents has hampered studies aimed at understanding the mechanism for eliciting improved antibody titers with molecular detail.

With the need for antigens (and vaccines) to present immunologically-relevant and structurally-defined epitopes, some even stabilized in their native conformation through structure-based immunogen engineering[17,65,73], display of antigenic complexes on multivalent carriers can elicit improved epitope-specific antibody titers by orders of magnitude while providing a systematic platform to correlate functional outcomes with defined structural arrangements of the epitopes. These strategies have been pursued for recombinant protein vaccines, but the first generation of antigen-tailored protein nanoparticles presented in this study represent a new avenue for scaffolding entire antigenic complexes in a way that allows for optimizing the desired response as a function of antigen geometry, supporting the increasingly prominent modality for vaccine development.

The success in the described geometrically-targeted homotrimers and subsequent two-component nanoparticles showcase the capabilities of computational protein design on the nanometer scale from the ground up. Because the employed design methods had previously proven their robustness, the design problem manifested as an investigation of sequential compatibility between previously developed methods[6,7,11]. The symmetric docking and RPX-scoring method were previously described for generating a novel set of *de novo* cyclic oligomers[11], however these initial designs were produced in a configuration-agnostic manner with no intent of any particular biological function. The selective geometric control at the oligomer design stage does allow for preorganization of the viral trimer domains that may ordinarily be conformationally flexible. Of note, the designed DS-Cav1 fusion complex presented in this study displayed enhanced ELISA signal directly out of cell supernatant when compared to the corresponding monovalent antigen (**Appendix Figure 2.5**), which may indicate enhanced expression yields, conformational stability, or some combination of the two. Given the inherent instability generally associated with viral fusion proteins, this observation suggests the feasibility of stabilizing arbitrary domains to the fusion particles for a variety of biological applications. This study demonstrates, both at the oligomer and nanoparticle design stages, the capability of generating biologically relevant materials in a tailored manner through symmetric complex design and scaffolding.

This approach also has several unique features from a biomaterials standpoint: first, the *de novo* repeat proteins are exceptionally stable, conferring stability to a corresponding nanoparticle and conceivably antigen stabilization; second, the initial design step of trimeric units enables for complete control over the positioning of fusion termini amongst a broad set of

trimeric configurations; finally, the sequences have low antigenic complexity by virtue of their short amino acid sequence modules. This latter feature in particular may eventually prove to be advantageous for minimizing adaptive immune repertoire diversity against the underlying scaffold.

Generation of custom-designed display platforms are undoubtedly of interest in the fields of immunology and vaccinology. It is conceivable to imagine these platforms being used in immunogen-scaffolding studies, and provide a basis for understanding how the tunable arrangement of multivalent antigens presented to lymphocytes plays a part in the immune response. Particularly, the degree and nature of receptor cross-linking could be investigated to determine both the intensity and diversity in the antibody repertoire elicited through adaptive immunity. The diversity in unique geometric arrangements of antigens placeable on the presented nanoparticles comprise a unique molecular toolkit for probing the immune response in such a way, and has the granularity to study the system even at the level of signaling activation as a function of B-cell receptor clustering and signal activation.

Figures

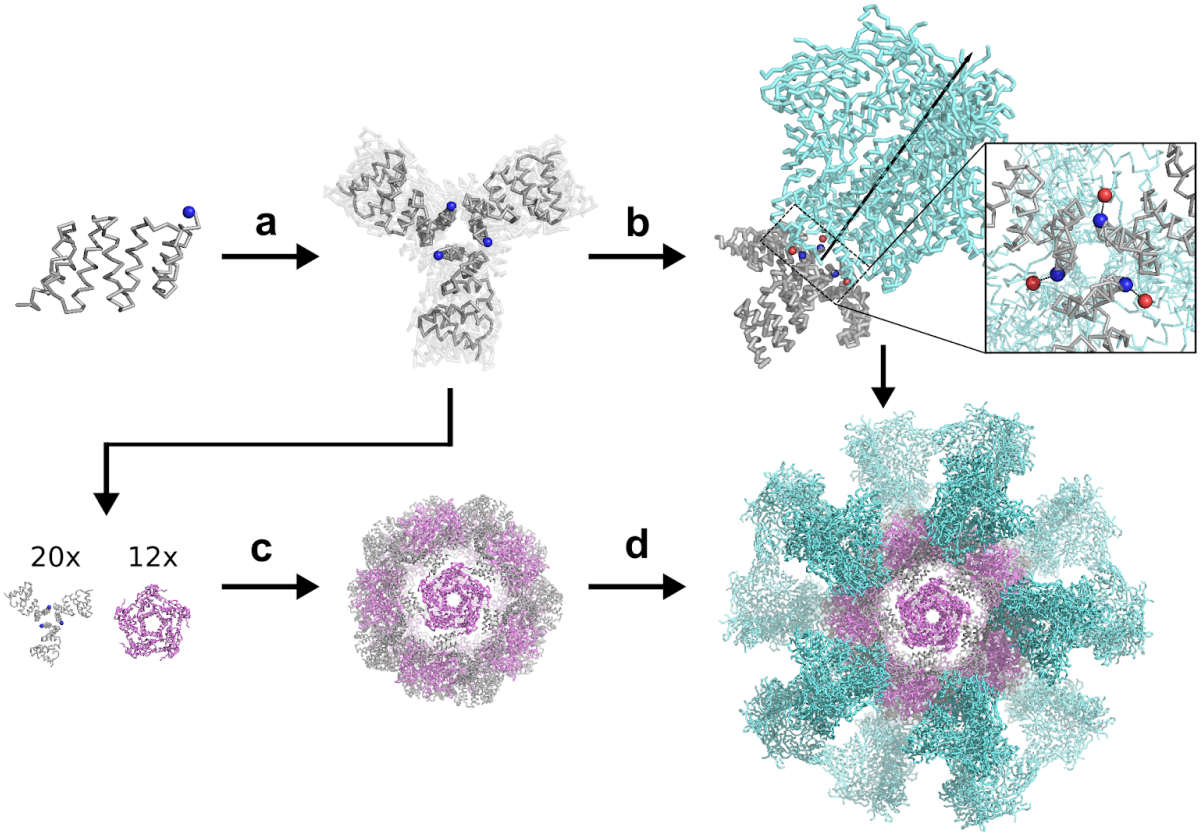


Figure 2.1. Design schematic for multivalent antigen-displaying nanoparticles. a, Computational sampling and designability assessment of trimeric docks from monomeric scaffold. **b,** Selection of trimer docks for Rosetta design and experimental screening based on termini alignment to antigenic target complex. **c,** Pairing of selected trimers with partner oligomers for two-component nanoparticle design. **d,** Replacement of designed trimeric component with antigen-fused counterpart yields a multivalent particle.

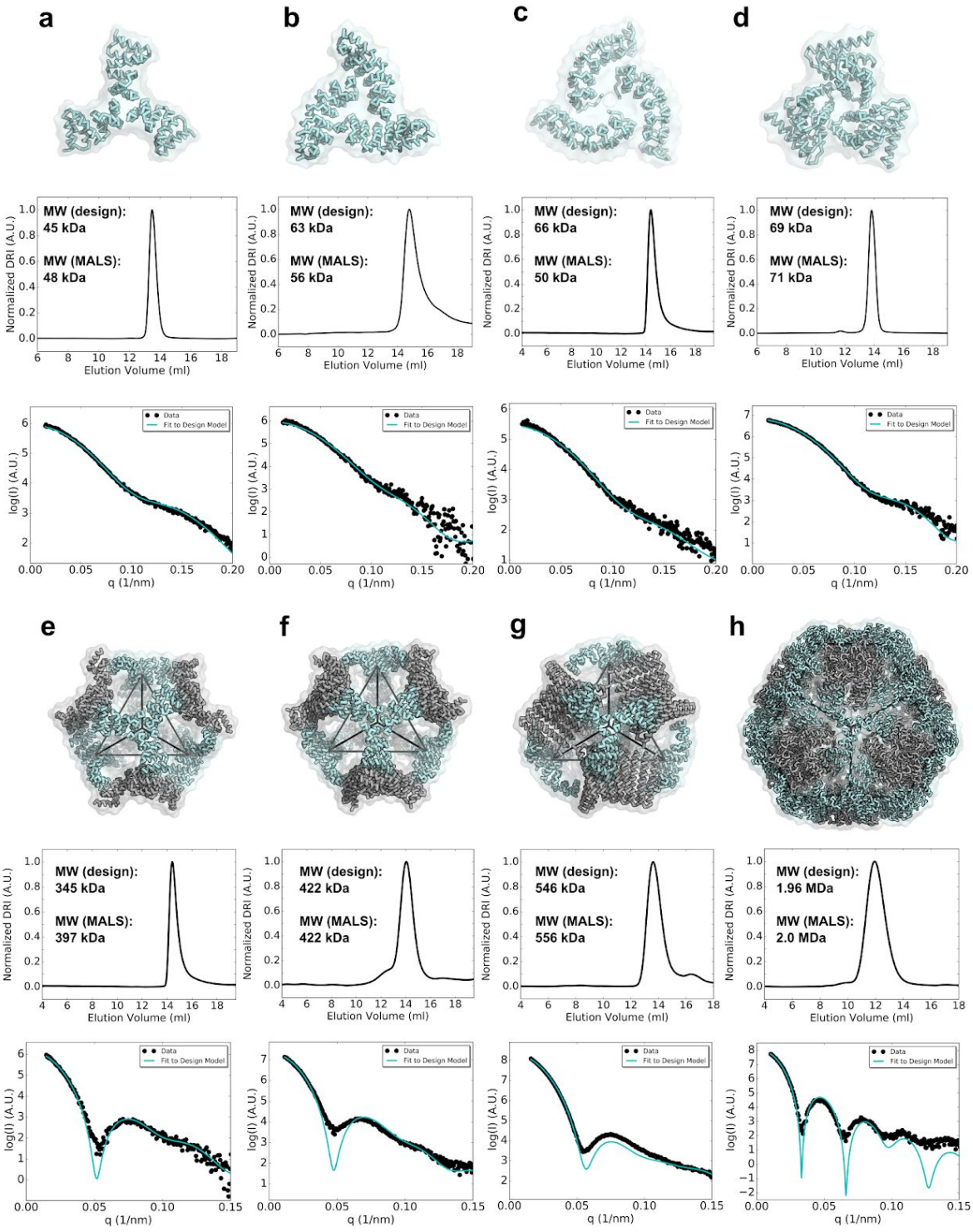


Figure 2.2. Biophysical characterization of antigen-tailored trimers and two-component nanoparticles. Top rows, designed models. Middle rows, size-exclusion chromatograms and calculated molecular weights from multi-angle light scattering. Bottom rows, small-angle X-ray scattering comparisons between experimental data and curve from model. **a,** 1na0C3_2. **b,** 3ltjC3_1v2. **c,** 3ltjC3_11. **d,** HR04C3_5v2. **e,** T33_dn2. **f,** T33_dn5. **g,** T33_dn10. **h,** I53_dn5.

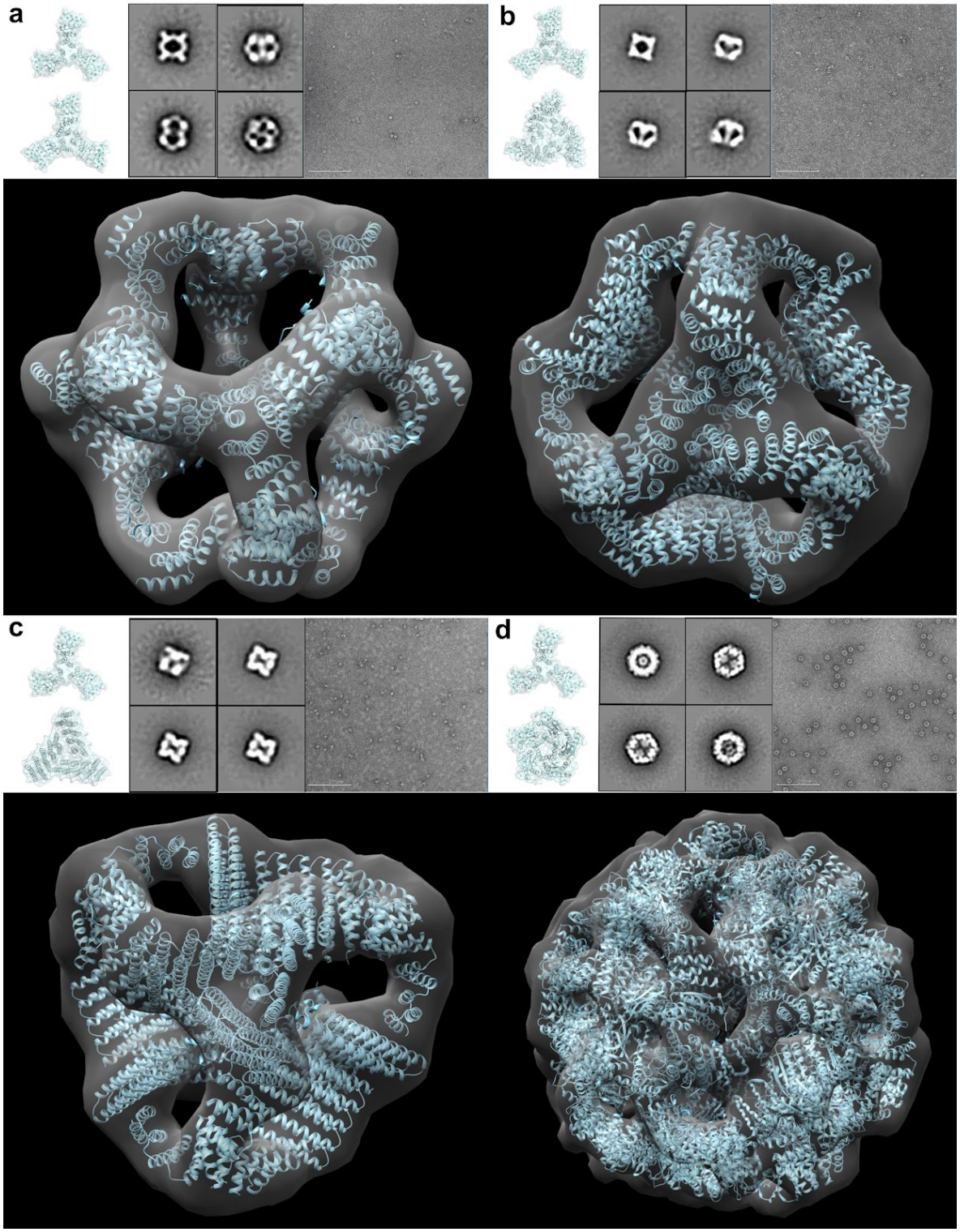


Figure 2.3. NS-EM reconstructions of antigen-tailored nanoparticles. Top lefts, designed trimeric components. Top centers, two-dimensional projection class averages from experimental dataset. Top rights, electron micrograph field image for assembled nanoparticles. Bottoms, three-dimensional reconstruction of electron density for each designed nanoparticle. **a,** T33_dn2. **b,** T33_dn5. **c,** T33_dn10. **d,** I53_dn5.

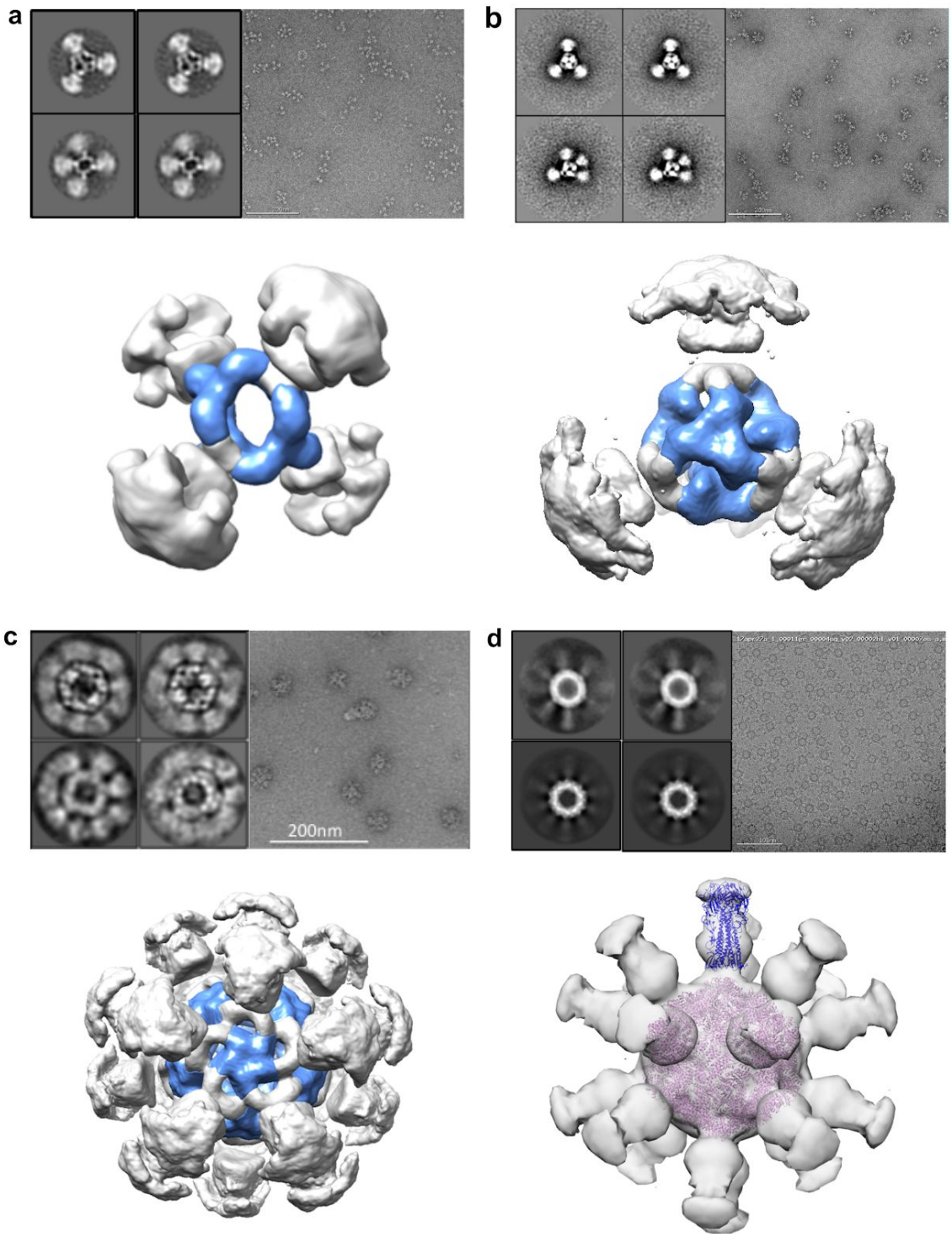


Figure 2.4. Structural validation of antigen-fused particles. Top lefts, two-dimensional class averages with 500-1000 particles picked per class. Top rights, field images for assembled particles. Bottoms, three-dimensional particle reconstructions using over 4000 picked particles each. Nanoparticle designs and fused antigens: **a,** T33_dn2-SOSIP. **b,** T33_dn10-SOSIP. **c,** I53_dn5-SOSIP. **d,** I53_dn5-HA.

Tables

Table 2.1. Summary of the experimental results for the antigen-tailored symmetric homotrimeric proteins and subsequent two-component nanoparticles.

Design	Antigen(s) Targeted	Experimental Molecular Weight (kDa)	Target Molecular Weight (kDa)	SAXS *X* value	Resolution, r.m.s.d. design to crystal structure (Å)
1na0C3_2	HA, SOSIP, DS-Cav1	48	45	1.4	2.6, 1.4
3ltjC3_1v2	SOSIP, DS-Cav1	56	63	1.1	2.3, 0.8
3ltjC3_11	SOSIP, DS-Cav1	50	66	1.6	--
HR04C3_5v2	SOSIP	71	69	1.5	--
T33_dn2	HA, SOSIP, DS-Cav1	397	345	4.8	--
T33_dn5	HA, SOSIP, DS-Cav1	422	422	1.7	--
T33_dn10	HA, SOSIP, DS-Cav1	546	556	2.3	--
I53_dn5	HA, SOSIP, DS-Cav1	2000	1960	1.2	--

Chapter 3. Tunable Synthetic Cell Signaling Agonists

3.1 Introduction to Cytokine Signaling

Cytokines are secreted proteins that mediate a wide range of functions, principally in the immune system, but they also exert many other homeostatic functions that impact mammalian physiology[18,74,75]. As one example, hematopoiesis, the process by which blood cells are formed, is influenced at both early and late stages by a variety of different cytokines that engage and activate specific cell surface receptors expressed on different lineages of cells[76]. While there exist almost 50 different cytokines and cytokine receptors encoded within the human genome, they generally signal through a similar paradigm of receptor dimerization, which results in transphosphorylation of JAK/TYK kinases, STAT activation, and subsequent propagation of a phosphorylation cascade to induction of gene expression programs[75,77]. A wealth of literature from structural biology of cytokine-receptor complexes, and chimeric receptor functional experiments has demonstrated that a diverse range of receptor dimer geometries are compatible with signaling[78–80]. That being said, the actionable value of this natural topological variation for drug discovery has not been explored because of the absence of a suitable ligand engineering platform to exploit this parameter.

Given their powerful actions, many natural cytokines have been investigated as therapeutic agents, but most have severe shortcomings such as toxicity or lack of efficacy, which are contributed by their pleiotropic effects on many different cells and by stimulation of multiple downstream signaling responses[20,21]. This is not surprising given that cytokines evolved to regulate highly complex and finely balanced *in vivo* homeostatic functions, and not to be used as

drugs. Nevertheless, cytokines evoke tantalizing possibilities as therapeutics if their actions can be harnessed, as exemplified by recent advances for cytokines such as Interleukin-2 and IL-10 through half-life extension[26,81], which still does not alter the signal of the natural cytokine. Recent studies, using both natural and engineered ligands to both cytokine and receptor tyrosine kinase class receptors, have begun to reveal that signal strength can be modulated downstream of dimeric receptors in a manner analogous to GPCRs[27,82–84]. Furthermore, natural and synthetic mutants of cytokines such as IL-2, hGH, SCF and EPO have been found to exhibit partial and biased signaling properties[25,27,85]. By analogy, tunable signaling through GPCRs has been exploited as a drug discovery strategy using medicinal chemistry, which is not feasible for type-I transmembrane receptors that bind to protein ligands through large extracellular domains (ECDs), such as cytokines. Thus, new ligand engineering strategies are needed to exploit the potential of rheostat-like signaling behavior for dimeric cytokine receptors in a manner analogous to GPCR medicinal chemistry.

One mechanistic finding that merits attention is that the overall topology of the cytokine receptor dimer can influence the downstream signal[78,86,87]. For example, intact antibodies that bind with two ‘arms’ to receptor extracellular domain exhibit highly variable agonistic properties[88,89]. The flexibility of the Fab connection to the Fc has precluded a structural understanding of the relationship between dimer topology and signaling by agonistic antibodies. Furthermore, diabodies, which are less flexible than antibodies, can act as surrogate cytokine agonists[90,91]. However, antibody-based surrogate agonists are not ideal for this strategy because the precise topology of the dimer cannot be controlled.

Here we offer one solution to this problem by integrating a previous protein binding scaffold technology, termed DARPins (Designed Ankyrin Repeat Protein)[24], with protein design and engineering approaches to create a self-assembling, rigidly-connected dimeric scaffold that enables precise control of the relative orientations of the receptor ECDs in a dimeric complex (**Figure 3.1A**). We focused our efforts on the erythropoietin receptor (EpoR), which plays a critical role in erythroid lineage commitment and differentiation to produce mature red blood cells[25]. We find that systematic variation of angular and distance parameters of the agonists showed clear evidence of partial and biased agonism of the EpoR signal. Furthermore, this variation in topology has allowed us to selectively activate EpoR signaling and erythropoiesis in a hematopoietic stage-specific and tunable manner, and thus decouple the activities of a specific cytokine receptor in a particular cellular context. Collectively, these studies offer a general strategy for tuning cytokine, and potentially other receptor, signaling outputs through topological control.

3.2 Selected Binder Against the Erythropoietin Receptor

Our overall strategy was to engineer a high-affinity protein binding module to a cytokine receptor ECD and subsequently design the module to form a range of dimeric geometries that would compel the bound receptor to dimerize in corresponding alternative topologies (**Figure 3.1A**). We focused on human EpoR as our cytokine receptor target for topological tuning. The EPO/EpoR system has been extremely well characterized with respect to molecular mechanisms of signaling and function, and also offered readily performed functional assays in order to test the effects of our ligands in a physiological context. We chose DARPins as the binding scaffold because of their structural modularity[24], as well as previous success in designing self-assembling

DARPin oligomers[11]. A DARPin molecule typically consists of "X" number of ankyrin repeats flanked by N- and C-terminal capping regions, resulting in "NXC" DARPins. Each ankyrin repeat is formed by 33 amino acids and consists of a β-turn followed by two anti-parallel α-helices. Repeat proteins provide an ideal scaffold for the generation of high affinity ligands and size variations due to their modular nature. We developed two yeast-displayed DARPin libraries – N2C and N3C (**Figure 3.1B; Appendix Figure 3.1**). The variable positions in the ankyrin repeats are based on the consensus sequence[92,93], which allows modifications at seven positions per ankyrin repeat (depicted by blue residues in **Figure 3.1B**), whereas the rest of the backbone residues are crucial to maintaining the DARPin structure. Yeast-displayed DARPin libraries were screened against EpoR ECD resulting in a set of low affinity ligands with dissociation constants (K_d) in the μM range (**Figure 3.1C; Appendix Figure 3.2**). To further enhance the affinity, a combination of random mutagenesis by error-prone PCR and DNA-shuffling techniques were employed, which resulted in a $>10^4$-fold increase in affinity of individual clones (**Figure 3.1C,D**). The highest affinity clone, N3C - E2, was expressed in *E. coli* and its affinity measured by surface plasmon resonance (SPR, see **Appendix Methods**). E2 binds to EpoR ECD with a 1.89 nM K_d providing a >1000-fold increase in binding affinity relative to 8A9, the highest affinity first-generation EpoR binding DARPin (**Figure 3.1D**).

We determined the crystal structure of the highest affinity N3C clone, E2 (**Figure 3.1E-G; Appendix Table 3.1**). E2 binds to the EpoR membrane-distal D1 domain with a 1:1 stoichiometry, at a site different but overlapping with the EPO binding site. E2 engages the top of the D1 domain using the β-turn loop regions while the helical region packs against the side of the D1 domain with intimate shape complementarity and extensive buried surface area.

3.3 Hybridization with Designed Dimer

Dimeric DARPin variants were generated by mutations on the non-binding face of the DARPin (formed by rear helices)[11], which create a homodimer interface, and result in the formation of non-covalent obligate homodimers, A_R3 and C_R3, where Rx indicates the number of ankyrin repeats, (**Figure 3.2A,B; Appendix Figure 3.3 and 3.4**). The A and C dimers differ by a shift in the relative orientation of the monomers across the C2 axis: dimer C_R3 has an offset of one repeat at the dimerizing interface compared to dimer A_R3, providing a twist of 30° in the relative orientation of the two dimerized EpoRs (**Appendix Figure 3.5**). The formation of the C_R3 DARPin dimer and a 2:2 C_R3/EpoR complex was confirmed by size exclusion chromatography, and X-ray structures determined to 1.2 Å and 3.2 Å resolution, respectively (**Figure 3.2B**, crystallography metrics in **Appendix Table 3.1**). We assessed the ability of the DARPin dimers to dimerize EpoR on the surface of live cells using single-molecule TIRF spectroscopy. Like EPO, both ligands showed ligand-dependent dimerization of EpoR, although C_R3 was less efficient than A_R3 (discussed below, **Figure 3.2C**).

3.4 Signaling Activity with Dimeric Binders

Dimerization of EpoR initiates a cascade of phosphorylation events, principally activation of STAT5, and ERK, and to a lesser extent STAT1 and STAT3. We analyzed the phosphorylation levels of various downstream effectors on UT7 cells[94] using phospho-flow cytometry. Monomeric DARPins do not dimerize EpoR thereby generating no pSTAT5 signal. Both DARPin dimers induce similar E_{max} for STAT5 and ERK. Interestingly dimer A_R3 exhibits a higher E_{max} for STAT1 and STAT3 signals, even higher than EPO, whereas dimer

C_R3 exhibits a lower E_{max}, thus acting as partial agonists for these second messengers (**Figure 3.2D**). Thus, even the relatively small orientational differences between dimers A_R3 and C_R3 result in apparent signaling bias.

3.5 Effects on Cell Signaling Through Extension Series

We chose to characterize the receptor dimerization topology using two parameters: scissor angle (rotation) and proximity (distance). We designed three different DARPin dimer extension series to interrogate signaling as a function of these parameters in a systematic manner. The "angle" series is intended to pivot the EpoR dimer in a scissor-like manner (**Figure 3.3**, top). The "distance" series is intended to separate the EpoR subunits in the dimer while not significantly changing the dimer angular relationship (**Figure 3.3**, middle). The "mono-extended" series differs from the angle and distance series in that both EpoR subunits are bound to one extended DARPin (monomeric) that contains two EpoR binding sites (**Figure 3.3**, bottom). Systematic and predefined variation in each of these parameters can be achieved by the insertion of non-binding 'P' repeats derived from the consensus ankyrin repeat sequence.

The E2 DARPin comprises three EpoR binding ankyrin repeats, denoted as "EEE", together with the N- and C-terminal capping regions. Thus, E2 is coded as NEEEC. In a repeat protein such as a DARPin, the ankyrin repeats serve as building blocks. Here, we use two types of blocks, E and P, to modify the architecture of each DARPin in two unique extension series. Furthermore, we can derive different dimer geometries by utilizing the non-planar, helical topology of the DARPin molecules[95]. To engineer the "angle" series, one or more P repeats can be inserted at the C-terminal end to provide EpoR-binding DARPins such as NEEEP..C, (**Figure 3.3**, top; **Appendix Figure 3.6**). The dimerizing residues (indicated by underline) are placed on

the same terminus as the P repeat insertions. As the added repeats grow along a helical path, it results in the generation of a series of complexes with variation in the angle between the two receptor chains. The resultant extended dimer is termed A_angle_Rx, where A indicates the designed dimer interface and "x" indicates the total number of ankyrin repeats. For example, NEEEPPC is termed A_angle_R5. To engineer the "distance" series, the non-binding P repeats are inserted at the *N*-terminus, for example NPP…EEEC (**Figure 3.3**, middle; **Appendix Figure 3.7 and 3.8**). This results in two helical DARPins growing in opposite directions with a 2-fold symmetry, as the dimerizing residues are also placed at the *N*-terminus. With the addition of more P repeats, the binding repeats move further apart, leading to a distance-based series, termed A_dist_Rx, where "dist" indicates distance. Finally, we designed a monomeric, extended DARPin with two binding sites included on the same DARPin chain (**Figure 3.3**, bottom; **Appendix Figure 3.8**), but with the two EpoR in inverted rotational orientations (i.e. "face-to-back" versus "face-to-face"). The two sets of EpoR binding repeats are positioned at the two termini, separated by a series of P repeat insertions that can be varied in number to control the distance between the two receptors, for example NEEEPPPP…EEEC. These monomeric extended DARPins are termed M_Rx, where "x" indicates the total number of ankyrin repeats.

Each of the dimers was expressed in *E. coli* and purified by size-exclusion chromatography, and bound to EpoR with similar affinities, as measured by SPR (**Appendix Figure 3.6 and 3.8**). To validate the designs and also to determine if any clashes are apparent between the EpoR receptors, we determined six crystal structures: the apo C_R3 dimer (1.2 Å resolution) as well as EpoR complexes with C_R3 (3.46-3.16 Å resolution), C_angle_R5 (3.0-2.0 Å resolution), A_angle_R5 (3.39-2.45 Å resolution), A_dist_R7 (5.1 Å resolution), and

M_R12 (4.58 – 3.14 Å resolution) (**Figure 3.3; Appendix Figure 3.9;** crystallography metrics in **Appendix Table 3.1 and 3.2**). The structures validate the DARPin dimer designs, albeit with minor rigid body variations, allowing us to model the subsequent different geometry series based on the known effects of P repeat insertion. In all of the structures determined, while the D1 domain of EpoR maintains a rigid contact with the DARPin, the D2 domain exhibits some segmental flexibility and adapts variable positions relative to D1 that are likely influenced by crystal packing. Thus, the flexibility of the D2 domain prevents us from knowing with certainty its exact location on the cell surface within each dimer, although the flexibility is constrained. We measured a range of 28° variation in D2 position (**Appendix Figure 3.10**), consistent with that observed in previously published structures of EpoR. Nevertheless, our goal was not to use these ligands as molecular rulers, but rather to systematically vary the receptor topologies, and extract correlative structure-activity relationships with signaling output.

The basic building block of the angle series is A_R3 with zero P repeats (NEEEC), and subsequent members with one or more P repeats added in; A_angle_R4 (NEEEPC) through A_angle_R7 (NEEEPPPPC) progressively tilt the EpoR dimer towards the membrane (**Figure 3.4A**). The dimer A_angle series exhibits a progressive reduction in E_{max} levels of STAT5 phosphorylation as the EpoR dimer scissor half-angle increases from approximately 38° (A_R3) to 72° (A_Angle_R9), **Figure 3.4A,B**. Dimers A_angle_R4 and A_angle_R5 act as partial agonists, whereas A_angle_R6 does not elicit a signal. Similarly, the C_angle series shows a stepwise decline in activation (**Appendix Figure 3.11A**). In addition to the reduction in E_{max}, a right shift in EC_{50} is also observed, despite the fact that all of the dimers have the same affinity for EpoR, as measured by SPR using soluble EpoR ECD (**Appendix Figure 3.6**). We speculate

that as the scissor angle between EpoRs becomes more extreme, strain is induced as a result of cell membrane constraints, and this is reflected in less efficient bivalent complex formation. We also analyzed a broad panel of downstream phosphorylated proteins, which showed a similar trend as the more limited analysis of STAT5/1/3 and ERK (**Figure 3.4C,D; Appendix Figure 3.11B**). At very high concentrations, a bias is observed for STAT1 and STAT3 by A_R3. A similar STAT1 bias is observed for C_angle_R4 and C_angle_R5 with STAT1 signal increased by ~10-fold at highest concentrations (**Appendix Figure 3.11**).

The basic building block of the distance series is A_R3 with zero P repeats (NEEEC), and subsequent members incorporate one or more P repeats; A_dist_R4 (NPEEEC) through A_dist_R9 (NPPPPPPEEEC) (**Figure 3.5A**). Contrary to the angle series where a reduction in pSTAT5 and pERK E_{max} is observed, the first few members of A_dist exhibit the same E_{max} despite the increased distance between the receptors (**Figure 3.5B**). Instead, the effects of the distance are reflected in a right shift in the EC_{50} values. Furthermore, a bias towards higher pSTAT1 and pSTAT3 E_{max} than EPO or other distances, is observed for A_dist_R5 (**Appendix Figure 3.12 and 3.13**). Analysis of a panel of downstream phosphorylated proteins reflected the above trends, though a strong bias is observed for STAT1 for the A_dist_R5 ligand (**Figure 3.5C,D**). For the distance series, the right shift of the EC_{50} was surprising, since we would not imagine strain on the dimer through pure translation of the receptors to wider inter-dimer distances unless there was receptor pre-association. While we do not see EpoR pre-dimerization by single molecule TIRF microscopy, increasing the distance may interfere with weak synergistic interactions between the receptor subunits, possibly mediated by the transmembrane domains[96].

In the native EPO/EpoR dimer structure, the receptors are "face-to-face." Having validated the design of a "face-to-back" orientation with the crystal structure of the DARPin M_R12/EpoR complex (**Figure 3.3**, bottom), we wished to assess "face-to-back" orientation whilst maintaining a close inter-dimer distance (**Figure 3.5E**). Thus, we compared the signaling response for A_dist_R6 ("face-to-face" dimer, distance 49 Å) and M_R11 ("face-to-back" dimer, distance 53 Å), where M_R11 is a mono-extended DARPin containing two EpoR binding sites separated by five ankyrin repeats, NEEEPPPPPEEEC. M_R11 delivers an extremely weak pSTAT5 signal (**Figure 3.5F**), suggesting that the face-to-face orientation appears to be favorable for the intracellular JAK2 to trans-phosphorylate within an EpoR dimer, thus more effectively initiating signaling.

3.6 Model Generation of Extension Series

The synthetic EpoR agonists described in this work have a unique property that differentiate them from previous efforts at generating synthetic type-I cytokine agonists, namely they have fairly well-defined and rigid conformations that provide a structural basis for each observed physiological output. This advantage was leveraged by generating a crystallographic structure-informed set of signalling complex models for each of the ANR, ACR, and CNR series. To best recapitulate both the backbone topology of the P repeat and the N-/C-terminal mode of EpoR binding (N- for CR series, C- for NR series), the backbone coordinates from the mono-extended crystal structure were used to construct each experimentally tested member of the extension series using the he Rosetta macromolecular modeling suite (see **Appendix Methods**). While most presented non-clashing complex models upon superposition of the E2-EpoR structure with the extended backbone, several cases showed significant clashes

between EpoR subunits that could not be physically feasible. We used a constrained Rosetta minimization protocol on each of the clashing complex models to generate a complete set of plausible models (see **Appendix Methods**)[97]. The models were then analyzed to quantify the relative orientation between the two EpoR ECD's, and measurements of distance (Å) and angle of spread (°) between ECDs were selected as descriptive parameters.

3.7 *In vivo* Effects with Designed EpoR Mimetics

Erythropoiesis begins with lineage commitment of hematopoietic stem and progenitor cells (HSPCs) into the erythroid lineage, and cells progressively differentiate and mature to become RBCs. This process of RBC development can be faithfully recapitulated *in vitro* by culturing human primary HSPCs with various cytokines over two to three weeks (**Figure 3.6A**)[98,99]. As a main driver of RBC development, the role of EPO during erythropoiesis is not only to upregulate erythroid-specific genes to promote cell maturation, but to also allow commitment by multipotential HSPCs to erythroid lineage differentiation and to prevent cell death by activating anti-apoptotic pathways[100,101]. Therefore, distinct stages of erythroid maturation (from early progenitors to late precursors) may utilize EPO signaling differently and may have variable reliance upon different pathways at specific stages.

Using DARPins that either increased the distance or altered interaction angles between EpoR dimer pairs, we found a graded reduction in overall erythroid differentiation and cellular expansion (**Figure 3.6B, Appendix Figure 3.14**). Looking at overall erythroid differentiation using flow cytometry allowed us to assess the kinetics of HSPC commitment into the erythroid lineage and the continuation of erythroid terminal maturation throughout the culture. In addition, proliferation is one of important hallmarks in erythropoiesis that pairs well with the overall

differentiation efficacy. Therefore, these analyses together in a time-dependent manner can give us further insights into stage-selective dependencies of specific EpoR activities in erythropoiesis. This range of erythroid differentiation efficacy stimulated by various DARPin-series is reflected in the percent of total CD235a^{+} cells throughout the culture. The majority of cells cultured with EPO, A_R3, A_angle_R4, C_R3, and A_dist_R5 differentiated into CD235a^{+} erythroid cells, while the later series (wider angle or farther distance of the EpoR dimer) impaired or failed to effectively promote erythroid commitment and differentiation (**Figure 3.6C**). Consistent with the pattern we observed with flow cytometry for cell surface markers, treatment of cells with agonists that resulted in delayed or impaired erythroid differentiation also resulted in reduced cellular proliferation (**Figure 3.6D**). Interestingly, the cells cultured with EPO, A_R3, or A_dist_R5, at maximally potent concentrations showed very similar erythroid differentiation efficacy on differentiation day 5 (early phase of erythropoiesis), and CD235a^{+}CD71^{+} was 31.7 ± 1.0%, 28.0 ± 1.6%, and 30.9 ± 0.6%, respectively (**Figure 3.6E**). However, on differentiation day 12, cells cultured with A_dist_R5 exhibited 1.7 ± 0.1% CD235a^{+}CD71^{-} cells, while cells with the recombinant EPO and A_R3 showed 28.7 ± 2.0% and 26.8 ± 1.6% in this population, respectively, indicating the impairment of terminal erythropoiesis in the late stage with A_dist_R5 (**Figure 3.6E**). Furthermore, cells cultured with A_dist_R5 demonstrated reduced proliferation, relative to EPO or A_R3, in the later stages of the erythroid differentiation culture (**Figure 3.6D**), demonstrating that A_dist_R5 specifically promotes early erythropoiesis with impairments at the later stages. In addition, C_R3 promoted early erythropoiesis poorly, but showed increased differentiation efficacy in the later stages, (**Appendix Figure 3.14**), suggesting that this ligand selectively promotes the later stage erythropoiesis while demonstrating impaired

stimulation at the earlier stages. Collectively, these results demonstrate how topological alterations in EpoR dimerization can selectively promote effective signaling at specific stages of human hematopoiesis/erythropoiesis but fail to effectively enable differentiation and proliferation at other stages. Critically, we could identify ligands that either fail to effectively promote the earlier stages of hematopoietic differentiation and erythroid lineage commitment, while enabling effective differentiation in the terminal stages or the reciprocal. This finding not only highlights the potential for designed ligands to tune hematopoietic differentiation, but also suggests that aspects of endogenous EpoR signaling may vary in a stage-specific manner during hematopoiesis.

Such compelling phenotypes led us to ask whether there may be biased signaling responses downstream of EpoR that underlie such effects (**Figure 3.6F**). Erythroid cells stimulated with A_dist_R5 showed higher activation of STAT1, STAT3, and STAT5 than with the recombinant EPO or A_R3 in the early stage of erythropoiesis (**Figure 3.6G**). However, erythroid cells stimulated with A_dist_R5 showed higher activation of STAT1 and STAT3, but lower STAT5 relative to the recombinant EPO or A_R3 in the late stage of erythropoiesis (**Figure 3.6G**). Furthermore, cells stimulated with C_R3 showed low activation of STAT5 on differentiation day 5. However, while the level of activation of STAT5 was still lower with C_R3 relative to EPO or A_R3, the level of STAT5 activation between A_dist_R5 and C_R3 was not significantly different (p-value = 0.40) on differentiation day 12, a time point when C_R3 began to perform better than A_dist_R5 (**Figure 3.6B,G)**. This data suggests that biased signaling can be tuned by topological changes of EpoR in primary human erythroid cells. We would note that we have only surveyed a limited range of downstream effectors and thus may be

missing specific signals that mediate the observed phenotypic effects. Interestingly, our data suggests that this biased signaling is also dependent on the differentiation-stage specific context of the cells upon which they act. The same agonists can result in different downstream signaling depending on cell state. Using this property of the DARPins we studied here, highlights the ability to decouple signaling through a single cytokine pathway at various stages of hematopoietic differentiation – a phenomenon that has broad applications in tuning desired responses for specific ligands.

3.8 Discussion on Synthetic Agonist Design

While tunable signaling has been a powerful feature of GPCRs that can be exploited through medicinal chemistry for drug discovery of partial, inverse and biased agonists, tunable signaling has not been deeply explored for dimeric receptors that engage cytokines and growth factors. For GPCRs, signal modulation is achieved by small molecule ligands that induce different conformations of the GPCR helices within the membrane, resulting in differential activation of G-proteins and arrestins. Indeed, this strategy is emerging as a powerful means to decouple GPCR efficacy from toxicity[102]. However, a comparable approach does not exist for the large class of type-I transmembrane receptors, including cytokines and receptor tyrosine kinases, which engage their ligands with large ECDs, and signal by ligand-mediated dimerization. The platform we describe here offers a general strategy to probe the effects of topological variation of receptor dimerization on signaling and function for cytokines, and in principle, any type-I class receptor that signals as a dimer.

The strategy itself is mechanistically agnostic. The signaling effects we observe in response to the topologically varied ligands are likely due to two factors. First, the orientational

effects of the ECD propagate to the intracellular domains that associate with the JAK2 kinases, affecting their relative proximities and distances, which can alter the efficiency by which JAK2 acts on its substrates. In different receptor topologies, the JAK2/ICD complexes will be repositioned into orientations that are more, or less conducive for transphosphorylation of the opposing JAK2 in the dimer, as well as tyrosine sites on the EpoR ICD. This will subsequently result in altered recruitment and/or phosphorylation of STATs, and other adaptors that directly engage the JAK2 and ICD phospho-tyrosines, and these membrane-proximal phosphorylation differences will subsequently propagate to gene expression programs in the nucleus[103,104]. The second explanation of the observed effects likely relates to mechanical distortion: as the topologies are forced into more extreme ranges, strain is induced in the system resulting in fewer numbers of signaling complexes formed, and reduced receptor dimerization as observed by single molecule assays, leading to lower E_{max}, right-shifted EC_{50}s and disproportionately affecting some second messengers more than others. Supporting this interpretation, there is a systematic pattern of reduced E_{max} that correlates with increasing dimer angle. For the distance series, there is no change in E_{max} until a sharp threshold, at which point it decreases abruptly. This could reflect a 'rupture' distance at which point dimerized JAK2 molecules can no longer trans-phosphorylate efficiently. This "rupture" happens earlier with the angle series than with the distance series because the contortions needed to recover are greater. Possibly, both the angle and distance series interfere with weak, cooperative interactions between the transmembrane domains of EpoR that have been reported[96].

As we have demonstrated for red blood cell biogenesis, this approach is particularly suited to examining how specific properties of cytokine signaling pathways may have

differentiation stage-specificity in hematopoiesis or other aspects of cell differentiation. This approach could also be used to identify signaling outputs of clinical interest in a particular cytokine receptor system. For example, the scaffold we describe here can be used as a preclinical exploratory tool to determine to what degrees of agonism are optimal for a given cytokine in a particular therapeutic indication. While DARPins are only beginning to be explored as therapeutics, dimer topology can be used to explore therapeutic opportunities of tuned receptor systems in mouse models as a precursor to engineering of more conventional topologically tunable protein scaffolds.

The approach we describe is general, and easily applied to other receptor systems since it involves a relatively rapid process of 1- selection of DARPins against a recombinant receptor ECD target, 2- reformatting the DARPins into dimers, and 3- inserting repeats into either the N- or C-terminal ends of the DARPins in order to systematically vary the angle and distance of the bound ECDs. The approach is applicable to heterodimeric receptor systems through re-engineering the DARPins to form heterodimers. One limitation of the approach is that one cannot control where the DARPin binds to the receptor ECD, which becomes an issue when certain DARPin/receptor configurations result in clashes between the two receptors in particular orientations. One solution to this in the future could be to leverage new computational protein interface design methods to rationally select binders in geometrically privileged configurations for oligomeric signaling complex formation. The approach is also not limited to receptor ECDs, as the topological scaffolds can be used to introduce systematic variations in angles and distances between any two proteins bound to the DARPin dimer, which allows for exploration of functional effects of topological variation to many important biological systems.

Figures

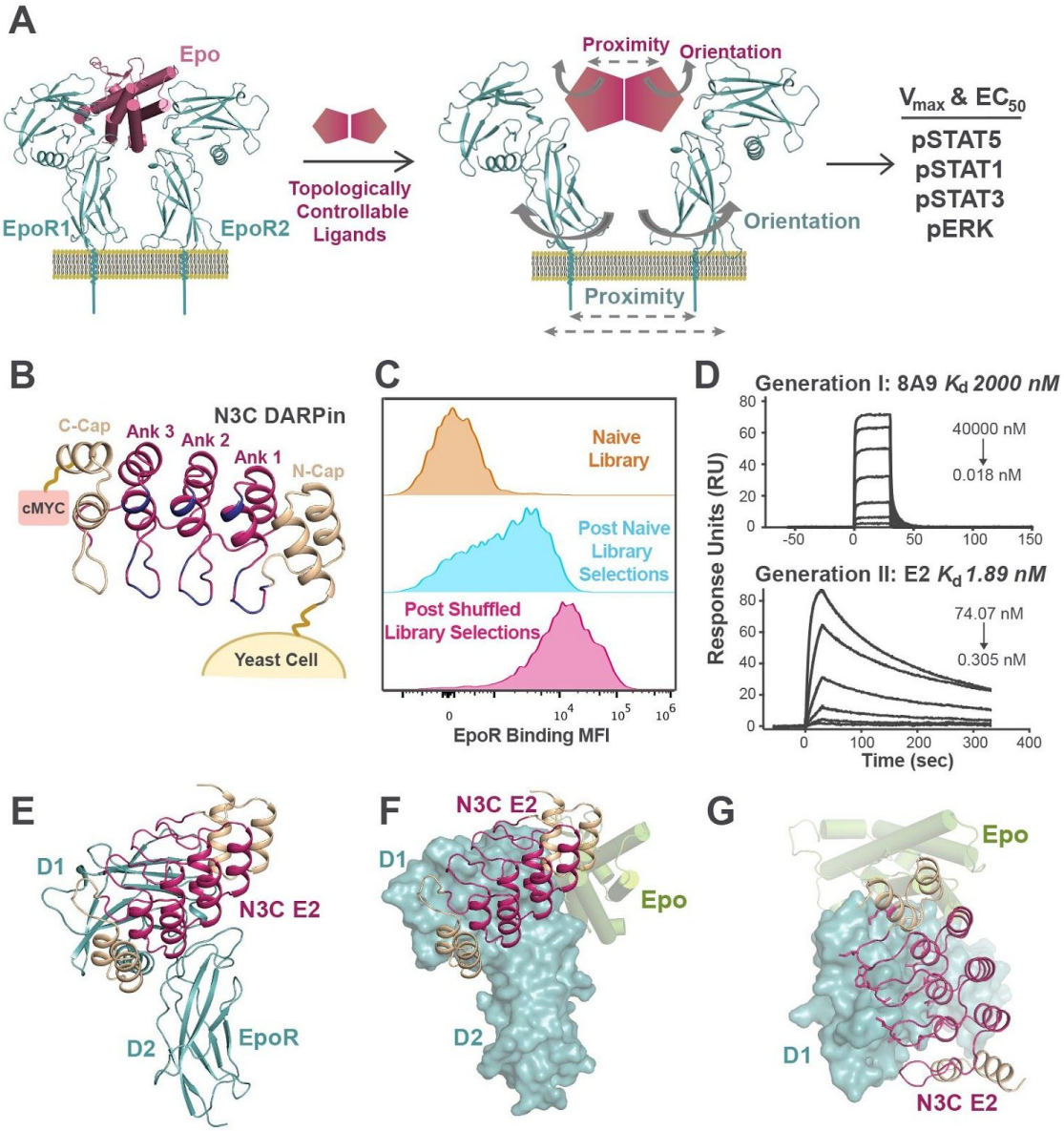


Figure 3.1. Engineering and characterization of high affinity DARPin binding to EpoR. A, An overview of our topological control strategy using EpoR as a model system to investigate modulation of signaling responses. **B,** Schematic representation of the yeast-displayed N3C DARPin construct used for library generation. The randomized positions are shown in blue. **C,** Histograms of yeast-displayed naïve N3C library, pool of DARPin clones after *in vitro* evolution, and pool of shuffled library clones after *in vitro* evolution binding to EpoR. MFI, mean fluorescence intensity. **D,** Representative surface plasmon resonance sensograms for pre- and post-shuffled library N3C DARPin clones binding to EpoR. K_d, Dissociation constant. **E,** The structure of N3C E2 binding to EpoR is shown. **F,** The structure of E2/EpoR (PDB 6MOE) is aligned with EPO/EpoR to indicate relative binding sites of the two ligands. EPO is depicted

as green cylinders. **G,** The intimate shape complementarity of E2 binding to EpoR is depicted with the side chains shown for key interacting E2 residues.

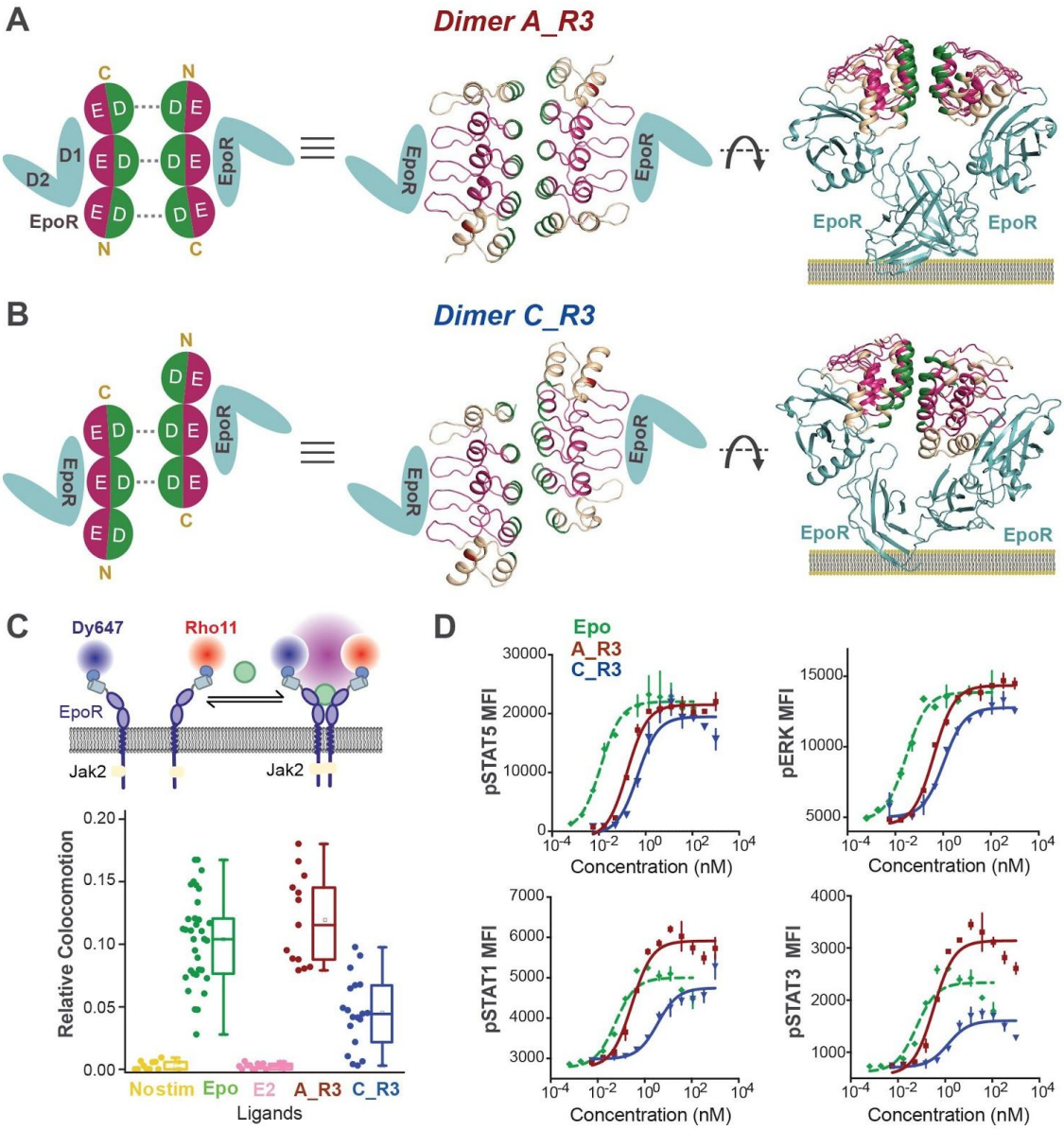


Figure 3.2. Dimerization scaffolds for E2 resulting in agonism. **A,B,** Point mutations (shown in green) inserted on the rear face of the DARPin lead to the formation of non-covalent homodimers with two different geometries. A schematic representation is shown for formation of the 2:2 DARPin:EpoR complex. An N3C DARPin containing three ankyrin repeats is represented by three spheres. The binding face of the N3C E2 is shown as E (pink) and the dimerizing face is shown as D (green), whereas the two EpoR domains D1 and D2 are shown in teal. The dimer interface for C_R3 (PDB 6MOG) is offset by one repeat providing additional twist in the EpoR geometry in C_R3/EpoR (PDB 6MOH). **C,** Ligand-induced dimerization of EpoR on the surface of cells, as observed by single molecule TIRF. **D,** Dose response curves for

STAT1/3/5 and ERK activation performed in UT7 cells stimulated with EPO, A_R3, or C_R3 for 15 mins. Data are (mean ± SD) for two independent replicates.

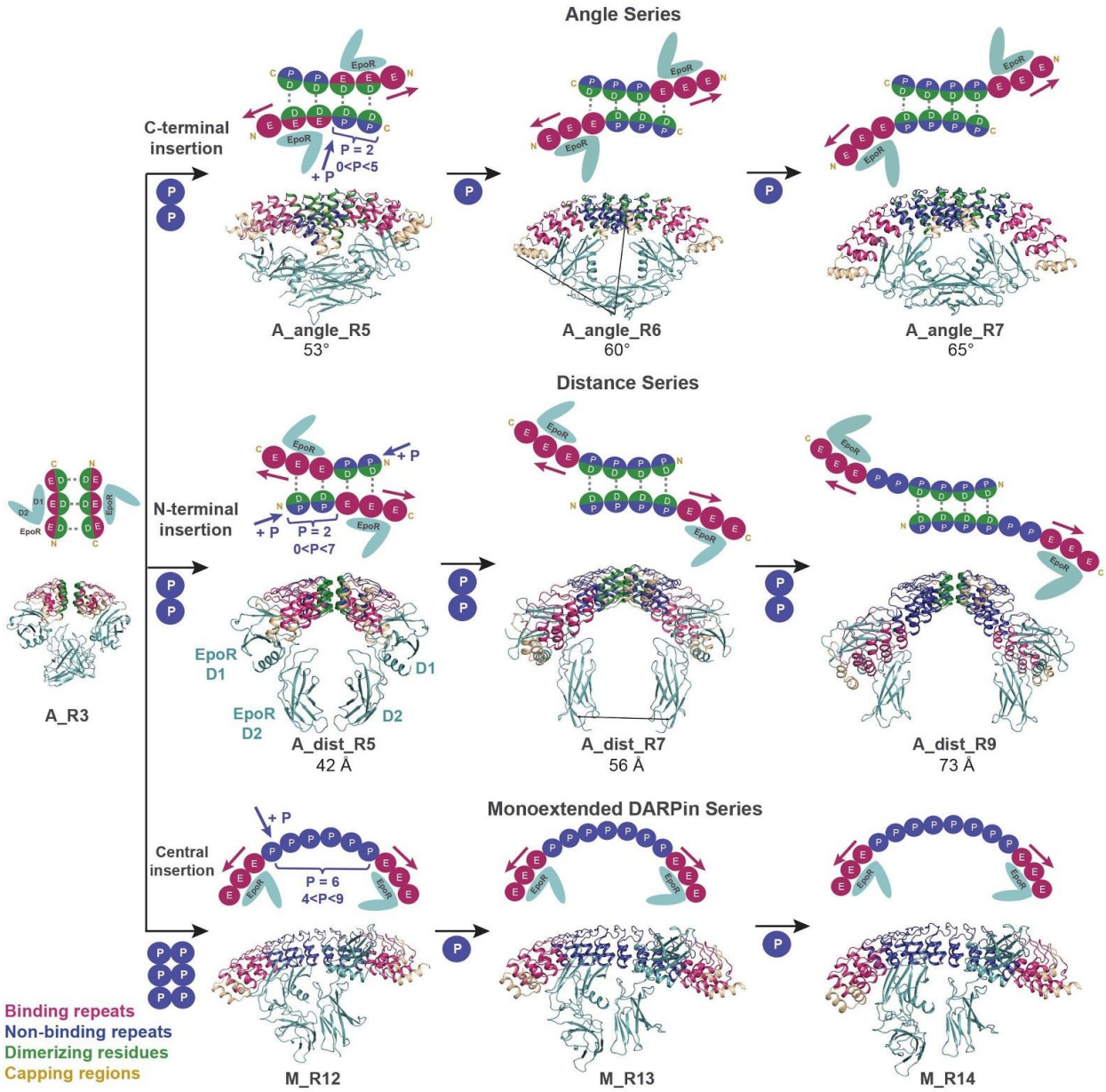


Figure 3.3. Topological control of EpoR geometry. Three different series are designed to systematically modify the different components of receptor dimerization, specifically angle and rotation (A_angle_Rx), distance (A_dist_Rx), and inversion of geometry (M_Rx). The insertion (indicated by blue arrows) of non-binding 'P' ankyrin repeats (blue) at different termini of non-planar DARPins leads to different receptor dimer topologies. The relative movement of the two receptors with the insertion of P repeats is indicated by pink arrows. The series are generated by step-wise insertion of one P repeat, and the nomenclature 'Rx' indicates the total number of ankyrin repeats. The cartoon representations are shown along with the crystal structures for one

representative member per series – A_angle_R5 (PDB 6MOJ), A_dist_R7 (PDB 6MOK), and M_R12 (PDB 6MOL).

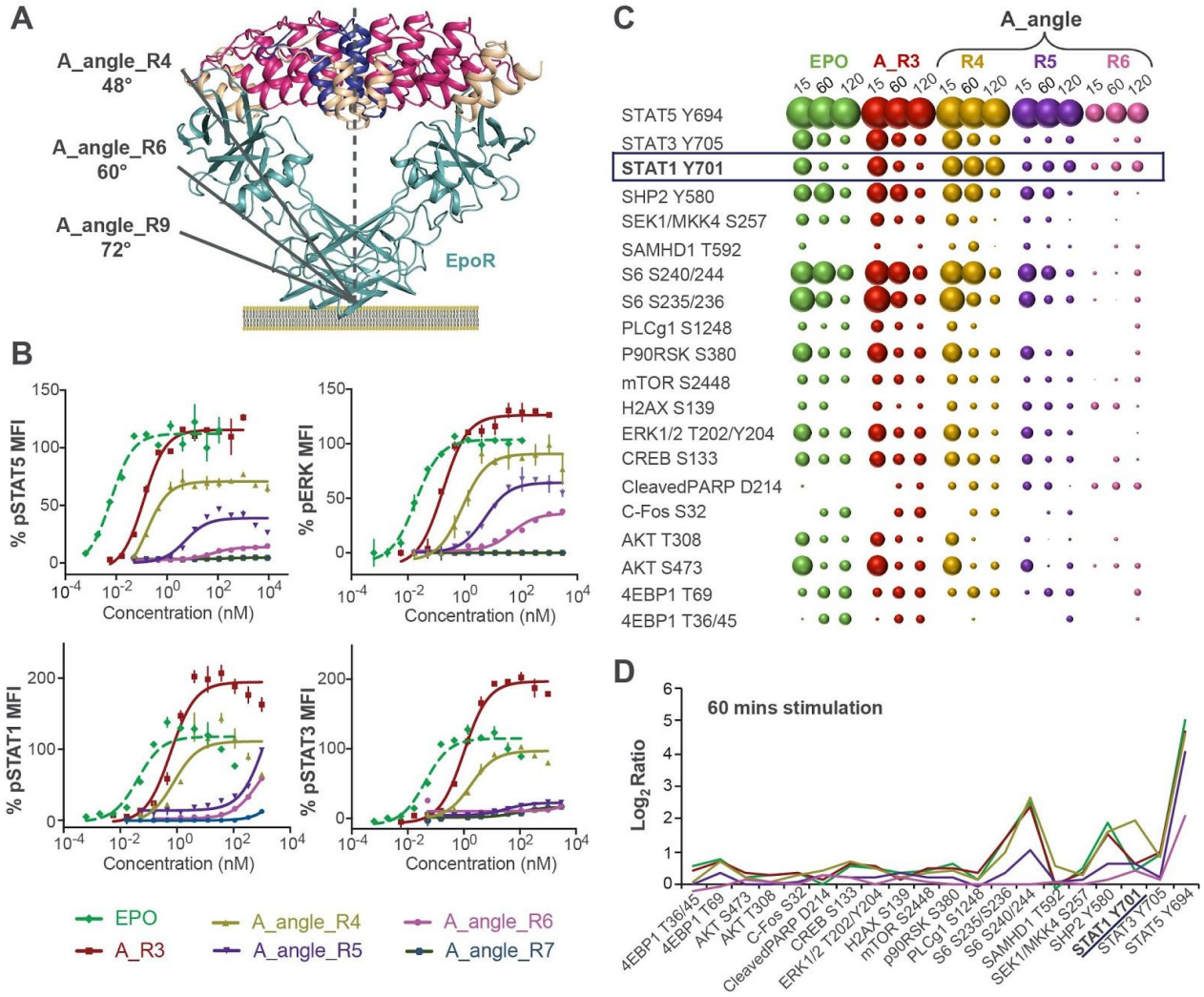


Figure 3.4. Signaling responses induced by variation in EpoR dimer angle. **A,** Addition of P repeats at the C-terminus leads to increase in the angle between the EpoR ECDs, as shown. **B,** Signaling bias observed for the dose response curves for the angle series. Percentage of pSTAT1/3/5 and pERK induced by the various DARPin ligands relative to EPO in UT7 cells. Data are (mean ± SD) for two independent replicates. **C,** Bubble plot representation of the various downstream pathways activated by EPO or the DARPin ligands in UT7 cells with 15, 60, or 120 min stimulation. The size of the bubble correlates to the signal strength. **D,** The Log_2 ratio of the signal generated by the ligands to unstimulated cells at 60 mins is shown.

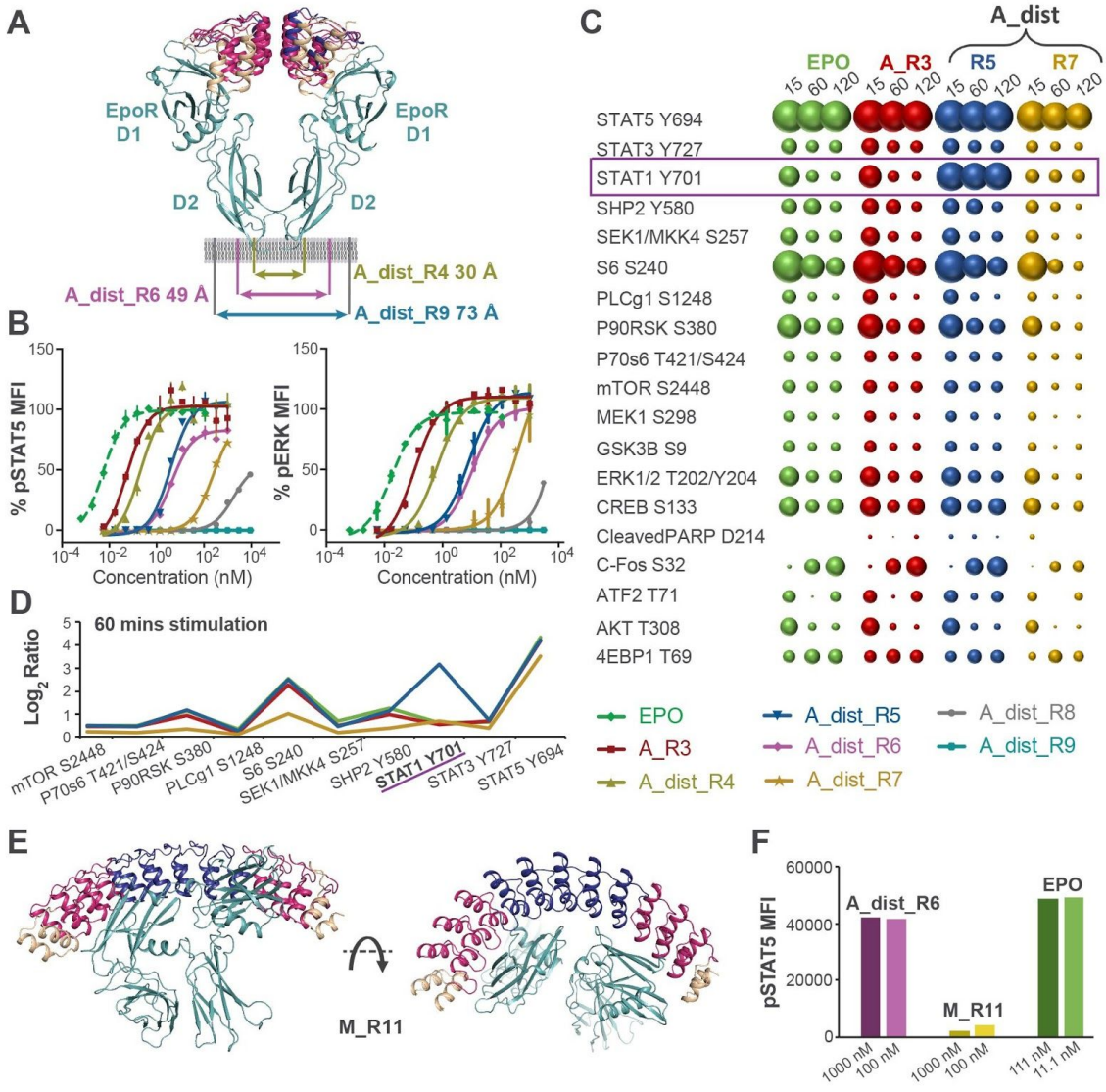


Figure 3.5. Signaling responses induced by variation in EpoR ECD proximity. **A,** Addition of P repeats at the *N*-terminus leads to increase in the distance between the EpoR ECDs, as shown. **B,** Signaling bias observed for the dose response curves for the distance series. Percentage of pSTAT5 and pERK induced by the various DARPin ligands relative to EPO in UT7 cells. Data are (mean ± SD) for two independent replicates. **C,** Bubble plot representation of the various downstream pathways activated by EPO or the DARPin ligands in UT7 cells with 15, 60, or 120 min stimulation. The size of the bubble correlates to the signal strength. **D,** The Log_2 ratio of the signal generated by the ligands to unstimulated cells at 60 mins, is shown. **E,** The model for M_R11 binding to EpoR is shown. Inversion of the relative geometry of the EpoR

ECDs for the mono-extended M_R11, which possesses a similar inter-ECD distance to A_dist_R6, leads to a complete loss of pSTAT5 activity **F**.

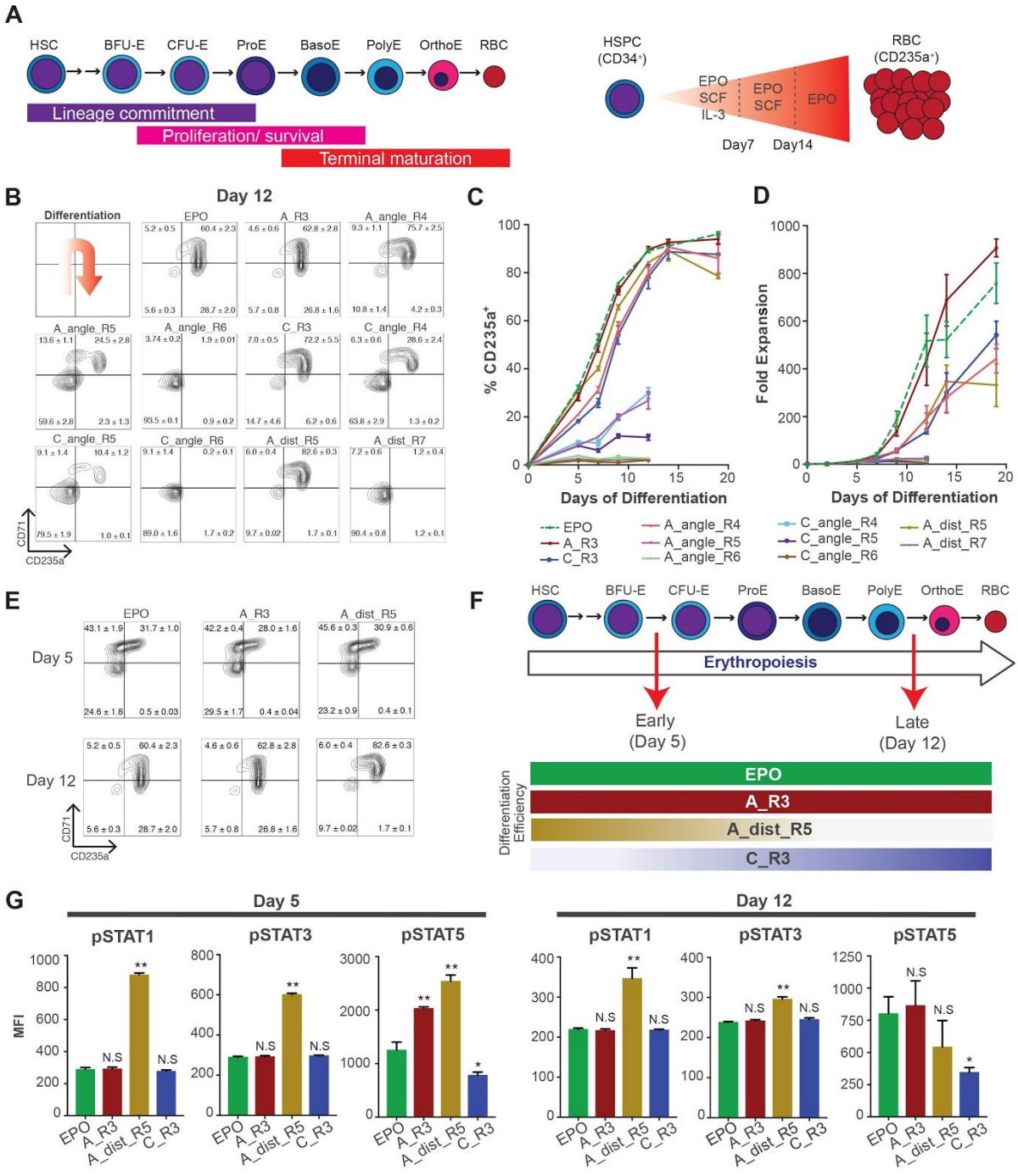


Figure 3.6. Tuning hematopoiesis through topological control of EPO receptor signaling. A, Overview of erythropoiesis and *in vitro* differentiation of human primary hematopoietic cells. **B,** Erythroid differentiation from $CD34^+$ HSPC with synthetic EPO ligands is shown.

Differentiation has been assessed by flow cytometry with the markers CD71 and CD235a at day 12 of differentiation. Means ± SEM for three independent experiments are shown. **C,** Total $CD235a^+$ cells and **D,** overall cell proliferation during erythroid differentiation are shown. Means ± SEM for three independent experiments (N=2 for A_dist_R5) are shown. **E,** Erythroid differentiations by EPO, A_R3, or A_dist_R5 at different time points are shown with flow cytometry analysis. Means ± SEM for three independent experiments are shown. **F,** Experimental scheme of analyzing downstream signaling stimulated by EPO, A_R3, A_dist_R5, and C_R3 during erythropoiesis. **G,** Total Mean Fluorescence Intensity (MFI) of pSTAT1/3/5 stimulated with EPO, A_R3, A_dist_R5, or C_R3 at early (day 5 of differentiation) and late (day 12 of differentiation) erythropoiesis are shown. Means ± SEM for three independent experiments are shown. ** p-value <0.01, * p-value <0.05.

SUPPLEMENTARY FILES

Attached are files including:

Cyclic_design_models.zip (includes desiged homooligomer models from C2-C5 symmetry)

Cyclic_docking_design_examples.zip (includes example command lines in README file, and motif database and scores for docking stage, RosettaScripts two-sided interface design .xml file, and required flags files)

BIBLIOGRAPHY

1. Goodsell, D. S. & Olson, A. J. Structural Symmetry and Protein Function. *Annu. Rev. Biophys. Biomol. Struct.* **29,** 105–153 (2000).
2. Marsh, J. A. & Teichmann, S. A. Structure, dynamics, assembly, and evolution of protein complexes. *Annu. Rev. Biochem.* **84,** 551–575 (2015).
3. Jones, M. R., Seeman, N. C. & Mirkin, C. A. Nanomaterials. Programmable materials and the nature of the DNA bond. *Science* **347,** 1260901 (2015).
4. Pereira-Leal, J. B., Levy, E. D., Kamp, C. & Teichmann, S. A. Evolution of protein complexes by duplication of homomeric interactions. *Genome Biol.* **8,** R51 (2007).
5. Crick, F. H. C. & Watson, J. D. Structure of Small Viruses. *Nature* **177,** 473–475 (1956).
6. Bale, J. B. *et al.* Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science* **353,** 389–394 (2016).
7. King, N. P. *et al.* Accurate design of co-assembling multi-component protein nanomaterials. *Nature* **510,** 103–108 (2014).
8. King, N. P. *et al.* Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* **336,** 1171–1174 (2012).
9. Mackinnon, S. S., Malevanets, A. & Wodak, S. J. Intertwined associations in structures of homooligomeric proteins. *Structure* **21,** 638–649 (2013).
10. Hashimoto, K. & Panchenko, A. R. Mechanisms of protein oligomerization, the critical role of insertions and deletions in maintaining different oligomeric states. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 20352–20357 (2010).
11. Fallas, J. A. *et al.* Computational design of self-assembling cyclic protein homo-oligomers. *Nat.*

Chem. **9,** 353–360 (2017).

12. Bachmann, M. F. & Zinkernagel, R. M. NEUTRALIZING ANTIVIRAL B CELL RESPONSES. *Annu. Rev. Immunol.* **15,** 235–270 (1997).
13. Burton, D. R. Antibodies, viruses and vaccines. *Nat. Rev. Immunol.* **2,** 706–713 (2002).
14. Frietze, K. M., Peabody, D. S. & Chackerian, B. Engineering virus-like particles as vaccine platforms. *Curr. Opin. Virol.* **18,** 44–49 (2016).
15. Wang, H. *et al.* Asymmetric recognition of HIV-1 Envelope trimer by V1V2 loop-targeting antibodies. *Elife* **6,** (2017).
16. Kadam, R. U. *et al.* Potent peptidic fusion inhibitors of influenza virus. *Science* **358,** 496–502 (2017).
17. McLellan, J. S. *et al.* Structure-based design of a fusion glycoprotein vaccine for respiratory syncytial virus. *Science* **342,** 592–598 (2013).
18. Akdis, M. *et al.* Interleukins (from IL-1 to IL-38), interferons, transforming growth factor β, and TNF-α: Receptors, functions, and roles in diseases. *J. Allergy Clin. Immunol.* **138,** 984–1010 (2016).
19. Rider, P., Carmi, Y. & Cohen, I. Biologics for Targeting Inflammatory Cytokines, Clinical Uses, and Limitations. *Int. J. Cell Biol.* **2016,** 9259646 (2016).
20. Moraga, I., Spangler, J., Mendoza, J. L. & Garcia, K. C. Multifarious determinants of cytokine receptor signaling specificity. *Adv. Immunol.* **121,** 1–39 (2014).
21. Ozaki, K. & Leonard, W. J. Cytokine and cytokine receptor pleiotropy and redundancy. *J. Biol. Chem.* **277,** 29355–29358 (2002).
22. Vazquez-Lombardi, R. *et al.* Challenges and opportunities for non-antibody scaffold drugs. *Drug Discov. Today* **20,** 1271–1283 (2015).
23. Jacobson, K. A. New paradigms in GPCR drug discovery. *Biochem. Pharmacol.* **98,** 541–555 (2015).

24. Plückthun, A. Designed ankyrin repeat proteins (DARPins): binding proteins for research, diagnostics, and therapy. *Annu. Rev. Pharmacol. Toxicol.* **55,** 489–511 (2015).

25. Kim, A. R. *et al.* Functional Selectivity in Cytokine Signaling Revealed Through a Pathogenic EPO Mutation. *Cell* **168,** 1053–1064.e15 (2017).

26. Sockolosky, J. T. *et al.* Selective targeting of engineered T cells using orthogonal IL-2 cytokine-receptor complexes. *Science* **359,** 1037–1042 (2018).

27. Mitra, S. *et al.* Interleukin-2 activity can be fine tuned with engineered receptor signaling clamps. *Immunity* **42,** 826–838 (2015).

28. Kalliolias, G. D. & Ivashkiv, L. B. TNF biology, pathogenic mechanisms and emerging therapeutic strategies. *Nat. Rev. Rheumatol.* **12,** 49–62 (2016).

29. Ali, M. H. & Imperiali, B. Protein Oligomerization: How and Why. *ChemInform* **36,** (2005).

30. Nishi, H., Hashimoto, K., Madej, T. & Panchenko, A. R. Evolutionary, physicochemical, and functional mechanisms of protein homooligomerization. *Prog. Mol. Biol. Transl. Sci.* **117,** 3–24 (2013).

31. Fletcher, J. M. *et al.* A basis set of de novo coiled-coil peptide oligomers for rational protein design and synthetic biology. *ACS Synth. Biol.* **1,** 240–250 (2012).

32. Smock, R. G., Yadid, I., Dym, O., Clarke, J. & Tawfik, D. S. De Novo Evolutionary Emergence of a Symmetrical Protein Is Shaped by Folding Constraints. *Cell* **164,** 476–486 (2016).

33. Voet, A. R. D. *et al.* Computational design of a self-assembling symmetrical β-propeller protein. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 15102–15107 (2014).

34. Stranges, P. B., Machius, M., Miley, M. J., Tripathy, A. & Kuhlman, B. Computational design of a symmetric homodimer using β-strand assembly. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 20562–20567 (2011).

35. Der, B. S. *et al.* Metal-Mediated Affinity and Orientation Specificity in a Computationally Designed

Protein Homodimer. *J. Am. Chem. Soc.* **134,** 375–385 (2011).

36. Mou, Y., Huang, P.-S., Hsu, F.-C., Huang, S.-J. & Mayo, S. L. Computational design and experimental verification of a symmetric protein homodimer. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 10714–10719 (2015).
37. Huang, P.-S. *et al.* De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. *Nat. Chem. Biol.* **12,** 29–34 (2015).
38. Lin, Y.-R. *et al.* Control over overall shape and size in de novo designed proteins. *Proc. Natl. Acad. Sci. U. S. A.* **112,** E5478–85 (2015).
39. Koga, N. *et al.* Principles for designing ideal protein structures. *Nature* **491,** 222–227 (2012).
40. Parmeggiani, F. *et al.* A general computational approach for repeat protein design. *J. Mol. Biol.* **427,** 563–575 (2015).
41. Huang, P.-S. *et al.* High thermodynamic stability of parametrically designed helical bundles. *Science* **346,** 481–485 (2014).
42. Brunette, T. J. *et al.* Exploring the repeat protein universe through computational protein design. *Nature* **528,** 580–584 (2015).
43. Main, E. R. G., Xiong, Y., Cocco, M. J., D'Andrea, L. & Regan, L. Design of stable alpha-helical arrays from an idealized TPR motif. *Structure* **11,** 497–508 (2003).
44. Pechmann, S., Levy, E. D., Tartaglia, G. G. & Vendruscolo, M. Physicochemical principles that regulate the competition between functional and dysfunctional association of proteins. *Proceedings of the National Academy of Sciences* **106,** 10159–10164 (2009).
45. Levy, E. D. & Teichmann, S. Structural, evolutionary, and assembly principles of protein oligomerization. *Prog. Mol. Biol. Transl. Sci.* **117,** 25–51 (2013).
46. Bahadur, R. P., Chakrabarti, P., Rodier, F. & Janin, J. Dissecting subunit interfaces in homodimeric proteins. *Proteins* **53,** 708–719 (2003).

47. Janin, J., Bahadur, R. P. & Chakrabarti, P. Protein–protein interaction and quaternary structure. *Q. Rev. Biophys.* **41,** (2008).

48. Leaver-Fay, A. *et al.* ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487,** 545–574 (2011).

49. Chen, R. & Weng, Z. Docking unbound proteins using shape complementarity, desolvation, and electrostatics. *Proteins* **47,** 281–294 (2002).

50. Pierce, B., Tong, W. & Weng, Z. M-ZDOCK: a grid-based approach for Cn symmetric multimer docking. *Bioinformatics* **21,** 1472–1478 (2004).

51. Schneidman-Duhovny, D., Inbar, Y., Nussinov, R. & Wolfson, H. J. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.* **33,** W363–7 (2005).

52. Lyskov, S. & Gray, J. J. The RosettaDock server for local protein-protein docking. *Nucleic Acids Res.* **36,** W233–W238 (2008).

53. Gray, J. J. *et al.* Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J. Mol. Biol.* **331,** 281–299 (2003).

54. Lu, M., Dousis, A. D. & Ma, J. OPUS-PSP: an orientation-dependent statistical all-atom potential derived from side-chain packing. *J. Mol. Biol.* **376,** 288–301 (2008).

55. Zheng, W., Schafer, N. P., Davtyan, A., Papoian, G. A. & Wolynes, P. G. Predictive energy landscapes for protein-protein association. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 19244–19249 (2012).

56. DiMaio, F., Leaver-Fay, A., Bradley, P., Baker, D. & André, I. Modeling symmetric macromolecular structures in Rosetta3. *PLoS One* **6,** e20450 (2011).

57. Fleishman, S. J. *et al.* RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite. *PLoS One* **6,** e20161 (2011).

58. Leaver-Fay, A. *et al.* Scientific benchmarks for guiding macromolecular energy function improvement. *Methods Enzymol.* **523,** 109–143 (2013).

59. Schneidman-Duhovny, D., Hammel, M. & Sali, A. FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res.* **38,** W540–4 (2010).

60. Schneidman-Duhovny, D., Hammel, M., Tainer, J. A. & Sali, A. Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophys. J.* **105,** 962–974 (2013).

61. Boyken, S. E. *et al.* De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity. *Science* **352,** 680–687 (2016).

62. Mateu, M. G. Assembly, stability and dynamics of virus capsids. *Arch. Biochem. Biophys.* **531,** 65–79 (2013).

63. Sanders, B., Koldijk, M. & Schuitemaker, H. Inactivated Viral Vaccines. in *Vaccine Analysis: Strategies, Principles, and Control* 45–80 (2014).

64. Sanders, R. W. & Moore, J. P. Native-like Env trimers as a platform for HIV-1 vaccine design. *Immunol. Rev.* **275,** 161–182 (2017).

65. Jardine, J. *et al.* Rational HIV immunogen design to target specific germline B cell receptors. *Science* **340,** 711–716 (2013).

66. Nascimento, I. P. & Leite, L. C. C. Recombinant vaccines and the development of new vaccine strategies. *Braz. J. Med. Biol. Res.* **45,** 1102–1111 (2012).

67. Kanekiyo, M. *et al.* Self-assembling influenza nanoparticle vaccines elicit broadly neutralizing H1N1 antibodies. *Nature* **499,** 102–106 (2013).

68. Georgiev, I. S. *et al.* Two-Component Ferritin Nanoparticles for Multimerization of Diverse Trimeric Antigens. *ACS Infect Dis* **4,** 788–796 (2018).

69. Bayliss S Wang D (Meads C.). *Palivizumab for Immunoprophylaxis of Respiratory Syncytial Virus (RSV) Bronchiolitis in High-Risk Infants and Young Children: Systematic Review and Additional Economic Modelling of Subgroup Analyses*. (2011).

70. Lee, K. K. & Gui, L. Dissecting Virus Infectious Cycles by Cryo-Electron Microscopy. *PLoS*

Pathog. **12,** e1005625 (2016).

71. Ozorowski, G. *et al.* Effects of Adjuvants on HIV-1 Envelope Glycoprotein SOSIP Trimers. *J. Virol.* **92,** (2018).
72. Kimanius, D., Forsberg, B. O., Scheres, S. H. & Lindahl, E. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *Elife* **5,** (2016).
73. Ofek, G. *et al.* Elicitation of structure-specific antibodies by epitope scaffolds. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 17880–17887 (2010).
74. Delgoffe, G. M., Murray, P. J. & Vignali, D. A. A. Interpreting mixed signals: the cell's cytokine conundrum. *Curr. Opin. Immunol.* **23,** 632–638 (2011).
75. Spangler, J. B., Moraga, I., Mendoza, J. L. & Christopher Garcia, K. Insights into Cytokine–Receptor Interactions from Cytokine Engineering. *Annu. Rev. Immunol.* **33,** 139–167 (2015).
76. Kaushansky, K. Lineage-Specific Hematopoietic Growth Factors. *N. Engl. J. Med.* **354,** 2034–2045 (2006).
77. Stroud, R. M. & Wells, J. A. Mechanistic diversity of cytokine receptor signaling across cell membranes. *Sci. STKE* **2004,** re7 (2004).
78. Syed, R. S. *et al.* Efficiency of signalling through cytokine receptors depends critically on receptor orientation. *Nature* **395,** 511–516 (1998).
79. Socolovsky, M., Fallon, A. E. & Lodish, H. F. The prolactin receptor rescues EpoR-/- erythroid progenitors and replaces EpoR in a synergistic interaction with c-kit. *Blood* **92,** 1491–1496 (1998).
80. Livnah, O. *et al.* Functional mimicry of a protein hormone by a peptide agonist: the EPO receptor complex at 2.8 A. *Science* **273,** 464–471 (1996).
81. Mumm, J. B. *et al.* IL-10 elicits IFNγ-dependent tumor immune surveillance. *Cancer Cell* **20,** 781–796 (2011).

82. Ho, C. C. M. *et al.* Decoupling the Functional Pleiotropy of Stem Cell Factor by Tuning c-Kit Signaling. *Cell* **168,** 1041–1052.e18 (2017).

83. Thomas, C. *et al.* Structural linkage between ligand discrimination and receptor activation by type I interferons. *Cell* **146,** 621–632 (2011).

84. Freed, D. M. *et al.* EGFR Ligands Differentially Stabilize Receptor Dimers to Specify Signaling Kinetics. *Cell* **171,** 683–695.e18 (2017).

85. Rowlinson, S. W. *et al.* An agonist-induced conformational change in the growth hormone receptor determines the choice of signalling pathway. *Nat. Cell Biol.* **10,** 740–747 (2008).

86. Ballinger, M. D. & Wells, J. A. Will any dimer do? *Nat. Struct. Biol.* **5,** 938–940 (1998).

87. Staerk, J. *et al.* Orientation-specific signalling by thrombopoietin receptor dimers. *EMBO J.* **30,** 4398–4413 (2011).

88. Kai, M. *et al.* Switching constant domains enhances agonist activities of antibodies to a thrombopoietin receptor. *Nat. Biotechnol.* **26,** 209–211 (2008).

89. Zhang, H. *et al.* Selecting agonists from single cells infected with combinatorial antibody libraries. *Chem. Biol.* **20,** 734–741 (2013).

90. Nakano, K. *et al.* Effective screening method of agonistic diabodies based on autocrine growth. *J. Immunol. Methods* **347,** 31–35 (2009).

91. Moraga, I. *et al.* Tuning cytokine receptor signaling by re-orienting dimer geometry with surrogate ligands. *Cell* **160,** 1196–1208 (2015).

92. Binz, H. K., Stumpp, M. T., Forrer, P., Amstutz, P. & Plückthun, A. Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *J. Mol. Biol.* **332,** 489–503 (2003).

93. Binz, H. K. *et al.* High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat. Biotechnol.* **22,** 575–582 (2004).

94. Komatsu, N. *et al.* Establishment and characterization of a human leukemic cell line with megakaryocytic features: dependency on granulocyte-macrophage colony-stimulating factor, interleukin 3, or erythropoietin for growth and survival. *Cancer Res.* **51,** 341–348 (1991).

95. Li, J., Mahajan, A. & Tsai, M.-D. Ankyrin Repeat: A Unique Motif Mediating Protein−Protein Interactions†. *Biochemistry* **45,** 15168–15178 (2006).

96. Ebie, A. Z. & Fleming, K. G. Dimerization of the erythropoietin receptor transmembrane domain in micelles. *J. Mol. Biol.* **366,** 517–524 (2007).

97. Nivón, L. G., Moretti, R. & Baker, D. A Pareto-optimal refinement method for protein design scaffolds. *PLoS One* **8,** e59004 (2013).

98. Giani, F. C. *et al.* Targeted Application of Human Genetic Variation Can Improve Red Blood Cell Production from Stem Cells. *Cell Stem Cell* **18,** 73–78 (2016).

99. Hu, J. *et al.* Isolation and functional characterization of human erythroblasts at distinct stages: implications for understanding of normal and disordered erythropoiesis in vivo. *Blood* **121,** 3246–3253 (2013).

100. Metcalf, D. Hematopoietic cytokines. *Blood* **111,** 485–491 (2008).

101. Testa, U. Apoptotic mechanisms in the control of erythropoiesis. *Leukemia* **18,** 1176–1199 (2004).

102. Winpenny, D., Clark, M. & Cawkill, D. Biased ligand quantification in drug discovery: from theory to high throughput screening to identify new biased μ opioid receptor agonists. *Br. J. Pharmacol.* **173,** 1393–1403 (2016).

103. Waters, M. J. & Brooks, A. J. JAK2 activation by growth hormone and other cytokines. *Biochem. J* **466,** 1–11 (2015).

104. Ferrao, R. D., Wallweber, H. J. & Lupardus, P. J. Receptor-mediated dimerization of JAK2 FERM domains is required for JAK2 activation. *Elife* **7,** (2018).

APPENDIX

Appendix Methods

Repeat protein scaffolds. Repeat proteins, comprised of recurring 20-50 residue stretches, are ideal for use in protein-based material design due to their high stability and capability to have altered lengths and curvatures by varying the number of repeating modules. Listed below are the RCSB Protein Data Bank entries for selected scaffolds. An additional a set of scaffolds is provided in which experimental small-angle X-ray scattering data agreed with the computational model.

X-ray Structures (PDB ID)		SAXS Validated Models
ank1	(4GPM)	tpr1
ank3	(4GMR)	KP16
ank4	(4HB5)	KP17
3ltj	(3LTJ)	K64s
1na0	(1NA0)	hl1
2fo7	(2FO7)	HR00
arm8	(4HXT)	
HR04	(5CWB)	
HR07	(5CWD)	
HR08	(5CWF)	
HR10	(5CWG)	
HR14	(5CWH)	
HR54	(5CWL)	
HR64	(5CWM)	
HR71	(5CWN)	

Symmetry definition file. In order to model symmetric complexes within the Rosetta framework, we implemented symmetry definition files to generate the specified symmetry starting from a single subunit of each docked configuration. These files are needed to properly calculate all of the score terms in symmetric poses. Example cyclic symmetry definition files used in this study are provided in the **Supplementary Files**.

Two-body Asymmetric Docking. Asymmetric docking was performed for chains A and B of each oligomer. Patchdock was used to generate 7 alternative starting configurations that were each used together with the designed binding mode as a starting points for a RosettaDock calculation using BOINC. Each oligomer was characterized by a value ($\Delta E = \min(E_{local}) - \min(E_{global})$) corresponding to the difference in energy of the lowest energy state sampled during the local section of the algorithm minus the lowest energy state sampled during the global section of the algorithm. Negative values indicate that the starting conformation corresponds to the lowest energy state found during docking while positive values indicate that alternative lower energy conformations were found.

Size exclusion chromatography. IMAC elution samples for each designed protein were concentrated down using a 10,000 MWCO protein concentrator (Novagen) and fractionated by size on an AKTA pure chromatography system using a Superdex 75 10/300 GL, Superdex 200 10/300 GL, or Superose 6 10/300 GL gel filtration column (GE Life Sciences) in 25mM Tris

150mM Nacl pH 8 (TBS) unless otherwise specified. Sizing profiles were noted based on absorption at 220 nm or 280 nm wavelength light for each fraction. Molecular weights for predominant species in each protein trace were estimated by comparison to the corresponding protomer chromatogram.

Protein Expression and Purification. Synthetic genes for designed proteins were optimized for *E. coli* expression and assembled from purchased genes (Genscript) ligated into the pET21-NESG (homooligomers) or pET28b-NESG (heterodimers and nanoparticles) vector at restriction sites NdeI and XhoI or NcoI and XhoI respectively. These plasmids were cloned into BL21 (DE3) or Lemo21 (DE3) *E. coli* competent cells. Transformants were inoculated and grown in either LB or TB medium with either 100 mg/L carbenicillin or 100μg/L kanamycin at 37 °C until an OD_{600} of 0.5-0.8. Isopropyl-thio-β-D-galactopyranoside (IPTG) was then added at a concentration of 0.5-1mM to induce protein expression, or cultures were auto-induced using galactose as an inducing reagent for expression among other salts and trace metals according to the Studier protocols. Expression proceeded for 20 hours at 18-37°C until the cell cultures were harvested by centrifugation. Cell pellets were resuspended in TBS, 5mM imidazole, DNase, and EDTA-free protease inhibitors (Pierce) and lysed by sonication or microfluidization. Each protein was then purified from lysate by Ni^{2+} immobilized metal affinity chromatography with Ni-NTA Superflow resin (Qiagen or GE). Resin with bound cell lysate was washed with 15 column volumes of TBS with 40 mM imidazole. Proteins were eluted with five column volumes of 200-400 mM imidazole for further purification by size exclusion chromatography.

Size exclusion chromatography with multi-angle light scattering. Fractions containing single predominant species from the initial round of size exclusion chromatography were concentrated down with 10,000 MWCO protein concentrators (Novagen) to a concentration of 1.0-2.0 mg mL^{-1}. 100 uL of each sample was then run through a high-performance liquid chromatography system (Agilent) using (unless otherwise noted) a Superdex 200 10/300 GL gel filtration column (GE Life Sciences) at an elution rate of 0.50 mL min^{-1} in TBS. These fractionation runs were coupled to a multi-angle light scattering detector (Wyatt) in order to determine the absolute molecular weights for each designed protein. The following equation derived from the Rayleigh-Debye-Gans theory of light scattering was used in the ASTRA software to calculate the molecular weight of the major species present in each sample:

$$\frac{K^* c}{R(\theta, c)} = \frac{1}{M_w P(\theta)} + 2A_2 c$$

where:

- $R(\theta,c)$ is the excess Rayleigh ratio of the solution as a function of scattering angle θ and concentration c. It is directly proportional to the intensity of the scattered light in excess of the light scattered by the pure solvent.
- c is the solute concentration.
- M_w is the weight-averaged solute molar mass.
- A_2 is the second virial coefficient in the virial expansion of the osmotic pressure.
- K^* is the constant $4\pi^2(dn/dc)^2 n_0^2/N_a\lambda_0^4$.
- N_a is Avogadro's number. This number always appears when concentration is measured in g/mL and molar mass in g/mol.
- $P(\theta)$ describes the angular dependence of the scattered light, and can be related to the rms radius.
- n_0 is the index of refraction of the solvent
- λ_0 is the vacuum wavelength of the laser

Small-angle X-ray Scattering. Designed proteins that predominantly formed the target oligomeric species were re-expressed and purified for low-resolution structure determination

while in solution by small-angle X-ray scattering (SAXS). A purified elution sample and concentrated sample of each protein were sent for data collection at the SIBYLS High Throughput SAXS Advanced Light Source in Berkeley, California. A beam exposure time of between 0.5-2.0 seconds was used to obtain diffraction data, which we represent in plots of log intensity (I) vs. q. Data was collected at 11kEV/1.125A with a 2 meter beamstop. Experimental diffraction data was then analyzed with the java-based application, Scatter. Minimum q values (q_{min}) and experimental radii of gyration (Rg) were determined by Guinier analysis. Data resolution, reflected by maximum q value (q_{max}), was determined by a characteristic asymptote in signal intensity described by Porod's Law. Refined data sets and corresponding designed model .pdb files were input to the FoXS web server to compute the agreement (evaluated as χ) between the experimental and model-computed profiles.

Generation of extension ensemble and determination of SAXS-suggested model. A set of designed homooligomers, one each of C2 and C4 symmetry, that had been structurally validated by X-ray diffraction crystallography and/or SAXS were selected as candidates for extension. Because the repeating units of the initial scaffolds were not perfectly superimposable, each unique repeat unit (aside from N- and C- capping repeats) was propagated to generate several models with two additional repeat units (three for C2 oligomer, two for C4 oligomer). 100 trajectories of a Rosetta protocol that previously showed to conformationally sample the local energy landscape was then performed on each extended model. The total extension set was then input to FoXS with an experimentally-obtained profile to determine an ensemble of models that agreed within a threshold to the data.

Electron microscopy and preparation of samples. Nanoparticle samples were diluted to 20-50 μg/ml and loaded onto the carbon-coated 400-mesh Cu grid (glow-discharged). Grids were negatively stained with 2% (w/v) uranyl-formate for 60 seconds, and data collection was performed on a Tecnai Spirit Electron microscope operating at 120keV, magnification of 52,000x with a pixel size of 2.05Å at the specimen plane, an electron dose of 25 e-/Å^2, and defocus value of -1.5μm. Formation of monodisperse particles was apparent, and micrographs were recorded on a Tietz 4k x 4k TemCam-F416 CMOS camera using the Leginon automated imaging interface. Data processing was performed in the Appion data processing suite.

Engineering the extended DARPins. The modular nature of the DARPin was utilized towards the generation of the extended DARPins. The DARPin molecules are non-planar, and each repeat provides a 2-3° rotation relative to the preceding repeat. Thus, it is crucial to maintain the overall curvature of the DARPin to avoid a loss in binding affinity while inserting more ankyrin repeats. The sequence of the P repeat was carefully designed to maintain the curvature observed in the N3C E2 monomer, and the A_R3 and C_R3 dimers. The framework residues in the P repeat were based on the sequence of the consensus repeat. The positions that were randomized in the library design (denoted as X, **Appendix Figure 3.1**) were assigned residues from a Frizzled8-binding DARPin. As a negative control, a DARPin dimer comprising of only non-binding P repeats was designed (NPPPPC) to eliminate the possibility of off-target (Frizzled8) effects. The control dimer shows no affinity towards EpoR or Frizzled8 (**Appendix**

Figure 3.6). The E2 DARPin is comprised of three EpoR binding repeats, denoted as "EEE", and the N- and C-terminal capping regions.

To engineer the "angle" series, one or more P repeats was inserted at the C-terminus to provide DARPins such as NEEEP..C. Furthermore, the dimerizing repeats (indicated by underline) are placed on the same terminus as the P repeat insertions for maximum impact. In dimer A_R3/C_R3 (NEEEC), the dimerizing residues span over the three ankyrin repeats and the two capping regions. To generate A_angle_R4, one P repeat was inserted at the C-terminal end (NEEEPC) and the set of dimerizing residues were also shifted by one repeat. Similarly, the design for A_angle_R5 can summarized as NEEEPPC.

P repeat sequence: DAAGMTPLHLAAANGHLEIVEVLLKYGADVNAK

To engineer the "distance" series, the non-binding P repeats were inserted at the *N*-terminus, for example NPP…EEEC. Insertion of repeats at the *N*-terminus proved challenging, as illustrated in **Appendix Figure 3.9**. A straightforward design based on the angle series such as NPEEEC displayed a significant loss in binding affinity towards EpoR. Three possible reasons were identified. First, an analysis of the models revealed a clash between residues of the inner helices from the first binding repeat and the preceding P repeat (**Appendix Figure 3.7A**). No such clashes were observed in the E2 DARPin where the *N*-terminal capping region helix precedes the first binding repeat (**Appendix Figure 3.7B**). Second, an analysis of the sequence of the E2 *N*-terminal capping region revealed a Glu to Lys mutation in the inner helix, which resulted from the error-prone mutagenesis performed during affinity maturation. The Lys residue

is well positioned to interact with EpoR. Third, insertion of the P repeat implies existence of β-turn loops preceding the loop region from the first binding repeat. In E2, A_R3 (NEEEC) and the angle series, no β-turn loops precede the loop region for the first ankyrin repeat. The newly inserted loops could introduce destabilizing interactions. To address these issues, a modified repeat (P') was used as the repeat preceding the first E repeat such as NP..P'EEEC. In the P' repeat, the sequence of the inner helix was modified from HLAAANG to HKAARAG. The 'KAARA' sequence is derived from the inner helix of the *N*-terminal capping region and installed in the P' repeat based on structural symmetry of the two helices. This eliminates the clashes between the helices and also re-positions the Lys residue to interact with EpoR. Furthermore, the loop region with the sequence DAAGM was shortened to Gly, to facilitate the turn while eliminating any destabilizing interactions, (**Appendix Figure 3.7C**). Here, we show that the loops can also be shortened without affecting the structure for A_dist_R7 (**Appendix Figure 3.7D**). The dimerizing residues are now maintained at the *N*-terminal repeats. To generate A_dist_R4, P' repeat and the dimerizing residues were inserted at the *N*-terminal end (NP'EEEC). Whereas, to generate A_dist_R5, one P followed by one P' repeat was inserted, NPP'EEEC. Similarly, the design for A_dist_R6 can summarized as NPPP'EEEC. Also, the sequence of the P repeat used for generating the distance series is different from the P repeat used for the angle series (C-terminal insertions). But both the sequences are referred to as 'P' repeats for the sake of simplicity.

P repeat sequence: DAAGGTPLHEAARAGHLEIVEVLLKYGADVNAV

P' repeat sequence: DGTPLHKAARAGHLEIVEVLLKYGADVNAV

To engineer the "mono-extended" DARPin series, two sets of EpoR binding repeats were included on the same DARPin molecule. This series is based on monomeric extended DARPins, and thus named "mono-extended." To generate a series, the two sets of binding repeats were positioned at either termini, and the P repeats are inserted in the middle, for example NEEEPP…EEEC. The second set of EpoR binding repeats are positioned similar to the distance series. Thus, the P' approach was adopted for the generation of the mono-extended series, as well. The insertion of one to four repeats in the middle showed severe clashes in the models generated. As a result, the first member of the series consisted of five non-binding repeats, M_R11, NEEEPPPPP'EEEC. Similarly, the design for M_R12 can be summarized as NEEEPPPPPP'EEEC with a total of six non-binding repeats inserted.

P repeat sequence: DAAGGTPLHEAARAGHLEIVEVLLKYGADVNAV

P' repeat sequence: DGTPLHKAARAGHLEIVEVLLKYGADVNAV

Surface Plasmon Resonance. Dissociation constants (K_d) for EpoR-binding DARPin ligands were determined by surface plasmon resonance (SPR) using the BIAcore T100 instrument (GE Healthcare). First, biotinylated EpoR was captured on a streptavidin-coated (SA) sensor chip (GE Healthcare) with immobilization density in the range of 100 resonance units (RU). Similarly, a control flow cell was also prepared with an off-target protein such as biotinylated-IFNAR1 or biotinylated-Fzd8 for reference subtraction. The binding kinetics were performed at 25 ºC with a flow rate of 50 µL/min. EpoR binding DARPin ligands were serially

diluted in HBS buffer supplemented with 0.005% P20 surfactant (GE Healthcare) and injected over the SA chip. Dissociation kinetics were monitored for 200-500 seconds, as needed, based on the ligand affinity. For high affinity ligands, the chip was regenerated with 4 M $MgCl_2$ after each ligand injection. The binding kinetics and dissociation constants were determined using the BIAcore T100 evaluation software. The sensograms obtained were either fit to the 1:1 kinetic binding model or to the steady-state affinity model. The kinetic binding curves for each ligand were generated by plotting the time-dependent response units in Prism 7 (GraphPad).

Phospho-Flow Signaling Assays. UT7/EPO cells were starved overnight for ~18 h in base media without EPO. The cells were stimulated with EPO or the DARPin ligands for 15 mins at 37 ºC, followed by fixation with paraformaldehyde (Electron Microscopy Sciences) for 10 mins at room temperature. Next, the cells were permeabilized for intracellular staining by treatment with methanol (Fisher) for 30 mins at -20 ºC. The cells were then incubated with the desired antibodies at a 1:50 dilution for 1 hour at room temperature. The levels of the various phosphorylated proteins correspond to the measured values of mean fluorescence intensity (MFI). The background fluorescence of the unstimulated samples was subtracted from all the readouts. Data was acquired using CytoFlex, flow cytometer instrument (Beckman Coulter). The MFI values were normalized to the MFI of EPO at 37.04 nM and plotted in Prism 7 (GraphPad). All the antibodies used were purchased from BD Biosciences: Alexa Fluor 488 Mouse Anti-Stat1 (pY701), Alexa Fluor 647 Mouse Anti-Stat3 (pY705), Alexa Fluor 488 Mouse Anti-Stat3 (pY705), Alexa Fluor 647 Mouse Anti-Stat5 (pY694), and Alexa Fluor 488 Mouse Anti-ERK1/2 (pT202/pY204).

Phosphorylated Protein Analysis. UT7/EPO cells were starved overnight and stimulated with saturating concentrations of EPO and DARPin ligands for 15, 60, and 120 mins at 37 ºC, followed by fixation with paraformaldehyde for 10 mins at room temperature. Next, the cells were washed and resuspended in 0.5% bovine serum albumin in PBS, and stored at -80 ºC. The cells were then prepared for staining according to standard protocols. The cells were stained with a panel of antibodies (Primity Bio Pathway Phenotyping service) and analyzed on LSRII Flow cytometer (Becton Dickinson). The levels of phosphorylated protein were calculated as the Log (base 2) ratio of the MFI of stimulated samples to unstimulated samples.

Generation of EpoR-Bound Extension Models. The synthetic EpoR agonists described in this work have a unique property that differentiate them from previous efforts at generating synthetic type-I cytokine agonists, namely they have fairly well-defined and rigid conformations that provide a structural basis for each observed physiological output. This advantage was leveraged by generating a crystallographic structure-informed set of signaling complex models for each of the A_dist_Rx, A_angle_Rx, and C_angle_Rx series. To best recapitulate both the backbone topology of the P repeat and the N-/C-terminal mode of EpoR binding, the backbone coordinates from the M_R12/EpoR crystal structure were used to construct each experimentally tested member of the extension series using the Rosetta macromolecular modeling suite. While most presented non-clashing complex models upon superposition of the E2-EpoR structure with the extended backbone, several cases showed significant clashes between EpoR subunits that could

not be physically feasible. We used a constrained minimization protocol on each of the clashing complex models to generate a complete set of plausible models.

Backbone and symmetric complex generation. Backbone coordinates for each DARPin extension length were derived from the M_R12/EpoR structure while keeping capping modules intact. For A/C_angle, the N-terminal EpoR binding mode was preserved from M_R12. For A_dist, the C-terminal EpoR binding mode was preserved which included the P' shortened loop modification which was impactful for binding. Each DARPin/EpoR complex was aligned to chain A of dimers A (pdb id 5HRY) and C (pdb id 5KBA). The synthetic signaling complex was generated in the Rosetta macromolecular modeling suite using a symmetry definition file to generate the C2 symmetry starting from each aligned DARPin/EpoR complex. The experimentally tested sequences were threaded to each backbone using the Rosetta packing algorithm with the default talaris2014 score function followed by side-chain energy minimization.

Relax of EpoR clashing models. Models that presented clashes between the membrane-proximal regions of the EpoR ECD were identified by visual inspection. An analysis of the EpoR conformation in all the crystal structures obtained singled residues 113-120 as a hinge region which allows for conformational flexibility between the binding domain and the membrane proximal domain. A Rosetta relax trajectory in search of a state with an energy minimum allowing backbone torsion angles to be modeled flexibly in this region while keeping the rest of the construct fixed using coordinate constraints was carried out for each clashing

model. In the case of dimer A_R3, the first attempt with Rosetta relax to relieve the severe clashes between EpoR subunits was unsuccessful without having to make significant moves to the head region or rearranging the designed dimer interface orientation, so the EpoR conformation from A_angle_R5 was used instead for the relax trajectory. The relaxed output EpoR conformation from dimer A_R3 was used to rescue a non-clashing model for the A_angle_R4 complex. Although the generated non-clashing models do not represent unique solutions, they demonstrate the steric plausibility of a given complex.

Structural analysis of complex models. Although the synthetic agonists themselves are rather rigid, the EpoR ECD has some intrinsic flexibility between the D1 and D2 domains giving rise to uncertainty in measurements derived from models. Indeed, multiple EpoR conformers were observed in the crystal structures obtained for this work (**Appendix Figure 3.10**). Because the DARPin-EpoR D1 interaction is rigid, we made all distance and angle measurements using Val119 of EpoR, at the hinge between D1 and D2. Distance measurements between EpoR ECDs were made between the Cαs of the residue Val119. Scissor half angle measurements were taken using Val119 of EpoR as the vertex with rays extending through Asp2 at the N-terminal cap of the DARPins and through an alanine at the start of the C-terminal cap of the DARPin (due to varying insertion lengths in DARPins with X repeats, this was residue number 39 + 33X). Although the measurements presented (**Figures 3.3 and 3.4**) represent the general trend of complex geometries within the extension series, these measurements do not directly correspond to the relationship between the transmembrane or intracellular portions of EpoR in the cellular context.

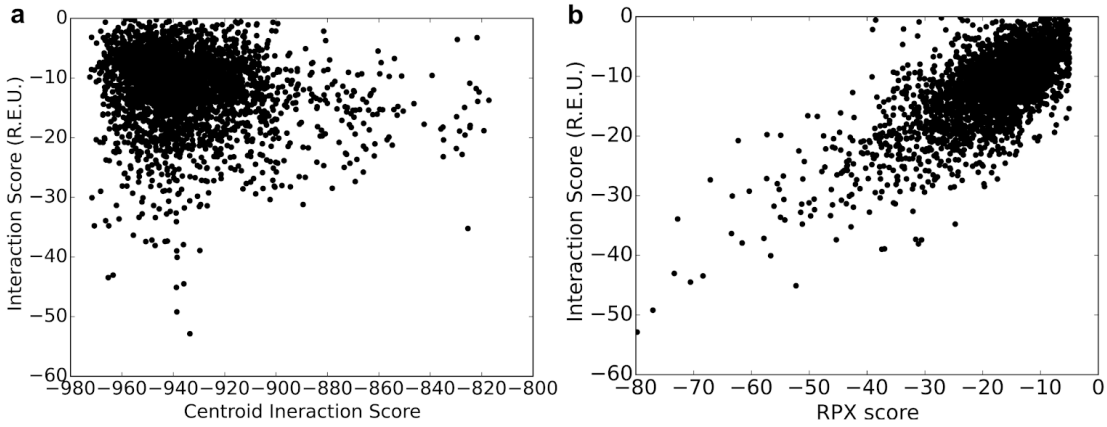
Single molecule fluorescence imaging. Experiments on single molecule level were performed by total internal fluorescence (TIRF) microscopy. An inverted microscope (Olympus IX71) equipped with a triple-line total internal reflection condenser (Olympus) and a back-illuminated electron multiplied (EM) CCD camera (iXon DU897D, Andor Technology) has been used. The sample was TIR illuminated with a 150x magnification objective with a numerical aperture of 1.45 (UAPO 150 x /1.45 TIRFM Olympus). The here presented experiments were performed at room temperature in presence of medium without phenol red. Additionally, each sample was supplemented with an oxygen scavenger and a redox-active photo protectant to minimize photo bleaching. For cell surface labeling of mXFP-tagged EpoR equal concentration of DY647- and Rho11-labeled Nanobodies (NBs) were added to the sample and incubated at least for 5 min. High equilibrium binding of the labeled NBs was ensured by keeping these in bulk solution. Rho11 was excited by 561 nm laser (CrystaLaser) at 0.95 mW (~32 W/cm^2) and DY647 by a 642 nm laser (Omicron) at 0.65 mW (~22W/cm^2). For fluorescent detection a spectral image splitter (DualView, Optical Insight) with a 640 DCXR dichroic beam splitter (Chroma) combined with a 580/40 (Semrock, Rho11 detection) and a 690/70 (Chroma, DY647 detection) bandpass filters has been used dividing each emission channel into 512x256 pixel. For each cell image stacks of 150 frames were recorded with temporal resolution of approximately 32 ms/frame.

Human Primary Cell Culture and Analysis. $CD34^+$ hematopoietic stem and progenitor cells from mobilized peripheral blood were purchased from the Fred Hutchinson Cancer Research

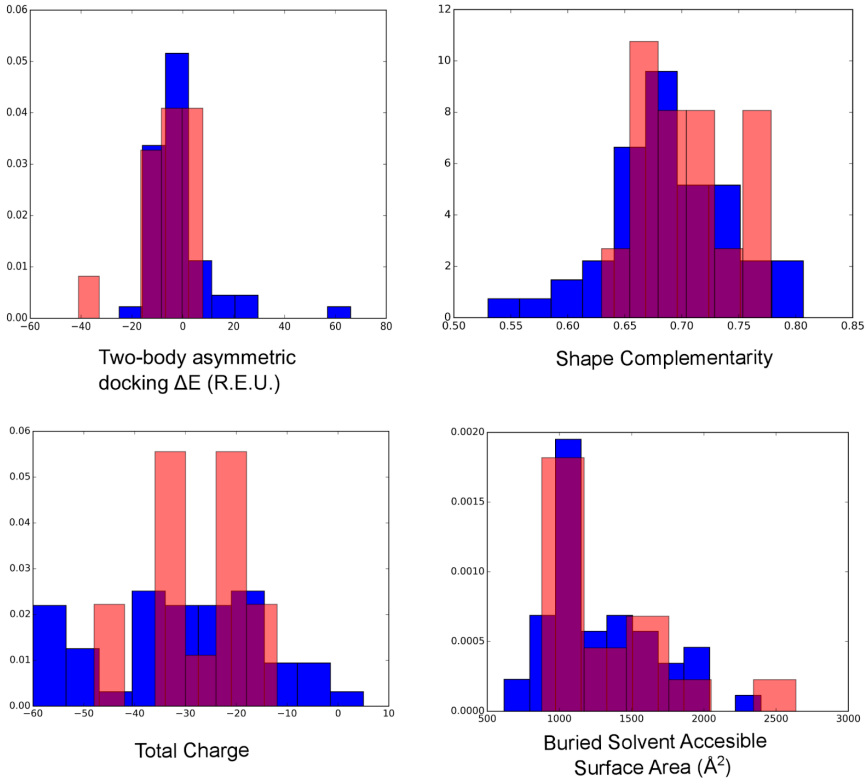
Center. Cells were cultured with one of following ligands: EPO (3 U/mL), A_R3 (10 nM), A_angle_R4 (1 μM), A_angle_R5 (1 μM), A_angle_R6 (1 μM), C_R3 (10 nM), C_angle_R4 (100 nM), C_angle_R5 (100 nM), C_angle_R6 (100 nM), A_dist_R5 (333 nM), and A_dist_R7 (1 μM). Cell concentration was measured using an automated cell counter (Beckman Coulter). To evaluate erythroid differentiation, cells were stained with CD71-Fluorescein Isothiocyanate (FITC), CD11b/ CD41a-Phycoerythrin (PE), and CD235a-Allophycocyanin (APC). Propidium Iodide (PI; 1:1,000) was used to discriminate live and dead cells. All antibodies were used at a 1:20 dilution, unless otherwise noted.

For phospho-STAT staining, differentiated primary erythroid cells were starved in cytokine-free media (1% BSA in IMDM) for 4 hours and stimulated either with EPO (3U/mL), A_R3 (10 nM), or A_dist_R5 (333 nM) for 1 hour. Treated cells were fixed and permeabilized then stained with Alexa Fluor-647 Mouse Anti-STAT5 (pY694; 1:20 dilution), Alexa Fluor-647 Mouse Anti-STAT3 (pY705; 1:20 dilution), or Alexa Fluor-647 Mouse Anti-STAT1 (pY701; 1:20 dilution). Data were acquired using either BD Accuri C6 Cytometer or Canto instruments (Becton, Dickinson and Company) and all analysis was performed using FlowJo.

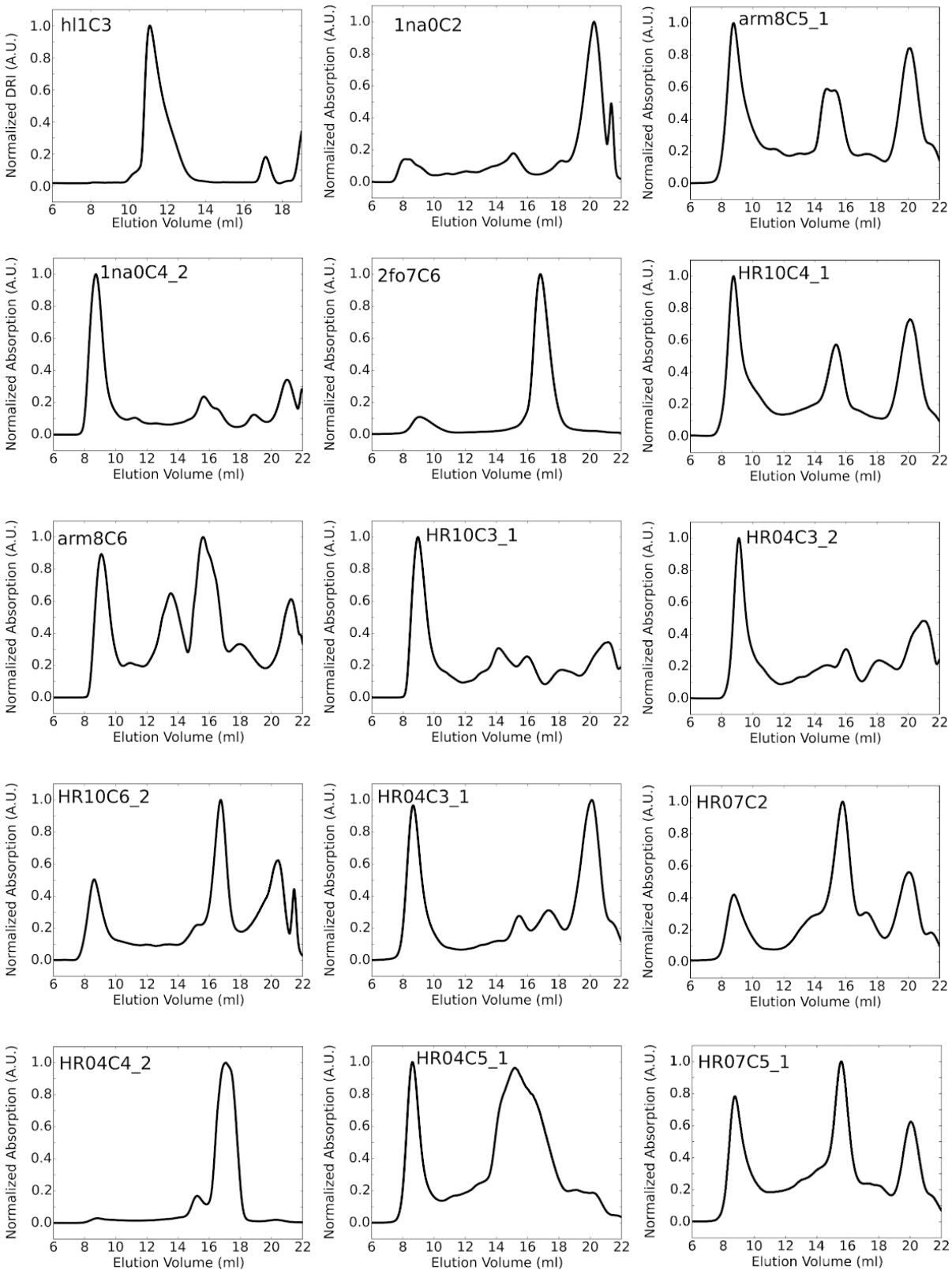
Appendix Figures

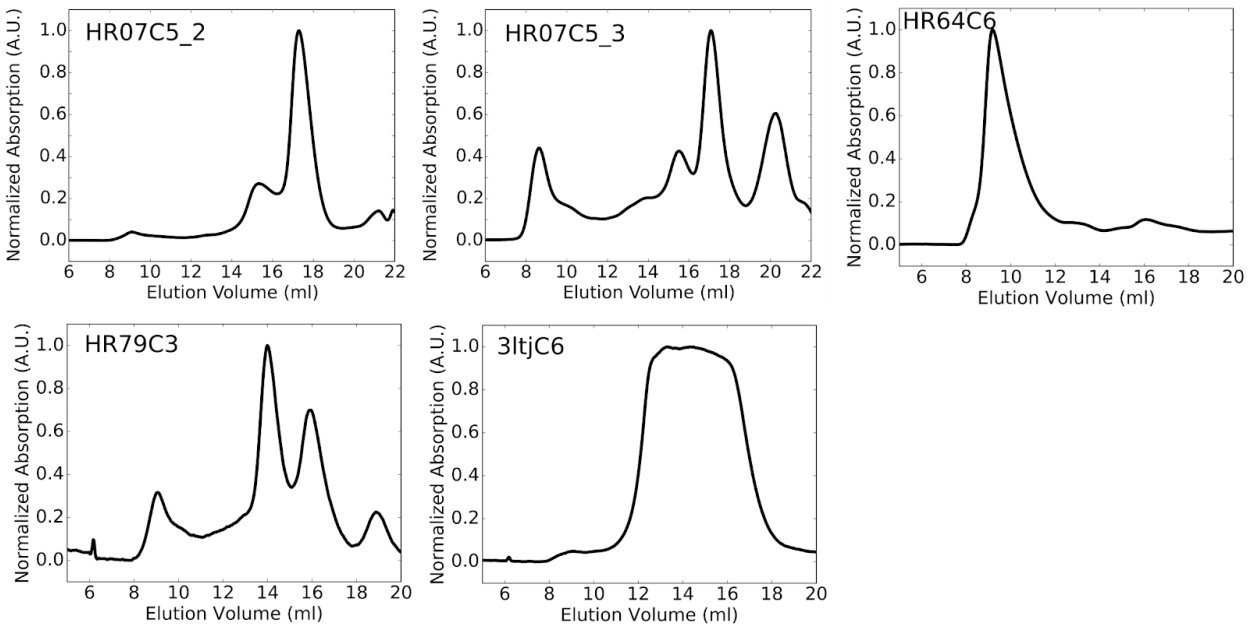


Appendix Figure 1.1. Comparison of the full atom and coarse-grained computed binding energies for a set of designed repeat protein oligomers. **a**, Rosetta full atom energy function and a sequence-agnostic Rosetta centroid score function. **b**, Rosetta full atom energy function and RPX score as utilized to design the homooligomers.

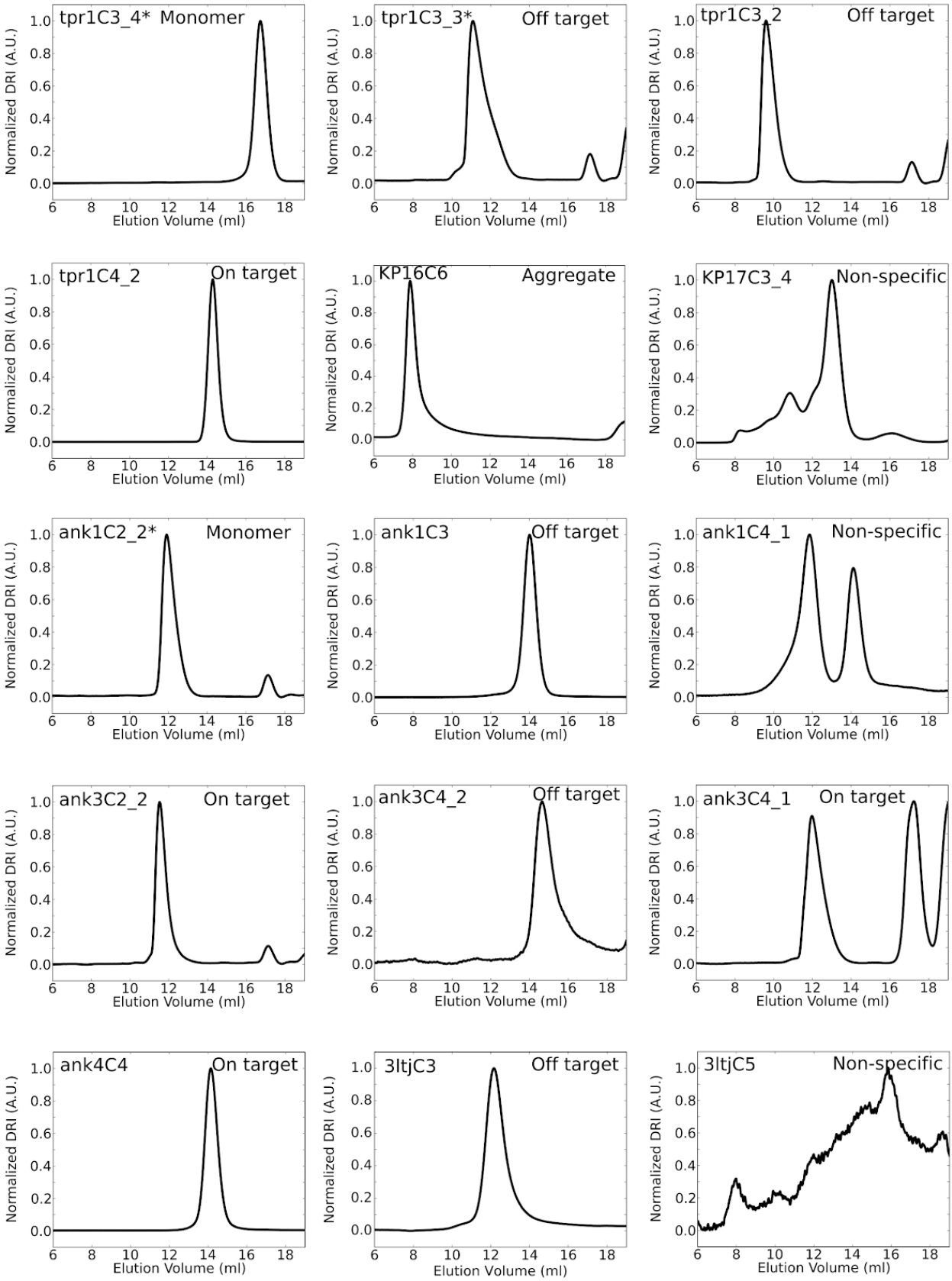


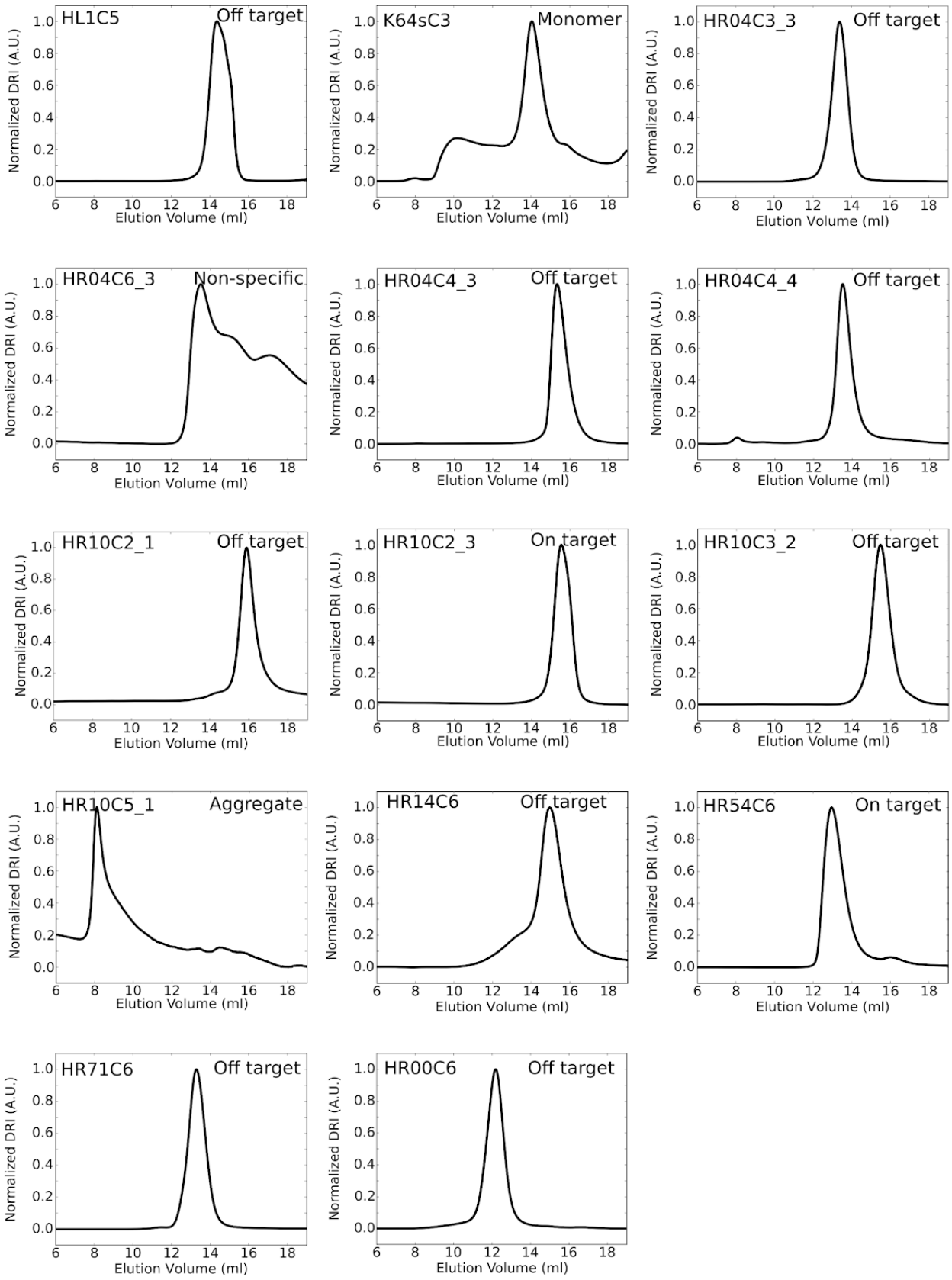
Appendix Figure 1.2. Histograms of computational metrics for designs that expressed solubly. Structurally validated designs are shown in red and unsuccessful designs in blue. See **Appendix Methods** for description on two-body asymmetric docking.


hl1C3
1na0C2
arm8C5_1
1na0C4_2
2fo7C6
HR10C4_1
arm8C6
HR10C3_1
HR04C3_2
HR10C6_2
HR04C3_1
HR07C2
HR04C4_2
HR04C5_1
HR07C5_1
Normalized DRI (A.U.)
Normalized Absorption (A.U.)
Elution Volume (ml)



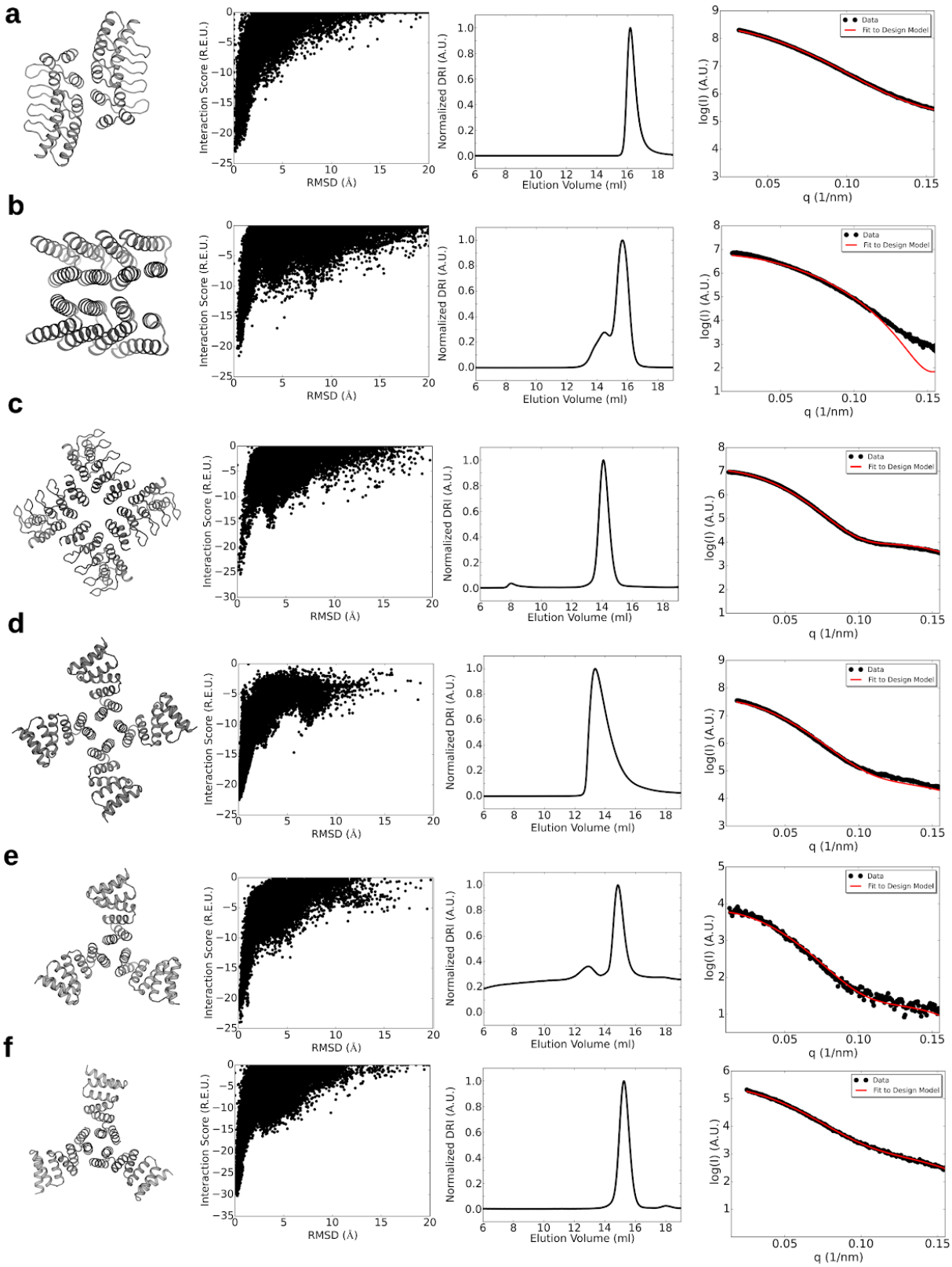
Appendix Figure 1.3. SEC profiles for failed designs. Primary size exclusion chromatograms obtained from a Superdex 200 gel filtration column for expressed proteins directly after purification by immobilized Ni^{2+} affinity chromatography. Designs listed here either obviously formed non-specific assemblies and/or did not appear to form the target oligomeric species based on the column retention time of their corresponding monomeric scaffolds. 5 designs presented here had monodisperse profiles with retention times that were estimated to form off-target oligomeric species, whereas 15 designs had polydisperse or aggregated profiles.

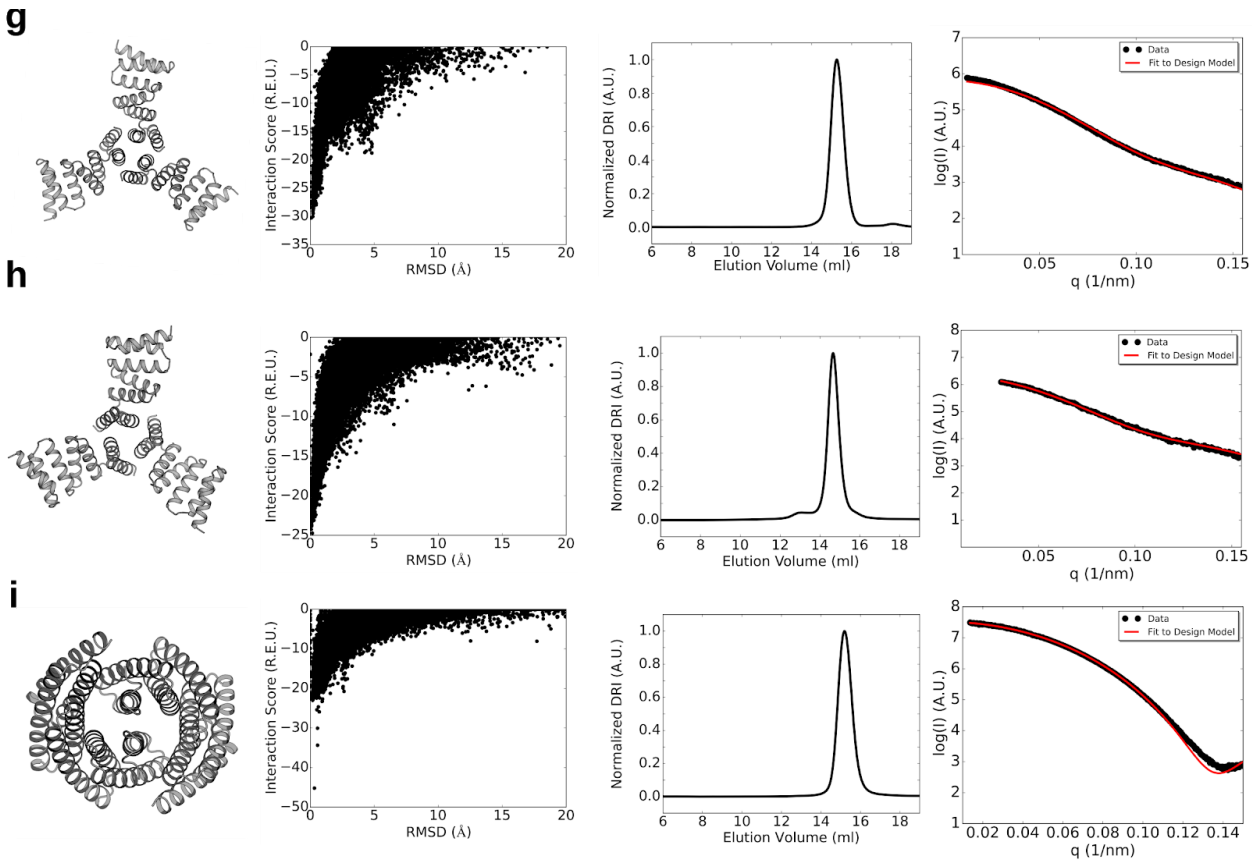

tpr1C3_4* Monomer
tpr1C3_3* Off target
tpr1C3_2 Off target
tpr1C4_2 On target
KP16C6 Aggregate
KP17C3_4 Non-specific
ank1C2_2* Monomer
ank1C3 Off target
ank1C4_1 Non-specific
ank3C2_2 On target
ank3C4_2 Off target
ank3C4_1 On target
ank4C4 On target
3ltjC3 Off target
3ltjC5 Non-specific
Normalized DRI (A.U.)
Elution Volume (ml)



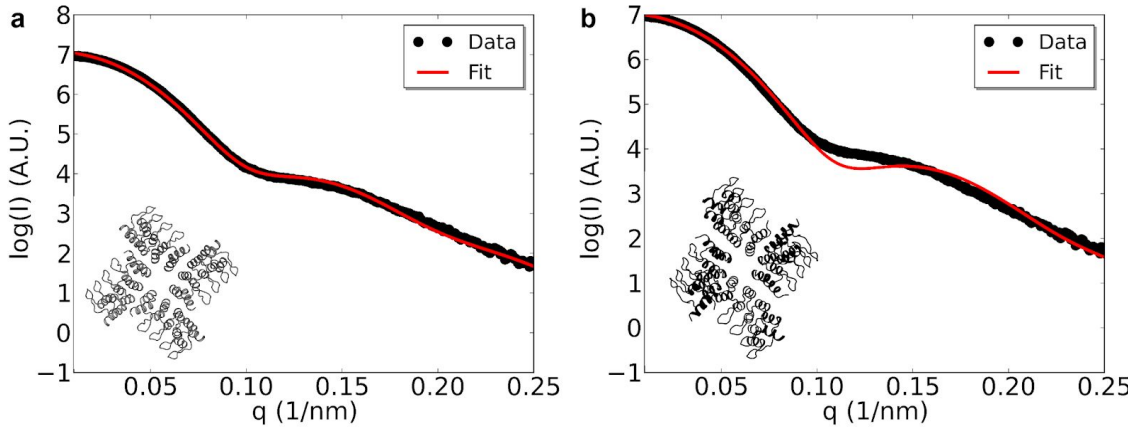
* Size-exclusion chromatogram produced using a Superdex 75 10/300 GL gel filtration column

Appendix Figure 1.4. SEC-MALS chromatograms for failed designs. Predominant oligomeric species for each design were collected by fractionation from a primary size exclusion run, and 29 sizing profiles are presented here from the subsequent round of high-performance size exclusion chromatography from a Superdex 200 gel filtration column. 21 designs including those in **Figure 1.2** and **Appendix Figure 1.5** had monodisperse profiles with calculated molecular weights within 13% of the target oligomeric species (on target). 6 designs shown here had χ values between computed and experimental SAXS profiles of greater than 3.1, and were thus considered to adopt off-target structures but were in the correct oligomeric state. 17 designs formed monodisperse species in off-target oligomeric states (monomer or off target), 4 designs displayed polydisperse profiles indicating multiple oligomeric states (non-specific), and 2 designs formed soluble aggregates (aggregate).

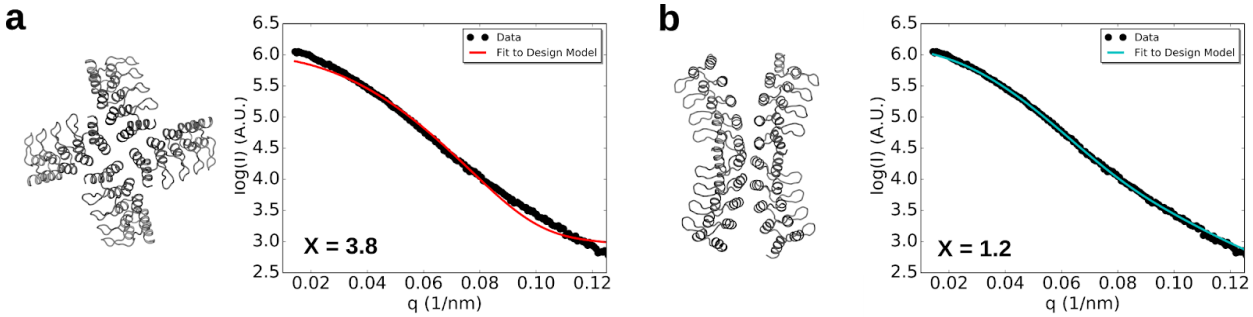

a
b
c
d
e
f
Interaction Score (R.E.U.)
RMSD (Å)
Normalized DRI (A.U.)
Elution Volume (ml)
log(I) (A.U.)
q (1/nm)
Data
Fit to Design Model



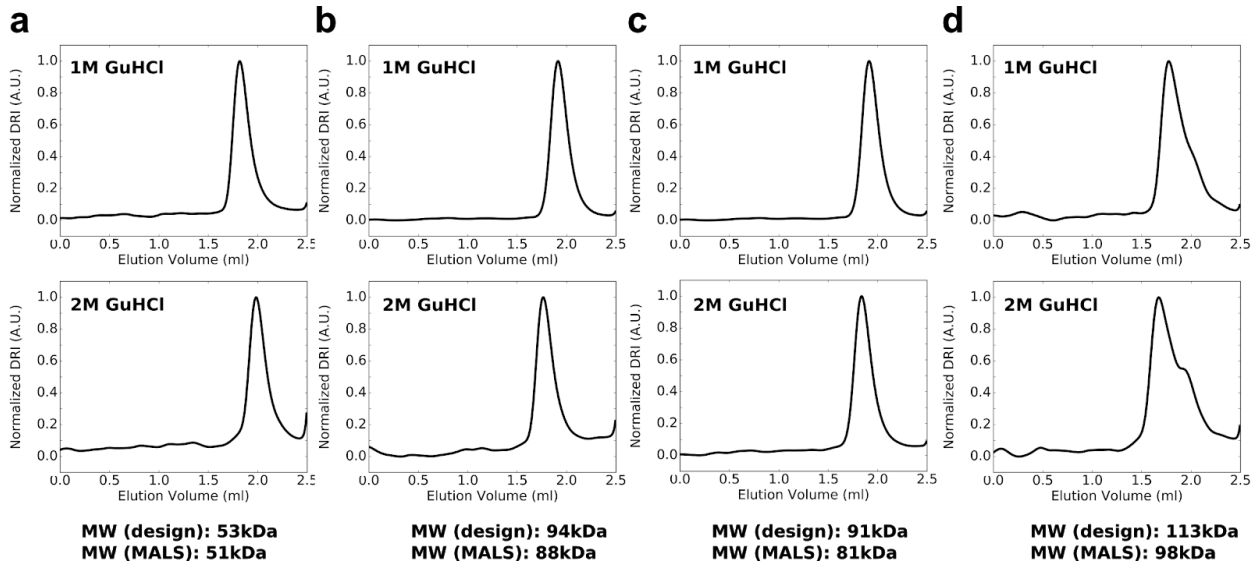
Appendix Figure 1.5. Assessment of the solution conformation of selected cyclic oligomers. From left to right: computational model, symmetric docking funnel, SEC chromatogram used for molecular weight determination and SAXS scattering profiles, measured (black dots) and computed from the model (red line). **a,** ank1C2_1. **b,** HR10C2_2. **c,** ank1C4_2. **d,** 1na0C4_1. **e,** 1na0C3_1. **f,** 1na0C3_3. **g,** 1na0C3_5. **h,** 1na0C3_7. **i,** HR81C2.



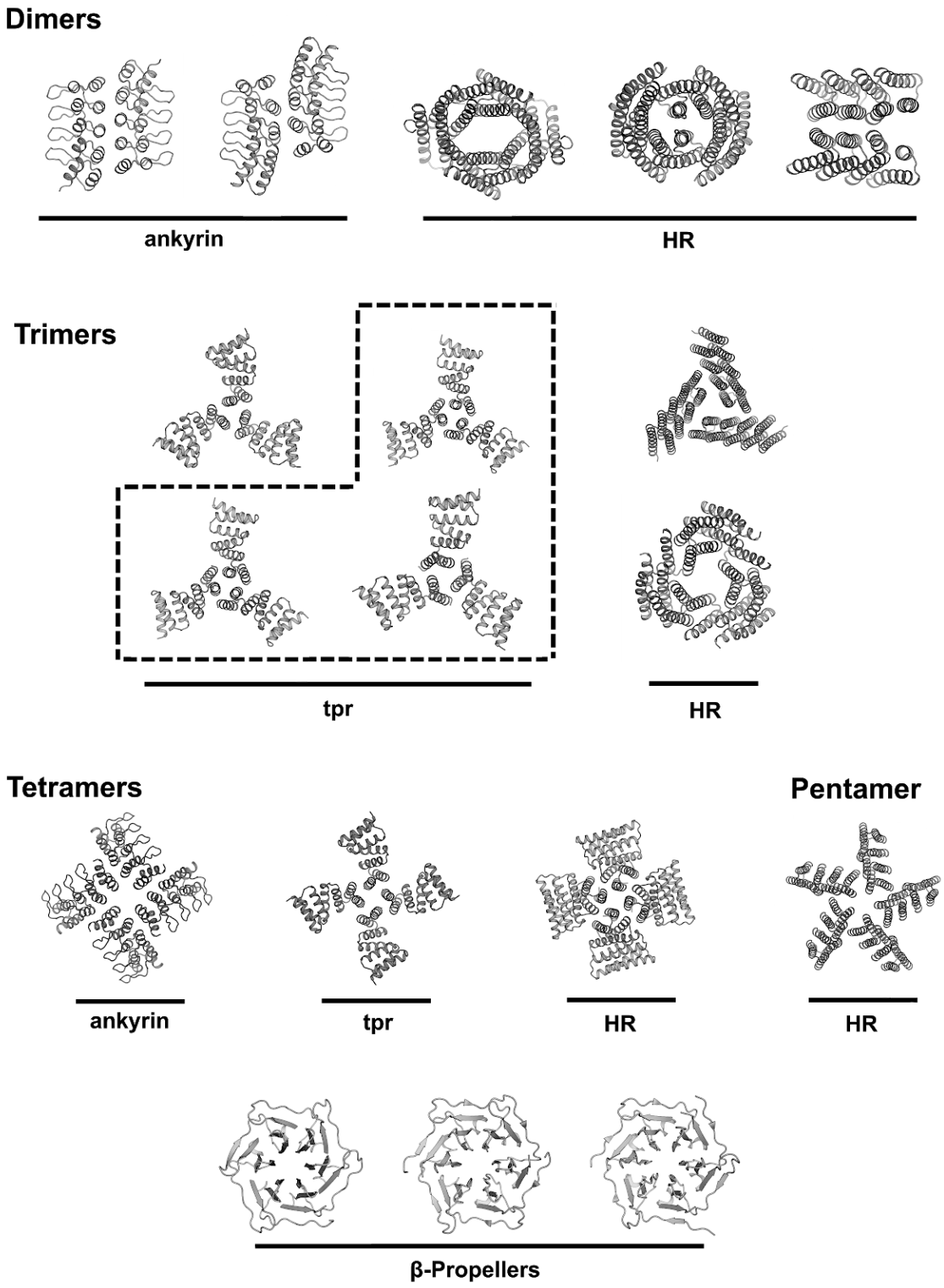
Appendix Figure 1.6. Solution conformation of design ank1C4_2. **a**, Comparison between the experimental SAXS scattering profile (black dots) and C4 symmetric design model (red line). **b**, Comparison between the experimental SAXS scattering profile (black dots) and C2 tetramer found in the X-ray structure.



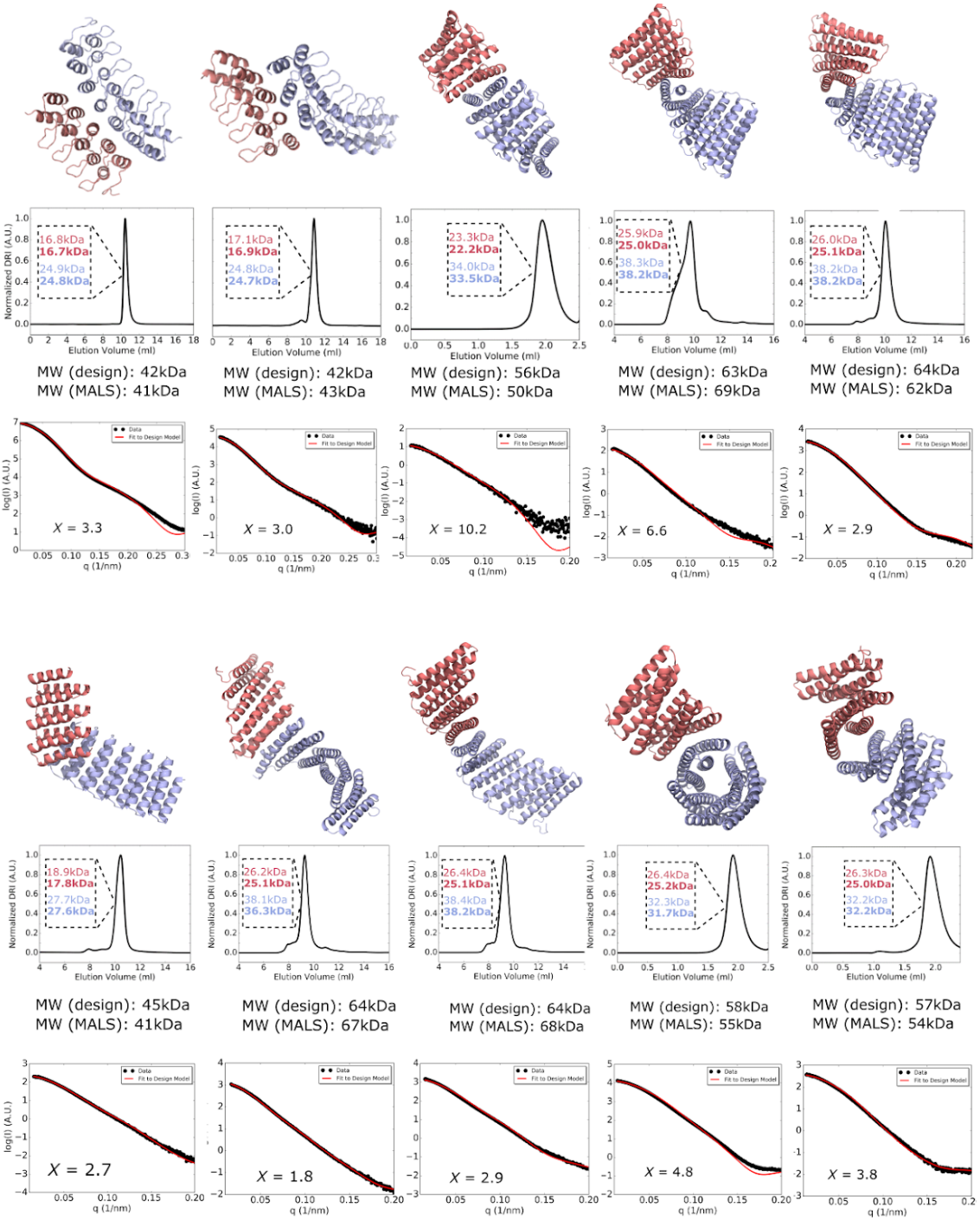
Appendix Figure 1.7. Solution-state SAXS profile comparison for ank4C4 design model and corresponding crystal structure. **a,** Comparison between the experimental SAXS scattering profile (black dots) and C4 symmetric design model (red line). **b,** Comparison between the experimental SAXS scattering profile (black dots) and D2 symmetric crystal structure (cyan line).



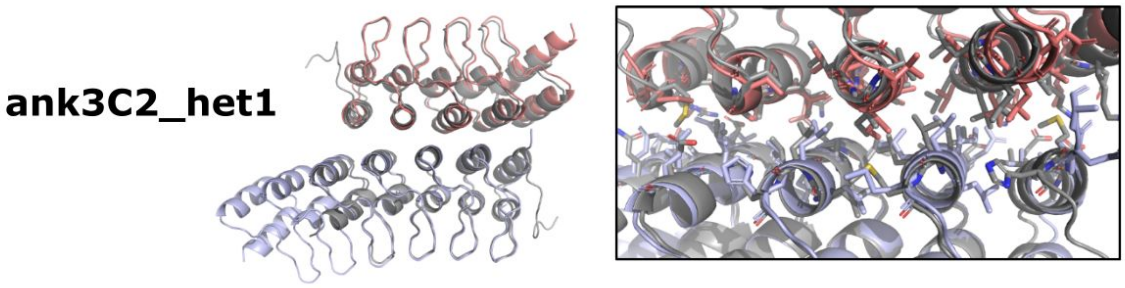
Appendix Figure 1.8. SEC-MALS of selected designs in 1M and 2M GuHCl. Secondary size exclusion profiles obtained from a Superdex 200 5/150 GL gel filtration column coupled to multi-angle light scattering in 25mM Tris 150mM NaCl pH 8.2 1M GuHCl (top row) and 2M GuHCl (bottom row). "MW (design)" refers to the molecular weight of the target oligomer, whereas "MW (MALS)" refers to the experimentally determined molecular weight for the predominant species. **a,** HR81C2. **b,** HR00C3_2. **c,** HR04C4_1. **d,** HR10C5_2.



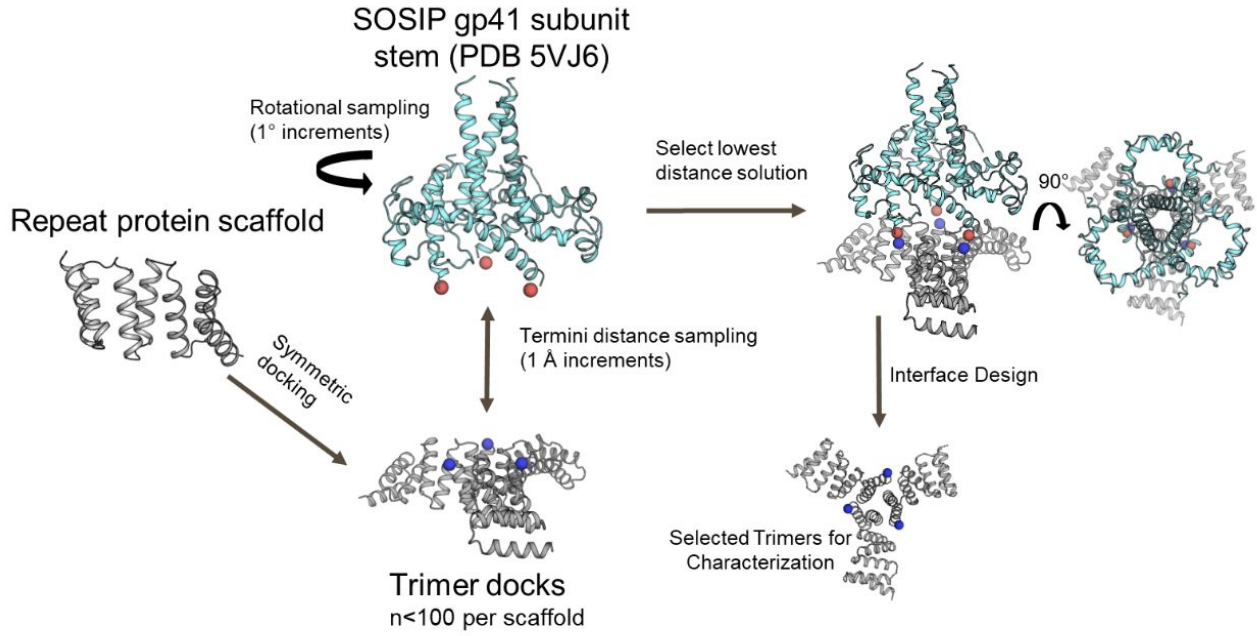
Appendix Figure 1.11. Structural comparison of the 15 validated designs. Designs are grouped by repeat scaffold family and oligomerization state; β-propellers (PDB ID from left to right 3ww9, 3ww8, 3wwb) are shown for structural diversity comparison. Boxed designs share < 2.5Å r.m.s.d.



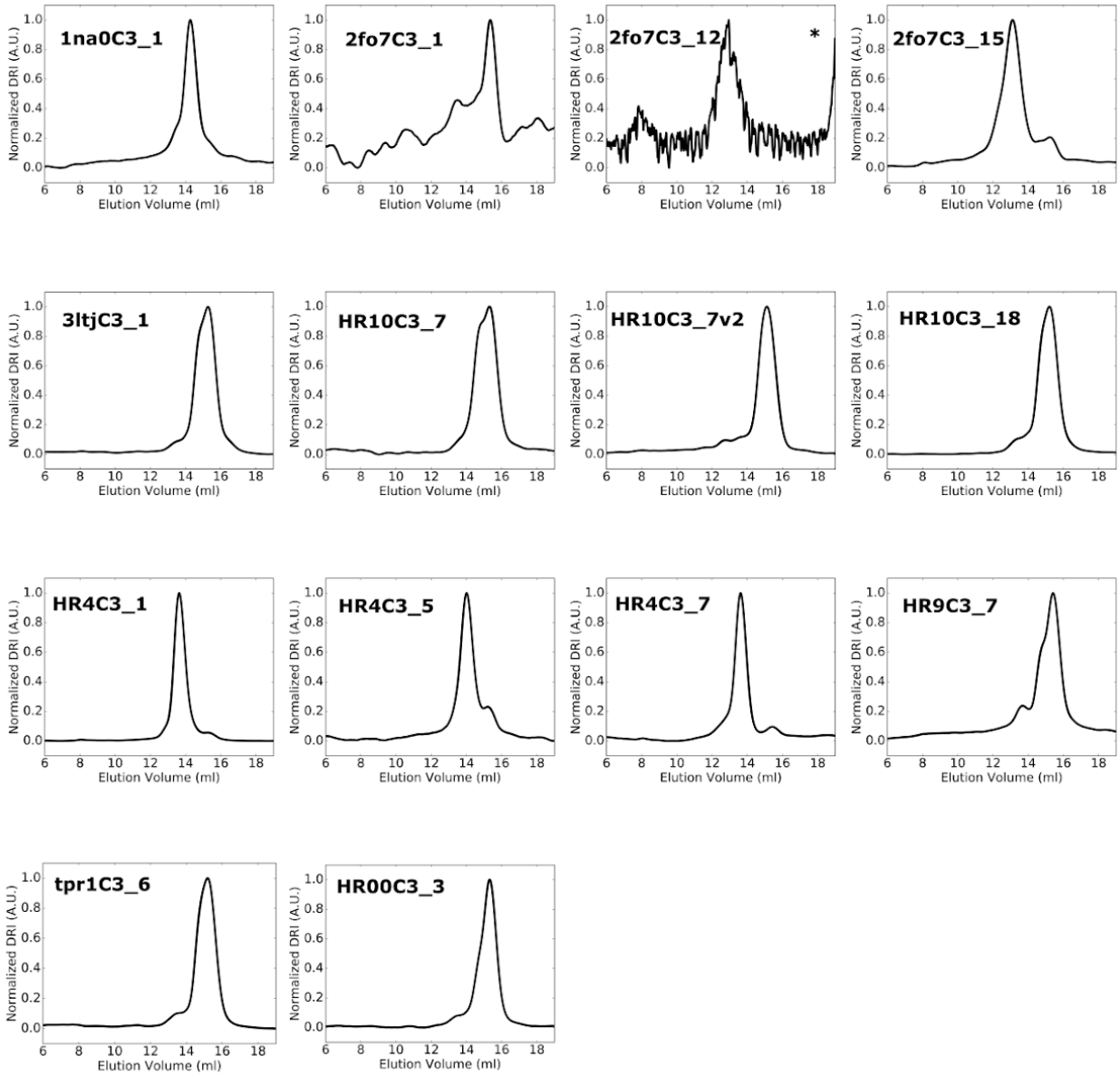
Appendix Figure 1.12. Solution conformation assessment of designed pseudosymmetric heterodimers. Top and bottom panels from left to right: ank3_asym1, ank1_asym1v1, HR04_asym08, HR18_asym12, HR18_asym13, HR49_asym24, HR79_asym02, HR79_asym04, HR81_asym05, HR81_asym09 design models depicting chain A (n repeats) in red and chain B (n+2 repeats) in blue; SEC-MALS chromatograms with molecular weights of the two most predominant species by LC-MS (predicted values in bold, experimental values in normal font); SAXS profile comparisons between experimental scattering profiles at 1-5 mg/ml (black) and computed profile from the computational model (red).



Appendix Figure 1.13. Comparison between the experimentally determined crystal structures and corresponding models of a designed heterodimer based off the ankyrin fold. Crystal structure is shown in color and the computational model in gray. Left, pseudosymmetric computational model and crystal structure superposition (0.6Å r.m.s.d. between heavy atoms); Right, superposition showing side chain rotamers at the designed interface.

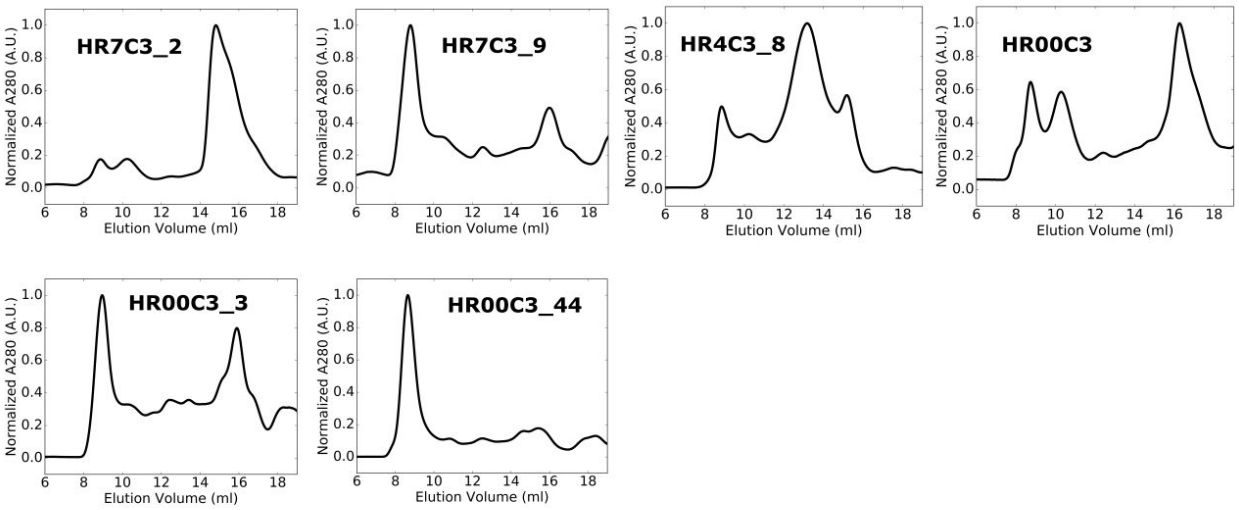


Appendix Figure 2.1. *In silico* sampling method to select C3 trimers (N-terminal residue labeled in blue) capable of scaffolding target antigens. BG505 SOSIP (C-terminal residue 664 labeled in red) trimeric glycoprotein subunits gp41 (stem) shown in schematic for illustration.

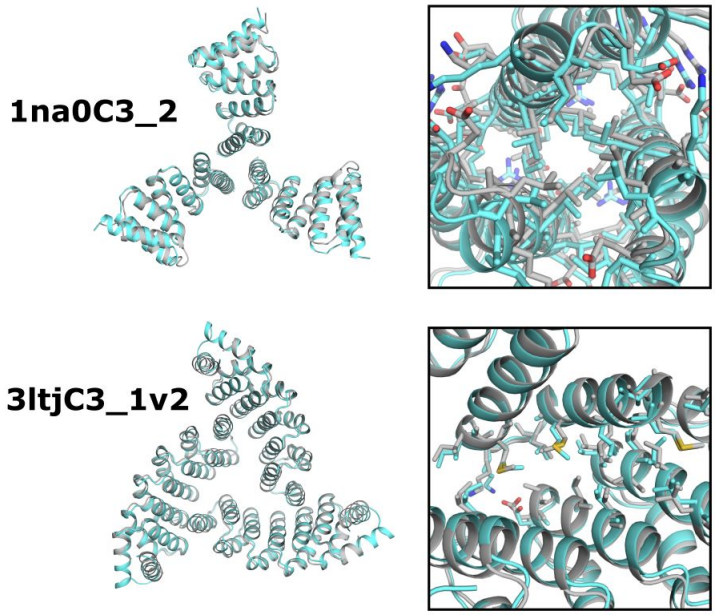


* Chromatogram obtained on Superdex 75 10/300 GL gel filtration column.

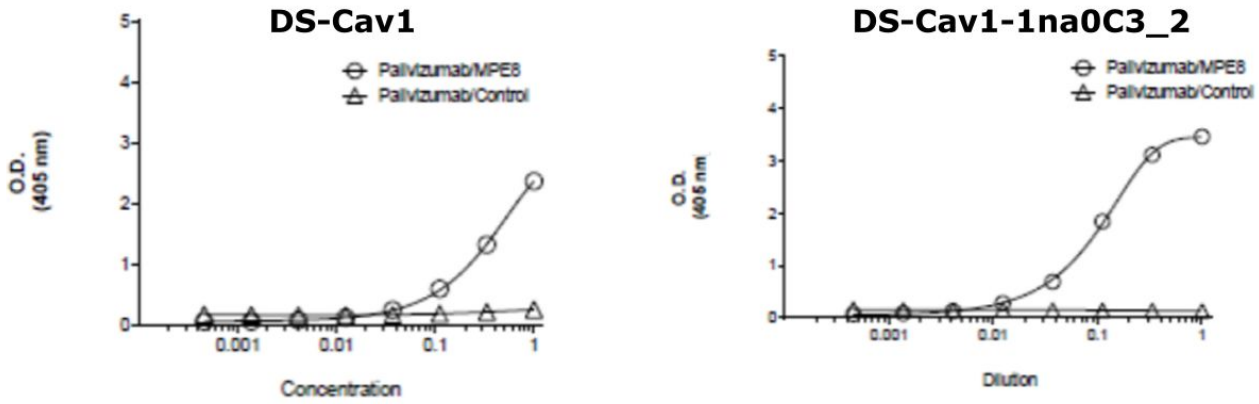
Appendix Figure 2.2. SEC-MALS chromatograms for failed designs, occupying an unintended oligomerization state or supramolecular configuration. Predominant oligomeric species for each design were collected by fractionation from a primary size exclusion run, and 14 sizing profiles are presented here from the subsequent round of high-performance size exclusion chromatography from a Superdex 200 gel filtration column.



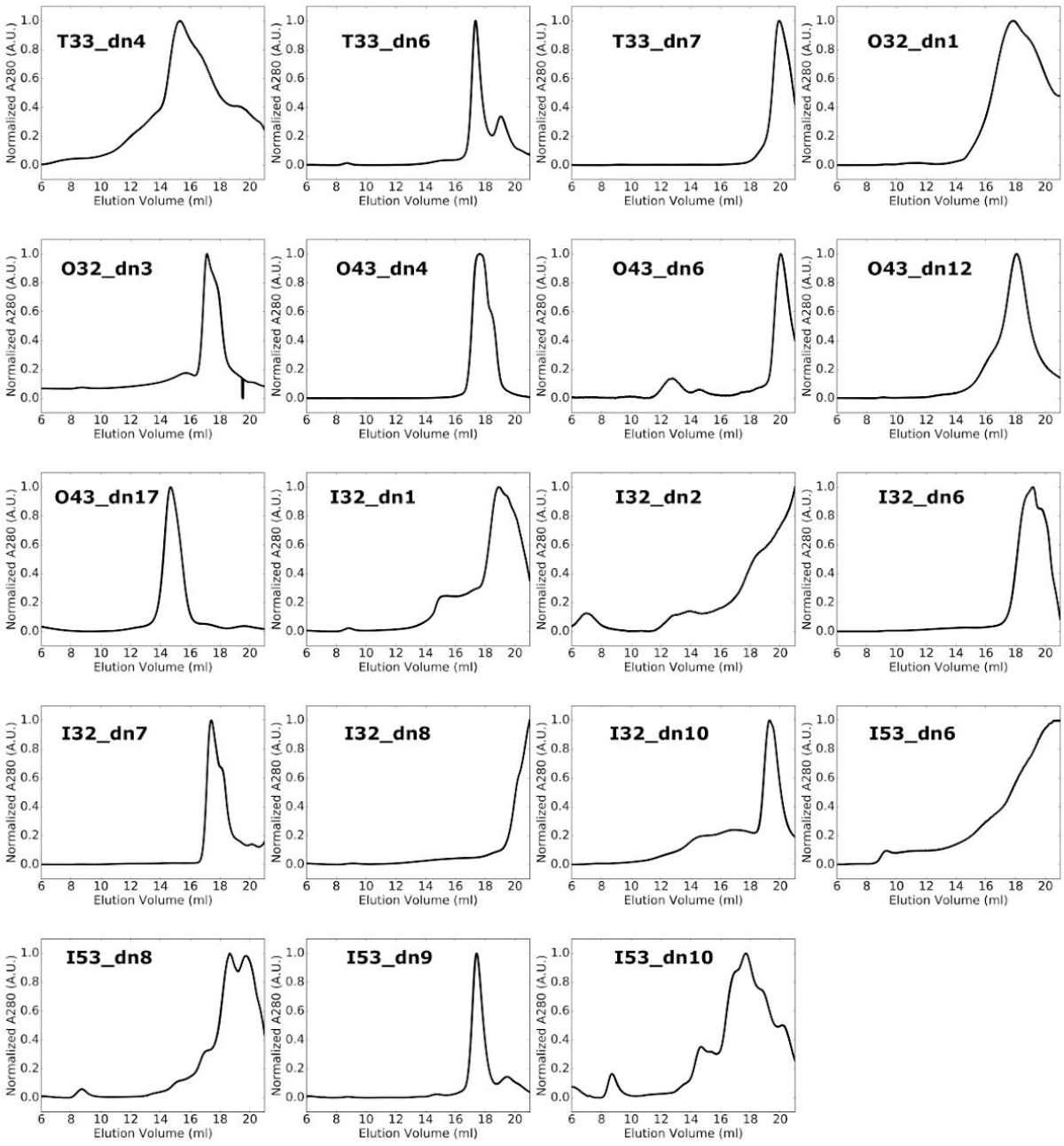
Appendix Figure 2.3. SEC profiles for designed antigen-scaffolding trimers. Primary size exclusion chromatograms obtained from a Superdex 200 gel filtration column for soluble proteins directly after purification by immobilized Ni^{2+} affinity chromatography. 2 designs displayed polydisperse profiles indicating formation of off-target oligomers, whereas 4 designs formed soluble aggregate.



Appendix Figure 2.4. Comparison between the experimentally determined crystal structures and corresponding models of designed trimers. Crystal structures are shown in cyan and models in gray. Left column, full model and crystal structure superposition; Right column, superposition showing chains and side chains at the designed interface.



Appendix Figure 2.5. ELISA of DS-Cav1 with foldon domain and DS-Cav1 fused to designed trimer 1na0C3_2. Proper folding of DS-Cav1 was assessed with conformationally-specific antibody Palivizumab. ELISA fluorescence signal is plotted as a function of cell supernatant dilution.



Appendix Figure 2.6. SEC profiles for designed two-component nanoparticles. Primary size exclusion chromatograms obtained from a Superose 6 gel filtration column for soluble proteins directly after purification by immobilized Ni^{2+} affinity chromatography. Designs presented here seemed to form off-target assemblies based on retention volume, occupy polydisperse complex states, or preserve their state as individual oligomeric components indicated by doublets late in the chromatogram.

A *Ankyrin Consensus Sequence*

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
	β Turn						α Helix 1										α Helix 2															
D	X	X	G	X	T	P	L	H	L	A	A	X	X	G	H	L	E	I	V	E	V	L	L	K	N	G	A	D	V	N	A	X

B *N2C Library Design*

N	ANKYRIN 1						ANKYRIN 2					
32	2	3	5	13	14	33	2	3	5	13	14	33
R	N	A	A	R	A	R	N	N	A	R	A	
S	D	R	L	N	N	N	D	S	L	S	N	
T	H	G	M	Q	D	K	H	T	M	T	D	
	I	L	S	H	E	S	I	Y	S	W	E	
	L	M	T	I	I	T	L		T		I	
	P	S	V	L	K		P		V		K	
	Y	T		K	T		Y				T	
	V	W		M	V		V				V	
		V		S								

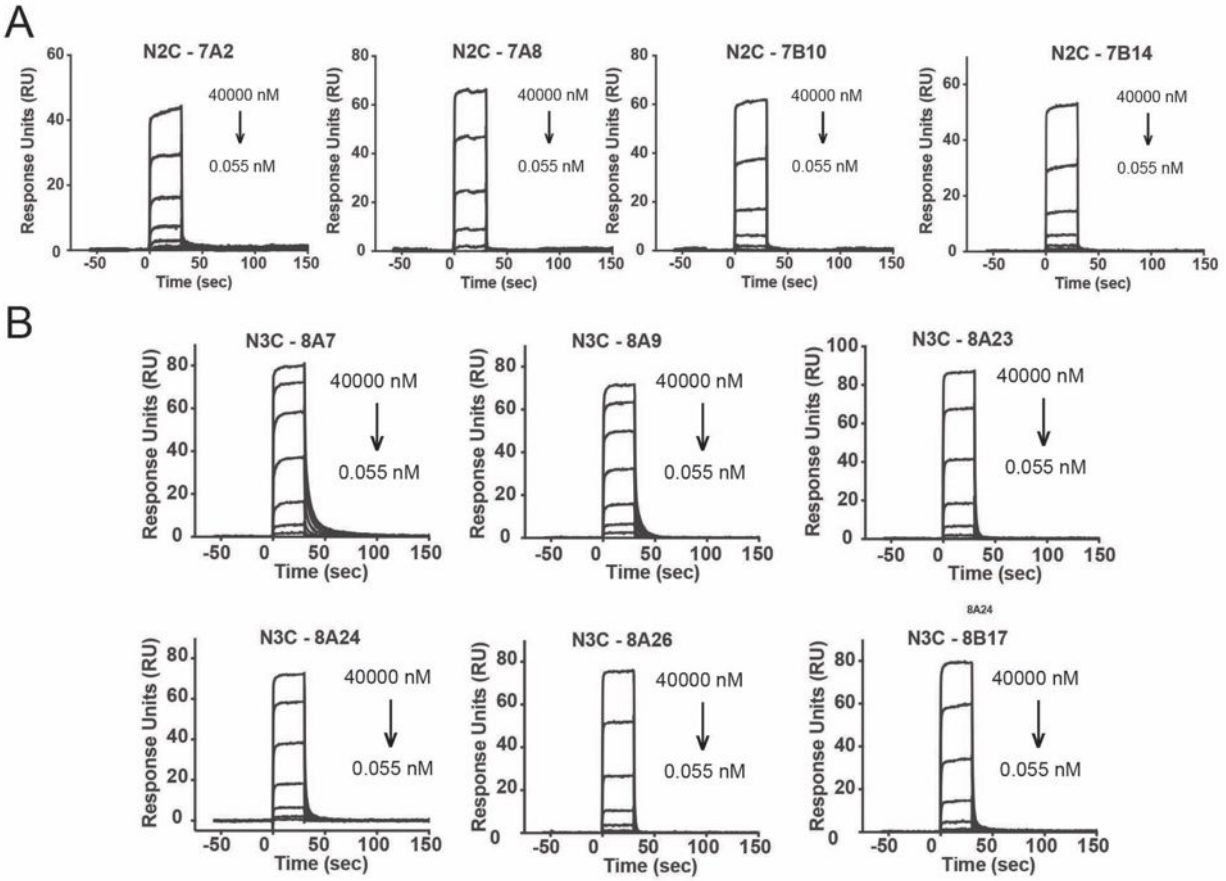
Theoretical diversity	2.1×10^{10}
Library diversity	1.8×10^{9}

C *N3C Library Design*

N	ANKYRIN 1						ANKYRIN 2						ANKYRIN 3				
32	2	3	5	13	14	33	2	3	5	13	14	33	2	3	5	13	14
R	D	N	A	Q	A	R	N	N	A	N	R	Y	D	N	A	R	N
S	H	S	F	I	N	N	Q	S	L	I	N	S	H	S	M	S	I
T	L	T	S	L	D	K	H	T	M	S	I		L	T	T	T	K
	V	Y	V	K	I	S	I	Y	S	T	K		V	Y	V	W	M
					T	T	L		T		M						
					V		K		V		S						
							M										

Theoretical diversity	5.2×10^{11}
Library diversity	1.0×10^{9}

Appendix Figure 3.1. A, The sequence of the Ankyrin consensus repeat, which was used for designing the libraries. 'X' indicates the variable positions in the consensus sequence. Library designs for the yeast-displayed N2C (**B**) and N3C (**C**) DARPin constructs. The design is based on the DARPin consensus sequence. The amino acid mutations at each position are listed under their respective X position.

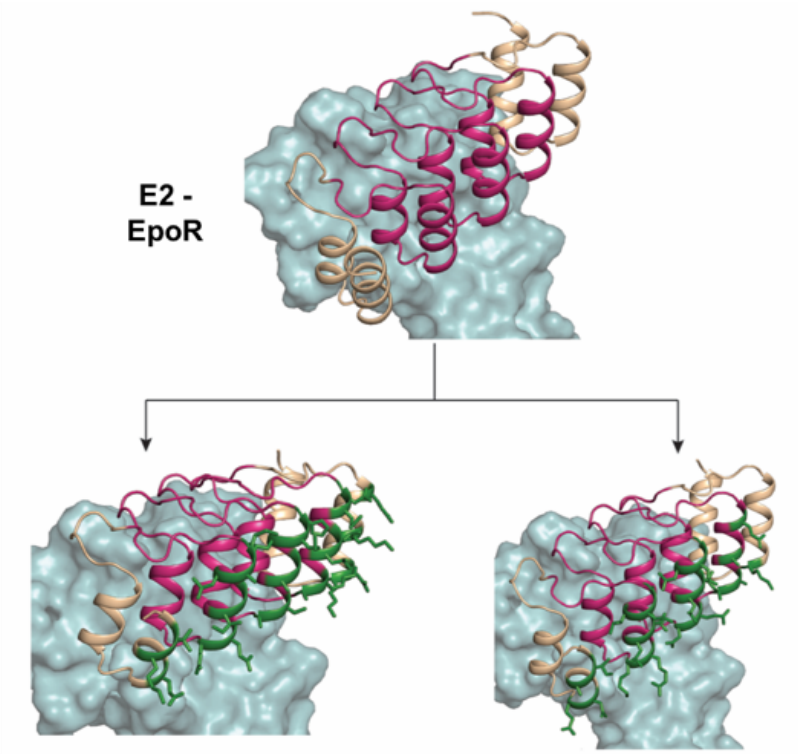


C

	K_d Values (μM)
N2C – 7A2	10.8
N2C – 7A8	10.2
N2C – 7B10	18.5
N2C – 7B14	20

	K_d Values (μM)
N3C – 8A7	1.87
N3C – 8A9	2
N3C – 8A23	6.3
N3C – 8A24	5
N3C – 8A26	11.7
N3C – 8B17	7.8

Appendix Figure 3.2. The kinetic binding curves for EpoR binding DARPin N2C (**A**) and N3C (**B**) clones selected from the first-generation yeast-displayed DARPin library. **C**, The K_d values represent SPR measurements and were derived from the steady-state affinity model using the Biacore evaluation software.



E2

MGSDLGKKLL KAARAGQDDE VRILMANGAD VNATDIWDAT PLHLAALIGH
LEIVEVLLKN GADVNASDIT GTTPLHLAAT MGHLEIVEVL LKYGADVNAY
DLNGATPLHL AARMGHVEIVEVLLKYGADV NAQDKFGKTA FDISIDNGNE
DLAEILQ

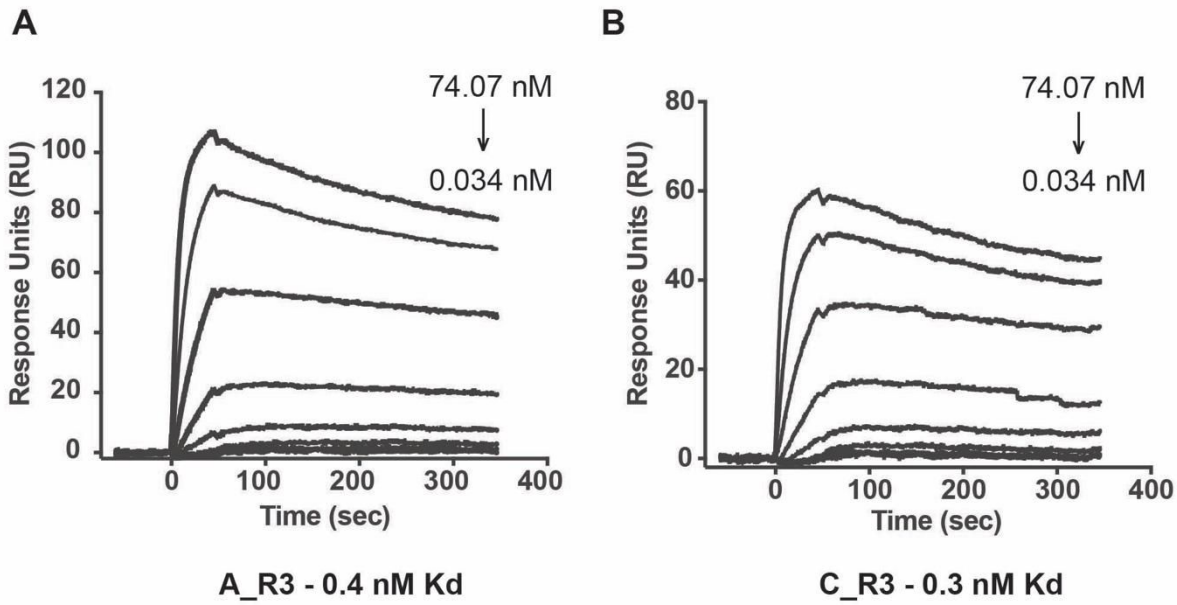
A_R3

MGSDLGKKLL KAARAGQKDE VKILMENGAD VNATDIWDAT PLHLAALIGH
AKIVLLLLEQ GADVNASDIT GTTPLHLAAT MGHAVIVALL LMHGADVNAY
DLNGATPLHL AARMGHEEIV ILLLAMGADV NAQDKFGKTA FDISIDNGNE
ELAKVLQDH

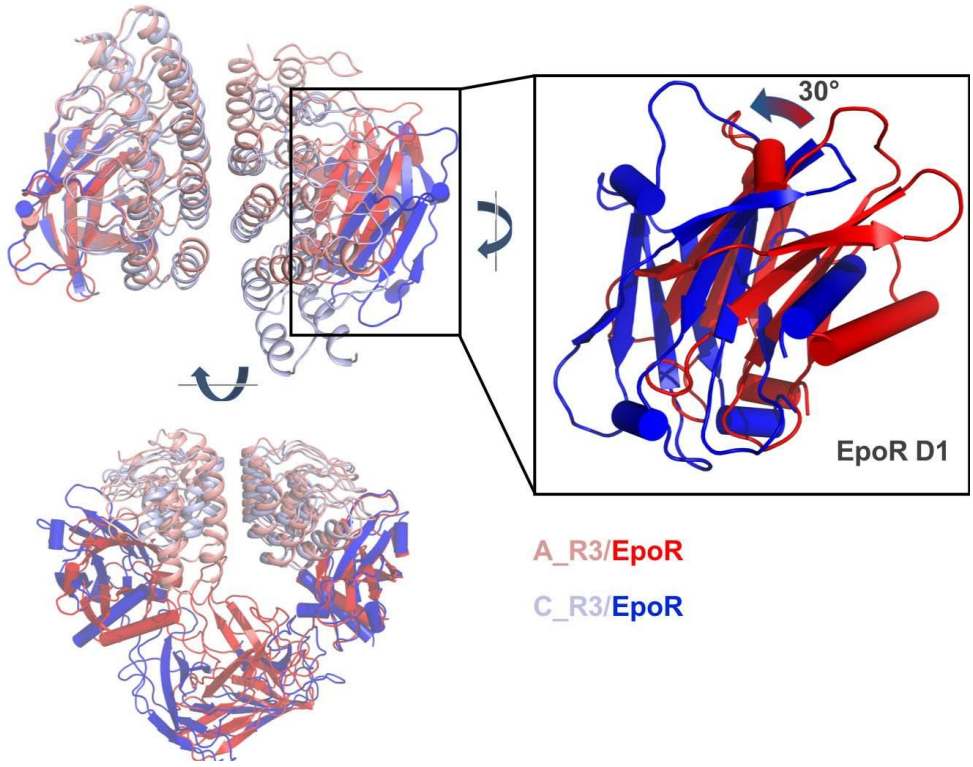
C_R3

MGSDLGKKLL KAARAGQKDE VRILMANGAD VNATDIWDAT PLHLAALIGH
AEIVAVLLEN GADVNASDIT GTTPLHLAAT MGHDEIVLIL LLKGADVNAY
DLNGATPLHL AARMGHKRIV LVLILAGADV NAQDKFGKTA FDISIDNGNE
DLAKILQKQ

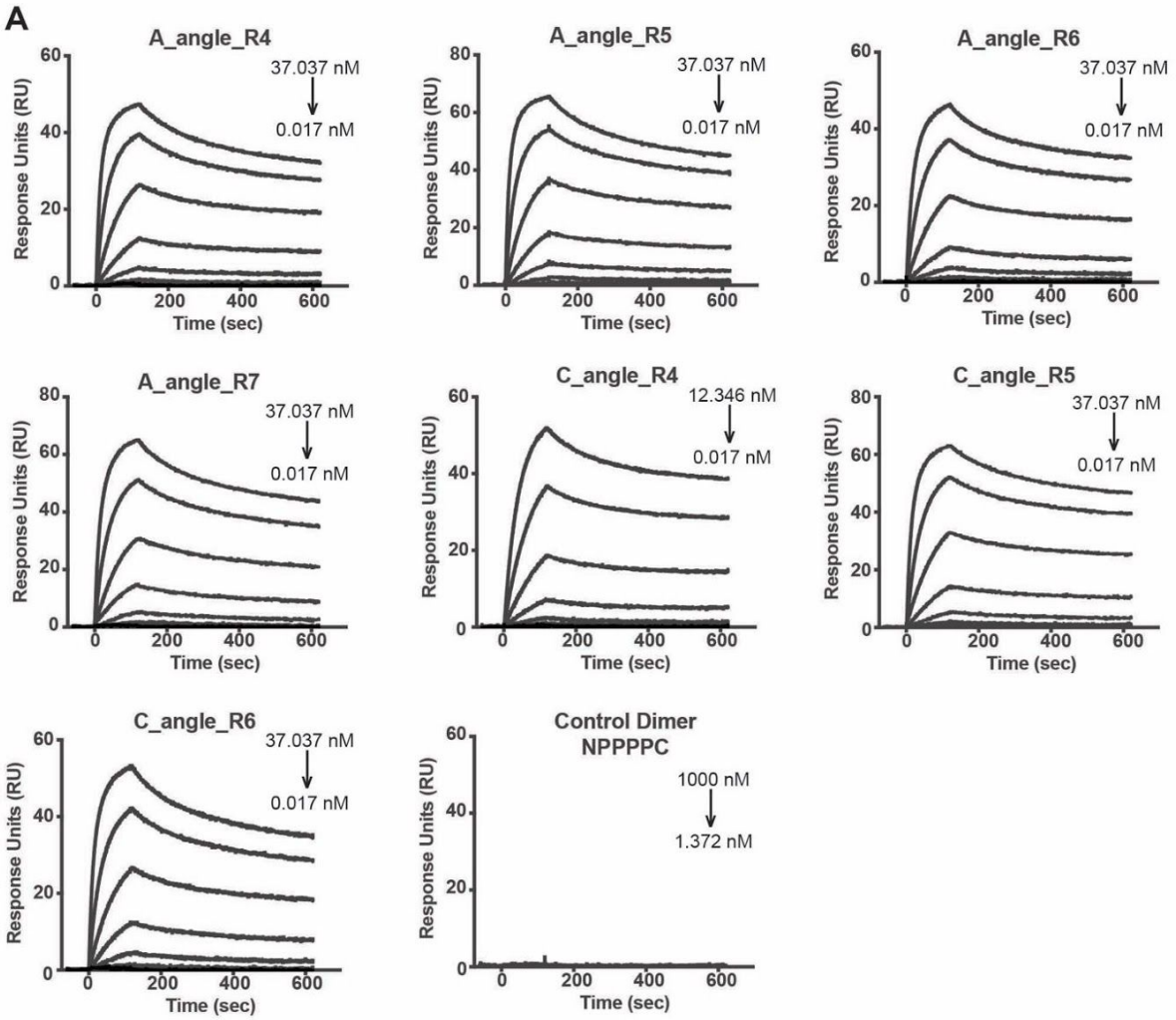
Appendix Figure 3.3. Dimerizing residues (green) are inserted on the rear, unobstructed face of the E2 DARPin to generate two different dimer geometries, A_R3 and C_R3 (top). The sequences for the E2 monomeric DARPin ligand and the dimer scaffolds, A_R3 and C_R3 (bottom). The binding residues are indicated by red whereas the dimerizing residues are shown in green.



Appendix Figure 3.4. The kinetic binding curves and K_d values for dimers A_R3 and C_R3, as measured by SPR.



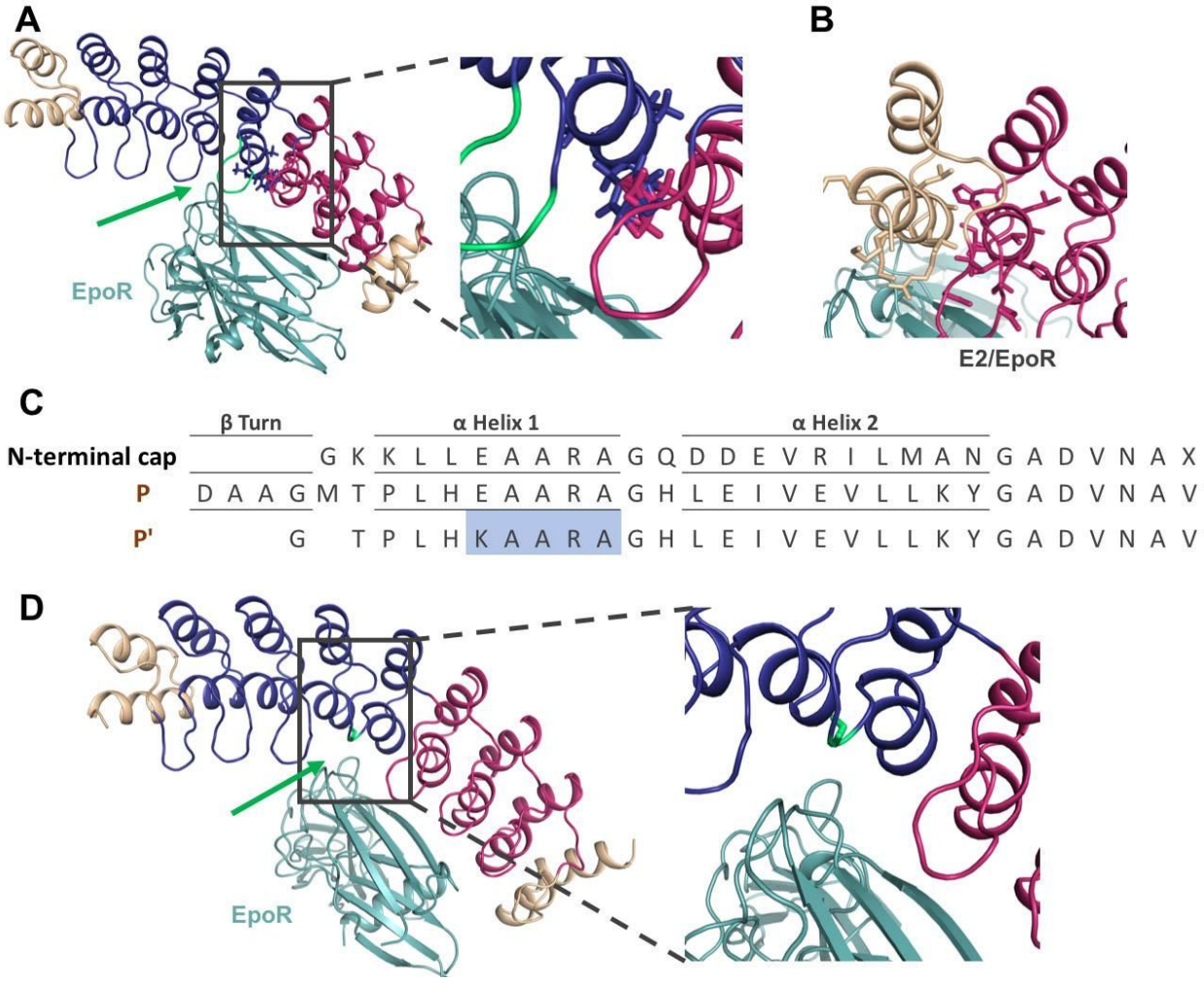
Appendix Figure 3.5. Relative twist of the A_R3 and C_R3 dimers. A model of the A_R3/EpoR (salmon/red) complex was aligned to a D1 domain from the C_R3/EpoR crystal structure (light blue/blue). The second D1 domain is related by a 30° left-handed twist between the A_R3 model and the C_R3 structure (inset). D2 domains are omitted for clarity in the inset.



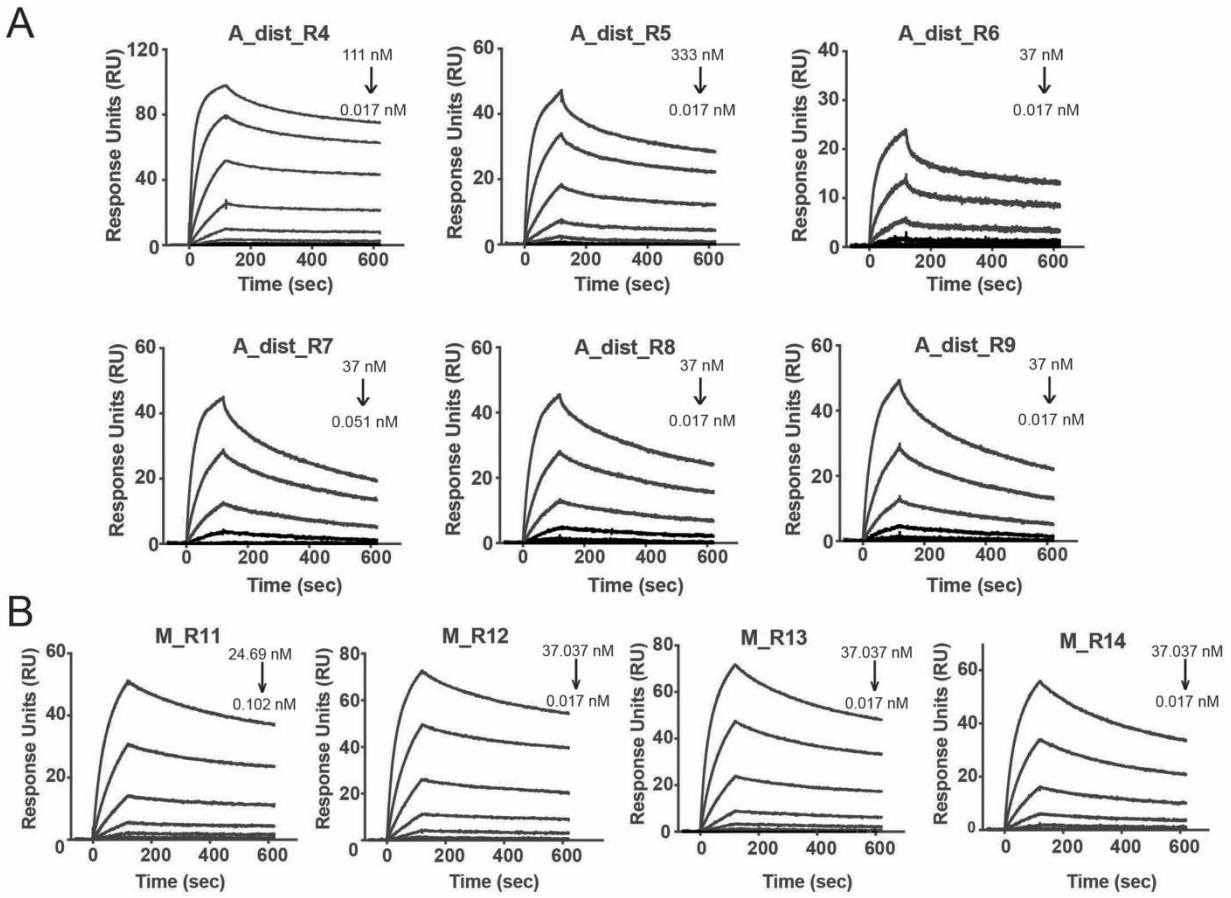
B

	K_d Values (nM)		K_d Values (nM)
A_angle_R4	0.375	C_angle_R4	0.202
A_angle_R5	0.328	C_angle_R5	0.360
A_angle_R6	0.474	C_angle_R6	0.510
A_angle_R7	0.540	Control Dimer	0

Appendix Figure 3.6. A, Kinetic binding curves for dimers in the A_angle_Rx and C_angle_Rx series as obtained from SPR measurements. **B**, The binding affinities for the A_angle_Rx and C_angle_Rx dimers remain in the same range with the addition of one or more P repeats. A control dimer made entirely of non-binding ankyrin repeats, NPPPPC, shows no significant binding affinity to EpoR, as expected.



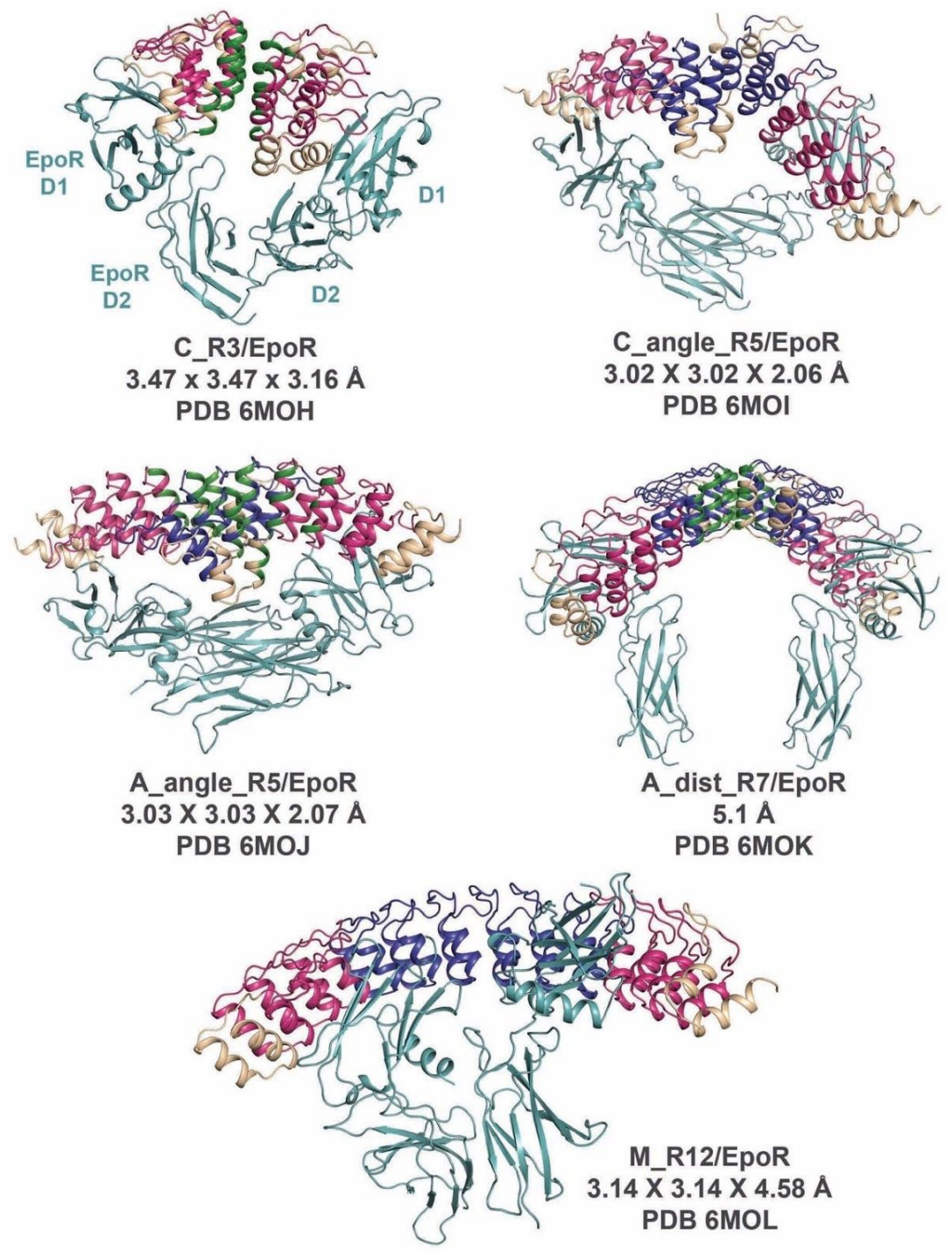
Appendix Figure 3.7. Optimization of extended DARPin design with internal binding repeats. Inserting P repeats before the binding repeats lead to a complete loss in binding affinity. Analysis of the models revealed a clash (**A**) between the inner helices of the first binding repeat and the preceding P repeat. No such clashes were observed with the inner α-helix of *N*-terminal capping region in E2/EpoR bearing the sequence 'KAARA' (**B**). Furthermore, the Lys residue in the 'KAARA' region is an error-prone mutation, which is well-positioned to interact with EpoR. Additionally, the β-turn loop (shown with a green arrow) preceding the binding repeats could pose hindrance to the loop region thereby forcing a change in the curvature. **C**, The series is modified to the format NPP…P'EEEC, where the Ankyrin repeat preceding the binding repeats, P', has been modified to incorporate the 'KAARA' sequence in the α-helix region based on structural symmetry with the *N*-terminal capping region. The possible hindrance from the preceding loop is relieved by shortening the β-turn sequence 'DAAGM' and replacing it with a Gly residue, which is suitable for turns due to its small size. **D**, Structure of A_dist_R7 (NPPPP'EEEC) with the shortened loop region. The NPP…P'EEEC format reinstates the affinity of the dimers in this series.



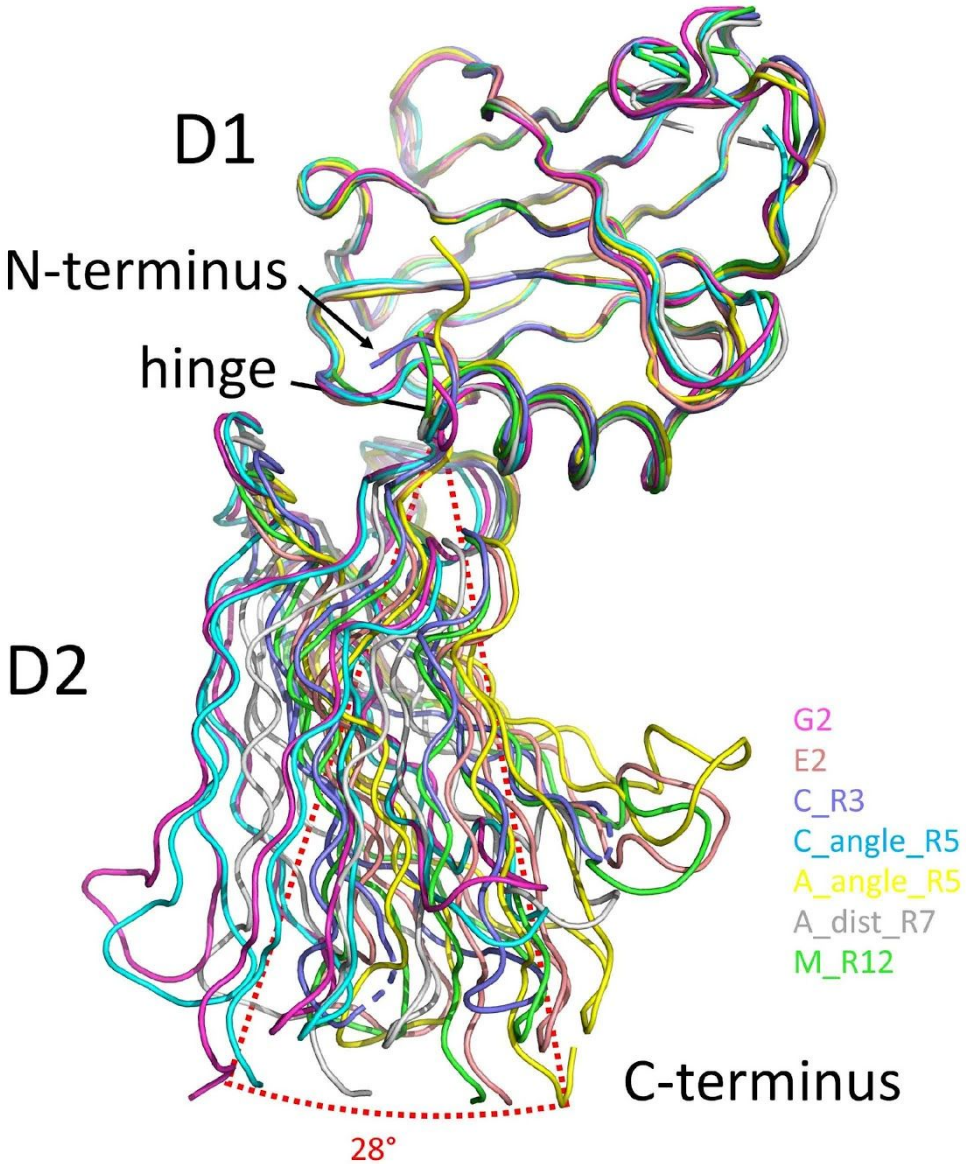
C

	K_d Values (nM)		K_d Values (nM)
A_dist_R4	0.716	M_R11	0.638
A_dist_R5	5.27	M_R12	0.560
A_dist_R6	0.963	M_R13	0.940
A_dist_R7	1.88	M_R14	1.500
A_dist_R8	1.7		
A_dist_R9	2.55		

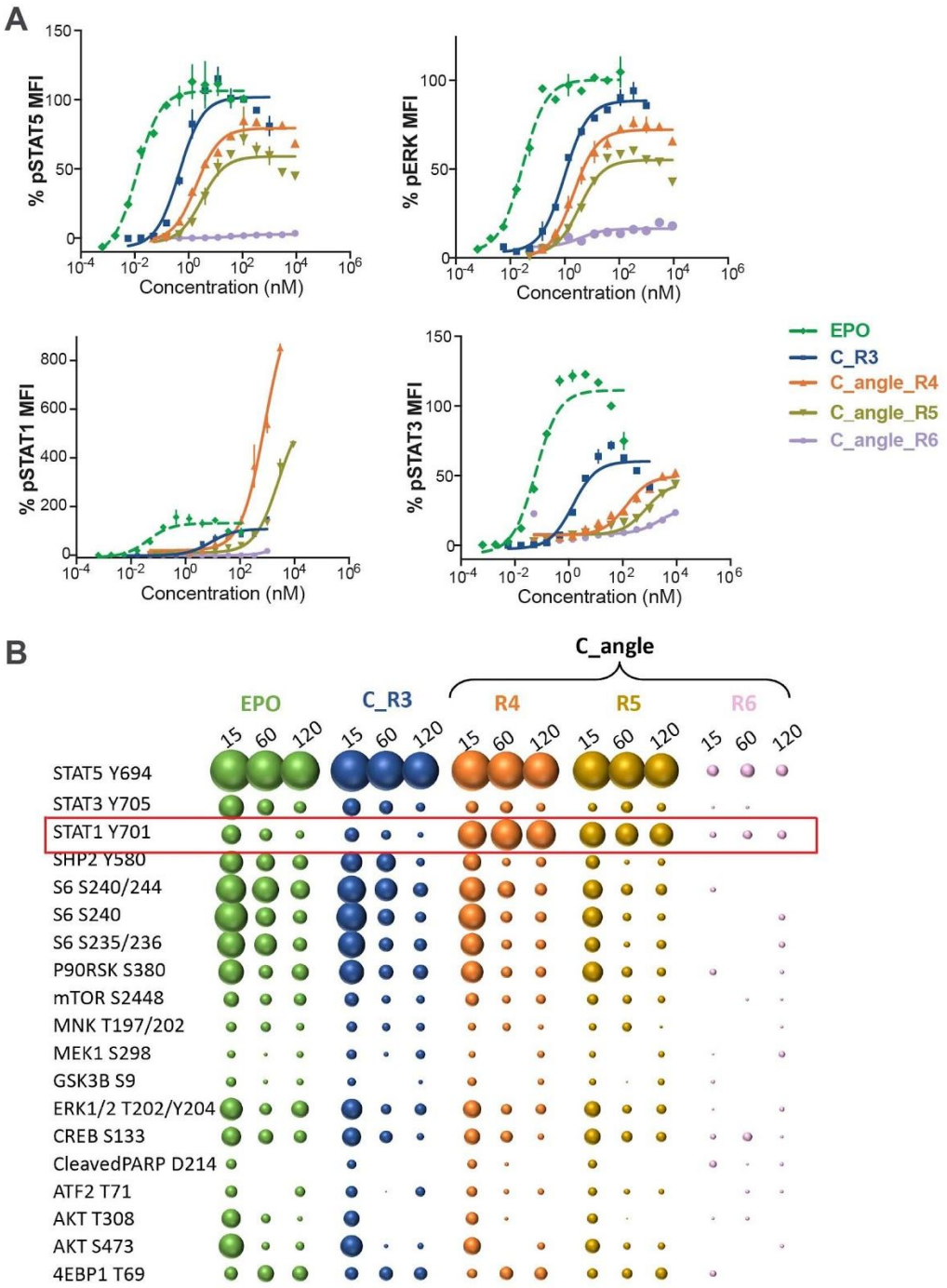
Appendix Figure 3.8. Kinetic binding curves for DARPins in the A_dist series (**A**) and the monoextended series (**B**), as obtained from SPR measurements. **C**, The binding affinities for the DARPins dimers remain in the same range with the addition of one or more P repeats.



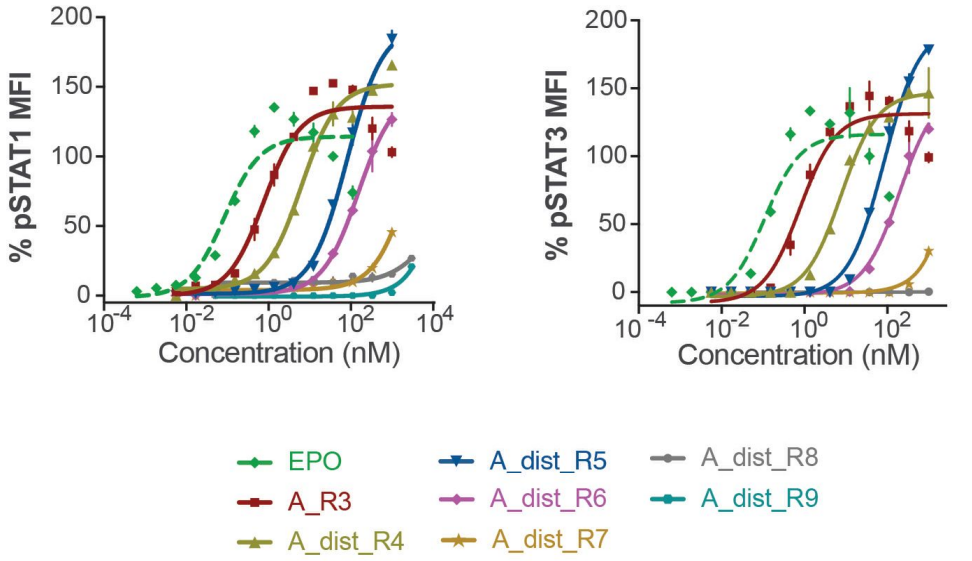
Appendix Figure 3.9. Structures of the various DARPin/EpoR complexes determined to validate the designs. The non-binding P repeats in the DARPin dimer are shown in blue whereas the three binding ankyrin repeats are shown in pink. The capping regions are depicted in wheat color.



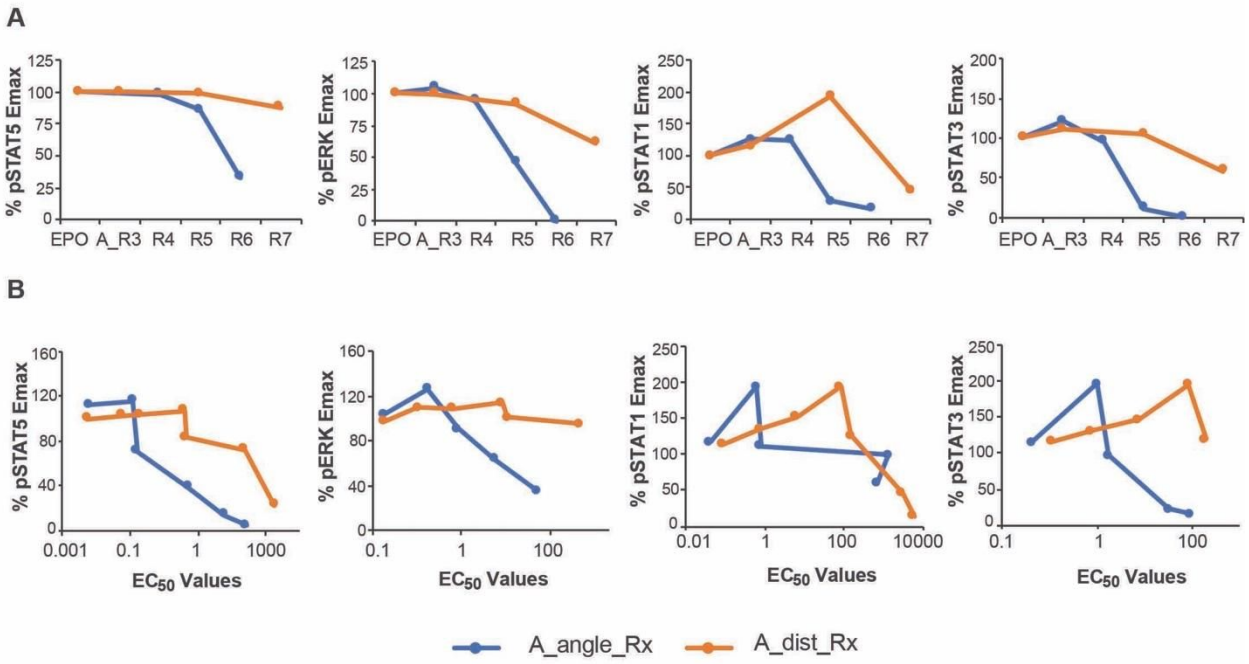
Appendix Figure 3.10. Structural alignment of the D1 domain of EpoR from the G2 monomer (magenta), E2 monomer (salmon), C_R3 (purple), C_angle_R5 (cyan), A_angle_R5 (yellow), A_dist_R7 (gray), and M_R12 (green) structures. The range of rotation of domain D2 is measured as an angle with a vertex at Val 119 in the hinge region and rays passing through Asp 223 at the C terminus.



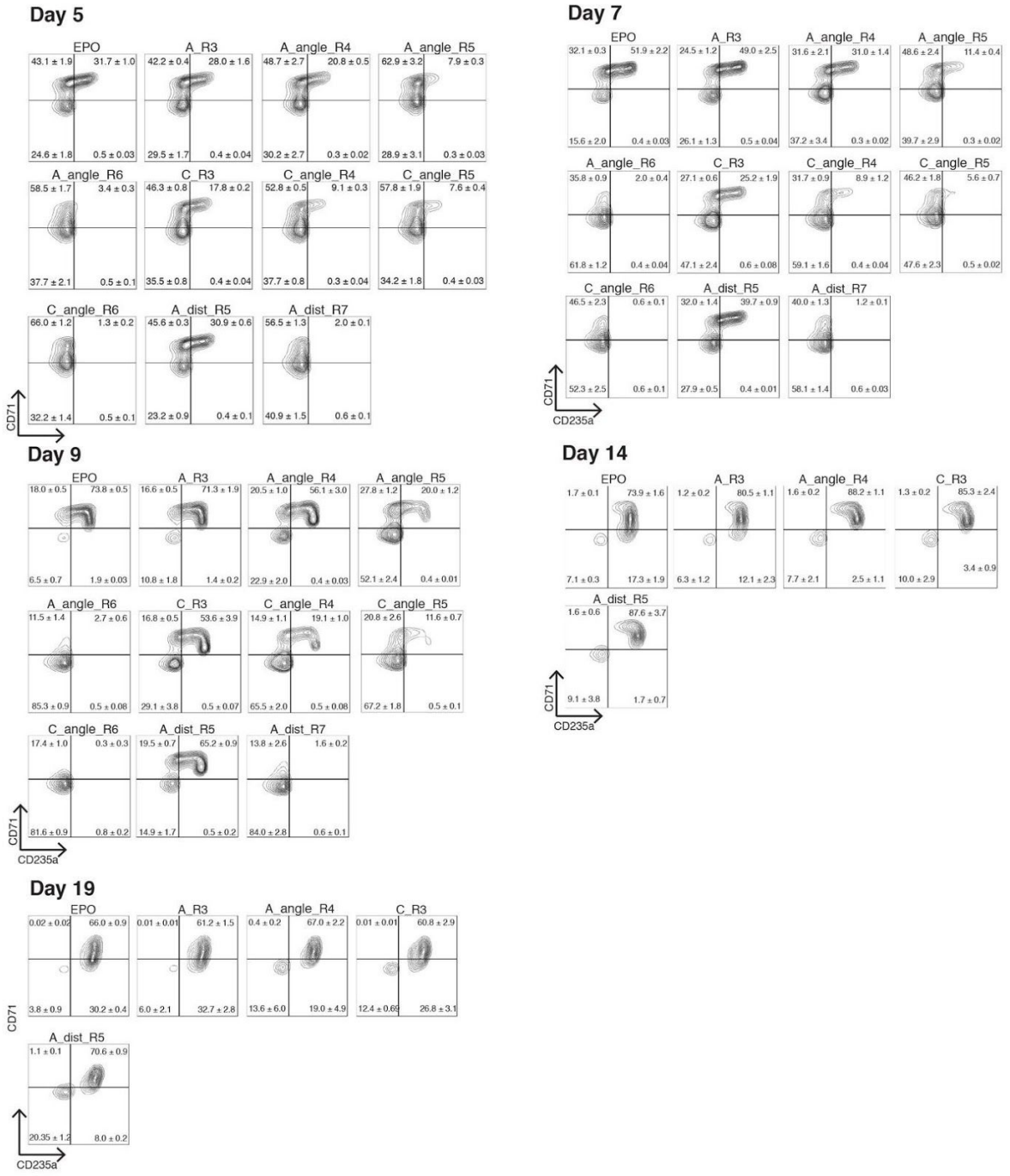
Appendix Figure 3.11 A, The activation of EpoR downstream effectors STAT5/1/3 and ERK by the C_angle series was analyzed by measuring the levels of the corresponding phosphorylated proteins after 15 mins of ligand stimulation on UT7 cells. **B**, The signaling pathways activated by EPO and the C_angle series DARPin dimers with stimulation times of 15, 60 and 120 mins, as indicated. The size of the bubble is proportional to the levels of the phosphorylated proteins.



Appendix Figure 3.12. The activation of EpoR downstream effectors STAT1 and STAT3 by the A_dist_Rx series DARPins was analyzed by measuring the levels of the corresponding phosphorylated proteins after 15 mins of ligand stimulation on UT7 cells.



Appendix Figure 3.13. A, B, The general trends in E_{max} and EC_{50} values for DARPin dimers corresponding to the A_angle and A_dist series. **B**, The ligands are plotted from left to right in the following order – EPO, A_R3, R4 through R7 for A_angle, and R4 through R8 for A_dist. EC_{50} values have been omitted for ligands that do not generate a saturating signal.



Appendix Figure 3.14. Alternation of erythropoiesis via topological control of EPO receptor signaling. Erythroid differentiation from $CD34^{+}$ HSPC with synthetic EPO ligands is shown. Differentiation has been assessed by flow cytometry with the markers CD71 and CD235a at various time points during erythroid differentiation. Means ± SEM for three independent experiments (N=2 for A_dist_R5) are shown.

Appendix Tables

Appendix Table 1.1. List of all homooligomer and heterodimer designs tested and corresponding amino acid sequences including initiating methionine and $(His)_6$ tag. Designs that expressed solubly with the predicted polypeptide molecular weights are in bold and under their name the the biophysical characterization methods used to characterize them.

Design	Sequence
1na0C2 (SEC)	**MHHHHHHGSNSAEAWYNLGNAYYKQGDYDEAIEYYQKALELDPNNAEA WYNLGNAYYKQGDYDEAIEYYKKALRLDPNNAEAWYNLGNAYYKQGDY EQAILAYVIALALDPNNAEAKQNLGNAEQKEK**
ank1C2_1 (SEC-MALS, SAXS)	**MHHHHHHGSWGSSELGKRLIEAAENGNKDRVKDLIENGADVNASDSDGR TPLHHAAENGHAEVVALLIEKGADVNAKDSDGRTPLHHAAENGHDEVVLI LLLKGADVNAKDSDGRTPLHHAAENGHKRVVLVLILAGADVNTSDSDGRT PLDLAREHGNEEVVKALEKQ**
ank1C2_1_r5 (SEC-MALS, SAXS)	**MHHHHHHGSWGSSELGKRLIEAAENGNKDRVKDLIENGADVNASDSDGR TPLHHAAENGHKEVVKLLISKGADVNAKDSDGRTPLHHAAENGHKEVVK LLISKGADVNAKDSDGRTPLHHAAENGHAEVVALLIEKGADVNAKDSDGR TPLHHAAENGHDEVVLILLLKGADVNAKDSDGRTPLHHAAENGHKRVVL VLILAGADVNTSDSDGRTPLDLAREHGNEEVVKALEKQGGW**
ank1C2_2 (SEC-MALS)	**MSELGKRLIEAAENGNKDRVKDLIENGADVNASDSDGRTPLHHAAENGHK EVVNLLISKGADVNAKDSDGRTPLHHAAENGHLRVVLKLISKGADVNAKD SDGRTPLHHAAENGHKEVVLALIAAGADVNTSDSDGRTPLDLAREHGNEE VVKELEDDLEHHHHHH**
ank3C2_1 (SEC-MALS, SAXS)	**MSELGKRLIEAAENGNKDRVKDLLENGADVNASDSDGKTPLHLAAENGH AKVVLLLLEQGADPNAKDSDGKTPLHLAAENGHAVVVALLLMHGADPNA KDSDGKTPLHLAAENGHEEVVILLLAMGADPNTSDSDGRTPLDLAREHGN EEVVKVLEDHGGWLEHHHHHH**
ank3C2_2 (SEC-MALS, SAXS)	**MSELGKRLIEAAENGNKDRVKDLLENGADVNASDSDGKTPLHLAAENGH KDVVELLLRQGADPNAKDSDGKTPLHLAAENGHKVVVMLLLSQGADPNA KDSDGKTPLHLAAENGHEDVVLLLLLMGADPNTSDSDGRTPLDLAREHGN EDVVEALKAAGGWLEHHHHHH**
HR10C2_1 (SEC-MALS, SAXS)	**MSSTKEELRELLVKVVVENAKRKGDDTEEAREAAREAFELVREAAERAGI DSTVVLVAAALLIMSVVAAAQSAGYDISEAARAAAEAFKRVAEAAKRAGIT SSSVLSLAIALISLVVSNAQSEGYDISEAARAAAEAFKRVAEAAKRAGITKEE TLMAAIAEILARVAEAAEERNDISEAARQAAEEFRKKAEELKGSWLEHHH HHH**
HR10C2_2 (SEC-MALS, SAXS)	**MSSEKEELRERLVKIVVENAKRKGDDTEEAREAARAAFEIVRAAAKLAGI DSSEVLELAIRLIKEVVENAQREGYDIAVAAIAAAVAFAVVAVAAAAADITS SEVLELAIRLIKEVVENAQREGYVILLAALAAAAAFVVVAAAAKRAGITSSE TLKRAIEEIRKRVEEAQREGNDISEAARQAAEEFRKKAEELKGSWLEHHH HHH**
HR10C2_3 (SEC-MALS, SAXS)	**MSEAKETLREILVTIVVLAAKITGDDTEEAREAAREAFELVREAAERAGIDS VAVLMAAILLIAAVVVNAAAEGYDISEAARAAAEAFKRVAEAAKRAGITSS**

	EVLQLAINLIAEVVENAQARGYDISEAARAAAEAFKRVAEAAKRAGITSSE TLAAAILAILARVVEAAAEDNDISEAARQAAEEFRKKAEELKGSWLEHHH HHH
HR07C2 (SEC)	**MHHHHHHGSWGSTKEDARSTCEKAARKAAESNDEEVAKQAAKDCLEVA KQAGMPTKEAARSFCEAAARAAAESNDEEVAKIAAKACLEVAKQAGMPT KEAAASFCEAAARAAAESNDEEVAKIAAKACLEVAKQAGMPTKTAALAFC AAALRAALESNDEEVAKIAMKACREVAKQAGMD**
HR81C2 (SEC-MALS, SAXS)	**MGEREELSELAERILQKARKLSEEARERGDLKELALALILEALAVLLLAIA ALLRGNSEEAERASEKAQRVLEEARKVSEEAREQGDDEVLALALIAIALAV LALALVACCRGNSEEAERASEKAQRVLEEARKVSEEAREQGDDEVLALAL IAIALAVLALAIVACCRGNKEEAERAAEDAIKVAMEALEVLLSAVEQGDLK VALAAVIAILLAIAALLMVIMCKGGWLEDLEHHHHHH**
HR79C2 (SEC-MALS, SAXS)	**MGREDELRARLLILLAELAAERADIAAERTGDPRVRELARELKRLAQEAA EEVKRDPSSSDVNEALKLIVEAIEAAVRALEAAERTGDPEVRELARELVRL AVEAAEEVQRNPSSSDVNEALKLIVEAIEAAVRALEAAERTGDPEVRELAR ELVRLAVEAAEEVQRNPSSKEVDMALLLILIAILEAVLSLLRAERSGDPEKR EKARERVREAVERAEEVQRDPSGWLEHLEHHHHHH**
1na0C3_1 (SEC-MALS, SAXS)	**MERAEMAAIVGDAIYIMGLYRLAIKMYLIALKLDPNNAEAWYNLGNAYY KQGDYDEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYDEAIEYYQKA LELDPNNAEAKQNLGNAKQKQGLEHHHHHH**
1na0C3_2	MTEAEEMYNLGNKAYKEGDEDVAIAAYIAALSHDKNNAEAWYNLGNAAYK AGREKTAEEAYQKALELDPNNAEAWYNLGNAYYKQGDYDEAIEYYQKALEL DPNNAEAKQNLGNAKQKQGLEHHHHHH
1na0C3_3 (SEC-MALS, SAXS)	**MNLAEKMYKAGNAMYRKGQYTIAIIAYTLALLKDPNNAEAWYNLGNAAY KKGEYDEAIEAYQKALELDPNNAEAWYNLGNAYYKQGDYDEAIEYYQKA LELDPNNAEAKQNLGNAKQKQGLEHHHHHH**
1na0C3_4	MNSWLQWAVRGLMAKYEGIYREAIEAYQKALELDPNNDMAWLLLGAAYYK QGDYDEAIEYYQKALELDPDNAKAWYKLGNAYYKQGDYDEAIEYYQKALEL DPNNAEAKQNLGNAKQKQGLEHHHHHH
1na0C3_5 (SEC-MALS, SAXS)	**MNEAELAYDLGNEAYKDGEYRLAAVAYVLALAVDPNNAEAWYNLGNAY YKQGRYDKAIKYYQKALELDPNNAEAWYNLGNAYYKQGDYDEAIEYYQK ALELDPNNAEAKQNLGNAKQKQGLEHHHHHH**
1na0C3_6	MRAAEKAYKLGNEEYKQGRYQDAIALYLMALIEDPNNAEAWYNLGNAYYKQ GKYKAAIAAYEMALEKDPNNAEAWYNLGNAYYKQGDYDEAIEYYQKALELD PNNAEAKQNLGNAKQKQGLEHHHHHH
1na0C3_7 (SEC-MALS, SAXS)	**MNSAEAMYKMGNAAYKQGDYILAIIAYLLALEKDPNNAEAWYNLGNAAY KQGDYDEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYDEAIEYYQKA LELDPNNAEAKQNLGNAKQKQGLEHHHHHH**
1na0C3_8	MNSAELAYALGNLAYRLGDYEIAITAYQIALELDPNNAEAWYNLGNAYYKQG DYDEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYDEAIEYYQKALELDP NNAEAKQNLGNAKQKQGLEHHHHHH

tpr1C3_1	MAEKAKKLGKEEEKLGRLQDAIALYLMALIEDPNDAEAWKELGKVYEKLGK YKAAIAAYEMALEKDPNDAEAWKELGKVLEKLGRLDEAAEAYKKAIELDPND AEAWKELGKVLEKLGRLDEAAEAYKKAIELDPNDLEHHHHHH
tpr1C3_2 (SEC-MALS)	**MAKIAMLLGRVAEQMGQLEIAAKAYKLAIELDPNDAEAWKELGKVLEKL GRLDEAAEAYKKAIELDPNDAEAWKELGKVLEKLGRLDEAAEAYKKAIE LDPNDAEAWKELGKVLEKLGRLDEAAEAYKKAIELDPNDLEHHHHHH**
tpr1C3_3 (SEC-MALS)	**MAELAYDLGKEAEKDGELRLAAVAYVLALAVDPNDAEAWKELGKVLEK QGRLDKAAKYYKKAIELDPNDAEAWKELGKVLEKLGRLDEAAEAYKKAI ELDPNDAEAWKELGKVLEKLGRLDEAAEAYKKAIELDPNDLEHHHHHH**
tpr1C3_4 (SEC-MALS)	**MAQKAKRIGKKAEKKGQYLLAMLAYIQALHEDPNDAEAWKELGKVAEK DGELDEAAEAYKKAIELDPNDAEAWKELGKVLEKLGRLDEAAEAYKKAI ELDPNDAEAWKELGKVLEKLGRLDEAAEAYKKAIELDPNDLEHHHHHH**
tpr1C3_5	MAKLAMLVGKVAAQAGRLEMAAKLYKIAIELDPNDAEAWKELGKVLEKLGR LDEAAEAYKKAIELDPNDAEAWKELGKVLEKLGRLDEAAEAYKKAIELDPND AEAWKELGKVLEKLGRLDEAAEAYKKAIELDPNDLEHHHHHH
ank1C3 (SEC-MALS)	**MSDLGENLAMAAALGNKDRVKDLIENGADVNASIDLGLTPLHMAAEMGH KEVVKLLISKGADVNAKDSDGMTPLHHAARNGHKEVVKLLISKGADVNA KDSDGRTPLHHAAENGHAEVVKLLISKGADVNTSDSDGRTPLDLAREHGQ TLVVLLLRLQLEHHHHHH**
HL1C3 (SEC)	**MYENEAIIWAMQGDYTEAAKAAEKAGEYRMAATAWEKSGDYTEAAKAW EKAGDYTEAAKAWEKSGDYTEAAKAWEKAGDYTEAAKAWEKSGDYTEA AKAWEKAGDYTEAAKAWEKSGDYTEAAKAWEKAGDYTEAAKAWEKSG DYTEAAKAWEKAGDLEHHHHHH**
HR10C3_1 (SEC)	**MSSEKEELEERLVKIVVENAKRKGDDTEEARIAAAIAFALVMAAALMAGI DSSEVLELAIRLIKEVVENAQREGYDISEAALAAAEAFARVAEAAKRAGITS SEVLELAIRLIKEVVENAQREGQDISFAARAAATAFKLVALAAKRAGITSSE TLKRAIEEIRKRVEEAQREGNAIAAAALKAAAEFVKKAQELKGSWLEHHH HHH**
HR10C3_2 (SEC-MALS, SAXS)	**MHHHHHHHGSWGSSEKEELREILVAVVVANAKEKGDDTEEAREAAREAFE LVREAAERAGIDSSEVLLLAILLIIAVVILAAAMGYDISEAARAAAEAFKRV AEAAKRAGITSSEVLELAIRLIKEVVANAQREGYDISEAARAAAEAFKRVA EAAKRAGITSSTTLKMAIIIIRLAVEAAQEAGNDISEAARRAAEAFREAAEE KK**
HR04C3_1 (SEC)	**MSTCETAARLVAIMVELLKRLGVSEDEIAEIVALIISLVISTEKRSGSSYEVI CECVARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESGSSYEVICECV ARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESGSSYEVIKECVQRIV EEIVEALKRSGTSEDEINEIVRRVKSEVERTLKESGSWLEHHHHHH**
HR04C3_2 (SEC)	**MHHHHHHHGSWGSDECEEKARRVAEKVERLKRSGTSANEIAEEVAREISEV IRTLKESGSSYEVICECVARIVAEIVEALKRSGTMELLIALIVARVISEVIRTL KESGSSYEVICECVARIVAEIVEALKRSGTSEVEIAIIVAIVIAEVIRTLKESGS SAEVIMKCVQRIVEEIVEALKRSGTSEDEIREIVAMVLLVVAITLLESGA**

K64sC3 (SEC-MALS)	**MSEKESATLLALLRALAKLKAKDPEAQKSFREALGEALKKLGAASPKAIE AFAEALGIALESLGATDPEAIKAFAEALGAALKRLGATDPVAIVAFALALG LALEELGATDPEAIKAFAEALGAALKRLGATDPEAIQAFATALGKALKEL GATDPEAIKAFAEALGAALKRLGATDPEAIKAFAEALGKALKELGATDPE AIKAFAEALGAALKRLGATDPEAIKAFAEALGKALKELGATDPEAIKAFAE ALGAALKRLGATSPEAIKAFAEALGKALKELGATDPEAIKRFAERLGDELR KEGATDPERIKERAEREGRERKQEGKTDGSWLEHHHHHH**
KP16C3_1	MAKNELVDIAQAAIERQDRALVALAVSVVAMSNDKEVLIEIAKVAIDKQDENI VAAVVKLVAASNDKEVLIEIAKVAIDKQDENIVQQVVKIVAESNDKEVLIEIAK VAIDKQDENIVTSVVKIVAESNDKEVLIEIAKVAIDKQDENIVASVVKIVAESND KEVLIEIAKVAIDKQDENIVQQVVKIVAESNDKEVLIEIAKVAIDKQDENIVTSV VKIVAESNDEEVIQEIYKVAREKQDENIAQQIEEVLKEDNGSWLEHHHHHH
KP16C3_2	MHHHHHHGSWGSEKNKLVEEAKEAIENQDENKVQQIVEDVARSNDKEVLIEI AKVAIDKQDENIVASVVKIVAESNDKEVLIEIAKVAIDKQDENIVQQVVKIVAE SNDKEVLIEIAKVAIDKQDENIVTSVVKIVAESNDKEVLIEIAKVAIDKQDENIV ASVVKIVAESNDREVLIEIAKVAIDKQDENIVQQVVKIVAESNDKKVLIQIAQV AIDKQDENIVTSVVKIVAESDDAEVIMSILLVAILKKDENIAQQILEVLSEDE
KP17C3_1	MSNELALDIVALALSVSTETESIREIIEKLYKRDKELIKQAVAEALASVKDTEVI RVIIEVLYKEDKELIKQAVAEALKRVKDTEVIRVIIEVLYKEDKELIKQAVAEAL TSVTDTEVIRVIIEVLYKEDKELIKQAVAEALASVTDTEVIRVIIEVLYKEDKELI KQAVAEALKRVTDTEVIRVIIEVLYKEDKELIKQAVAEALTSVTDTEVIRVIIEV LYKEDERLIDEAVREALREVTDPEVQRIIKEVRNQQGSWLEHHHHHH
KP17C3_2	MHHHHHHGSWGSNEQRLQEVAERLREETDTERIRQIIEQLYKEDKELIKQAVA EALASVTDTEVIRVIIEVLYKEDKELIKQAVAEALKRVTDTEVIRVIIEVLYKED VELIADAVAEALTSVTDTEVIRVIIEVLYKEAEAAILAAVAEALASVTDTEVIRV IIEVLYKEDKELIKQAVADALKNVTDTEVIRVIIEVLYKEDKELIKQAVAEALTS VTDSEVIRVIIEVLYKEDERLIDEAVREAKREVKDPRVRLLIQAVRKRQ
KP17C3_3	MRNEANLSLVALLLSVATDTERIRQIIEQLYKEDKELIKQAVARALASVEDTEVI RVIIEVLYKEDKELIKQAVAEALKRVTDTEVIRVIIEVLYKEDKELIKQAVAEAL TSVTDEKVIVVIIQVLTKEDKELIKQAVAEALASVTDEQVIMVIIAVLFSLDKELI KQAVAEALKRVTDTEVIRVIIEVLYSLDKELIKQAVAEALTSVTDTEVIRVIIEVL YKEDERLIDEAVREALREVTDPEVQRIIKEVRNQQGSWLEHHHHHH
KP17C3_4 (SEC-MALS)	**MHHHHHHGSWGSNEQRLQEVAERLREETDTERIRQIIEQLYKEDKELIKQ AVAEALASVTDTEVIRVIIEVLYKEDKELIKQAVAEALKRVTDTEVIRVIIEV LYKEDKELIKQAVAEALTSVTDTEVIRVIIEVLYKEDKELIKQAVAEALASV TDTEVIRVIIEVLYKEDKDLIAEAVAEALKRVTDTEVIRVIIEVLYKEDKALI VLAVAEALTSVTDTEVIRVIIEVLYKEDEKAITLAVVLALVQVTDPRVMSII VSVALQQ**
KP17C3_5	MSNSVSLAVVIVLLSIETDTERIRRIIEKLYSRDKELIKQAVATALALVEDTEVIR VIIEVLYKEDKELIKQAVAEALKSVEETEVIRVIIEVLYKEDKELIKQAVAEALTS VTDTEVIRVIIEVLYKEDKELIKQAVAEALASVTDTEVIRVIIEVLYKEDKELIKQ AVAEALKRVTDTEVIRVIIEVLYKEDKELIKQAVAEALTSVTDTEVIRVIIEVLY KEDERLIDEAVREALREVTDPEVQRIIKEVRNQQGSWLEHHHHHH
HR00C3_1	MIEEVVAEMIDILAESSKKSIEELARQADNKTTEQQVAQSIEQIANLATTAIQLIE NLAKNLASEEFMARAISAIAELAKKAIEAIYRLADNHTTDTFMANAINAIANLA

	VTAILAIANLASNHTTEEFMARAISAIAELAKKAIEAIYRLADNHTTDKFMAQAI EAIALLATLAILAIALLASNHTTEEFMAKAISAIAELAKKAIEAIYRLADNHTSPT YIEKAIEAIEKIARKAIKAIEMLAKNITTEEYKEKAKSAIDEIREKAKEAIKRLED NRTLEHHHHHH
HR00C3_2 (SEC-MALS, SAXS)	**MIEEVVAEMIDILAESSKKSIEELARAADNKTTEKAVAEAIEEIARLATAAI QLIEALAKNLASEEFMARAISAIAELAKKAIEAIYRLADNHTTDTFMARAIA AIANLAVTAILAIAALASNHTTEEFMARAISAIAELAKKAIEAIYRLADNHTT DKFMAAAIEAIALLATLAILAIALLASNHTTEEFMAKAISAIAELAKKAIEAI YRLADNHTSPTYIEKAIEAIEKIARKAIKAIEMLAKNITTEEYKEKAKSAIDE IREKAKEAIKRLEDNRTLEHHHHHH**
tprxC3	MHHHHHHGSASAEAWKELGKVLEKLGRLDEAAEAYKKAIELDPNDAEAWKE LGKVLEKLGRLDEAAEAYKKAIELDPNDAEAWKELGKVLEKLGRLKEAAEAY RKAIELDPNDAEAWKELGKVLEKLRALLMAALAYKVAITLDPTSAEAWKELG KVLEKLGRLDEAAEAYKKAIELDPNDAEAWKELGKVLEKLGRLDEAAEAYKK AIELDPNDAEAWKELGKVLEKLGRLDEAAEAYKKAIELDPNDAEAWKELGKV LEKLGRLDEAAEAYKKAIELDPTSAEAWKELGKVLEKLGRLDEAAEAYKKAIE LDPNDAEAWKELGKVLEKLGRLDEAAEAYKKAIELDPKDAEAWKELGKVLE KLGRLDEAAEAYEKAIKLKPNDAEAWKELGKVLEKLGELDEAAEAYKKAIEL DPTSAEAWKELGKVLEKLGRLDEAAEAYKKAIELDPNDAEAWKELGKVLEKL GRLDEAAEAYKKAIELDPNDAEAWKELGKVLEKLGRLDEAAEAYKKAIELDP NDAEAWKELGIVLSELNRVEEALEALKKATELDD
HR08C3 (SEC-MALS, SAXS)	**MGDEMRKVMLALAIALVRALLNEDIEVAKEIARAADEIEEALRENNSDEM AKVMLALAKAVLLAAKNNDDRVAEVIALAAAEIVKALRRNNSDEMAKVM LALAKAVLLAAKNNDDEVAKEIAIAAMIIVIALRAENSDEMAKKMLELAK RVLDAAKNNDDETAREIAEQAAEELEAWLEHHHHHH**
HR04C3_3 (SEC-MALS)	**MGHHHHHHWDECEEKARRVAEKVERLKRSGTSEDEIAEEVAREISEVIRT LKESGSSYEVICECVARIVAEIVEALKRSGTSRKEIAEIVARVISEVIRTLKES GSSYEVICECVARIVAEIVEALKRSGTSAEEIALIVARVISEVIRTLKESGSSE EVILKCVARILLEILEALERSGTKKLKILIIALMVLIVIVLTILRSGE**
HR79C3 (SEC)	**MGESDEILAMLVIILALLAAILAMMAAKETGDPRVEELAEELVRLALEAAE EVERDPSSRDVLKALLLIILAIVVAVLALLAALITGDPEVRELARELVRLAV EAAEEVQRNPSSSDVNEALKLIVKAILAAVRALRAAERTGDPEVRELAREL VRLAVEAAEEVQRNPSSEEVNEALKKIVKAIREAVESLREAEESGDPEKRE KARERVREAVERAEEVQRDPSGWLEHLEHHHHHH**
3ltjC3 (SEC-MALS)	**MGHREHTDPLKVLLYIVILEAEKYLARAKAAEALGKIGDEEAVEPLIKALK DEDALVRAAAADALGQIGDERAVEPLIKALKDEDGLVRASAAVALGQIGD ERAVEPLIKALKDERAVVRVAAAIALGLIGDERAVEPLIKALKDERGKVRK AAARALGEIGGERVRAAMEKLAETGTGFARKVAVNYLETHKWLEHHHH HH**
ank1C4_1 (SEC-MALS)	**MSMLGKLLILAAELGLLVVVMLLISNGADVNASDSDGRTPLHHAAENGHK MVVIMLLIKGADVNAKDSDGRTPLHHAAENGHKEVVKELIEMGADVNAK DSDGRTPLHHAAENGHKEVVKLLISKGADVNTSDSDGRTPLDLAREHGNE EVVKLLEKQLEHHHHHH**
ank1C4_2 (SEC-MALS, SAXS)	**MSEDGELLILAAELGIAEAVRMLIEQGADVNASDDDGRTPLHHAAENGHL AVVLLLLLKGADVNAKDSDGRTPLHHAAENGHKTVVLLLILMGADVNAK**

	DSDGRTPLHHAAENGHKEVVKLLIRKGADVNTSDSDGRTPLDLAREHGNE EVVKLLEKQLEHHHHHH
ank3C4_2 (SEC-MALS)	**MTELGIALAIAALVGDKDRVKDLLENGADVNASARAGMTPLHLAALLGH KEVVKLLLSQGADPNAKDKKGKTPLHLAARNGHKEVVKLLLSQGADPNA KDKDGKTPLHLAAENGHVAVVHLLLEQGADPNTSDSDGRTPLDLAKEHG HEVVVTLLILKGGEIAHHDHDLEHHHHHH**
ank3C4_1 (SEC-MALS, SAXS)	**MNDLGMLLIMAAMEGNKDRVKDLLENGADVNASSKAGKTPLHLAAENG HKEVVKLLLSQGADPNAKDNEGKTPLHLAAENGHKEVVKLLLSQGADPN AKDKDGKTPLHLAAENGHAEVVKLLLSQGADPNTSDSDGRTPLDLAMEH KNIRVVLLLIMQGGDLMHHANELEHHHHHH**
ank4C4 (SEC-MALS, SAXS)	**MSTEGKMLIIAAREGMIIVVIVLLEKGADPNASDKDGRTPLHYAAENGHLI IVLLLLEKGADPNAKDSDGRTPLHYAAENGHKEIVEALLEHGADPNAKDS DGRTPLHYAAENGHKEIVKLLLSKGADPNTSDSDGRTPLDLAREHGNEEIV KLLEKQLEHHHHHH**
1na0C4_1 (SEC-MALS, SAXS)	**MTLARVAYILGAIAYAQGEYDIAITAYQVALDLDPNNAEAWYNLGNAYYK QGDYDEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYDEAIEYYQKAL ELDPNNAEAKQNLGNAKQKQGLEHHHHHH**
1na0C4_2 (SEC)	**MTAAEIAYNMGNAAYKEGDYDMAITAYQLALELDPNNAEAAYNLGNAYY KQGTYLMAILFYLIALLLDPNNAEAWYNLGNAYYKKGRYDEAIEYYQKAL ELDPNNAEAKQNLGNAKQKQGLEHHHHHH**
1na0C4_3	MHHHHHHGSNNAEAWYNLGNAYYKQGDASEAIVYYLLALVLDPNNAEAWY NLGNAAYKRGDYDEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYLEAIV YYILALELDPNNAEAKQNLGNAKQKEG
2fo7C4	MASMAYAIGNSFYRQGDYTMAMLMYILALLLDPRSAEAAYNLGNAAYKSGD YEAAIEAYQKALELDPRSAEAWYNLGNAYYKQGDYDEAIEYYQKALELDPRS AEAWYNLGNAYYKQGDYDEAIEYYQKALELDPRSLEHHHHHH
HR10C4_1 (SEC)	**MTVLAVILALILLIIVVANAKRKGDDTTEAALAALEAFALVLEAAERAGIDS SEVLELAIRLIKEVVENAQREGYDISEAARAAAEAFKRVAEAAKRAGITSSE VLELAIRLIKEVVENAQREGYDISEAARAAAEAFKRVAEAAKRAGITSSETL KRAIEEIRKRVEEAQREGNDISEAARQAAEEFRKKAEELKGSWLEHHHHH H**
HR10C4_2	MHHHHHHGSWGSSEKEELRLLLAVIVMIAAILKGDDSEEAREAAREAFELVRE AAERAGIDSSEVLELAIRLIEEVVENAEEEGYDISEAARAAAEAFKRVAEAAKR AGITSSEVLELAIRLIKEVVENAQREGYDISEAARAAAEAFKRVAEAAKRAGITS SEMLKMAIEIIRKMVEAAQRRGNDISEAARQAAEAVRKVAEVIK
HR04C4_1 (SEC-MALS, SAXS)	**MHHHHHHGSWGSDECEEKARRVAEKVERLKRSGTSEDEIAEEVAREISEV IRTLKESGSSYEVICECVARIVAEIVEALKRSGTSAVEIAKIVARVISEVIRTL KESGSSYEVICECVARIVAEIVEALKRSGTSAAIIALIVALVISEVIRTLKESG SSFEVILECVIRIVLEIIEALKRSGTSEQDVMLIVMAVLLVVLATLQLSGS**
HR04C4_1_r4 (SEC-MALS, SAXS)	**MHHHHHHGSWGSDECEEKARRVAEKVERLKRSGTSEDEIAEEVAREISEV IRTLKESGSSYEVICECVARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTL KESGSSYEVICECVARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESG**

	SSYEVICECVARIVAEIVEALKRSGTSAVEIAKIVARVISEVIRTLKESGSSYE VICECVARIVAEIVEALKRSGTSAAIIALIVALVISEVIRTLKESGSSFEVILEC VIRIVLEIIEALKRSGTSEQDVMLIVMAVLLVVLATLQLSGS
HR04C4_2 (SEC)	**MHHHHHHGSWGSDECEEKARRVAEKVERLKRSGTSEDEIAEEVAREISEV IRTLKESGSSYEVICECVARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTL KESGSSYVVICECVARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESG SSYVIIKACVEAIVREIVEALKRSGTSEIEIRLIVLRVKTEVERTLKESGS**
HR04C4_3 (SEC-MALS, SAXS)	**MHHHHHHGSWGSDECEEKARRVAEKVERLKRSGTSEDEIAEEVAREISEV IRTLKESGSSYEVICECVARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTL KESGSSYIVICLCVARIVAEIVEALKRSGTSAVEIALIVARVISEVIRTLKESG SSYSVIRMCVIIIVAEIIEALKRSGTSEVEINMIVKMVEAVVEQTLRESGE**
HR04C4_4 (SEC-MALS, SAXS)	**MHHHHHHGSWGSDECEEKARRVAEKVERLKRSGTSEDEIAEEVAREISEV IRTLKESGSSFEVICECVARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTL KESGSSFAVICICVARIVAEIVEALKRSGTSSIEIAEIVARVISEVIRTLKESGS SYIMIEACVKIIVIAIVEALKRSGTSETEIEIIVLLVKAIVRITLIESGS**
K64sC4	MLEKESARLLALLEALSKLGAEDPQAAESFERALGEALKSLGAASPEAIKAFAE ALGKALKELGATRPAAIKAFALALGLALAALGATDPEAIKAFAEALGKALKEL GATDPLAIAAFALALGAALKRLGATDPEAIKAFAEALGKALKELGATDPKAIK AFAEALGAALKRLGATDPEAIKAFAEALGKALKELGATDPEAIKAFAEALGAA LKRLGATDPEAIKAFAEALGKALKELGATDPEAIKAFAEALGAALKRLGATSPE AIKAFAEALGKALKELGATDPEAIKRFAERLGDELRKEGATDPERIKERAEREG RERKQEGKTDGSWLEHHHHHH
HR09C4	MYEDEAEEKALQVALRVLEEKVRKTSEDSIAELVALVISMVIRALKSSGSSYEV IAEIVARIVASIVEALKNAGTSEDEIAEIVARVISEVIRTLKESGSSYEVIAEIVARI VAEIVEALKRSKTSEDEIAEIVARVISEVIRTLKESGSSYEVIKEIVQRIVEEIVEA LKRSGTSEDEINEIVRRVKSEVERTLKESGSSLEHHHHHH
tpr1C4_1	MHHHHHHGSAEAWKELGKVLEKLGRLDEAAEAYKKAIELDPNDAEAWKELG KVLEKLGRLDEAAEAYKKAIELDPNDAEAWKELGKVLEKLGRLDEAAEAYKK AIKLDPNDAEAWKELGSVLEKLGRETMALLAYILALLLDPND
tpr1C4_2 (SEC-MALS, SAXS)	**MASSWVMLGLLLSLLNRLSLAAEAYKKAIELDPNDALAWLLLGSVLEKLK RLDEAAEAYKKAIELKPNDASAWKELGKVLEKLGRLDEAAEAYKKAIELD PEDAEAWKELGKVLEKLGRLDEAAEAYKKAIELDPNDLEHHHHHH**
2fo7C5_1	MAEAWYNLGNAYYKQGDYDEAIEYYQKALELDPRSAEAWYNLGNAYYKQG DYDEAIEYYQKALELDPRSAEAWYNLGNAYYKQGDYDEALKYYLIALILNPD SAEAARMTGMAALTQGDVKLAVLLLLLALQLNPNNLEHHHHHH
2fo7C5_2	MAEAWYNLGNAYYKQGDYDEAIEYYQKALELDPRSAEAWYNLGNAYARLG RLDEAIEYYQKALELDPRSAEAWYNLGLAYALLGDYDEAIEYYQKALELLPSL AEALYTLGLAYELQGDYDEAIAYKVLALLLKPRELEHHHHHH
3ltjC5 (SEC-MALS)	**MHHHHHHHKELLEVALYILLLKSDEEKKRMAAADALGKIGAELAVEPLIKA LKDERKVVATAAAAALGAIGDERAVEPLIKALKDEEGAVRLSAAVALGKI GDERAVEPLIKALKDERKVVRVAAAFALGEIGDERAVEPLIKALKDEEGM VRQSAADALGEIGGERVRAAMEKLAETGTGFARKVAVNYLETHK**

HL1C5 (SEC-MALS)	**MYQLIAMLMVLLGAYKLAAIAAEKAGLYLAAAVAWELSGDYTEAAKAW EKAGDYTEAAKAWEKSGDYTEAAKAWEKAGDYTEAAKAWEKSGDYTEA AKAWEKAGDYTEAAKAWEKSGDYTEAAKAWEKAGDYTEAAKAWEKSG DYTEAAKAWEKAGDLEHHHHHH**
tpr1C5	MAEAWKELGKVLEKLGRLDEAAEAYKKAIELDPNDAEAWKELGKVLEKLGR LDEAAEAYKKAIELDPNDAEAWKELGKVLEKLGRLDEAAEAYRMALLEDVT DAEAAMLLGRVLTKLGRLALALLAMILAVLLKPNALEHHHHHH
ank1C5	MSILGLMLIVAARNGNKDLVKRLIENGADVNASDSDGRTPLHHAAENGHAEV VELLISKGADVNAKDSDGRTPLHHAAENGHLIVVLLLISKGADVNAKDSDGRT PLHHAAENGHKEVVEALISAGADVNTSDSDGRTPLDLAREHGNEEVVKLLEK QGSWLEHHHHHH
arm8C5_1 (SEC)	**MNTVERAVKLLTSTDSRTQIAAALLLALIASGPASAIVLIVIAGGVEVLVKL LTSTDSEVQKEAARALANIASGPDVAIRAIVEAGGVEVLVKLLTSTDSEVQ KEAARALANIASGPDEAIKAIVDAGGVEVLVKLLTSTDSEVQKEAARALAN IASGPDEAIKAIVDAGGVEVLVKLLTSTDSEVQKEAARALANIASGPTSAIK AIVDAGGVEVLQKLLTSTDSEVQKEAQRALENIKSGSWLEHHHHHH**
arm8C5_2	MNDVEQLVKALTSTDSTLQMVAAMMLAEIASGPARAIALIVVAAGVEVLVKL LTSTDSEVQKEAARALANIASGPDLAIRAIVDAGGVEVLVKLLTSTDSEVQKEA ARALANIASGPDEAIKAIVDAGGVEVLVKLLTSTDSEVQKEAARALANIASGPD EAIKAIVDAGGVEVLVKLLTSTDSEVQKEAARALANIASGPTSAIKAIVDAGGV EVLQKLLTSTDSEVQKEAQRALENIKSGSWLEHHHHHH
HR10C5_1 (SEC-MALS)	**MSNTTQLLVEILVIIVVANARRKGDDTDEAKLAAIIAALLVIIAAERAGIDSS EVLELAIRLIKEVVENAQREGYDISEAARAAAEAFKRVAEAAKRAGITSSEV LELAIRLIKEVVENAQREGYDISEAARAAAEAFKRVAEAAKRAGITSSETLK RAIEEIRKRVEEAQREGNDISEAARQAAEEFRKKAEELKGSWLEHHHHHH**
HR10C5_2 (SEC-MALS, SAXS)	**MSAEKLMLMAKLIIIVAENAKRKGDDTLIAIMAAKLAFEIVRIAAEEAGIDS SEVLELAIRLIKEVVENAQREGYDISIAALAAAMAFALVAIAAKRAGITSSE VLELAIRLIKEVVENAQREGYDIAEAARAAAEAFKRVAEAAKRAGITSSET LKRAIEEIRKRVEEAQREGNDISEAARQAAEEFRKKAEELKGSWLEHHHH HH**
HR07C5_1 (SEC)	**MTKSTARITCMIAAIKAARENDEMVAVMAALVCLMVAEQAGMPTKEAAR SFCEAAARAAAISNDEEVAKIAAKACLEVAKAAGMPTKEAARSFCEAAAR AAAESNDEEVAKIAAKACLEVAKQAGMPTKEAARSFCEAAKRAAKESNDE EVEKIAKKACKEVAKQAGMPGSWLEHHHHHH**
HR07C5_2 (SEC)	**MHHHHHHGSWGSTKEDARSTCEKAARKAAESNDEEVAKQAAKDCLEVA KQAGMPTKEAARSFCEAAARAAAESNDEEVAKIAAKACLEVAKQAGMPT KEAARSFCAAAARAAAESNDEEVAKIAAKACLEVAKQAGMPTREAAAIFC VAARLAAMESNDEEVEKIAEKACLLVALQAGMP**
HR07C5_3 (SEC)	**MHHHHHHGSWGTKEDARSTCEKAARKAAESNDEMEAIAAALLCALVAK EAGMPTKEAARSFCEAAARAAAESNDEEVAKIAAKACLIVAKAAGMPTKE AARSFCEAAARAAAESNDVRVAVIAALACLEVAKQAGMPTKEAARSFCEA AKRAAKESNDEEVEKVAKMACVAVAIAAGMP**

HR04C5_1 (SEC)	**MRACEAEAMLIAAKVMLLKENGTSEDEIAEEVAREISEVIRTLKESGSSYL VICECVARIVAMIVEALKLSGTSEDEIAEIVARVISEVIRTLKESGSSYKKICV CVADIVAEIVEALKRSNTSEDEIAEIVARVISEVIRTLKESGSSYEVIAACVIAI VLAIIKALKRSGTSEDEINEIVRRVKSEVERTLKESGSWLEHHHHHH**
K64sC5	MKEKEAARTLALLLALSELRARDPLAQLLFDVALGVALDRLGAASPEAIKAFA EALGKALKELGATSPVAIAAFALALGAALKRLGATDPEAIKAFAEALGKALKE LGATDPEAIKAFAEALGAALKRLGATDPEAIKAFAEALGKALKELGATDPEAIK AFAEALGAALKRLGATDPEAIKAFAEALGKALKELGATDPEAIKAFAEALGAA LKRLGATDPEAIKAFAEALGKALKELGATDPEAIKAFAEALGAALKRLGATSPE AIKAFAEALGKALKELGATDPEAIKRFAERLGDELRKEGATDPERIKERAEREG RERKQEGKTDGSWLEHHHHHH
KP16C5_1	MHHHHHHGSWGSEKNKLVEEAKEAIENQDKKKVSQIVEDVARSNDKEVLIEI AKVAIDKQDANIVAVVVVLVAESNDKEVLIEIAKVAIDKQDERIVQLVVSIVAI SNDKEVLIEIAKVAIDKQDENIVTSVVRIVARSNDKEVLIEIAKVAIDKQDENIV ASVVKIVAESNDKEVLIEIAKVAIDKQDENIVQQVVKIVAESNDKEVLIEIAKVA IDKQDSNIVTSVVKIVAESNDEEVIQEIYKVAREKQDREIAVSIIVVLILDD
KP16C5_2	MHHHHHHGSWGSEKNKLVEEAKEAIENQDENKVEQIVEDVARSNDKEVLIEIA KVAIDKQDEEIVAAVVIIVAESNDKEVLIEIAKVAIDKQDENIVTSVVAVVAVSE DKEVLIEIAKVAIDKQDENIVTSVVSVVASSNDKEVLIEIAKVAIDKQDENIVAS VVKIVAKSNDKEVLIEIAKVAIDKQDENIVQQVVKIVAESNDKEVLIEIAKVAID KQDKRIVRSVVKIVAESNDKEVIQEIYKVAREKQDLTIAVSIMMVAIEDD
2fo7C6 (SEC)	**MAEALYNMGNKYYKQGDYEVAIIAYQAALELDPRSAEAWYNLGNAYYKQ GDYDEAIEYYQKALELDPRSAEAAYNLGNAYYKQGDYQEAAIAYLLALAL DPRSAEAWYNLGNAMYKMGIYDSAIEYYQKALELDPRSLEHHHHHH**
arm8C6 (SEC)	**MNEVEKLVKLLTSTDSVLQMMAAAMLANIASGPAEAIARIVLAGGVKVLV KLLTSTDSAVQKLAARALANIASGPDLAILAIVDAGGVEVLVKLLTSTDSEV QKEAARALANIASGPDEAIKAIVDAGGVEVLVKLLTSTDSEVQKEAARALA NIASGPDEAIKAIVDAGGVEVLVKLLTSTDSEVQKEAARALANIASGPTSAI KAIVDAGGVEVLQKLLTSTDSEVQKEAQRALENIKSGSWLEHHHHHH**
HR10C6_1	MSKEKELLRLILIAIVIIAALRKGDDAQVAAMAARVAFLLVRMAAEIAGIDSSE VLELAIRLIKEVVENAQREGYDIAAAALAAALAFMRVAEAAKRAGITSSEVLEL AIRLIKEVVENAQREGYEITEAARAAAEAFKRVAEAAKRAGITSSETLKRAIEEI RKRVEEAQREGNDISEAARQAAEEFRKKAEELKGSWLEHHHHHH
HR10C6_2 (SEC)	**MSAVKQALRMRLIMIVIENAKRKGDDTRLAEKAAEAAFEIVREAAERAGI DSSEVLELAIRLIKEVVENAQREGYDISKAALAAASAFMRVAEAAKRAGITS SEVLELAIRLIKEVVENAQREGYDISEAARAAAEAFKRVAEAAKRAGITSSE TLKRAIEEIRKRVEEAQREGNDISEAARQAAEEFRKKAEELKGSWLEHHH HHH**
HR04C6_1	MHHHHHHGSWGSDECEEKARRVAEKVERLKRSGTSEDEIAEEVAREISEVIRT LKESGSSYDVICECVARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESGST YIIICLCVAIIVAEIVEALKRSGTSTIEIAIIVARVISEVIRTLKESGSSYLIIKACVQV IVLMIVLALKRSGTEEVQIKLIVLAVKLAVEKTLKESGS
HR04C6_2	MRVCEAIAEAIAILVERLKRSGTSEDEIAEEVARIISEVIRRLKESGSSYEVICECV ARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESGSSEDVICICVARIVAEIV

	EALKRSGTSEDEIAEIVARVISEVIRTLKESGSSYLVILACVILIVKEIVEALKRSG TSEDEINEIVRRVKSEVERTLKESGSWLEHHHHHH
HR04C6_3 (SEC-MALS)	**MGKRCEEKARLVARLVRIAKEDGTSEDEIAEEVAREISRVIRRLKESGSEY EVICECVARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESGSDYMVIC ECVARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESGSKRIVIKLCVE LIVEEIVEALKRSGTSEDEINEIVRRVKSEVERTLKESGSWLEHHHHHH**
HR07C6_1	MHHHHHHGSWGSTKEDARSTCEKAARKAAESNDTNVALNAAARCLVVAKQ AGMPTKEAARSFCEAAARAAAESNDEEVAKIAAQACAAVAMAAGMPAKEAA RSFCEAAARAAAESNDIEVAQIAAKACLEVAKQAGMPTKEAARSFCEAAKRA AKESNDEEVLVVAIVACLAVALQAGMP
KP16C6 (SEC-MALS)	**MHHHHHHGSWGSEKNKLVEEAKEAIENQDVVKVVQIVVKVAKSNDKEVL IEIAKVAIDKQDRRIVAAVVVAVALSNDKEVLIEIAKVAIDKQDENIVRAVV SIVALSNDKEVLIEIAKVAIDKQDENIVTSVVKIVAESNDKEVLIEIAKVAIDK QDENIVASVVKIVAESNDKEVLIEIAKVAIDKQDENIVQQVVKIVAESNDKE VLIEIAKVAIDKQDSNIVSSVVEIVAESNDEEVIQEIYKVAREKQDEVIIVLIL AVLVADE**
HR00C6 (SEC-MALS, SAXS)	**MTEEKIAKEISLVAESAKEMIEELARMADNVETETVVDEVIEAIAKLARLAI KTIEDLAKNLASEEFMARAISAIAELAKKAIEAIYRLADNHTTDTFMAKAIE AIAELAKEAIKAIADLAKNHTTESFMARAIKAIAELAKKAIEAIYRLADNHT TDTFMAKAIEAIAELAKEAIKAIADLAKNHTTEMFMVAAIVAIALLAASAIT AIYRLADNHTSPTYIEKAIEAIEKIARTAIKAIEDLAKNITTEEAKSIAKVAIR IIRALAKSAIKKLEDNRTLEHHHHHH**
3ltjC6 (SEC)	**MGHHHHRDPAKVLLYIERLEKDSDEVRMEAALALGKIGDERAVAPLIKAL KDERRAVRAAAALALGLIGDERAVEPLIKALKDEDGRVRQMAAVALGEIG DERAVEPLIKALKDEDEFVRIAAAFALGEIGDERAVEPLIKALKDEDGWVR RSAADALGEIGGERVRAAMEKLAETGTGFARKVAVNYLETHKLEHHHHH H**
HR71C6 (SEC-MALS)	**MGPEILLELAKESLERARRASERGREEMFRAAAELALEIAKLLVEVAEVEG DPELVLEAAKVALRVAELAAKNGDKEVFKKAAESALEVAKRLVEVASKE GDPELVARAAKVALRVAELAAENGDKEVFKKAAESALEVAKRLVEVASK EGDVALVIVALVVAMAVRLLAKVQGDEEVYEKARETAREVREEAKRVIEE KRGWLEHHHHHH**
HR64C6 (SEC)	**MGPEDELKRVEKLVKEAEELLRQAKEKGSRLDLLAALITAIAAALEAKKV LEQAEKEGDPEVALRAVELVVRVAELLLRIAKESGAELVLEIALIVAEEAA RLAKRVLELAEKQGDPIVAALAVILVLRVALLLLRIAKESGSEEALERALR VAEEAARLAKRVLELAEKQGDPLVAMIAVIAVRIVAEILEEIARESGSEEAK ERAERVREEARELQERVKELRERWLEHHHHHH**
HR54C6 (SEC-MALS, SAXS)	**MGTEDERRELEKVALAALLAAVLGNTEEVLKQIIRALIIALESGTTEAVKL ALEVVAEVAIEAARRGNTDAVRVALAAALLIARRSGTTEAVKLALEVVAR VAIEAARRGNTDAVREALEVALEIARESGTEEAVRLALEVVKRVSDEAKK QGNEDAVKEAEEVRKKIEEESGGWLLEHHHHHH**
HR14C6 (SEC-MALS)	**MGDSEDENELVKVMAEIAKEARIKELVILMVALLAELAKESTDSKLVNRIV KQLAEVAKEATDKELVIYIVKILAELAKQSTDRILVIMIVLQLALVAIEATD**

	KELVIYIVKILAELAKQSTDSRLVRVIVMLLRIVARKATDKELVEHIEKILE ELKKQSWLEHHHHHH
ank1_asymJF1v1A	MGSARGRLLIQAARNGVKRIVKQLIRQGADVNASDSDGRTPLHHAAENGHKE VVKLLISKGADVNAKDSDGRTPLHHAAENGHKEVVKLLISKGADVNAKDSDG RTPLHHAAENGHKEVVKLLISKGADVNTSDSDGRTPLDLAREHGNEEVVKLLE KQGGW
ank1_asymJF1v1B	MSDDGEILIMAAENGNKDMVKLAIELGADVNASDSDGRTPLHHAAENGHKEV VKLLISKGADVNAKDSDGRTPLHHAAENGHKEVVKLLISKGADVNAKDSDGR TPLHHAAENGHKEVVKLLISKGADVNAKDSDGRTPLHHAAENGHKEVVKLLI SKGADVNAKDSDGRTPLHHAAENGHKEVVKLLISKGADVNTSDSDGRTPLDL AREHGNEEVVKLLEKQGGWLEHHHHHH
ank1_asymJF1v2A	MGSKLGRALIRAAQQGAKKVVKLLIKAGADVNASDSDGRTPLHHAAENGHKE VVKLLISKGADVNAKDSDGRTPLHHAAENGHKEVVKLLISKGADVNAKDSDG RTPLHHAAENGHKEVVKLLISKGADVNTSDSDGRTPLDLAREHGNEEVVKLLE KQGGW
ank1_asymJF1v2B	MSDNGEALIAAASNGMSEVVEILIELGADVNASDSDGRTPLHHAAENGHKEVV KLLISKGADVNAKDSDGRTPLHHAAENGHKEVVKLLISKGADVDSDGRTPLH HAAENGHKEVVKLLISKGADVNAKDSDGRTPLHHAAENGHKEVVKLLISKGA DVNAKNAKDSDGRTPLHHAAENGHKEVVKLLISKGADVNTSDSDGRTPLDLA REHGNEEVVKLLEKQGGWLEHHHHHH
ank1_asymJF1v3A	MGSNLGRALIIAAKNGMKRAVKIMIRRGADVNASDSDGRTPLHHAAENGHKE VVKLLISKGADVNAKDSDGRTPLHHAAENGHKEVVKLLISKGADVNAKDSDG RTPLHHAAENGHKEVVKLLISKGADVNTSDSDGRTPLDLAREHGNEEVVKLLE KQGGW
ank1_asymJF1v3B	MSDEGELLIIAAENGVKELVKLLIENGADVNASDSDGRTPLHHAAENGHKEVV KLLISKGADVNAKDSDGRTPLHHAAENGHKEVVKLLISKGADVDSDGRTPLH HAAENGHKEVVKLLISKGADVNAKDSDGRTPLHHAAENGHKEVVKLLISKGA DVNAKNAKDSDGRTPLHHAAENGHKEVVKLLISKGADVNTSDSDGRTPLDLA REHGNEEVVKLLEKQGGWLEHHHHHH
ank1_asymJF1v4A	MGSKRGRALIMAAKKGIKRLVKLLIQAGADVNASDSDGRTPLHHAAENGHKE VVKLLISKGADVNAKDSDGRTPLHHAAENGHKEVVKLLISKGADVNAKDSDG RTPLHHAAENGHKEVVKLLISKGADVNTSDSDGRTPLDLAREHGNEEVVKLLE KQGGW
ank1_asymJF1v4B	MSEDGEALIIAAQNGEKEIVKLLIEAGADVNASDSDGRTPLHHAAENGHKEVV KLLISKGADVNAKDSDGRTPLHHAAENGHKEVVKLLISKGADVDSDGRTPLH HAAENGHKEVVKLLISKGADVNAKDSDGRTPLHHAAENGHKEVVKLLISKGA DVNAKNAKDSDGRTPLHHAAENGHKEVVKLLISKGADVNTSDSDGRTPLDLA REHGNEEVVKLLEKQGGWLEHHHHHH
ank1_asym1v1A	**MGWGSELGKRLIEAAENGNKDRVKDLIENGADVNASDSDGRTPLHHAAE NGHAKVVALLIKKGADVNAKDSDGRTPLHHAAENGHLIVVMILLLKGAD VNAKDSDGRTPLHHAAENGHVRVVLALILAGADVNTSDSDGRTPLDLARE HGNERVVKALEKQ**

ank1_asym1v1B **(SEC, SAXS, LC-MS)**	**MHHHHHHWGSELGKRLIEAAENGNKDRVKDLIENGADVNASDSDGRTPL HHAAENGHKEVVKLLISKGADVNAKDSDGRTPLHHAAENGHKEVVKLLIS KGADVNAKDSDGRTPLHHAAENGHAEVVALLIDKGADVNAKDSDGRTPL HHAAENGHLEVVAILLLQGADVNAKDSDGRTPLHHAAENGHLAVVLLLII AGADVNTSDSDGRTPLDLAREHGNEEVVEALEEQ**
ank1_asym1v2A	MGWGSELGKRLIEAAENGNKDRVKDLIENGADVNASDSDGRTPLHHAAENGH AKVVALLIKKGADVNAKDSDGRTPLHHAAENGHLRVVMILLRKGADVNAKD SDGRTPLHHAAENGHKRVVLMLILIGADVNTSDSDGRTPLDLAREHGNEKVV KALERQ
ank1_asym1v2B	MHHHHHHWGSELGKRLIEAAENGNKDRVKDLIENGADVNASDSDGRTPLHH AAENGHKEVVKLLISKGADVNAKDSDGRTPLHHAAENGHKEVVKLLISKGAD VNAKDSDGRTPLHHAAENGHAEVVALLIDKGADVNAKDSDGRTPLHHAAEN GHLEVVLILVAEGADVNAKDSDGRTPLHHAAENGHVDVVLVLILVGADVNTS DSDGRTPLDLAREHGNEEVVEALEED
ank3_asym1A	**MGSELGKRLIEAAENGNKNRVKDLLENGADVNASDSDGKTPLHLAAENG HLKVVLLLLLQGADPNAKDSDGKTPLHLAAENGHAAVVAALLIVGADPNA KDSDGKTPLHLAAENGHKQVVILLLAVGADPNTSDSDGRTPLDLAREHGN EQVVRVLQAMGGW**
ank3_asym1B **(SEC, SAXS, LC-MS)**	**MSELGKRLIEAAENGNVERVQDLLENGADVNASDSDGKTPLHLAAENGHA EVVLLLLEQGADPNAKDSDGKTPLHLAAENGHVVVVALLLLHGADPNAK DSDGKTPLHLAAENGHEAVVILLLLMGADPNAKDSDGKTPLHLAAENGH KEVVEVLLMHGADPNAKDSDGKTPLHLAAENGHKEVVKLLLSQGADPNT SDSDGRTPLDLAREHGNEEVVKLLEKQGGWLEHHHHHH**
ank3_asym2v1A	MGSELGKRLIEAAENGNKRRVMRLLLRGADVNASDSDGKTPLHLAAENGHRQ VVALLLLQGADPNAKDSDGKTPLHLAAENGHKQVVRLLLKAGADPNAKDSD GKTPLHLAAENGHKEVVKLLLSQGADPNTSDSDGRTPLDLAREHGNEEVVKL LEKQGGW
ank3_asym2v1B	MSELGKRLIEAAENGNEDRVLELLMDGADVNASDSDGKTPLHLAAENGHEEV VMALILQGADPNAKDSDGKTPLHLAAENGHKSVVKLLLEVGADPNAKDSDGK TPLHLAAENGHKEVVKLLLSQGADPNAKDSDGKTPLHLAAENGHKEVVKLLL SQGADPNAKDSDGKTPLHLAAENGHKEVVKLLLSQGADPNTSDSDGRTPLDL AREHGNEEVVKLLEKQGGWLEHHHHHH
ank3_asym2v2A	MGSELGKRLIEAAENGNKRRVMRLLLRGADVNASDSDGKTPLHLAAENGHRQ VVALLLLQGADPNAKDSDGKTPLHLAAENGHKQVVRLLLKAGADPNAKDSD GKTPLHLAAENGHKEVVKLLLSQGADPNTSDSDGRTPLDLAREHGNEEVVKL LEKQGGW
ank3_asym2v2B	MSELGKRLIEAAENGNEDRVLELLMDGADVNASDSDGKTPLHLAAENGHEEV VMALILQGADPNAKDSDGKTPLHLAAENGHKSVVKLLLEIGADPNAKDSDGK TPLHLAAENGHKEVVKLLLSQGADPNAKDSDGKTPLHLAAENGHKEVVKLLL SQGADPNAKDSDGKTPLHLAAENGHKEVVKLLLSQGADPNTSDSDGRTPLDL AREHGNEEVVKLLEKQGGWLEHHHHHH
ank3_asym2v3A	MGSELGKRLIEAAENGNKRRVMKLLMRGADVNASDSDGKTPLHLAAENGHR QVVALLLLQGADPNAKDSDGKTPLHLAAENGHKQVVRLLLKQGADPNAKDS

	DGKTPLHLAAENGHKEVVKLLLSQGADPNTSDSDGRTPLDLAREHGNEEVVK LLEKQGGW
ank3_asym2v3B	MSELGKRLIEAAENGNEDRVLELLANGADVNASDSDGKTPLHLAAENGHSEV VMALLLQGADPNAKDSDGKTPLHLAAENGHKEVVKLLLETGADPNAKDSDG KTPLHLAAENGHKEVVKLLLSQGADPNAKDSDGKTPLHLAAENGHKEVVKLL LSQGADPNAKDSDGKTPLHLAAENGHKEVVKLLLSQGADPNTSDSDGRTPLD LAREHGNEEVVKLLEKQGGWLEHHHHHH
ank3_asym2v4A	MGSELGKRLIEAAENGNRKRVMRLLAKGADVNASDSDGKTPLHLAAENGHR RVVALLLLQGADPNAKDSDGKTPLHLAAENGHKQVVRLLLKTGADPNAKDS DGKTPLHLAAENGHKEVVKLLLSQGADPNTSDSDGRTPLDLAREHGNEEVVK LLEKQGGW
ank3_asym2v4B	MSELGKRLIEAAENGNEDRVLELLMNGADVNASDSDGKTPLHLAAENGHEDV VMALLLQGADPNAKDSDGKTPLHLAAENGHKNVVELLLDQGADPNAKDSDG KTPLHLAAENGHKEVVKLLLSQGADPNAKDSDGKTPLHLAAENGHKEVVKLL LSQGADPNAKDSDGKTPLHLAAENGHKEVVKLLLSQGADPNTSDSDGRTPLD LAREHGNEEVVKLLEKQGGWLEHHHHHH
ank3_asym2v5A	MGSELGKRLIEAAENGNRRRVMQLLMRGADVNASDSDGKTPLHLAAENGHR RVVALLLLQGADPNAKDSDGKTPLHLAAENGHKQVVRLLLKNGADPNAKDS DGKTPLHLAAENGHKEVVKLLLSQGADPNTSDSDGRTPLDLAREHGNEEVVK LLEKQGGW
ank3_asym2v5B	MSELGKRLIEAAENGNTDRVLELLANGADVNASDSDGKTPLHLAAENGHEEV VMALLLQGADPNAKDSDGKTPLHLAAENGHKSVVKLLLETGADPNAKDSDG KTPLHLAAENGHKEVVKLLLSQGADPNAKDSDGKTPLHLAAENGHKEVVKLL LSQGADPNAKDSDGKTPLHLAAENGHKEVVKLLLSQGADPNTSDSDGRTPLD LAREHGNEEVVKLLEKQGGWLEHHHHHH
ank3_asym4v1A	MGSELGKRLIEAAENGNKDRVKDLLRNGADVNASDSDGKTPLHLAAENGHLR VVLLLLAKGADPNAKDSDGKTPLHLAAENGHNKVVLALALAGADPNAKDSD GKTPLHLAAENGHKVVVAILLIVGADPNTSDSDGRTPLDLAREHGNEEVVKLL ERMGGW
ank3_asym4v1B	MSELGKRLIEAAENGNKDRVEDLLENGADVNASDSDGKTPLHLAAENGHMEV VILLLAQGADPNAKDSDGKTPLHLAAENGHDAVVLLLAIQGADPNAKDSDGK TPLHLAAENGHKDVVMVLLMAGADPNAKDSDGKTPLHLAAENGHKEVVELL LEMGADPNAKDSDGKTPLHLAAENGHKEVVKLLLSQGADPNTSDSDGRTPLD LAREHGNEEVVKLLEKQGGWLEHHHHHH
ank3_asym4v2A	MGSELGKRLIEAAENGNKDRVKDLLKNGADVNASDSDGKTPLHLAAENGHLR VVILLAAKGADPNAKDSDGKTPLHLAAENGHVKVVLVLIAIGADPNAKDSDG KTPLHLAAENGHKMVVMVLLMAGADPNTSDSDGRTPLDLAREHGNEEVVKL LEKQGGW
ank3_asym4v2B	MSELGKRLIEAAENGNKDRVKDLLDNGADVNASDSDGKTPLHLAAENGHLEV VLLLAAQGADPNAKDSDGKTPLHLAAENGHDAVVLVLLIIGADPNAKDSDGK TPLHLAAENGHKDVVAALLMQGADPNAKDSDGKTPLHLAAENGHKEVVELL LEQGADPNAKDSDGKTPLHLAAENGHKEVVKLLLSQGADPNTSDSDGRTPLD LAREHGNEEVVKLLEKQGGWLEHHHHHH

ank3_asym4v3A	MGSELGKRLIEAAENGNKDRVKDLLRNGADVNASDSDGKTPLHLAAENGHVR VVILLLMKGADPNAKDSDGKTPLHLAAENGHKKVVLLLLIQGADPNAKDSDG KTPLHLAAENGHKQVVAALLAAGADPNTSDSDGRTPLDLAREHGNEEVVKLL EKQGGW
ank3_asym4v3B	MSELGKRLIEAAENGNKDRVEDLLDNGADVNASDSDGKTPLHLAAENGHSEV VLLLAMQGADPNAKDSDGKTPLHLAAENGHDIVVLILILLGADPNAKDSDGKT PLHLAAENGHEDVVAVLLMAGADPNAKDSDGKTPLHLAAENGHKEVVELLL EIGADPNAKDSDGKTPLHLAAENGHKEVVKLLLSQGADPNTSDSDGRTPLDLA REHGNEEVVKLLEKQGGWLEHHHHHH
ank3_asym6v1A	MGSELGKRLIEAAENGNKRRVMNLLANGADVNASDSDGKTPLHLAAENGHVI VVILLVARGADPNAKDSDGKTPLHLAAENGHKKVVLILLMKGADPNAKDSDG KTPLHLAAENGHKEVVKLLLSQGADPNTSDSDGRTPLDLAREHGNEEVVKLLE KQGGW
ank3_asym6v1B	MSELGKRLIEAAENGNDDRVMELLQNGADVNASDSDGKTPLHLAAENGHEIV VLLLLAEGADPNAKDSDGKTPLHLAAENGHEDVVLILLLEGADPNAKDSDGK TPLHLAAENGHKEVVKLLLSQGADPNAKDSDGKTPLHLAAENGHKEVVKLLL SQGADPNAKDSDGKTPLHLAAENGHKEVVKLLLSQGADPNTSDSDGRTPLDL AREHGNEEVVKLLEKQGGWLEHHHHHH
ank3_asym6v2A	MGSELGKRLIEAAENGNKRRVINLLNNGADVNASDSDGKTPLHLAAENGHVIV VMLLLARGADPNAKDSDGKTPLHLAAENGHKRVVALLLARGADPNAKDSDG KTPLHLAAENGHKEVVKLLLRQGADPNTSDSDGRTPLDLAREHGNEEVVKLL EKQGGW
ank3_asym6v2B	MSELGKRLIEAAENGNAERVMDLLENGADVNASDSDGKTPLHLAAENGHSIV VILLLMEGADPNAKDSDGKTPLHLAAENGHEEVVVLLLDEGADPNAKDSDGK TPLHLAAENGHKEVVKLLLSQGADPNAKDSDGKTPLHLAAENGHKEVVKLLL SQGADPNAKDSDGKTPLHLAAENGHKEVVKLLLSQGADPNTSDSDGRTPLDL AREHGNEEVVKLLEKQGGWLEHHHHHH
ank4_asym2A	MGWGSELGKRLIEAAENGNKDRVKDLLENGADPNASDSDGRTPLHYAAENG HKQIVLLLLRKGADPNAKDSDGRTPLHYAAENGHVKIVLLLAMKGADPNAKD SDGRTPLHYAAENGHRMIVLILLIIGADPNTSDSDGRTPLDLAREHGNEQIVRLL ERQ
ank4_asym2B	MHHHHHHWGSELGKRLIEAAENGNKDRVKDLLENGADPNAKDSDGRTPLHY AAENGHKEIVKLLLSKGADPNAKDSDGRTPLHYAAENGHKEIVKLLLSKGADP NASDSDGRTPLHYAAENGHKEIVLLLLDKGADPNAKDSDGRTPLHYAAENGH SEIVLLLLIEGADPNAKDSDGRTPLHYAAENGHDIIVLILLMTGADPNTSDSDGR TPLDLAREHGNEEIVTLLEEQ
HR04_asym06A	MGDECEEKARRVAEEVERLKREGVSEKIIAELVALMIAIVILLLKMDGSSYEVI CECVARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRILKTSGSSYEVICECVARI VAEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESGSSYEVIKECVQRIVEEIVEA LKRSGTSEDEINEIVRRVKSEVERTLKESGSGWGS
HR04_asym06B	MDECEEKARRVAERVEELKREGVSEDIIALMVALMIAIVILLLKISGSSYEVICE CVARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTLKKSGSSYEVICECVARIV AEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESGSSYEVIKECVQRIVEEIVEAL KRSGTSEDEIAEIVARVISEVIRTLKESGSSYEVIKECVQRIVEEIVEALKRSGTSE

	DEIAEIVARVISEVIRTLKESGSSYEVIKECVQRIVEEIVEALKRSGTSEDEINEIV RRVKSEVERTLKESGSGWLEHHHHHH
HR04_asym08A	**MGRSCEEIARMIAIAVERAKRSGLSEDKIAEMVALVIAAVIRVLKESGSSYE VICECVARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESGSSYEVICE CVARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESGSSYEVIKECVQ RIVEEIVEALKRSGTSEDEINEIVRRVKSEVERTLKESGSGWGS**
HR04_asym08B (SEC, SAXS, LC-MS)	**MDDCELIARIVAVAVEIAKRRGLSEDDIAIMVALLISAVIRVLKEKGSSYEVI CECVARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESGSSYEVICECV ARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESGSSYEVICECVARIV AEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESGSSYEVICECVARIVAEIV EALKRSGTSEDEIAEIVARVISEVIRTLKESGSSYEVIKECVQRIVEEIVEALK RSGTSEDEINEIVRRVKSEVERTLKESGSGWLEHHHHHH**
HR04_asym10A	MGWGDECEEKARRVAEKVERLKRSGTSEDEIAEEVAREISEVIRTLKESGSSYE VICECVARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESGSSYEVICECVA RIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESGSSYEVIKECVQRIVEEIV EALKRSGTDQDEIDLIVIRVELEVERTLIESGSGS
HR04_asym10B	MHHHHHHWGDECEEKARRVAEKVERLKRSGTSEDEIAEEVAREISEVIRTLKE SGSSYEVICECVARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESGSSYEV ICECVARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESGSSYEVICECVAR IVAEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESGSSYEVICECVARIVAEIVE ALKRSGTSEEEIAEIVARVISEVIRTLKESGSSYEVIKECVQRIVEEIVEALKRSGT SEVEIRLIVKRVLLEVRRTLRESGS
HR04_asym23A	MGMSSEKEELRERLVKIVVENAKRKGDDTEEAREAAREAFELVREAAERAGID SSEVLELAIRLIKEVVENAQREGYDISEAARAAAEAFKRVAEAAKRAGITSSEV LELAIRLIKEVVENAQREGYDISEAARAAAEAFKRVAEAAKRAGITSSEVLMLA IILIRIMVEEAQREGNDISEAARKAAELVRKAAEVLKGS
HR04_asym23B	MHHHHHHWGMSSEKEELRERLVKIVVENAKRKGDDTEEAREAAREAFELVRE AAERAGIDSSEVLELAIRLIKEVVENAQRKGDDTEEAREAAREAFELVREAAER AGIDSSEVLELAIRLIKEVVENAQRKGDDTEEAREAAREAFELVREAAERAGID SSEVLELAIRLIKEVVENAQREGYDISEAARAAAEAFKRVAEAAKRAGITSSEV LELAIRLIKEVVENAQREGYDISEAARAAAEAFKRVAEAAKRAGITSSEILMMA IIMIRIMVEEAQREGNDISEAARKAAELLRKAAEVLK
HR04_asym46A	MGMSSEKEELRERLVKIVVENAKRKGDSTRLAEIAAIAAFLIVRVVAERAGIDS SEVLELAIRLIKEVVENAQREGYDIVKAAMLAALAFAMVAMAAKRAGITSSEV LELAIRLIKEVVENAQREGYDILRAAVAAAEAFLRVAEAAKRAGITSSETLKRA IEEIRKRVEEAQREGNDISEAARQAAEEFRKKAEELKGWGS
HR04_asym46B	MMSSEKEELRERLVKIVVENAKRKGDVTELAEAAAMVAFAMVRIAAERAGID SSEVLELAIRLIKEVVENAQREGYDIAKAALAAALAFVAVAIAAKRAGITSSEV LELAIRLIKEVVENAQREGYDILRAAEAAAEAFERVAEAAKRAGITSSETLKRAI EEIRKRVEEAQREGYDISEAARAAAEAFERVAEAAKRAGITSSETLKRAIEEIRK RVEEAQREGYDISEAARAAAEAFERVAEAAKRAGITSSETLKRAIEEIRKRVEE AQREGNDISEAARQAAEEFRKKAEELKGWLEHHHHHH
HR04_asym54A	MGESELKEKLREILVKIVVENAKRKGDDTEEAREAAREAFELVREAAERAGID SVLVLILAILLIKEVVENAQREGYDISEAARAAAEAFKRVAEAAKRAGITSLVV

	LLLAIRLIKEVVENAQREGYDISEAARAAAEAFKRVAEAAKRAGITEEATLIRAI LEILKRVEEAQREGNDISEAARQAAEEFRKKAEELKGWGS
HR04_asym54B	MMSELKEKLRERLVKIVVENAKRKGDDTEEAREAAREAFELVREAAERAGIDS VLVLILAIRLIKEVVENAQREGYDISEAARAAAEAFKRVAEAAKRAGITSLVVL LLAIRLIKEVVENAQREGYDISEAARAAAEAFKRVAEAAKRAGITSSATLIRAIL EILKRVEEAQREGYDISEAARAAAEAFKRVAEAAKRAGITSSETLKRAIEEIRKR VEEAQREGYDISEAARAAAEAFKRVAEAAKRAGITSSETLKRAIEEIRKRVEEA QREGNDISEAARQAAEEFRKKAEELKGWLEHHHHHH
HR05_asym03A	MGSEEKELARKLLVMICILNANLKGDDKEEAREAAREAFELVREAAERAGIDS REVLLLAIRLIKECVENAQREGYDISEACRAAAEAFKRVAEAAKRAGITSSEVL ELAIRLIKECVENAQREGYDISEACRAAAEAFKRVAEAAKRAGITSSETLKRAIE EIRKRVEEAQREGNDISEACRQAAEEFRKKAEELKRRGWGS
HR05_asym03B	MSERKELAREMLVVICILNAKEKGDDKEEAREAAREAFELVREAAERAGIDSD EVLLLAIRLIKECVENAQREGYDISEACRAAAEAFKRVAEAAKRAGITSSEVLE LAIRLIKECVENAQREGYDISEACRAAAEAFKRVAEAAKRAGITSSEVLELAIRL IKECVENAQREGYDISEACRAAAEAFKRVAEAAKRAGITSSEVLELAIRLIKECV ENAQREGYDISEACRAAAEAFKRVAEAAKRAGITSSETLKRAIEEIRKRVEEAQ REGNDISEACRQAAEEFRKKAEELKRRGWLEHHHHHH
HR05_asym25A	MGSSEKEELRERLVKICVENAKRKGDDTSLAEAAARMAFMLVRLMAELAGID SSEVLELAIRLIKECVENAQREGYDIAEACIAAALAFMIVAVAALLAGITSSEVL ELAIRLIKECVENAQREGYDISEACRAAAEAFKRVAEAAKRAGITSSETLKRAIE EIRKRVEEAQREGNDISEACRQAAEEFRKKAEELKRRGWGS
HR05_asym25B	MSSEKEELRERLVKICVENAKRKGDDTRLARLAARMAFIIVRLMAEMAGIDSS EVLELAIRLIKECVENAQREGYDILEACMAAALAFMAVAIAALVAGITSSEVLE LAIRLIKECVENAQREGYDISEACRAAAEAFKRVAEAAKRAGITSSEVLELAIRL IKECVENAQREGYDISEACRAAAEAFKRVAEAAKRAGITSSEVLELAIRLIKECV ENAQREGYDISEACRAAAEAFKRVAEAAKRAGITSSETLKRAIEEIRKRVEEAQ REGNDISEACRQAAEEFRKKAEELKRRGWLEHHHHHH
HR05_asym62A	MGKEEKAELLLRLMKICVENAKRKGDDTREAAIAAVAAFMLVREAAERAGID SSEVLELAIRLIKECVENAQREGYDISEACRAAAEAFKRVAEAAKRAGITSSEVL ELAIRLIKECVENAQREGYDISEACRAAAEAFKRVAEAAKRAGITSSETLKRAIE EIRKRVEEAQREGNDISEACRQAAEEFRKKAEELKRRGWGS
HR05_asym62B	MKEEIAELLVRLVEICVENAKRKGDDTEEAAKAAILAMMEVREAAERAGIDSS EVLELAIRLIKECVENAQREGYDISEACRAAAEAFKRVAEAAKRAGITSSEVLE LAIRLIKECVENAQREGYDISEACRAAAEAFKRVAEAAKRAGITSSEVLELAIRL IKECVENAQREGYDISEACRAAAEAFKRVAEAAKRAGITSSEVLELAIRLIKECV ENAQREGYDISEACRAAAEAFKRVAEAAKRAGITSSETLKRAIEEIRKRVEEAQ REGNDISEACRQAAEEFRKKAEELKRRGWLEHHHHHH
HR07_asym05A	MGSKDLARLTCIRAALEAAESNDEEVAKQAAKDCLEEAKQAGMPTKEAARSF CEAAARAAAESNDEEVAKIAAKACLEVAKQAGMPTKEAARSFCEAAARAAAE SNDEEVAKIAAKACLEVAKQAGMPTKEAARSFCEAAKRAAKESNDEEVEKIA KKACKEVAKQAGMPGWGS
HR07_asym05B	MTKELARLVCIEAAARAAESNDEEVAKQAAKDCLEVAKQAGMPTKEAARSFC EAAARAAAKSNDEEVAKIAAKACLEVAKQAGMPTKEAARSFCEAAARAAAES

	NDEEVAKIAAKACLEVAKQAGMPTKEAARSFCEAAARAAAESNDEEVAKIAA KACLEVAKQAGMPTKEAARSFCEAAARAAAESNDEEVAKIAAKACLEVAKQ AGMPTKEAARSFCEAAKRAAKESNDEEVEKIAKKACKEVAKQAGMPGWLEH HHHHH
HR08_asym01A	MGDEMVEVIIALLRAVILAIKNKDKEVAKEILIALAEIMEALEENNSDEMAKVM LALAKAVLLAAKNNDDEVAREIARAAAEIVEALRENNSDEMAKVMLALAKA VLLAAKNNDDEVAREIARAAAEIVEALRENNSDEMAKKMLELAKRVLDAAK NNDDETAREIARQAAEEVEAGWGS
HR08_asym01B	MDEMVKVIIALLKAVILALENDDEEVAREILLALMEIMEALKENNSDEMAKVM LALAKAVLLAAKNNDDEVAREIARAAAEIVEALRENNSDEMAKVMLALAKA VLLAAKNNDDEVAREIARAAAEIVEALRENNSDEMAKVMLALAKAVLLAAK NNDDEVAREIARAAAEIVEALRENNSDEMAKVMLALAKAVLLAAKNNDDEV AREIARAAAEIVEALRENNSDEMAKKMLELAKRVLDAAKNNDDETAREIARQ AAEEVEAGWLEHHHHHH
HR08_asym05A	MGWGDEMKKVMEALKKAVELAKKNNDDEVAREIERAAKEIVEALRENNSDE MAKVMLALAKAVLLAAKNNKDIVARIIAIAAALIVQALRENNSDEMAKVMLA LAKAVLLAAKNNDDMVAMAIAIAAMVIVLALMKNNSDEMAKKMLELAKRV LDAAKNNDVRTALLIAVQALLEVMAGS
HR08_asym05B	MHHHHHHWGDEMKKVMEALKKAVELAKKNNDDEVAREIERAAKEIVEALR ENNSDEMAKVMLALAKAVLLAAKNNDDEVAREIERAAKEIVEALRENNSDEM AKVMLALAKAVLLAAKNNDDEVAREIERAAKEIVEALRENNSDEMAKVMLA LAKAVLLAAKNNDDEVARVIAAAAAMIVIALRENNSDEMAKVMLALAKAVL LAAKNNDDLVALAIALAAMVIVMALLENNSDEMAKKMLELAKRVLDAAKNN DEETAMIIAAQALAEVMA
HR08_asym09A	MGDEMKLVMMALKLMVRIAKKDNDDEVAREIERAAKEIVEALRENNSDEMA KVMLALAKAVLLAAENNDDEVAREIARAAAEIVEALRENNSDEMAKVMLAL AKAVLLAAKNNDDEVAREIARAAAEIVEALRENNSDEMAKKMLELAKRVLD AAKNNDDETAREIARQAAEEVEAGWGS
HR08_asym09B	MDEMKLVMMMLKLMVRLAKKDNDDEVAREIERAAKEIVEALRENNSDEMA KVMLALAKAVLLAAENNDDEVAREIARAAAEIVEALRENNSDEMAKVMLAL AKAVLLAAKNNDDEVAREIARAAAEIVEALRENNSDEMAKVMLALAKAVLL AAKNNDDEVAREIARAAAEIVEALRENNSDEMAKVMLALAKAVLLAAKNND DEVAREIARAAAEIVEALRENNSDEMAKKMLELAKRVLDAAKNNDDETAREI ARQAAEEVEAGWLEHHHHHH
HR08_asym14A	MGDEMKKVMEALEKAVELARKNNDDVVAMMIMIAAMMIRLALKKNNSDEM AKVMLALAKAVLLAAKNNDDEVAMIIALAAAMIVAALLLNNSDEMAKVMLA LAKAVLLAAKNNDDEVAREIARAAAEIVEALRENNSDEMAKKMLELAKRVLD AAKNNDDETAREIARQAAEEVEAGWGS
HR08_asym14B	MDEMKKVMEALRKAVELAKKNNDDMVAMVIMIAALLIILALRENNSDEMAK VMLALAKAVLLAAKNNDDEVAILIALVAAMIVLALLENNSDEMAKVMLALA KAVLLAAKNNDDEVAREIAKAAAEIVIALMKNNSDEMAKKMLELAKRVLDA AKNNDDETAREIARQAAEEVEKNNSDEMAKKMLELAKRVLDAAKNNDDETA REIARQAAEEVEKNNSDEMAKKMLELAKRVLDAAKNNDDETAREIARQAAEE VEAGWLEHHHHHH

HR14_asym09A	MGDSEEVNERVKQLAEKAKEATDKEEVIEIVKELAELALESTDSELVNEIVKQL AEVAKEATDKELVIYIVIILAILALLSTDSELVNEIVKQLAEVAKEATDKQLVIII VLILALLALLSTDSELVNEIVKQLEEVAKEATDKLLVELIELILEILKRVSGS
HR14_asym09B	MHHHHHHWGDSEEVNERVKQLAEKAKEATDKEEVIEIVKELAELAMKSTDSE LVNELVKQLAEVAKEATDKELVIYIVKILAELAKQSTDSELVNELVKQLAEVA KEATDKELVIYIVKILAELAKQSTDSELVNEIVKQLAEVAKEATDKELVIYIVLI LAILAIMSTDSELVNEIVKQLAEVAKEATDKLLVVLIVIILALLAILSTDSELVNEI VKQLEEVAKEATDKLLVEMIRAILEILKRIS
HR14_asym25A	MGDSEEVNERVKQLAEKALRATDKEEVIEIVKELAELAKQSTDSELVIEIVKQL VMVALVATDKELVIYIVKILAELAKQSTDSLLVIVIVLALAIVAMIATDKELVIY IVKILAELAKQSTDSLLVNLIVMMLTAVARIATDKELVEHIEKILEELKKQSGS
HR14_asym25B	MHHHHHHWGDSEEVNERVKQLAEKALEATDKEEVIEIVKELAELAKQSTDSE LVNEIVKQLAEVAKEATDKELVIYIVKILAELAKQSTDSELVNEIVKQLAEVAK EATDKELVIYIVKILAELAKQSTDSELVLEIVIQLVIVAMLATDKELVIYIVKILA ELAKQSTDSKLVLLIVLALALVAILATDKELVIYIVKILAELAKQSTDSLLVIMIV MMLIAVALIATDKELVEHIEKILEELKKQS
HR14_asym52A	MGDSEEVNRLVEILARLAKEADDKSVVIEIVEILAKLAKQSTDSELVNEIVKQL AEVAKEATDKELVIYIVKILAELAKQSTDSELVNEIVKQLAEVAKEATDKELVI YIVKILAELAKQSTDSELVNEIVKQLEEVAKEATDKELVEHIEKILEELKKQSG WGS
HR14_asym52B	MDSEDRNRLVKLLAELAKMARDKEVVIRIVKILAKLAKQSTDSELVNEIVKQL AEVAKEATDKELVIYIVKILAELAKQSTDSELVNEIVKQLAEVAKEATDKELVI YIVKILAELAKQSTDSELVNEIVKQLAEVAKEATDKELVIYIVKILAELAKQSTD SELVNEIVKQLAEVAKEATDKELVIYIVKILAELAKQSTDSELVNEIVKQLEEVA KEATDKELVEHIEKILEELKKQSGWLEHHHHHH
HR18_asym08A	MGWGDIEKLCKKAESEAREARSKAEELRQRHPDSQAARDAQKLASQAEEAVK LACELAQEHPNADIAKLCIRAASEAAEAASKAAELAQRHPDSQAARDAIKLAS QAAEAVKLACELAQEHPNADIAKLCILAASEAAEAASKAAELAQRHPDSQAAR DAIKLASQAAEAVKLACELAQEHPNAEIAKICILLAIIAALLAIIAAMEAQDHPD SQRAREEIKEASQKAEEVKERCERGS
HR18_asym08B	MHHHHHHWGDIEKLCKKAESEAREARSKAEELRQRHPDSQAARDAIKLASQA AEAVKLACELAQEHPNADIAKLCIKAASEAAEAASKAAELAQRHPDSQAARD AIKLASQAAEAVKLACELAQEHPNADIAKLCIKAASEAAEAASKAAELAQRHP DSQAARDAIKLASQAAEAVKLACELAQEHPNADIAKLCIKAASEAAEAASKAA ELAQRHPDSQAARDAIKLASQAAEAVKLACELAQEHPNADIAKLCIKAASEAA QAASKAAELAQRHPDSQAARDAIKLASQAAEAVKLACELAQEHPNASIAILCIL LAIIAALLAVIAAMEAQKHPDSQKAREEIKLASQLAEEVKERCER
HR18_asym12A	**MGWGDIEKLCKKAESEAREARSKAEELRQRHPDSQAARDAQKLASQAEE AVKLACELAQEHPNADIAKLCIKAASEAAEAASKAAELAQRHPDSQAARD AIKLASQAAEAVKLACELAQEHPNADIAKLCIKAASEAAEAASKAAELAQR HPDSQAARDAIKLASQAAEAVKLACELAQEHPNAEIAKLCISAASVAALVA SIAAEMAQRHPDSQIARLIIKAASQIAEIVKRICERGS**
HR18_asym12B	**MHHHHHHWGDIEKLCKKAESEAREARSKAEELRQRHPDSQAARDAQKL ASQAEEAVKLACELAQEHPNADIAKLCIKAASEAAEAASKAAELAQRHPDS**

(SEC, SAXS, LC-MS)	**QAARDAQKLASQAEEAVKLACELAQEHPNADIAKLCIKAASEAAEAASKA AELAQRHPDSQAARDAQKLASQAEEAVKLACELAQEHPNADIAKLCIKAA SEAAEAASKAAELAQRHPDSQAARDAIKLASQAAEAVKLACELAQEHPNA DIAKLCIKAASEAAEAASKAAELAQRHPDSQAARDAIKLASQAAEAVKLAC ELAQEHPNAEIAKICILAASAAALLASAAAMMAQRHPDSQMARKLIIAASR MAELVKRTCER**
HR18_asym13A	**MGWGDIEKLCKKAESEAREARSKAEELRQRHPDSQAARDAQKLASQAEE AVKLACELAQEHPNADIAKLCIKAASEAAEAASKAAELAQRHPDSQAARD AIKLASQAAEAVKLACELAQEHPNADIAKLCIKAASEAAEAASKAAELAQR HPDSQAARDAIKLASQAAEAVKLACELAQEHPNAEIAKMCILAASAAALM ASIAAILAQRHPDSQIARDLIRLASELAEMVKRMCERGS**
HR18_asym13B	**MHHHHHHWGDIEKLCKKAESEAREARSKAEELRQRHPDSQAARDAQKL ASQAEEAVKLACELAQEHPNADIAKLCIKAASEAAEAASKAAELAQRHPDS QAARDAIKLASQAAEAVKLACELAQEHPNADIAKLCIKAASEAAEAASKAA ELAQRHPDSQAARDAIKLASQAAEAVKLACELAQEHPNADIAKLCIKAASE AAEAASKAAELAQRHPDSQAARDAIKLASQAAEAVKLACELAQEHPNADI AKLCIKAASEAAEAASKAAELAQRHPDSQAARDAIKLASQAAEAVKLACE LAQEHPNADKAKLCILLASAAALLASIAAMLAQRHPDSQEARDMIRIASEL**
(SEC, SAXS, LC-MS)	**AELVKEICER**
HR49_asym03A	MGWGSEEEQERIRRILKEARKSGTEESLRQAIEDVAQLAKKSQDSEVLEEAIRVI LRIAKESGSEEALMLAILAVALIAMNAQDSEVLEEAIRVILRIAKESGSEEALML AMLAVMIIAKVAQDPRVLEEAIRVIRQIAEESGSEAARKAAELIEKAIRDGS
HR49_asym03B	MHHHHHHWGSEEEQERIRRILKEARKSGTEESLRQAIEDVAQLAKKSQDSEVL EEAIRVILRIAKESGSEEALRQAIRAVAEIAKEAQDSEVLEEAIRVILRIAKESGSE EALRQAIRAVAEIAKEAQDSEVLEEAIRVILRIAKESGSEKALILAILAVALIAME AQDSEVLEEAIRVILRIAKESGSRMALMLALMAVLIIAIVARDPRVLEEAIRVIR QIAEESGSELARRMAELAEEAIRI
HR49_asym12A	MGSLEEAKRIIRILKEARKSGTEESLRQAIEDVAQLAKKSQDSLVLLIAILVILIIA VESGSEEALRQAIRAVAEIAKEAQDSIVLIIAILVILMIAIISGSEEALRQAIRAVA EIAKEAQDPRVLEVAIRMIRVIAELSGSEEARRQAERAEEEIRRGWGS
HR49_asym12B	MRSEEIKRIIRILKEARKSGTEESLRQAIEDVAQLAKKSQDSLVLMVAILVILLIA LESGSEEALRQAIRAVAEIAKEAQDSLVLILAILVIVMIAIISGSEEALRQAIRAVA EIAKEAQDSEVLEEAIRVILRIAKESGSEEALRQAIRAVAEIAKEAQDSEVLEEAI RVILRIAKESGSEEALRQAIRAVAEIAKEAQDPRVLEVAIMMIRLIAELSGSEEA RRQAERAEEEIRRGWLEHHHHHH
HR49_asym24A	**MGSIEEVLRIIEILARAELSGTEESLRQAIEDVAQLAKKSQDSLVLIVAIVVI MIIAKLSGSEEALRQAIRAVAEIAKEAQDSKVLEEAIRAILRIAKESGSEEAL RQAIRAVAEIAKEAQDPRVLEEAIRVIRQIAEESGSEEARRQAERAEEEIRR GWGS**
HR49_asym24B	**MSMDEMMRILRILMEAIKSGTEESLRQAIEDVAQLAKKSQDSEVLILAIAVI LIIAKLSGSEEALRQAIRAVAEIAKEAQDSEVLEEAIRLILRIAKESGSEEAL RQAIRAVAEIAKEAQDSEVLEEAIRLILRIAKESGSEEALRQAIRAVAEIAKE AQDSEVLEEAIRLILRIAKESGSEEALRQAIRAVAEIAKEAQDPRVLEEAIRV**
(SEC, SAXS, LC-MS)	**IRQIAEESGSEEARRQAERAEEEIRRGWLEHHHHHH**

HR49_asym27A	MGWGSEEEQERIRRILKEARKSGTEESLRQAIEDVAQLAKKSQDSEVLEEAIRVI LRIAKESGSEEALRQAILAVMEIALEAQDSEVLEEAIRVILRIAKESGSERALIIAII MVLLIAISAQDPRVLEEAIRVIRQIAEESGSKEARMLAQLAEEIIRRGS
HR49_asym27B	MHHHHHHWGSEEEQERIRRILKEARKSGTEESLRQAIEDVAQLAKKSQDSEVL EEAIRVILRIAKESGSEEALRQAIRAVAEIAKEAQDSEVLEEAIRVILRIAKESGSE EALRQAIRAVAEIAKEAQDSEVLEEAIRVILRIAKESGSEEALRQAILAVMEIAK EAQDSEVLEEAIRVILRIAKESGSELALMLALLAVMVIALVAQDPRVLEEAIRVI RQIAEESGSEVARIIAEIIEEAIRR
HR53_asym06A	MGNDEKEVLKMILEAAEAAARAPDPEILKAAVRAAEMVVEVRPGSNLAKKAL EIILRAAEELAKLPDPEALKEAVKAAEKVVREQPGSNLAKKALEIILRAAEELA KLPDPEALKEAVKAAEKVVREQPGSELAKKALEIIERAAEELKKSPDPEAQKEA KKAEQKVREERPGGWGS
HR53_asym06B	MNREKKMLKVILEIARALAKAPDPELLKAAVRAAEMVVRMRPGSNLAKKALE IILRAAEELAKLPDPEALKEAVKAAEKVVREQPGSNLAKKALEIILRAAEELAK LPDPEALKEAVKAAEKVVREQPGSNLAKKALEIILRAAEELAKLPDPEALKEAV KAAEKVVREQPGSNLAKKALEIILRAAEELAKLPDPEALKEAVKAAEKVVREQ PGSELAKKALEIIERAAEELKKSPDPEAQKEAKKAEQKVREERPGGWLEHHHH HH
HR53_asym35A	MGEDLKEELKLLLKEAEELAKSPDPEVLILAVLIAEVVVLMKPGSNLAKKALEI ILRAAEELAKLPDPEALKEAVKAAEKVVREQPGSNLAKKALEIILRAAEELAKL PDPEALKEAVKAAEKVVREQPGSELAKKALEIIERAAEELKKSPDPEAQKEAK KAEQKVREERPGGWGS
HR53_asym35B	MRDVKEKLKLLLKIAEKLAKSPRPEALILAVLIAEAVVRMKPGSNLAKKALEII LRAAEELAKLPDPEALKEAVKAAEKVVREQPGSNLAKKALEIILRAAEELAKLP DPEALKEAVKAAEKVVREQPGSNLAKKALEIILRAAEELAKLPDPEALKEAVK AAEKVVREQPGSNLAKKALEIILRAAEELAKLPDPEALKEAVKAAEKVVREQP GSELAKKALEIIERAAEELKKSPDPEAQKEAKKAEQKVREERPGGWLEHHHHH H
HR54_asym03A	MGTEEVRRLLEAVARAAIEAARRGDTRLVRLMLLLALLIATRSGTTEAVKLAL EVVARVAIEAARRGNTDAVREALEVALEIARESGTTEAVKLALEVVARVAIEA ARRGNTDAVREALEVALEIARESGTEEAVRLALEVVKRVSDEAKKQGNEDAV KEAEEVRKKIEEESGWGS
HR54_asym03B	MTEIERRMLEILARAAIEAARRGDTKIVRIVLQMALIIARSSGTTEAVKLALEVV ARVAIEAARRGNTDAVREALEVALEIARESGTTEAVKLALEVVARVAIEAARR GNTDAVREALEVALEIARESGTTEAVKLALEVVARVAIEAARRGNTDAVREAL EVALEIARESGTTEAVKLALEVVARVAIEAARRGNTDAVREALEVALEIARESG TEEAVRLALEVVKRVSDEAKKQGNEDAVKEAEEVRKKIEEESGWLEHHHHHH
HR54_asym24A	MGTRDELRELAEVLARAVIAAADGNTDEVREQLQRALEIARESGTTEAVKAAL IVVAMVAIMAAERGNTDAVREALEVALEIARESGTTEAVKLALEVVARVAIEA ARRGNTDAVREALEVALEIARESGTEEAVRLALEVVKRVSDEAKKQGNEDAV KEAEEVRKKIEEESGWGS
HR54_asym24B	MTEDELAELAKVLARAVVAASRGNTDEVREQLQRALEIARESGTTDAVELALI VVAMVAIEAARRGNTDAVREALEVALEIARESGTTEAVKLALEVVARVAIEAA RRGNTDAVREALEVALEIARESGTTEAVKLALEVVARVAIEAARRGNTDAVRE

	ALEVALEIARESGTTEAVKLALEVVARVAIEAARRGNTDAVREALEVALEIARE SGTEEAVRLALEVVKRVSDEAKKQGNEDAVKEAEEVRKKIEEESGWLEHHHH HH
HR64_asym08A	MGPEDELKRVEKKVKEAEELLRQAKEKGSRKDLMEAVITAAAALLLALLVLL KAEKEGDPEVALRAVELVVRVAELLLRIAKESGSEELLEMALLVAIIAAMLAK LVLELAEKQGDPEVALRAVELVVRVAELLLRIAKESGSEEALERALRVAEEAA RLAKRVLELAEKQGDPEVARRAVELVKRVAELLERIARESGSEEAKERAERVR EEARELQERVKELRERGWGS
HR64_asym08B	MPEDELKRVEKRVKEAEELLRQAKEKGSEEDLMKAVITALMAALLALIVLML AEREGDPEVALRAVELVVRVAELLLRIAKESGSEKLLEIALLVAILAAMLAKLV LELAEKQGDPEVALRAVELVVRVAELLLRIAKESGSEEALERALRVAEEAARL AKRVLELAEKQGDPEVALRAVELVVRVAELLLRIAKESGSEEALERALRVAEE AARLAKRVLELAEKQGDPEVALRAVELVVRVAELLLRIAKESGSEEALERALR VAEEAARLAKRVLELAEKQGDPEVARRAVELVKRVAELLERIARESGSEEAKE RAERVREEARELQERVKELRERGWLEHHHHHH
HR64_asym29A	MGPERALKAVELIVRAAEMLLEAAKEVGAEELLELALRVAEIAAKIAKKVLEQ AEKEGDPEVALRAVELVVRVAELLLRIAKESGSEEALERALRVAEEAARLAKR VLELAEKQGDPEVALRAVELVVRVAELLLRIAKESGSEEALERALRVAEEAAR LAKRVLELAEKQGDPEVARRAVELVKRVAELLERIARESGSEEAKERAERVRE EARELQERVKELRERGWGS
HR64_asym29B	MPERMLKAVEKIVIAAELLLRAAKEVGSEELLEVALRVAEIAAKIAKKVLEQA EKEGDPEVALRAVELVVRVAELLLRIAKESGSEEALERALRVAEEAARLAKRV LELAEKQGDPEVALRAVELVVRVAELLLRIAKESGSEEALERALRVAEEAARL AKRVLELAEKQGDPEVALRAVELVVRVAELLLRIAKESGSEEALERALRVAEE AARLAKRVLELAEKQGDPEVALRAVELVVRVAELLLRIAKESGSEEALERALR VAEEAARLAKRVLELAEKQGDPEVARRAVELVKRVAELLERIARESGSEEAKE RAERVREEARELQERVKELRERGWLEHHHHHH
HR71_asym06A	MGPEEMLEAAKLALEIARMASELGIEEIFREAAQLALEIAKRLVEQAKKEGDPE LVLEAAKVALRVAELAAKNGDKEVFKKAAESALEVAKRLVEVASKEGDPELV LEAAKVALRVAELAAKNGDKEVFKKAAESALEVAKRLVEVASKEGDPELVEE AAKVAEEVRKLAKKQGDEEVYEKARETAREVKEELKRVREEKGWGS
HR71_asym06B	MPEKILEMAKIALEIARIASEEGDEKIFREAARLALRLAKELVEQAKKEGDPELV LEAAKVALRVAELAAKNGDKEVFKKAAESALEVAKRLVEVASKEGDPELVLE AAKVALRVAELAAKNGDKEVFKKAAESALEVAKRLVEVASKEGDPELVLEAA KVALRVAELAAKNGDKEVFKKAAESALEVAKRLVEVASKEGDPELVLEAAKV ALRVAELAAKNGDKEVFKKAAESALEVAKRLVEVASKEGDPELVEEAAKVAE EVRKLAKKQGDEEVYEKARETAREVKEELKRVREEKGWLEHHHHHH
HR71_asym09A	MGPEEILIRALLSVLRAMIAILRGDEEEFRKAAEKALELAKRLVEQAKKEGDPK LVLLAALVALIVAMLAALAGDKEVFKKAAESALEVAKRLVEVASKEGDPELV LEAARVALNVAILAADMGDKEVFKKAAESALEVAKRLVEVASKEGDPELVEE AAKVAEEVRKLAKKQGDEEVYEKARETAREVKEELKRVREEKGWGS
HR71_asym09B	MPREILMEAMESLLRAALAIIRGDEEEFRKAAEKALELAKRLVEQAKKEGDPR LVLMAALVALMVAALAAMNGDKEVFKKAAESALEVAKRLVEVASKEGDPEL VLEAAKVALMVAIMAAILGDKEVFKKAAESALEVAKRLVEVASKEGDPELVL EAAKVALRVAELAAKNGDKEVFKKAAESALEVAKRLVEVASKEGDPELVLEA

	AKVALRVAELAAKNGDKEVFKKAAESALEVAKRLVEVASKEGDPELVEEAAK VAEEVRKLAKKQGDEEVYEKARETAREVKEELKRVREEKGWLEHHHHHH
HR71_asym27A	MGWGPEEILERAKESLERAREASERGDEEEFRKAAEKALELAKRLVEQAKKEG DPELVLEAAKVALRVAELAAKNGDKEVFKKAALSALEVALRLVIVAVMERDP ELVLEAAKVALRVAELAAKNGDKEVFKLAALVALMVALVLVLVAAKEGDPE LVEEAAKVAEEVRKLAKKQGDQIVYELARVIATLVKILLKEVREEKGS
HR71_asym27B	MHHHHHHWGPEEILERAKESLERAREASERGDEEEFRKAAEKALELAKRLVE QAKKEGDPELVLEAAKVALRVAELAAKNGDKEVFKKAAESALEVAKRLVEV ASKEGDPELVLEAAKVALRVAELAAKNGDKEVFKKAAESALEVAKRLVEVAS KEGDPELVLEAAKVALRVAELAAKNGDKEVFKKAAESALQVALRLVIVAMKE GDPELVLEAAKVALRVAELAAKNGDKEVFKLAAIVALLVALMLVIVAIKEGDP ELVEEAAKVAEEVRKLAKKQGDEIVYELARIIAELVKVVLKEVREEK
HR71_asym65A	MGWGPEEILERAKESLVRAMEASMRGDEEEFRKAAEKALELAKRLVEQAKKE GDPELVLEAARVALRVAELAARNGDKEVFKKAAESALEVAKRLVEVASKEGD PELVLEAAKVALQVALLAARNGDKEVFKKAAESALEVAKRLVEVASKEGDPR LVRMAAIVALLVRLLARAQGDEEVYEKARETAREVKEELKRVREEKGS
HR71_asym65B	MHHHHHHWGPEEILERAKESLERAREASERGDEEEFRKAAEKALELAKRLVE QAKKEGDPELVLEAAKVALRVAELAAKNGDKEVFKKAAESALEVAKRLVEV ASKEGDPELVLEAAKVALVVAMLAAMNGDKEVFKKAAESALEVAKRLVEVA SKEGDPRLVLMAAIVALLVALLAAANGDKEVFKKAAESALEVAKRLVEVASK EGDPELVLEAAKVALRVAELAARNGDKEVFKKAAESALEVAKRLVEVASKEG DPRLVEMAAIVALLVRLLADAQGDEEVYEKARETAREVKEELKRVREEK
HR71_asym67A	MGPMEIILRAMLSLLRAVAASMLGDEEEFRKAAEKALELAKRLVEQAKKEGD PMLVALAALIALLVAELAAKNGDKEVFKKAAESALEVAKRLVEVASKEGDPK LVLMAARVALRVAELAAKNGDKEVFKKAAESALEVAKRLVEVASKEGDPEL VEEAAKVAEEVRKLAKKQGDEEVYEKARETAREVKEELKRVREEKGWGS
HR71_asym67B	MPLEIALRAMKSIMRALMASMLGDEEEFRKAAEKALELAKRLVEQAKKEGDP NLVALAALLALLVAELAAKNGDKEVFKKAAESALEVAKRLVEVASKEGDPKL VLLAARVALRVAELAAKNGDKEVFKKAAESALEVAKRLVEVASKEGDPKLVL EAARVALRVAELAAKNGDKEVFKKAAESALEVAKRLVEVASKEGDPKLVLEA ARVALRVAELAAKNGDKEVFKKAAESALEVAKRLVEVASKEGDPELVEEAAK VAEEVRKLAKKQGDEEVYEKARETAREVKEELKRVREEKGWLEHHHHHH
HR76_asym01A	MGPELLEWIKEALEVAEEVVDVIKRAEQEGNPDLRDSARELLLAVLEAIAEAL KQGNPELVEWVARAAKVAAEVIKVAIQAEKEGNRDLFRAALELVRAVIEAIEE AVKQGNPELVEWVARAAKVAAEVIKVAIQAEKEGNRDLFRAALELVRAVIEAI EEAVKQGNPELVERVARLAKKAAELIKRAIRAEKEGNRDERREALERVREVIE RIEELVGWGS
HR76_asym01B	MPELLDWIKRAIRVAREVVKVAKRAEEEGNPDLRDSAKELLLAVVKAMAEAL RQGNPELVEWVARAAKVAAEVIKVAIQAEKEGNRDLFRAALELVRAVIEAIEE AVKQGNPELVEWVARAAKVAAEVIKVAIQAEKEGNRDLFRAALELVRAVIEAI EEAVKQGNPELVEWVARAAKVAAEVIKVAIQAEKEGNRDLFRAALELVRAVI EAIEEAVKQGNPELVEWVARAAKVAAEVIKVAIQAEKEGNRDLFRAALELVR AVIEAIEEAVKQGNPELVERVARLAKKAAELIKRAIRAEKEGNRDERREALERV REVIERIEELVGWLEHHHHHH

HR76_asym03A	MGPELEEWIRRALDVAREVVRVAKRALDEGNPKLVLSALELLVAVLEAIMEA KKQGNPELVEWVARAAKVAAEVIKVAIQAEKEGNRDLFRAALELVRAVIEAIE EAVKQGNPELVEWVARAAKVAAEVIKVAIQAEKEGNRDLFRAALELVRAVIE AIEEAVKQGNPELVERVARLAKKAAELIKRAIRAEKEGNRDERREALERVREVI ERIEELVGWGS
HR76_asym03B	MPELEEWIRRALDVLREVLEVIERALDEGNPVLALSAARLAVAVAKAIEEAKK QGNPELVEWVARAAKVAAEVIKVAIQAEKEGNRDLFRAALELVRAVIEAIEEA VKQGNPELVEWVARAAKVAAEVIKVAIQAEKEGNRDLFRAALELVRAVIEAIE EAVKQGNPELVEWVARAAKVAAEVIKVAIQAEKEGNRDLFRAALELVRAVIE AIEEAVKQGNPELVEWVARAAKVAAEVIKVAIQAEKEGNRDLFRAALELVRA VIEAIEEAVKQGNPELVERVARLAKKAAELIKRAIRAEKEGNRDERREALERVR EVIERIEELVGWLEHHHHHH
HR76_asym06A	MGWGPELEEWIRRAKEVAKEVEKVAQRAEEEGNPDLRDSAKELRRAVEEAIE EAKKQGNPRLVMAVAIAALVAAEVIKVAIQAEKEGNRDLFRAALELVRAVIEA IEEAVKQGNPRLVMAVALAALIAAIVIKVAIQAEKEGNRDLFRAALELVRAVIE AIEEAVKQGRPRLVMIVALLALMAAMLIRRAIRAEKEGNRDERREALERVREV IERIEELVGS
HR76_asym06B	MHHHHHHWGPELEEWIRRAKEVAKEVEKVAQRAEEEGNPDLRDSAKELRRA VEEAIEEAKKQGNPELVEWVARAAKVAAEVIKVAIQAEKEGNRDLFRAALEL VRAVIEAIEEAVKQGNPELVEWVARAAKVAAEVIKVAIQAEKEGNRDLFRAAL ELVRAVIEAIEEAVKQGNPELVAAVAMAAKIAAEVIKVAIQAEKEGNRDLFRA ALELVRAVIEAIEEAVKQGNPELVLVVAMAAMIAAIVIEVAIQAEKEGNRDLFR AALELVRAVIEAIEEAVKQGDPELVMLVAMLALIAAILIEREIRAEKEGNRDER REALERVREVIERIEELV
HR79_asym02A	**MGSSDEEEARELIIRAVEAARRAIEAARRTGDPRVMELAAELAMLALLAAI EVLKDPSSSDVNEALKLIVEAIEAAVRALEAAERTGDPEVRELARELVRLA VEAAEEVQRNPSSSDVNEALKLIVEAIEAAVRALEAAERTGDPEVRELARE LVRLAVEAAEEVQRNPSSEEVNEALKKIVKAIQEAVESLREAEESGDPEKR EKARERVREAVERAEEVQRDPSGWGS**
HR79_asym02B **(SEC, SAXS, LC-MS)**	**MSSDEEEAVRLVIRAAEAAIRALEAALRTGDPRVLELALELVALALEAAEE VLRDPSSSDVNEALKLIVEAIEAAVRALEAAERTGDPEVRELARELVRLAV EAAEEVQRNPSSSDVNEALKLIVEAIEAAVRALEAAERTGDPEVRELAREL VRLAVEAAEEVQRNPSSSDVNEALKLIVEAIEAAVRALEAAERTGDPEVRE LARELVRLAVEAAEEVQRNPSSSDVNEALKLIVEAIEAAVRALEAAERTGD PEVRELARELVRLAVEAAEEVQRNPSSEEVNEALKKIVKAIQEAVESLREA EESGDPEKREKARERVREAVERAEEVQRDPSGWLEHHHHHH**
HR79_asym04A	**MGRSDDEEALELIKRALEAAERAMEAAERTGDPRVEKLALELIKLAMEAA LEVIRDPSSSDVNEALKLIVEAIEAAVRALEAAERTGDPEVRELARELVRLA VEAAEEVQRNPSSSDVNEALKLIVEAIEAAVRALEAAERTGDPEVRELARE LVRLAVEAAEEVQRNPSSEEVNEALKKIVKAIQEAVESLREAEESGDPEKR EKARERVREAVERAEEVQRDPSGWGS**
HR79_asym04B	**MRSDEDEALKLILRAIEAAKRAMEAAERTGDPRVEELALELVRLAMEAAL EVIRDPSSSDVNEALKLIVEAIEAAVRALEAAERTGDPEVRELARELVRLAV EAAEEVQRNPSSSDVNEALKLIVEAIEAAVRALEAAERTGDPEVRELAREL VRLAVEAAEEVQRNPSSSDVNEALKLIVEAIEAAVRALEAAERTGDPEVRE LARELVRLAVEAAEEVQRNPSSSDVNEALKLIVEAIEAAVRALEAAERTGD**

(SEC, SAXS, LC-MS)	**PEVRELARELVRLAVEAAEEVQRNPSSEEVNEALKKIVKAIQEAVESLREA EESGDPEKREKARERVREAVERAEEVQRDPSGWLEHHHHHH**
HR79_asym05A	MGWGSSDEEEARELIERAKEAAERAQEAAERTGDPRVRELARELKRLAQEAA EEVKRDPSSSDVNEALKLIVEAIEAAVRALEAAERTGDPEVRELARELVRLAVE AAEEVQRNPSSSDVNEALKLIVEAIEAAVRALEAAERTGDPLVRMMAIALVIM AVEAAEEVQRNPSSEEVNEALKKIVKAIQEAVESLREAEESGDPEKREEALLRV AIAVARAARVLTDPSGS
HR79_asym05B	MHHHHHHWGSSDEEEARELIERAKEAAERAQEAAERTGDPRVRELARELKRL AQEAAEEVKRDPSSSDVNEALKLIVEAIEAAVRALEAAERTGDPRVRELAREL VRLAVEAAEEVQRNPSSSDVNEALKLIVEAIEAAVRALEAAERTGDPRVRELA RELVRLAVEAAEEVQRNPSSDVNEALKLIVEAIEAAVRALEAAERTGDPRVRE LARELVRLAVEAAEEVQRNPSSSDVNEALKLIVEAIEAAVRALEAAERTGDPLV RVLAIILVLLAVRAAEEVQRNPSSEEVNEALKKIVKAIQEAVESLREAEESGDPK KREKARLRVVEAVLEAAKVMKDPS
HR79_asym24A	MGSSDEEEARELIERAKEAAERAQEAAERTGDPLVVILALILKRLAQAAAREV KRDPSSSDVNEALKLIVEAIEAAVRALEAAERTGDPEVRTLAIMLVAIAVAAAI LVLANPSSSDVNEALKLIVEAIEAAVRALEAAERTGDPEVRELARELVRLAIEA AIEVMRNPSSEEVNEALKKIVKAIQEAVESLREAEESGDPEKREKARERVREAV ERAEEVQRDPSGWGS
HR79_asym24B	MSSDEEEARELIERAKEAAERAQEAAERTGDPIVIILAIILEILAQLAAEEVKRDP SSSDVNEALKLIVEAIEAAVRALEAAERTGDPEVRELAILLVAIAVAAALEVML DPSSSDVNEALKLIVEAIEAAVRALEAAERTGDPEVRELARELVRLAIEAAIEV MKNPSSSDVNEALKLIVEAIEAAVRALEAAERTGDPEVRELARELVRLAIEAAE EVMRNPSSSDVNEALKLIVEAIEAAVRALEAAERTGDPEVRELARELVRLAIEA AEEVMRNPSSEEVNEALKKIVKAIQEAVESLREAEESGDPEKREKARERVREA VERAEEVQRDPSGWLEHHHHHH
HR81_asym01A	MGELRLLSEIMELILQMAREISEEAREKGNLKMLALALILEALAVLALAIVAKQ RGNSEEAERASEKAQRVLEEARKVSEEAREQGDDEVLALALIAIALAVLALAE VACCRGNSEEAERASEKAQRVLEEARKVSEEAREQGDDEVLALALIAIALAVL ALAEVACCRGNKEEAERAYEDARRVEEEARKVKESAEEQGDSEVKRLAELAE ELAREARRHVQECRGWGS
HR81_asym01B	MELEAMSRIAELILEVARMISEEAREEGDLKILAMALIVEALAVLALAAVARER GNSEEAERASEKAQRVLEEARKVSEEAREQGDDEVLALALIAIALAVLALAEV ACCRGNSEEAERASEKAQRVLEEARKVSEEAREQGDDEVLALALIAIALAVLA LAEVACCRGNSEEAERASEKAQRVLEEARKVSEEAREQGDDEVLALALIAIAL AVLALAEVACCRGNKEEAERAYEDARRVEEEARKVKESAEEQGDSKVKELAE IAEQLAREARRHVQECRGWLEHHHHHH
HR81_asym05A	**MGELERESMEAARRLVKAIKRSIEARERGDLKELAEALIEEARAVQELAR VACERGNSEEAELASLLAEIVLKLARRVSEEAREQGDDEVLALALIAIALA VLALAEVACCRGNSEEAERASEKAQRVLEEARKVSEEAREQGDDEVLALA LIAIALAVLALAEVACCRGNKEEAERAYEDARRVEEEARKVKESAEEQGD SEVKRLAEEAEQLAREARRHVQECRGWGS**
HR81_asym05B	**MELERESLEALKRLAEAMERSLDARERGDLKELAEALIEEARAVQELARV ACERGNSEEAELASLLARIVLKIARMVSEEAREQGDDEVLALALIAIALAVL ALAEVACCRGNSEEAERASEKAQRVLEEARKVSEEAREQGDDEVLALALI**

(SEC, SAXS, LC-MS)	**AIALAVLALAEVACCRGNSEEAERASEKAQRVLEEARKVSEEAREQGDDEVLALALIAIALAVLALAEVACCRGNKEEAERAYEDARRVEEEARKVKESAEEQGDSEVKRLAEEAEQLAREARRHVQECRGWLEHHHHHH**
HR81_asym06A	MGWGELERESEEAERRLQEARKRSEEARERGDLKELAEALIEEARAVQELARVACERGNSEEAERASEKAQRVLEEARKVSQEAIEQGDDEVLALALIAIALAVLALAEVACCRGNSEEAERASEKALRVLMEALMVALLADLQGDDEVLALALIAIALAVLALAEVACCRGNKEEAERAYLMARIVEEIARKVKESAEEQGDSEVKRLAEEAEQLAREARRHVQECRGS
HR81_asym06B	MHHHHHHWGELERESEEAERRLQEARKRSEEARERGDLKELAEALIEEARAVQELARVACERGNSEEAERASEKAQRVLEEARKVSREAIEQGDDEVLALALIAIALAVLALAEVACCRGNSEEAERASEKAQRVLEEARKVSREAIEQGDDEVLALALIAIALAVLALAEVACCRGNSEEAERASRKALLVLAEALLVAILASIQGDDEVLALALIALALAVLALAEVACCRGNKEEAELAYLLARAIEEIARIVKESAEEQGDSEVKRLAEEAEQLAREARRHVQECR
HR81_asym09A	**MGELERESEEAERRLQEAIKRSMEALERGDLKELAEALIEEARAVQELARVACERGNSEEAERASILALIVLAAARVVSALAEEQGDDEVLALALIAIALAVLALAEVACCRGNSEEAERASEKAQRVLEEARKVSEEAREQGDDEVLALALIAIALAVLALAEVACCRGNKEEAERAYEDARRVEEEARKVKESAEEQGDSEVKRLAEEAEQLAREARRHVQECRGWGS**
HR81_asym09B **(SEC, SAXS, LC-MS)**	**MELERESEEAERRLLEAIKRSVEALERGDLKELAEALIEEARAVQELARVACERGNSREAELASLLALLVLLIAILVSLDARKQGDDEVLALALIAIALAVLALAEVACCRGNSEEAERASEKAQRVLEEARKVSEEAREQGDDEVLALALIAIALAVLALAEVACCRGNSEEAERASEKAQRVLEEARKVSEEAREQGDDEVLALALIAIALAVLALAEVACCRGNKEEAERAYEDARRVEEEARKVKESAEEQGDSEVKRLAEEAEQLAREARRHVQECRGWLEHHHHHH**
HR00_asym07A	MGWGTEEKIAKEISRIAEESKKRIEELARKADNKTTETEVDKAIEKIAKLAREAIKRIEDLAKNLASEEFMARAISAIAELAKKAIEAIYRLADNHTTDTFMAKAIEAIAELAKEAIKAIADLAKNHTTEEFMARAISAIAELAKKAIEAIYRLADNHTTDTFMAKAIEAIAELAKEAIKAIADLAKNHTTEEFMAKAISAIAELAKKAIEAIYRLADNHTSPKYIMRAILAIVMIAVTAIIAIIDLAKNITTTEYIRKAISAIMEILEKALEAIKRLEDNMTGS
HR00_asym07B	MHHHHHHWGTEEKIAKEISRIAEESKKRIEELARKADNKTTETEVDKAIEKIAKLAREAIKRIEDLAKNLASEEFMARAISAIAELAKKAIEAIYRLADNHTTDTFMAKAIEAIAELAKEAIKAIADLAKNHTTEEFMARAISAIAELAKKAIEAIYRLADNHTTDTFMAKAIEAIAELAKEAIKAIADLAKNHTTEEFMAKAISAIAELAKKAIEAIYRLADNHTTDTFMAKAIEAIAELAKEAIKAIADLAKNHTTEEFMAKAISAIAELAKKAIEAIYRLADNHTSPMYILEALLAIAMIAATALAAIVNLAENITTDEYIRKAISALMEIIEKAIEAMRRLEDNLT
HR00_asym21A	MGTEEKIAKRISKLAEAAKKMIEMIARLADNITEERLVDMAIEMIAEIAKIIIKIIEELAKNLASEEFMARAISAIAELAKKAIEAIYRLADNHTTDTFMAKAIEAIAELAKEAIKAIADLAKNHTTEEFMARAISAIAELAKKAIEAIYRLADNHTTDTFMAKAIEAIAELAKEAIKAIADLAKNHTTEEFMAKAISAIAELAKKAIEAIYRLADNHTSPTYIEKAIEAIEKIARTAIKAIEDLAKNITTEEYKEKAKSAIDEIREKAKEAIKRLEDNRTGWGS

HR00_asym21B	MTEEKIAKKISEAAEIAKKVIEILAEVADEIRTEELVDRAIEIIAKMARMAIKLIE DLAKNLASEEFMARAISAIAELAKKAIEAIYRLADNHTTDTFMAKAIEAIAELA KEAIKAIADLAKNHTTEEFMARAISAIAELAKKAIEAIYRLADNHTTDTFMAKA IEAIAELAKEAIKAIADLAKNHTTEEFMAKAISAIAELAKKAIEAIYRLADNHTSP TYIEKAIEAIEKIARTAIKAIEDLAKNHTTEEFMAKAISAIAELAKKAIEAIYRLA DNHTSPTYIEKAIEAIEKIARTAIKAIEDLAKNHTTEEFMAKAISAIAELAKKAIE AIYRLADNHTSPTYIEKAIEAIEKIARTAIKAIEDLAKNITTEEYKEKAKSAIDEIR EKAKEAIKRLEDNRTGWLEHHHHHH
HR00_asym51A	MGTEEKIAKLISELAEMSKKLIEMIARLADEITEENLVDIAIELIAKIARMAIKAIE LLAKNLASEEFMARAISAIAELAKKAIEAIYRLADNHTTDTFMAKAIEAIAELA KEAIKAIADLAKNHTTEEFMARAISAIAELAKKAIEAIYRLADNHTTDTFMAKA IEAIAELAKEAIKAIADLAKNHTTEEFMAKAISAIAELAKKAIEAIYRLADNHTSP TYIEKAIEAIEKIARTAIKAIEDLAKNITTEEYKEKAKSAIDEIREKAKEAIKRLED NRTGWGS
HR00_asym51B	MTEEKIAKIISELAEMSKEMIELAARAADRLTREDMVDRIIELIAALARLAILIIEL LAKNLASEEFMARAISAIAELAKKAIEAIYRLADNHTTDTFMAKAIEAIAELAK EAIKAIADLAKNHTTEEFMARAISAIAELAKKAIEAIYRLADNHTTDTFMAKAIE AIAELAKEAIKAIADLAKNHTTEEFMAKAISAIAELAKKAIEAIYRLADNHTSPT YIEKAIEAIEKIARTAIKAIEDLAKNHTTEEFMAKAISAIAELAKKAIEAIYRLAD NHTSPTYIEKAIEAIEKIARTAIKAIEDLAKNHTTEEFMAKAISAIAELAKKAIEAI YRLADNHTSPTYIEKAIEAIEKIARTAIKAIEDLAKNITTEEYKEKAKSAIDEIREK AKEAIKRLEDNRTGWLEHHHHHH

Appendix Table 1.2. Summary of experimental and computed molecular weights used to assess the oligomeric configuration of the designed homooligomers in solution. MW design refers to the total expected molecular weight for designed oligomer, MW mon is the molecular weight of the protomer, MW MALS is the experimentally determined molecular weight by multi-angle light scattering. The designs in bold have the correct oligomerization state and exhibit discrepancies between the experimental and computational quantities <= 13% for the molecular weight.

* Elution volume reported for a Superdex 75 10/300 GL gel filtration column.

Design	MW design (kDa)	MW mon (kDa)	MW MALS (kDa)	Oligomerization State	Elution Volume (mL)
ank1C2_1	**36.2**	**18.1**	**31.5**	**1.7**	**16.0**
ank1C2_2	35.8	17.9	25.5	1.4	11.9*
ank1C3	53.4	17.8	28.2	1.6	17.7
ank1C4_1	71.6	17.9	38.8 179.3	2.2 10.0	14.1 11.9
ank1C4_2	**71.6**	**17.9**	**68.9**	**3.8**	**14.1**
ank3C2_1	**35.6**	**17.8**	**32.0**	**2.0**	**11.2**

ank3C2_2	**35.8**	**17.9**	**30.1**	**1.7**	**16.1**
ank3C4_1	**35.8**	**17.9**	**75.0** **16.0**	**4.2** **0.9**	**12.0** **17.2**
ank3C4_2	35.4	17.7	55.0	3.1	14.7
ank4C4	**72.4**	**18.1**	**72.4**	**4.0**	**14.1**
1na0C3_1	**44.4**	**14.8**	**46.1**	**3.1**	**15.2**
1na0C3_3	**44.1**	**14.7**	**45.6**	**3.0**	**15.3**
1na0C3_5	**44.2**	**14.7**	**44.0**	**3.0**	**15.3**
1na0C3_7	**44.1**	**14.7**	**46.5**	**3.2**	**14.6**
1na0C4_1	**58.0**	**14.5**	**55.1**	**3.8**	**13.3**
tpr1C3_2	48.6	16.2	16.3	1.0	16.7
tpr1C3_3	48.9	16.3	57.4	3.5	11.0*
tpr1C3_4	48.6	16.2	30	1.9	10.0*
tpr1C4_2	**64.8**	**16.2**	**65.5**	**4.0**	**14.4**
3ltjC3	65.4	21.8	237.7	10.9	12.2
3ltjC5	103.9	20.8	----	----	----
KP16C6	176	29.3	270	92.2	8.0
KP17C3_4	90.10	30.3	87.0	2.9	13.0
K64sC3	105.9	35.3	38.2	1.1	14.0
HL1C5	90.5	18.1	66.0	3.6	14.4
HR00C3_2	**93.6**	**31.2**	**81.0**	**2.6**	**13.9**
HR00C6	190.2	31.7	237.0	7.5	12.2
HR04C3_3	66.9	22.3	108.4	4.9	13.4
HR04C4_1	**90.4**	**22.6**	**88.2**	**3.9**	**13.5**
HR04C4_3	91.2	22.8	43.3	1.9	15.3
HR04C4_4	90.8	22.7	119.5	5.3	13.5

HR04C6_3	140.4	23.4	112.8	4.8	13.5
HR08C3	**60.9**	**20.3**	**56.5**	**2.8**	**15.0**
HR10C2_1	43.4	21.7	34.2	1.6	15.9
HR10C2_2	**44.9**	**22.4**	**42.6** **68.2**	**1.9** **3.1**	**15.6** **14.2**
HR10C2_3	**43.0**	**21.5**	**42.8**	**2.0**	**15.5**
HR10C3_2	64.5	21.5	44.2	2.1	15.5
HR10C5_1	113.7	22.7	370.0	165.9	8.1
HR10C5_2	**113.4**	**22.7**	**120.4**	**1.1**	**13.5**
HR14C6	116.4	19.4	53.8	2.8	15.0
HR54C6	**119.4**	**19.9**	**117.1**	**5.9**	**13.0**
HR71C6	139.2	23.2	111.9	4.8	13.3
HR79C2	**53.4**	**26.7**	**47.9**	**1.8**	**15.2**
HR81C2	**51.8**	**25.9**	**46.8**	**1.8**	**15.2**
ank1C2_1_r5	**50.7**	**25.4**	**48.0**	**1.9**	**15.7**
HR04C4_1_r4	**134.7**	**33.4**	**120.9**	**3.6**	**13.0**

Appendix Table 1.3. Summary of experimental and computed quantities used to assess the supramolecular configuration of the designed homooligomers in solution using SAXS and their closest homolog in the set determined by global multiple sequence alignment. The designs in this table are considered successful and exhibit discrepancies between the experimental and computational quantities <13% for the molecular weight, <11 % for the radius of gyration and a χ value <= 3.1.

Design	Rg model (Å)	Rg SAXS (Å)	χ SAXS	r.m.s.d. Crystal Struct.	Closest Homolog (Sequence Identity % / r.m.s.d.)
ank3C2_1	21.9	20.2	1.9	1.0 Å	ank1C2_1 (79 / 5.3 Å)
ank1C2_1	25	24.0	2.7	0.9 Å	ank3C2_1 (79 / 5.3 Å)
HR10C2_2	24.2	25.7	2.1	----	HR10C5_2 (71 / 10.4 Å)

HR79C2	22.5	23.4	2.4	----	HR81C2 (26 / 18.1 Å)
HR81C2	20.8	23.1	2.5	----	HR79C2 (26 / 18.1 Å)
1na0C3_1	27.9	28.0	1.2	----	1na0C4_1 (84 / 7.9 Å)
1na0C3_3	29.3	29.0	1.0	1.0 Å	1na0C3_7 (88 / 1.9 Å)
1na0C3_5	29.9	29.5	2.6	----	1na0C3_7 (84 / 2.5 Å)
1na0C3_7	26.2	25.4	1.0	----	1na0C3_3 (88 / 1.9 Å)
HR00C3_2	31.5	35.0	3.1	0.9 Å	HR10C2_2 (26 / 21.6 Å)
HR08C3	24.2	27.0	2.1	----	HR79C2 (20 / 21.9 Å)
1na0C4_1	31.2	30.7	2.3	-----	1na0C3_1 (84 / 7.9 Å)
ank1C4_2	32.2	29.2	2.9	1.1 Å	ank1C2_1 (81 / 13.0 Å)
HR04C4_1	33.2	36.0	2.8	-----	HR81C2 (18 / 17.8 Å)
HR10C5_2	38.9	43.3	3.1	-----	HR10C2_2 (71 / 10.4 Å)
ank1C2_1_r5	33.2	35.8	2.9	-----	-----
HR04C4_1_r4	40.9	41.9	1.0	-----	-----

Appendix Table 1.4. Summary of experimental and computed quantities used to assess the supramolecular configuration of the designed homooligomers in solution using SAXS. The designs in this table exhibit large discrepancies between the experimental and computational values for molecular weight and/or X-ray scattering profiles.

Design	Rg model (Å)	Rg SAXS (Å)	χ SAXS	r.m.s.d. xtal
ank3C2_2	24.9	21.6	5.1	------
HR10C2_1	20.14	23.0	2.7	------
HR10C2_3	24.5	30.0	22.5	------
HR10C3_2	24.5	29.1	6.6	------
ank3C4_1	31.1	31.1	23.0	------
ank4C4	32.2	36.4	3.8	> 10.0 Å
HR04C4_3	35.6	28.4	26.1	------

HR04C4_4	35.5	42.5	29.1	------
tpr1C4_2	21	23.1	8.8	------
HR00C6	50.2	46.1	2.2	------
HR54C6	33.4	48.9	15.6	------

Appendix Table 1.5. Data collection and refinement statistics for ank3C2_1, 1na0C3_3, and ank1C4_2. Statistics for the highest-resolution shell are shown in parentheses.

	ank3C2_1 (PDB ID 5HRY)	**1na0C3_3 (PDB ID 5HRZ)**	**ank1C4_2 (PDB ID 5HS0)**
Data Acquisition			
Space group	$P2_12_12_1$	R32	$P6_522$
Cell dimensions a, b, c (Å) α , β , γ (°)	106.3, 106.2,106.6 90.0, 90.0, 90.0	83.6, 83.6,141.9 90.0, 90.0, 120.0	110.5, 110.5,182.8 90.0, 90.0, 120.0
Resolution (Å)	74.24 – 2.00 (2.07-2.00)	64.48 – 2.15 (2.22-2.15)	84.75 – 2.40 (2.49-2.40)
R_{merge}	6.2 (46.0)	6.7 (67.6)	7.5 (67.7)
$CC_{1/2}$	0.999 (0.867)	0.999 (0.875)	0.999 (0.885)
$\langle I/\sigma_I \rangle$	11.9 (2.2)	20.9 (3.3)	17.6 (3.0)
Completeness (%)	94.0 (95.0)	99.7 (97.8)	99.8 (100)
Multiplicity	3.9 (3.7)	9.6 (9.2)	9.4 (10.0)
Wilson B-factor ($Å^2$)	34.1	45.9	54.1
Refinement			
Resolution range (Å)	75.24 - 2.00	64.48 - 2.15	84.75-2.40
No. of reflections	77065	10640	264243
R_{work}(%) / R_{free}(%)	0.21 / 0.25	0.18 / 0.21	0.18 / 0.21
Average B-factors ($Å^2$) Protein Water	53.8 53.8 28.2	48.3 48.2 55.9	66.5 66.1 54.5
r.m.s.d. deviations			

Bond length (Å) Bond angles (°)	0.010 1.15	0.010 0.94	0.004 0.76
Ramachandran favored (%)	99.7	99.2	98.1
Ramachandran outliers (%)	0.0	0.0	0.0

Appendix Table 1.6. Data collection and refinement statistics for ank1C2_1, HR00C3_2 and ank4C4. Statistics for the highest-resolution shell are shown in parentheses.

	ank1C2_1 (PDB ID 5KBA)	**HR00C3_2 (PDB ID 5K7V)**	**ank4C4 (PDB ID 5KWD)**
Data Acquisition			
Space group	C 1 2 1	P 3 2 1	$P2_12_12_1$
Cell dimensions a, b, c (Å) α , β , γ (°)	 93.37, 48.82, 139.27 90.0, 90.0, 98.9	 159.00,159.00,94.98 90.0, 90.0, 120.0	 77.7, 89.47, 99.80 90.0, 90.0, 90.0
Resolution (Å)	19.91 - 2.601 (2.694 - 2.601)	45.9 - 3.166 (3.28 - 3.166)	66.4-2.38
R_{merge}	0.139 (0.5507)	0.1486 (0.8645)	0.141 (0.643)
$CC_{1/2}$	0.994 (0.846)	0.997 (0.583)	.886(0.532)
$\langle I/\sigma_I \rangle$	14.52 (2.61)	13.32 (2.41)	3.0 (3.0)
Completeness (%)	98 (97)	98 (85)	97 (92)
Multiplicity	7.2 (5.9)	9.8 (7.0)	2.0 (2.0)
Wilson B-factor ($Å^2$)	41.38		38.2
Refinement			
Resolution range (Å)	19.90 - 2.60	47.49 - 3.16	66.40-2.75
No. of reflections	18970 (1564)	23539 (1994)	19498 (1248)
R_{work}(%)/ R_{free}(%)	0.20 / 0.24	0.20 / 0.22	0.23/0.28
Average B-factors ($Å^2$) Protein Water	59.51 59.96 36.93	80.67 80.89 46.96	50.5 50.6 40.0
r.m.s.d. deviations Bond length (Å) Bond angles (°)	 0.003 0.738	 0.003 0.702	 0.003 0.52

Ramachandran favored (%)	95.4	99.5	97.1
Ramachandran outliers (%)	0.5	0.2	0.3

Appendix Table 2.1. List of all designed homotrimers and pre-validated components tested with their corresponding amino acid sequences including initiating methionine and $(His)_6$ tag. Designs that expressed solubly are denoted in bold and under their name are the experimental methods used for characterization.

*Components from previously described designed homooligomers[11] or the Protein Data Bank (PDB IDs).

Design	Sequence
1na0C3_1 (SEC-MALS)	**MNIAEAAYRVGNKAYKKGRYELAILAYILAILLDPNNAEAWYNLGNAYYK EGEYDEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYDEAIEYYQKALE LDPNNAEAKQNLGNAKQKQGLEHHHHHH**
1na0C3_2 (SEC-MALS, SAXS)	**MEEAELAYLLGELAYKLGEYRIAIRAYRIALKRDPNNAEAWYNLGNAYYK QGDYDEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYDEAIEYYQKALE LDPNNAEAKQNLGNAKQKQGLEHHHHHH**
2fo7C3_1 (SEC-MALS)	**MAERLYKLGNKAYKRGEYILALIAYVVALRDDPRSAEAWYNLGNAAYKSG EYDEAIEYYQKALELDPRSAEAWYNLGNAYYKQGDYDEAIEYYQKALELD PRSAEAWYNLGNAYYKQGDYDEAIEYYQKALELDPRSLEHHHHHH**
2fo7C3_12 (SEC-MALS)	**MAEKAYNIGNAAYKEGEYRVAILAYMLALLADPRSAEALYNLGNAAYKEG DYKVAIAAYLLALDLDPRSAEAWYNLGNAYYKQGDYDEAIEYYQKALELD PRSAEAWYNLGNAYYKQGDYDEAIEYYQKALELDPRSLEHHHHHH**
2fo7C3_15 (SEC-MALS)	**MALRWLLLGILAMLLGAEELAIEAYQKALELEPRSAMAWLALGAAYYKE GDYDEAIEYYQKALELDPRSAAAWALLGNAYYKQGDYDEAIEYYQKALEN RPRSARAWYNLGNAYYKQGDYDEAIEYYQKALELDPRSLEHHHHHH**
3ltjC3_1 (SEC-MALS)	**MTDPLAVILYIAILKAEKSIARAKAAEALGKIGDERAVEPLIKALKDEDALV RAAAADALGQIGDERAVEPLIKALKDEEGLVRASAAIALGQIGDERAVEPLI KALKDERDLVRVAAAVALGRIGDERAVEPLIKALKDEEGEVREAAAIALGS IGGERVRAAMEKLAETGTGFARKVAVNYLETHKLEHHHHHH**
3ltjC3_1v2 (SEC-MALS, SAXS)	**MTDPMKVILYIAMLELEKYIMRAAAAYALGKIGDERAVEPLIKALKDEDAI VRAAAADALGQIGDERAVEPLIKALKDEDGAVRVSAAVALGQIGDERAVEP LIKALKDEDAVVRVAAAIALGLIGDERAVEPLIKALKDEKGKVREAAALAL GAIGGERVRAAMEKLAETGTGFARKVAVNYLETHKLEHHHHHH**
3ltjC3_11 (SEC-MALS, SAXS)	**MRREETDPLAVVMYRLNLRDDSYYVRRAAAYALGKIGDERAVEPLIKALK DEDAWVRRAAADALGQIGDERAVEPLIKALKDEDGWVRQSAAVALGQIGD ERAVEPLIKALKDEDWFVRAAAAAALGRIGDERAVEPLIKALKDEDEMVRE IAALALGMIGGERVRAAMEKLAETGTGFARKVAVNYLETHKLEHHHHHH**
DHR4C3_1 (SEC-MALS)	**MDICELEARLVALLVLLAKRAGADEDLIAELVAVMIMIVILRLKKSGSSYEV ICECVARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRALKRSGSSYEVICECV ARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESGSSYEVIKECVQRIVE EIVEALKRSGTSEDEINEIVRRVKSEVERTLKESGSLEHHHHHH**

DHR4C3_5 (SEC-MALS)	**MDECEEKARRVAEKVERLKRSGTSEDEIAEEVAREISEVIRTLKESGSEYKV ICRCVARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESGSKYKIICICV AIIVAEIVAALKRSGTSEDEIAEIVARVISEVIRTLKESGSSYEVIKQCVQAIVQ AIILALMKSGTEVEEILLIVLRVEEEVERTLKESGSLEHHHHHH**
DHR4C3_5v2 (SEC-MALS, SAXS)	**MDECEEKARRVAEKVERLKRSGTSEDEIAEEVAREISEVIRTLKESGSEYKV ICRCVARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESGSDYLIICVCV AIIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESGSSYEVIKECVQIIVL AIILALMKSGTEVEEILLILLRVKTEVRRTLKESGSLEHHHHHH**
DHR4C3_7 (SEC-MALS)	**MDECEKKARLVAILVIVAKALGAEEKLIALLVALEIVVVIIELKASGSSYEVI CECVARIVAEIVEALKRSGTSEDEIAEIVAKVIAAVIIVLKELGSSYEVICECV ARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESGSSYEVIKECVQRIVE EIVEALKRSGTSEDEINEIVRRVKSEVERTLKESGSLEHHHHHH**
DHR7C3_2 (SEC)	**MEKRIARELCELAAERAAESNDEREARIAAIECLLVAERAGMPTKEAARSF CEAAARAAAESNDEEVAKIAAKACLEVAKQAGMPTKEAARSFCEAAARAA AESNDEEVAKIAAKACLEVAKQAGMPTKEAARSFCEAAKRAAKESNDEEV EKIAKKACKEVAKQAGMPLEHHHHHH**
DHR7C3_8	MTEEDAARTCKKAARKAAESNDEEVAKQAAKDCLEVAKQAGMPTTIAAAIFC LAAARAAAESNDEEVAKIAAKACLEVAKQAGMPTKAAAIAFCIAAAMAAAES RDEEVAKIAAKACLEVAKQAGMPTKTAAALFMIAAIAAALRSEDEVVLAIAAL AIAEVLKQAGMPLEHHHHHH
DHR7C3_9 (SEC)	**MTKEMAAVLCMVLALKAAESNDEEKAKKAAKLCLIMADEAGMPTKEAAR SFCEAAAIAAAVSEDEEVAKIAAKACLEVAKQAGMPTKEAARSFCEAAASA AAILNEEEVAKIAAKACLEVAKQAGMPTKEAARSFCEAAKRAAKRSNDEE VEKIAKKACKEVAKQAGMPLEHHHHHH**
DHR10C3_7 (SEC-MALS)	**MSSEKEELRELLVAIVAVAAEDKGDDTEEAREAAREAFELVREAAERAGID SSEVLTLAILLILIVVLIAADAGYDISEAARAAAEAFKRVAEAAKRAGITSSEV LELAIRLIKEVVVNAAIRGYDISEAARAAAEAFKRVAEAAKRAGITSSKILK MAIILIRVMVKMAKERGKDISEAARQAAEIFRKAAERMRGSLEHHHHHH**
DHR10C3_7v2 (SEC-MALS)	**MSSEKEELRKMLVALVVVAAKEKGDDTEEAREAAREAFELVREAAERAGI DSSVVLALAILLILLVVLAAQMAGYDISEAARAAAEAFKRVAEAAKRAGITS SEVLELAIRLIKRVVLNAQIRGYDISEAARAAAEAFKRVAEAAKRAGITSSLL LKMAIVLIRVLVELAQESGADISEAARKAAEIMRRAAEDMRGSLEHHHHH H**
DHR10C3_8	MDECEEKARRVAEKVERLKRSGTSEDEIAEEVAREISEVIRTLKESGSEEVEICAC VARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESGSSYLVICMCVALIVAQ IVEALKRSGTSRKEIAEIVARVISEVIRTLKESGSSYEVIKECVERIVRAIVLALRES GTRITEIMAIVLAVLKEVLRTLKESGSLEHHHHHH
DHR10C3_18 (SEC-MALS)	**MKREKMELAKRLLKIVVENAKRKGDEEALAALAALLAFALVREAAERAGI DSSEVLELAIRLIKEVVENAQREGYRIALAALVAAMAFAVVAEAAKEAGITS SEVLELAIRLIKEVVENAQREGYEIVDAAMAAALAFARVAEAAKRAGITSSE TLKRAIEEIRKRVEEAQREGNDISEAAEQAAEEFRKKAEELKLEHHHHHH**
HR00C3 (SEC)	**MKEEKIAKLISLLAELSKKLIEIVARAADNKTTEEAVDIAILLIAIIARLAIRLI EMLAKNLASEEFMARAISAIAELAKKAIEAIYRLADNHTTDIRMLKAILAIA**

	ELAAEAIKAIADLAKNHTTEEFMARAISAIAELAKKAIEAIYRLADNHTTDLF MAIAIMAIAVLALLAIMAIADLAKNHTTEEFMAKAISAIAELAKKAIEAIYRL ADNHTSPDLIELAILAIEIIALAAIIAIEELAENITTEEYKEKAKSAIDEIREKA KEAIKRLEDNRTLEHHHHHH
HR00C3_3 (SEC)	**MKERLIAKLISVLAEASKILIRIAAKAADKLEREAAVILAIVLIAVAAIAAIAAI ALLAANLASEEFMARAISAIAELAKKAIEAIYRLADNHTEDEAMALAIEIIAI LALLAIVAIALLAANHTTEEFMARAISAIAELAKKAIEAIYRLADNHTTDTF MAKAIEAIAELAKEAIKAIAELAKNHTTEEFMAKAISAIAELAKKAIEAIYRL ADNHTSPTYIEKAIEAIEKIARTAIKAIEDLAKNITTEEYKEKAKSAIDEIREK AKEAIKRLEDNRTLEHHHHHH**
HR00C3_44 (SEC)	**MTEEKIAKEISRIAEESKKRIEELARKADNKTREAVVALAIAKIALLAREAIK RIEDLAKNLASEEFMARAISAIAELAKKAIEAIYRLADNHTKDVLMLVAIVAI AELAKEAIKAIADLAKNHTTEEFMARAISAIAELAKKAIEAIYRLADNHRLV AAMLLAIEAIAELAKEAIKAIADLAKNHTTEEFMAKAISAIAELAKKAIEAIY RLADNHRLPAAILLAALAILLIAVTAILAILILALNITTEEYKEKALSAIEEIVE KAEEAIDRLEDNLTLEHHHHHH**
tpr1C3_6 (SEC-MALS)	**MAEAWKELGKVLEKLGRLEEAAVAYLLAVIDDPNDAEAWKELGKVLEKL GELDAAAVAYEAAIELDPNDAEAWKELGKVLEKLGRLRRAALAYIKAIALD PNDAEAWKELGKVAEKLGRLKIAARIYKKAIELDPNDLEHHHHHH**
DS-Cav1-1na0C3_2	**MELLILKANAITTILTAVTFCFASGQNITEEFYQSTCSAVSKGYLSALRTGW YTSVITIELSNIKENKCNGTDAKVKLIKQELDKYKNAVTELQLLMQSTPATN NRARRELPRFMNYTLNNAKKTNVTLSKKRKRRFLGFLLGVGSAIASGVAV CKVLHLEGEVNKIKSALLSTNKAVVSLSNGVSVLTFKVLDLKNYIDKQLLPI LNKQSCSISNIETVIEFQQKNNRLLEITREFSVNAGVTTPVSTYMLTNSELLS LINDMPITNDQKKLMSNNVQIVRQQSYSIMCIIKEEVLAYVVQLPLYGVIDT PCWKLHTSPLCTTNTKEGSNICLTRTDRGWYCDNAGSVSFFPQAETCKVQS NRVFCDTMNSLTLPSEVNLCNVDIFNPKYDCKIMTSKTDVSSSVITSLGAIVS CYGKTKCTASNKNRGIIKTFSNGCDYVSNKGVDTVSVGNTLYYVNKQEGKS LYVKGEPIINFYDPLVFPSDEFDASISQVNEKINQSLAFIRKSDELLGSHHHH HHGGSGMEEAELAYLLGELAYKLGEYRIAIRAYRIALKRDPNNAEAWYNL GNAYYKQGDYDEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYDEAIE YYQKALELDPNNAEAKQNLGNAKQKQGLE**
ank1C2_1*	**MHHHHHHGSWGSSELGKRLIEAAENGNKDRVKDLIENGADVNASDSDGRT PLHHAAENGHAEVVALLIEKGADVNAKDSDGRTPLHHAAENGHDEVVLIL LLKGADVNAKDSDGRTPLHHAAENGHKRVVLVLILAGADVNTSDSDGRTP LDLAREHGNEEVVKALEKQ**
ank3C2_1*	**MSELGKRLIEAAENGNKDRVKDLLENGADVNASDSDGKTPLHLAAENGHA KVVLLLLEQGADPNAKDSDGKTPLHLAAENGHAVVVALLLMHGADPNAK DSDGKTPLHLAAENGHEEVVILLLAMGADPNTSDSDGRTPLDLAREHGNEE VVKVLEDHGGWLEHHHHHH**
1na0C3_3*	**MNLAEKMYKAGNAMYRKGQYTIAIIAYTLALLKDPNNAEAWYNLGNAAY KKGEYDEAIEAYQKALELDPNNAEAWYNLGNAYYKQGDYDEAIEYYQKAL ELDPNNAEAKQNLGNAKQKQGLEHHHHHH**
1na0C3_7*	**MNSAEAMYKMGNAAYKQGDYILAIIAYLLALEKDPNNAEAWYNLGNAAY KQGDYDEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYDEAIEYYQKA**

	LELDPNNAEAKQNLGNAKQKQGLEHHHHHH
HR00C3_2*	**MIEEVVAEMIDILAESSKKSIEELARAADNKTTEKAVAEAIEEIARLATAAIQ LIEALAKNLASEEFMARAISAIAELAKKAIEAIYRLADNHTTDTFMARAIAAI ANLAVTAILAIAALASNHTTEEFMARAISAIAELAKKAIEAIYRLADNHTTDK FMAAAIEAIALLATLAILAIALLASNHTTEEFMAKAISAIAELAKKAIEAIYRL ADNHTSPTYIEKAIEAIEKIARKAIKAIEMLAKNITTEEYKEKAKSAIDEIREK AKEAIKRLEDNRTLEHHHHHH**
1na0C4_1*	**MTLARVAYILGAIAYAQGEYDIAITAYQVALDLDPNNAEAWYNLGNAYYK QGDYDEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYDEAIEYYQKALE LDPNNAEAKQNLGNAKQKQGLEHHHHHH**
HR04C4_1*	**MHHHHHHGSWGSDECEEKARRVAEKVERLKRSGTSEDEIAEEVAREISEVI RTLKESGSSYEVICECVARIVAEIVEALKRSGTSAVEIAKIVARVISEVIRTLK ESGSSYEVICECVARIVAEIVEALKRSGTSAAIIALIVALVISEVIRTLKESGSS FEVILECVIRIVLEIIEALKRSGTSEQDVMLIVMAVLLVVLATLQLSGS**
2JFB (PDB ID)*	**MAVKGLGEVDQKYDGSKLRIGILHARWNRKIIDALVAGAVKRLQEFGVKE ENIIIETVPGSFELPYGSKLFVEKQKRLGKPLDAIIPIGVLIKGSTMHFEYICD STTHQLMKLNFELGIPVIFGVLTCLTDEQAEARAGLIEGKMHNHGEDWGA AAVEMATKFN**
2OBX (PDB ID)*	**MNQHSHKDYETVRIAVVRARWHADIVDQCVSAFEAEMADIGGDRFAVDVF DVPGAYEIPLHARTLAETGRYGAVLGTAFVVNGGIYRHEFVASAVIDGMM NVQLSTGVPVLSAVLTPHNYHDSAEHHRFFFEHFTVKGKEAARACVEILAA REKIAA**
2B98 (PDB ID)*	**MTKKVGIVDTTFARVDMASIAIKKLKELSPNIKIIRKTVPGIKDLPVACKKL LEEEGCDIVMALGMPGKAEKDKVCAHEASLGLMLAQLMTNKHIIEVFVHE DEAKDDKELDWLAKRRAEEHAENVYYLLFKPEYLTRMAGKGLRQGFEDA GPARE**

Appendix Table 2.2. List of all designed two-component nanoparticles tested and their corresponding amino acid sequences including initiating methionines and $(His)_6$ tags on single components. Designs that expressed solubly with expected number of components are denoted in bold and under all names in parentheses are the oligomers that the designed components originate from in **Appendix Table 2.1**.

Design	Sequence
T33_dn1A (1na0C3_3)	MGNLAEKMYKAGNAMYRKGQYTIAIIAYTLALLKDPNNAEAWYNLGNAAYKK GEYDEAIEAYQKALELDPNNAEAWYNLGNAYYKQGDYDEAIEYYERALELDPEN AEAALNLLEAKEKQG
T33_dn2B (1na0C3_2)	MEEAELAYLLGELAYKLGEYRIAIRAYRIALKRDPNNAEAWYNLGNAYYKQGDY REAIKYYAKALTLDPKNAEAWYNLGNAVYKQGDYRIAILFYRAALKLDPNNAEA KQNLGNAKQKQGLEHHHHHH

T33_dn2A (1na0C3_3)	**MGNLAEKMYKAGNAMYRKGQYTIAIIAYTLALLKDPNNAEAWYNLGNAAY KKGEYDEAIEAYQKALELDPNNAEAWYNLGNAYYKQGDYDEAIEYYKKALR LDPRNVDAIENLIEAEEKQG**
T33_dn2B (1na0C3_2)	**MEEAELAYLLGELAYKLGEYRIAIRAYRIALKRDPNNAEAWYNLGNAYYKQ GDYREAIRYYLRALKLDPENAEAWYNLGNALYKQGKYDLAIIAYQAALEEDP NNAEAKQNLGNAKQKQGLEHHHHHH**
T33_dn3A (1na0C3_7)	MGNSAEAMYKMGNAAYKQGDYILAIIAYLLALEKDPNNAEAWYNLGNAAYKQ GDYKEAILYYIRALQLDPNNAEAWYNLGNAFYKKGDYRVAIILYRMALKLDPNN AEAKQNLGNAKQKQGDIHHHHHH
T33_dn3B (1na0C3_2)	MEEAELAYLLGELAYKLGEYRIAIRAYRIALKRDPNNAEAWYNLGNAYYKQGDY DEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYEEAILYYLEALDLDPNNAEA AENLLNAVKKDE
T33_dn4A (1na0C3_7)	**MGNSAEAMYKMGNAAYKQGDYILAIIAYLLALEKDPNNAEAWYNLGNAAY KQGDYDEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYDEAIEYYEKALE LDPRNAEALKNLLEAKAKQD**
T33_dn4B (3ltjC3_1)	**MHHHHTDPLAVILYIAILKAEKSIARAKAAEALGKIGDERAVEPLIKALKDED ALVRAAAADALGQIGDERAVEPLIKALKDEEGLVRASAAIALGQIGDERAVRP LIKALADERDLVRVAAAVALGRIGDERAVKPLIIVLLDEEGEVREAAAIALGSI GGERVRAAMEKLAERGRGFARKVAVNYLETHKLEHHHHHH**
T33_dn5A 1na0C3_7	**MGNSAEAMYKMGNAAYKQGDYILAIIAYLLALEKDPNNAEAWYNLGNAAY KQGDYDEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYDEAIEYYEKALE LDPNNAEALKNLLEAIAEQD**
T33_dn5B (3ltjC3_1)	**MHHHHTDPLAVILYIAILKAEKSIARAKAAEALGKIGDERAVEPLIKALKDED ALVRAAAADALGQIGDERAVEPLIKALKDEEGLVRASAAIALGQIGDERAVQ PLIKALTDERDLVRVAAAVALGRIGDEKAVRPLIIVLKDEEGEVREAAAIALG SIGGERVRAAMEKLAERGTGFARKVAVNYLETHKLEHHHHHH**
T33_dn6A (1na0C3_2)	**MGEEAELAYLLGELAYKLGEYRIAIRAYRIALDEDPDNAEAWYNLGNAYYK QGDYREAILYYQMALRLDPNNAEAWYNLGNAYYKQGDYDRAIEYYQKALE LDPNNAEAKQNLGNAKQKQGDIHHHHHH**
T33_dn6B (3ltjC3_1)	**MHHHHTDPLAVILYIAILKAEKSIARAKAAEALGKIGDERAVEPLIKALKDED ALVRAAAADALGQIGDERAVEPLIKALKDEEGLVRASAAIALGQIGDKRAVR PLIRALKDERDLVREAAAVALGRIGDELAVEPLIKALKDEEGEVREAAAIALG SIGGEIVRMMMDKLAETGTGFARKVAVNYLETHK**
T33_dn7A (1na0C3_2)	**MGEEAELAYLLGELAYKLGEYRIAIRAYRIALKRDPNNAEAWYNLGNAYYK QGDYDEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYDEAIEYYRKALEL DPENEEALENLLNAKQKQGDIHHHHHH**
T33_dn7B (3ltjC3_1)	**MHHHHTDPLAVILYIAILKAEKSIARAKAAEALGKIGDERAVEPLIKALKDED ALVRAAAADALGQIGDERAVEPLIKALKDEEGLVRASAAIALGQIGDERAVEP LIKALKDERDLVRVAAAVALGRIGDERAVEPLIKALKDEEGEVREAAAIALGS IGGKRVRLAMLKLALEGTGFARKVAVNYLETHK**

T33_dn8A (1na0C3_2)	MGEEAELAYLLGELAYKLGEYRIAIRAYRIALKRDPNNAEAWYNLGNAYYKQGD YDEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYDEAIEYYRKALELDPENLE ALLNLLNAKDKRG
T33_dn8B (HR00C3_2)	MIEEVVAEMIDILAESSKKSIEELARAADNKTTEKAVAEAIEEIARLATAAIQLIEAL AKNLASEEFMARAISAIAELAKKAIEAIYRLADNHTTDTFMARAIAAIANLAVTAI LAIAALASNHTTEEFMARAIRAIAELAKKAIEAIYRLADNHTTDKFMAAAIEAIALL ATLAILAIALLASNHTTERFMAKAILAIAVLAKKAIEAIYRLADNHTSPTYIEKAIEA IEKIARKAIKAIEMLAKNITTEEYKEEAKSAIEIIRELARIAIRRLEDNRTLEHHHHH H
T33_dn9A (1na0C3_2)	MGEEAELAYLLGELAYKLGEYRIAIRAYRIALKRDPNNAEAWYNLGNAYYKQGD YDEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYDEAIEYYQKALELDPENLE AILNLGEAKLKQG
T33_dn9B (HR00C3_2)	MIEEVVAEMIDILAESSKKSIEELARAADNKTTEKAVAEAIEEIARLATAAIQLIEAL AKNLASEEFMARAISAIAELAKKAIEAIYRLADNHTTDTFMARAIAAIANLAVTAI LAIAALASNHTTEEFMARAISAIAELAKKAIEAIYRLADNHTTDKFMAAAIEAIALL ATLAILAIALLASNHTTEKFMAEAIIVIALLAVLAIMAIYRLADNHTSPTYIEKAIEAI EKIARKAIKAIEMLAKNITTEEYKEKAKSAIDLIRQLADIIIRKLEDNRTLEHHHHH H
T33_dn10A (1na0C3_2)	**MGEEAELAYLLGELAYKLGEYRIAIRAYRIALKRDPNNAEAWYNLGNAYYK QGDYDEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYDEAIEYYEKALEL DPENLEALQNLLNAMDKQG**
T33_dn10B (HR00C3_2)	**MIEEVVAEMIDILAESSKKSIEELARAADNKTTEKAVAEAIEEIARLATAAIQLI EALAKNLASEEFMARAISAIAELAKKAIEAIYRLADNHTTDTFMARAIAAIANL AVTAILAIAALASNHTTEEFMARAISAIAELAKKAIEAIYRLADNHTTDKFMAA AIEAIALLATLAILAIALLASNHTTEKFMARAIMAIAILAAKAIEAIYRLADNHT SPTYIEKAIEAIEKIARKAIKAIEMLAKNITTEEYKEKAKKIIDIIRKLAKMAIK KLEDNRTLEHHHHHH**
T33_dn11A (1na0C3_2)	MGEEAELAYLLGELAYKLGEYRIAIRAYRIALKRDPNNAEAWYNLGNAYYKQGD YDEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYDEAIEYYRKALELDKENIE ALLNLLNAKEKQD
T33_dn11B (HR00C3_2)	MIEEVVAEMIDILAESSKKSIEELARAADNKTTEKAVAEAIEEIARLATAAIQLIEAL AKNLASEEFMARAISAIAELAKKAIEAIYRLADNHTTDTFMARAIAAIANLAVTAI LAIAALASNHTTEEFMARAISAIAELAKKAIEAIYRLADNHTTDKFMAAAIEAIALL ATLAILAIALLASNHTTERFMAKAILAIAILAAKAIEAIYRLADNHTSPTYIEKAIEAI EKIARKAIKAIEMLAKNITTEEYKEEAKSAIEIIRLLAKAVIKRLQDNRTLEHHHHH H
O32_dn1A (1na0C3_3)	**MGELAEKMYKAGNAMYRKGQYTIAIIAYTLALLKDPNNAEAWYNLGNAAY KKGEYDEAIVAYVEALELDPNNAEAWYNLGNAYYKQGDYEEAIEYYQKALE LDPNNAEAKQNLGNAKQKQG**
O32_dn1B (ank1C2_1)	**MSRRGRLLIIAAENGNKDRVKDLIQRGADVNASDRRGRTPLHHAAENGHAE VVALLIEKGADVNAKDSDGRTPLHHAAENGHDEVVLILLLKGADVNAKDSD GRTPLHHAAENGHKRVVLVLILAGADVNTSDSDGRTPLDLAREHGNEEVVK ALEKQLEHHHHHH**

O32_dn2A (1na0C3_2)	MGEEAELAYLLGELAYKLGEYRIAIRAYRIALKRDPNNAEAWYNLGNAYYKQGD YDEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYDEAIEYYQKALELDPNNA EARKNLIIADLKQEDIHHHHHH
O32_dn2B (ank1C2_1)	MSELGEALILAAERGKKDRVKDLIEEGADVNASDSDGRTPLHHAAENGHAEVVA LLIEKGADVNAKDSDGRTPLHHAAENGHDEVVLILLLKGADVNAKDSDGRTPLH HAAENGHKRVVLVLILAGADVNTSDSDGRTPLDLAREHGNEEVVKALEKQ
O32_dn3A (3ltjC3_1)	**MGHHHHHHGWHHHHTDPLAVILYIAILKAEKSIARAKAAEALGKIGDERAV EPLIKALKDEDALVRAAAADALGQIGDERAVEPLIEALEDEEGLVRASAAIAL GQIGDERAVEPLILALADERDLVRVAAAVALGRIGDERAVEPLIVMLRDEEGE VREAAAIALGSIGGERVRAAMEELAERGRGFARKVAVNYLETHK**
O32_dn3B (ank1C2_1)	**MSELGKRLIEAAENGNKKRVKDLIENGADVNASDSDGRTPLHHAAENGHAE VVALLIEKGADVNAKDSDGRTPLHHAAENGHDEVVLILLLKGADVNAKDSD GRTPLHHAAENGHKRVVLVLILAGADVNTKDEEGDTPLALALEHGNREVIK ALLKQ**
O43_dn1A (1na0C4_1)	MGTLARVAYILGAIAYAQGEYDIAITAYQVALDLDPNNAEAWYNLGNAYYKQG DYDEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYLLAIVYYAKALILDPNN AEAKQNLGNAIQKQD
O43_dn1B (1na0C3_3)	MNLAEKMYKAGNAMYRKGQYTIAIIAYTLALLKDPNNAEAWYNLGNAAYKKG EYDEAIEAYQKALELDPNNAEAWYNLGNAYYKQGDYLEAIAYYAKALLLDPNN AEARQNLGNAMQKSELEHHHHHH
O43_dn2A (1na0C4_1)	MGTLARVAYILGAIAYAQGEYDIAITAYQVALDLDPNNAEAWYNLGNAYYKQG DYDEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYDEAILYYVKALVLDPNN AEAKQNLGNARQKQG
O43_dn2B (1na0C3_3)	MNLAEKMYKAGNAMYRKGQYTIAIIAYTLALLKDPNNAEAWYNLGNAAYKKG EYDEAIEAYQKALELDPNNAEAWYNLGNAYYKQGDYLEAILYYVKALKLDPNN AEAKQNLGNAEQKKDLEHHHHHH
O43_dn3A (1na0C4_1)	MGTLARVAYILGAIAYAQGEYDIAITAYQVALDLDPNNAEAWYNLGNAYYKQG DYDEAIKYYQKALELDPNNAEAWYNLGNAYYKQGDYVIAIALYQLALELDPNN AEAKQNLGNAEQKEGDIHHHHHH
O43_dn3B (1na0C3_7)	MNSAEAMYKMGNAAYKQGDYILAIIAYLLALEKDPNNAEAWYNLGNAAYKQG DYDEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYLEAIEYYIKALELDPNNE EARQNLLNAAKKIE
O43_dn4A (1na0C4_1)	**MGTLARVAYILGAIAYAQGEYDIAITAYQVALDLDPNNAEAWYNLGNAYYK QGDYDEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYEEAIEYYLKALEL DPNNAEARQNLRNAMQKEG**
O43_dn4B (1na0C3_7)	**MNSAEAMYKMGNAAYKQGDYILAIIAYLLALEKDPNNAEAWYNLGNAAYK QGDYDEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYLAAIIYYRRALEL DPNNAEAKQNLGNAEQKEGLEHHHHHH**

O43_dn5A (1na0C4_1)	MGTLARVAYILGAIAYAQGEYDIAITAYQVALDLDPNNAEAWYNLGNAYYKQG DYDEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYREALRYYIKALKLDPNN AEAKQNLGNALEKRG
O43_dn5B (1na0C3_7)	MNSAEAMYKMGNAAYKQGDYILAIIAYLLALEKDPNNAEAWYNLGNAAYKQG DYDEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYLVAIIYYLEALELDPNNA EAKQNLGNAKQKEGLEHHHHHH
O43_dn6A (1na0C4_1)	**MGTLARVAYILGAIAYAQGEYDIAITAYQVALDLDPNNAEAWYNLGNAYYK QGDYDEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYQEAIEYYARALRR DRRNKEAIENLINALQKED**
O43_dn6B (1na0C3_2)	**MEEAELAYLLGELAYKLGEYRIAIRAYRIALKRDPNNAEAWYNLGNAYYKQ GRYVRALIYYLRALLLDPENAEAWYNLGNAYYKKGDYDIAIVYYELALEDDP NNAEAKQLLGNAKQKQGLEHHHHHH**
O43_dn7A (1na0C4_1)	MGTLARVAYILGAIAYAQGEYDIAITAYQVALDLDPNNAEAWYNLGNAYYKQG DYDEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYDEAIEYYKKALRLDPNN EEAKQNLMNALQKQD
O43_dn7B (3ltjC3_1)	MHHHHTDPLAVILYIAILKAEKSIARAKAAEALGKIGDERAVEPLIKALKDEDALV RAAAADALGQIGDERAVEPLIKALKDEEGLVRASAAIALGQIGDERAVEPLIKALK DERDLVRVAAAVALGRIGDKKAVLPLIKALKDEEGEVREAAAIALGSIGGRLVRA MMELLAETGRGFARKVAVNYLETHKLEHHHHHH
O43_dn8A (HR04C4_1)	MGDECEEKARLLAELVETLKRSGTSEDEIAEDVARLISEMIRNLKESGSSYEVICEC VARIVAEIVEALKRSGTSAVEIAKIVARVISEVIRTLKESGSSYEVICECVARIVAEIV EALKRSGTSAAIIALIVALVISEVIRTLKESGSSFEVILECVIRIVLEIIEALKRSGTSEQ DVMLIVMAVLLVVLATLQLSGS
O43_dn8B (1na0C3_3)	MNLAEKMYKAGNAMYRKGQYTIAIIAYTLALLKDPNNAEAWYNLGNAAYKKG EYDEAIEAYQKALELEPNNAEAWYNLGNAYYKQGDYEEAIIYYLKALVLDPRNA EARQNLGNAKQKEGLEHHHHHH
O43_dn9A (HR04C4_1)	MGDECEELARIVAELVEKLKRSGTSEDEIAERVAREISEVIKLLKKSGSSYEVICEC VARIVAEIVEALKRSGTSAVEIAKIVARVISEVIRTLKESGSSYEVICECVARIVAEIV EALKRSGTSAAIIALIVALVISEVIRTLKESGSSFEVILECVIRIVLEIIEALKRSGTSEQ DVMLIVMAVLLVVLATLQLSGS
O43_dn9B (1na0C3_3)	MNLAEKMYKAGNAMYRKGQYTIAIIAYTLALLKDPNNAEAWYNLGNAAYKKG EYDEAIEAYQKALELDPENAEAWYNLGNAYYKQGEYLEALLYYLKALILDPNNA EAKQNLGNARQKQGLEHHHHHH
O43_dn10A (HR04C4_1)	MGDECERKARLVAKIVELLKRSGTSEDEIAEEVARLISLVIKVLKKSGSSYEVICEC VARIVAEIVEALKRSGTSAVEIAKIVARVISEVIRTLKESGSSYEVICECVARIVAEIV EALKRSGTSAAIIALIVALVISEVIRTLKESGSSFEVILECVIRIVLEIIEALKRSGTSEQ DVMLIVMAVLLVVLATLQLSGS
O43_dn10B (1na0C3_3)	MNLAEKMYKAGNAMYRKGQYTIAIIAYTLALLKDPNNAEAWYNLGNAAYKKG EYDEAIEAYQKALELDPENAEAWYNLGNAYYKQGDYAEAMLYYLKALLLDPNN AEAKQNLGNAEQKAGLEHHHHHH

O43_dn11A (HR04C4_1)	MGDECEEKAELVALLVEALKKLGTSEDEIAEEVAKEISRVIRRLKESGSSYEVICEC VARIVAEIVEALKRSGTSAVEIAKIVARVISEVIRTLKESGSSYEVICECVARIVAEIV EALKRSGTSAAIIALIVALVISEVIRTLKESGSSFEVILECVIRIVLEIIEALKRSGTSEQ DVMLIVMAVLLVVLATLQLSGS
O43_dn11B (1na0C3_7)	MNSAEAMYKMGNAAYKQGDYILAIIAYLLALEKDPNNAEAWYNLGNAAYKQG DYDEAIEYYQKALELDPENAEAWYNLGNAYYKQGDYELAIIFYKVALALDPNNA EAKQNLGNAKQKQGLEHHHHHH
O43_dn12A (HR04C4_1)	**MGDRCERRAKLVALKVELLKKDGTSEDEIAEEVAREISEVIRDLRKSGSSYEV ICECVARIVAEIVEALKRSGTSAVEIAKIVARVISEVIRTLKESGSSYEVICECVA RIVAEIVEALKRSGTSAAIIALIVALVISEVIRTLKESGSSFEVILECVIRIVLEIIE ALKRSGTSEQDVMLIVMAVLLVVLATLQLSGS**
O43_dn12B (1na0C3_2)	**MEEAELAYLLGELAYKLGEYRIAIRAYRIALKRDPNNAEAWYNLGNAYYKQ GDYDEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYDEAIEYYQKALELD PNNIKAELNLIIAEEKQGLEHHHHHH**
O43_dn13A (HR04C4_1)	MGDECEELARAVALVVEILKRSGTSEDEIAEEVARLISRVIRKLKESGSSYEVICEC VARIVAEIVEALKRSGTSAVEIAKIVARVISEVIRTLKESGSSYEVICECVARIVAEIV EALKRSGTSAAIIALIVALVISEVIRTLKESGSSFEVILECVIRIVLEIIEALKRSGTSEQ DVMLIVMAVLLVVLATLQLSGS
O43_dn13B (1na0C3_2)	MEEAELAYLLGELAYKLGEYRIAIRAYRIALKRDPNNAEAWYNLGNAYYKQGDY DEAIEYYQKALELDPNNAEAWYNLGNAYYKKGDYLIAILYYLVALTLDPNNAEA KQNLGNAKQKDGLEHHHHHH
O43_dn14A (HR04C4_1)	MGDKCEEMAELVAQLVELLKESGTSEDEIAEKVARLISKVIRKLKESGSSYEVICE CVARIVAEIVEALKRSGTSAVEIAKIVARVISEVIRTLKESGSSYEVICECVARIVAEI VEALKRSGTSAAIIALIVALVISEVIRTLKESGSSFEVILECVIRIVLEIIEALKRSGTSE QDVMLIVMAVLLVVLATLQLSGS
O43_dn14B (1na0C3_2)	MEEAELAYLLGELAYKLGEYRIAIRAYRIALKRDPNNAEAWYNLGNAYYKQGDY DEAIEYYMKALKLDPKNAEAWYNLGNAYYKQGDYLLAILIYEMALILDPNNAEA KQNLGNAKQKEGLEHHHHHH
O43_dn15A (HR04C4_1)	MGRKCELLARLVAMIVELLKESGTSEDEIAEEVAREISEVIRTLKEEGSSYEVICEC VARIVAEIVEALKRSGTSAVEIAKIVARVISEVIRTLKESGSSYEVICECVARIVAEIV EALKRSGTSAAIIALIVALVISEVIRTLKESGSSFEVILECVIRIVLEIIEALKRSGTSEQ DVMLIVMAVLLVVLATLQLSGS
O43_dn15B (1na0C3_2)	MEEAELAYLLGELAYKLGEYRIAIRAYRIALKRDPNNAEAWYNLGNAYYKQGDY DEAIEYYQKALELDPRNAEAWYNLGNAYYKQGDYLMAILIYQLALMLDPNNAE AKQNLGNAKQKRGLEHHHHHH
O43_dn16A (HR04C4_1)	MGEDCEELAELVAELVERLKRRGTSEDEIAEEVARIISEVIRMLKESGSSYEVICEC VARIVAEIVEALKRSGTSAVEIAKIVARVISEVIRTLKESGSSYEVICECVARIVAEIV EALKRSGTSAAIIALIVALVISEVIRTLKESGSSFEVILECVIRIVLEIIEALKRSGTSEQ DVMLIVMAVLLVVLATLQLSGS
O43_dn16B (1na0C3_2)	MEEAELAYLLGELAYKLGEYRIAIRAYRIALKRDPNNAEAWYNLGNAYYKQGDY KEAIKYYQKALKLDPNNAEAWYNLGNAYYKKGDYIMAILAYELALEEDPNNAE AKQNLGNAKQKQGLEHHHHHH

O43_dn17A (HR04C4_1)	**MGDECEEKARRVALKVLKLRLRGTSEDEIAEEVAREISKVIETLKESGSSYEV ICECVARIVAEIVEALKRSGTSAVEIAKIVARVISEVIRTLKESGSSYEVICECVA RIVAEIVEALKRSGTSAAIIALIVALVISEVIRTLKESGSSFEVILECVIRIVLEIIE ALKRSGTSEQDVMLIVMAVLLVVLATLQLSGS**
O43_dn17B (1na0C3_2)	**MEEAELAYLLGELAYKLGEYRIAIRAYRIALKRDPNNAEAWYNLGNAYYKQ GDYDEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYDEAIEYYQKALELD PNNEEAKIVLGLAKEEQELEHHHHHH**
I32_dn1A (1na0C3_3)	**MGNLAEKMYKAGNAMYRKGQYTIAIIAYTLALLKDPNNAEAWYNLGNAAY KKGEYDEAIEAYQKALELEPENAEALYNLGNAYYKQGEYDEAILYYLIALEL DPNNAEAKQNLGNAKQKQGDIHHHHHH**
I32_dn1B (ank1C2_1)	**MSRLGIRLIIAAIEGNKDRVKDLIENGADVNASDSVGRTPLHHAAENGHAEVV ALLIEKGADVNAKDSDGRTPLHHAAENGHDEVVLILLLKGADVNAKDRDGR TPLHHAAENGHKRVVLVLILAGADVNTSDSDGRTPLDLAREHGNEEVVKALE KQ**
I32_dn2A (1na0C3_3)	**MGDLAEKMYKAGNAMYRKGQYTIAIIAYTLALLKDPNNAEAWYNLGNAAY KKGEYDEAILAYLKALELDPNNAEAWYNLGNAFYKQGDYRMAIKYYQKALE LDPNNAEAKQNLGNAKQKQG**
I32_dn2B (ank1C2_1)	**MSELGELLIVAAENGNKKMVRDLIKNGADVNASDEDGRTPLHHAAENGHAE VVALLIEKGADVNAKDSDGRTPLHHAAENGHDEVVLILLLKGADVNAKDSD GRTPLHHAAENGHKRVVLVLILAGADVNTSDSDGRTPLDLAREHGNEEVVK ALEKQLEHHHHHH**
I32_dn3A (1na0C3_3)	MGRLAKKMYKAGNAMYRKGQYTIAIIAYTLALLKDPNNAEAWYNLGNAAYKK GEYAEAIVAYIKALELDPNNAEAWYNLGNALYKLGAYNAAIQVYQKALELDPNN AEAKQNLGNAKQKKG
I32_dn3B (ank1C2_1)	MKILGLALIAAARNGEKERVETLIEAGADVNASDDDGRTPLHHAAENGHAEVVA LLIEKGADVNAKDSDGRTPLHHAAENGHDEVVLILLLKGADVNAKDSDGRTPLH HAAENGHKRVVLVLILAGADVNTSDSDGRTPLDLAREHGNEEVVKALEKQLEHH HHHH
I32_dn4A (1na0C3_7)	MGKSAEAMYKMGNAAYKQGDYILAIIAYLLALEKDPKNAEAWYNLGNAAYKQ GDYEEAIRYYLKALLLDDNNAEAWYNLGNAYYKQGDYREAIMLYQKALELDPN NAEAKQNLGNAKQKQG
I32_dn4B (ank1C2_1)	MSELGKLLIMAAELGNKRLVKELIENGADVNASDSDGRTPLHHAAENGHAEVVA LLIEKGADVNAKDSDGRTPLHHAAEEGHDEVVLILLLKGADVNAKDSDGRTPLH HAAENGHKRVVLVLILAGADVNTSDSDGRTPLDLAREHGNEEVVKALEKQLEHH HHHH
I32_dn5A (1na0C3_7)	MGRSAEAMYKMGNAAYKQGDYILAIIAYLLALEKDPNNAEAWYNLGNAAYKQ GDYREAIRYYLKALALDPNNAEAWYNLGNAFYKQGDYNEAIEVYQKALELDPN NAEAKQNLGNAKQKQG
I32_dn5B (ank1C2_1)	MSELGRMLIEAAELGKKEIVKELIENGADVNASDSDGRTPLHHAAENGHAEVVAL LIEKGADVNAKDSDGRTPLHHAAENGHDEVVLILLLKGADVNAKDSDGRTPLHH AAENGHKRVVLVLILAGADVNTSDSDGRTPLDLAREHGNEEVVKALEKQLEHHH HHH

I32_dn6A (1na0C3_2)	**MGEEAELAYLLGELAYKLGEYRIAIRAYRIALKRDPNNAEAWYNLGNAYYK QGDYDEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYDEAIEYYQKALEL DPNNDEADDNLLNADQKQDDIHHHHHH**
I32_dn6B (ank1C2_1)	**MSRLGKKLIIAAERGNKDRVKDLIENGADVNASDEDGRTPLHHAAENGHAEV VALLIEKGADVNAKDSDGRTPLHHAAENGHDEVVLILLLKGADVNAKDSDG RTPLHHAAENGHKRVVLVLILAGADVNTSDSDGRTPLDLAREHGNEEVVKAL EKQ**
I32_dn7A (1na0C3_2)	**MGEEAELAYLLGELAYKLGEYRIAIRAYRIALKRDPNNAEAWYNLGNAYYK QGDYDEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYDEAIEYYQKALEL DPKNMEALLDLGNAKQKQKDIHHHHHH**
I32_dn7B (ank1C2_1)	**MSELGKDLIVAAALGNKDRVKDLIENGADVNASDRRGATPLHMAALNGHAE VVALLIEKGADVNAKDSDGRTPLHHAAENGHDEVVLILLLKGADVNAKDSD GRTPLHHAAENGHKRVVLVLILAGADVNTSDSDGRTPLDLAREHGNEEVVK ALEKQ**
I32_dn8A (1na0C3_2)	**MGEEAELAYLLGELAYKLGEYRIAIRAYRIALKRDPNNAEAWYNLGNAYYK QGDYDEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYLRAIAYYRKALEL DPNNAEAKQNLGNAKQKIE**
I32_dn8B (ank3C2_1)	**MELEGERLIEAAENGNKDRVKDLLENGALVNASDSDGKTPLHLAAENGHAK VVLLLLEQGAKPNAKDSDGKTPLHLAAENGHAVVVALLLMHGADPNAKDSD GKTPLHLAAENGHEEVVILLLAMGADPNTSDSDGRTPLDLAREHGNEEVVKV LEDHGGWLEHHHHALEHHHHHH**
I32_dn9A (HR00C3_2)	MGIEEVVAEMIDILAESSKKSIEELARAADNKTTEKAVAEAIEEIARLATAAIQLIE ALAKNLASEEFMADAISAIAELAKKAIEAIYRLADNHTTDTFMARAIAAIANLAVT AILAIAALASNHTTEQFMAIAIIAIAELAKKAIEAIYRLADNHTTDKFMAAAIEAIAL LATLAILAIALLASNHTTEAFMALAILLIAELAKKAIEAIYRLADNHTSPTYIEKAIE AIEKIARKAIKAIEMLAKNITTEEYKEKARAAILEIREKAKEAIKRLEDNRT
I32_dn9B (ank1C2_1)	MHHHHHHSELGKRLIEAAENGNKKRVLELIENGADVNASDSDGRTPLHHAAENG HAEVVALLIELGADVNAKDSDGRTPLHHAAENGHDEVVLILLLKGADVNAKDSD GRTPLHHAAENGHKRVVLVLILAGADVETSDSRGRTPLMLAVEHGNIEVALALLK QGW
I32_dn10A (HR00C3_2)	**MGHHHHHHWGIEEVVAEMIDILAESSKKSIEELARAADNKTTEKAVAEAIEEI ARLATAAIQLIEALAKNLASEEFMARAISAIAELAKKAIEAIYRLADNHTTDTF MARAIAAIANLAVTAILAIAALASNHTTEEFMARAISAIAELAKKAIEAIYRLA DNHTTDKFMAAAIEAIALLATLAILAIALLASNHTTEEFMAKAISAIARLAKKA ILAIYKLADNHTSPTYIEKAIEAIEKIARKAIKAIEMLAKNITTEEYKEKAKSAI DEIREIAKIAIKTLEDNRT**
I32_dn10B (ank1C2_1)	**MSEIGKRLIEAAENGNKERVKLLIELGADVNASDSDGRTPLHHAAENGHAEV VALLIEKGADVNAKDSDGRTPLHHAAENGHDEVVLILLLKGADVNAKDSDG RTPLHHAAENGHKRVVLVLILAGADVNTSDSDGRTPLDLAREHGNEEVVKAL EKQ**
I53_dn1A (2B98)	MGHHHHHHKKVGIVDTTFARVDMAIMAIIVLELRPRNIKIIRKTVPGIKDLPVACK KLLEEEGCDIVMALGMPGKAEKDKVCAHEASLGLMLAQLMTNKHIIEVFVHEDE AKDDRELDWLAKRRAEEHAENVYYLLFKPEYLTEMAGKGLRQGFEDAGP

I53_dn1B (HR00C3_2)	MIEEVVAEMIDILAESSKKSIRELAKAAKNKTTEKAVAEAIEEIARLATAAIQLIEA LAKNLASEEFMARAISAIAELAKKAIEAIYRLADKHKTDTFMARAIAAIANLAVTA ILAIAALASNHTTEEFMARAISAIAELAKKAIMAILLLALLHTTDKFMAAAIEAIAL LATLAILAIALLASNHTTEEFMAKAISAIAELAKKAIEAIYLLADLHTPVLYEDKAIE AIEKIARKAIKAIEMLAKNITTEEYKEKAKSAIDEIREKAKEAIKRLERNRE
I53_dn2A (2JFB)	MGRYDGSKLRIGILHARWNRSIILALVLGAIERLLEFGVRAKNIIIETVPGSFELPYG SKLFVEKQKRLGKPLDAIIPIGVLIKGSTMHFEYICDSTTHQLMKLNFELGIPVIFGV LTCLTDEQAEARAGLIDGKMHNHGEDWGAAAVEMATKFN
I53_dn2B (1na0C3_2)	MEEAELAYLLGELAYKLGEYRIAIRAYRIALKRDPNNAEAWYNLGNAYYKQGDY DEAIEYYRRALKLEPENAEAWYNLGNAYYKQGDYKEAIAYYLIALILDPNNAEA KQNLGNAEQKQDLEHHHHHH
I53_dn3A (2JFB)	MGKYDGSKLRIGILHARWNRAIIIALVLGALKRLLEFGVKAKNIIIETVPGSFELPY GSKLFVEKQKRLGKPLDAIIPIGVLIKGSTMHFEYICDSTTHQLMKLNFELGIPVIFG VLTCLTDEQAEARAGLIKGKMHNHGEDWGAAAVEMATKFN
I53_dn3B (1na0C3_2)	MEEAELAYLLGELAYKLGEYRIAIRAYRIALKRDPNNAEAWYNLGNAYYKQGDY DEAIEYYQEALELDPENAEAWYNLGNAYYKQGDYKEALAYYLLALELDPNNAE AEQNLGNAEQKRDLEHHHHHH
I53_dn4A (2JFB)	MGKYDGSKLRIGILHARWNVKIIIALILGAIKRLREFGVKRENIIIEIVPGSFELPYGS KLFVEKQKRLGKPLDAIIPIGVLIKGSTMHFEYICDSTTHQLMKLNFELGIPVIFGVL TCLTDEQAEARAGLIEGKMHNHGEDWGAAAVEMATKFN
I53_dn4B (1na0C3_2)	MEEAELAYLLGELAYKLGEYRIAIRAYRIALKRDPNNAEAWYNLGNAYYKQGDY DEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYDEAIEYYQKALELDPNNLD AVMNLLEASLKQELEHHHHHH
I53_dn5A (2JFB)	**MGKYDGSKLRIGILHARWNAEIILALVLGALKRLQEFGVKRENIIIETVPGSFE LPYGSKLFVEKQKRLGKPLDAIIPIGVLIKGSTMHFEYICDSTTHQLMKLNFE LGIPVIFGVLTCLTDEQAEARAGLIEGKMHNHGEDWGAAAVEMATKFN**
I53_dn5B (1na0C3_2)	**MEEAELAYLLGELAYKLGEYRIAIRAYRIALKRDPNNAEAWYNLGNAYYKQ GRYREAIEYYQKALELDPNNAEAWYNLGNAYYERGEYEEAIEYYRKALRLD PNNADAMQNLLNAKMREELEHHHHHH**
I53_dn6A (2JFB)	**MGDYDGSKLRIGILHARKNTEIIVALVIGAVERLEEFGVKRENIIIEIVPGSFEL PYGSKLFVEKQKRLGKPLDAIIPIGVLIKGSTMHFEYICDSTTHQLMKLNFEL GIPVIFGVLTCLTDEQAEARAGLIEGKMHNHGEDWGAAAVEMATKFN**
I53_dn6B (1na0C3_2)	**MEEAELAYLLGELAYKLGEYRIAIRAYRIALKRDPNNAEAWYNLGNAYYKQ GDYDEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYDEAIEYYKKALRLD PDNAKALLNLIEAILKQKLEHHHHHH**
I53_dn7A (2JFB)	MGKYDGSKLRIGILHARWNRRIILALVIGAIIRLLEFGVKEDNIIIETVPGSFELPYGS KLFVEKQKRLGKPLDAIIPIGVLIKGSTMHFEYICDSTTHQLMKLNFELGIPVIFGVL TCLTDEQAEARAGLIEGKMHNHGEDWGAAAVEMATKFN
I53_dn7B (3ltjC3_1)	MHHHHTDPLAVILYIAILKAEKSIARAKAAEALGKIGDERAVEPLIKALKDEDALV RAAAADALGQIGDERAVIPLLRALLDKEGLVRASAAIALGQIGDKRAVLILILALE

	DERDLVRVAAAVALGRIGDEKAVEPLIEALKDEEGEVREAAAIALGSIGGERVRA AMEKLAETGTGFARKVAVNYLETHKLEHHHHHH
I53_dn8A (2OBX)	**MGHHHHHHHHKDYETVRIAVVRARWHADIVRQCVMAFMKEMMRIGGRRFA VEVFDVPGAYEIPLHARTLAETGRYGAVLGTAFVVNGGIYRHEFVASAVIDG MMNVQLSTGVPVLSAVLTPHNYHDSAEHHRFFFEHFTVKGKEAARACVEILA ARERI**
I53_dn8B (1na0C3_2)	**MEEAELAYLLGELAYKLGEYRIAIRAYRIALKRDPNNAEAWYNLGNAYYKQ GDYDEAIEYYQKALELDPNNAEAWYNLGNAYYKQGDYDEAIEYYQKALELD PENEEAIDNLLEARQKQE**
I53_dn9A (2OBX)	**MGHHHHHHHHKDYETVRIAVVRARWHAEIVDVCVLAFEIEMLDIGGDRFAVD VFDVPGAYEIPLHARTLAETGRYGAVLGTAFVVNGGIYRHEFVASAVIDGMM NVQLSTGVPVLSAVLTPHNYHDSAEHHEFFFEHFTVKGKEAARACVEILAAR EKI**
I53_dn9B (HR00C3_2)	**MIEEVVAEMIDILAESSKKSIEELARAADNKTTEKAVAEAIEEIARLATAAIQLI EALAKNLASEEFMARAISAIAELAKKAIEAIYRLADNHTTDTFMARAIAAIANL AVTAILAIAALASNHTTEEFMARAISAIAELAKKAIAAIYRLADNHKTDKFMA AAIEAIALLATLAILAIALLASNHTTEEFMAKAIRAIAKLAKMAILVIYALAIM HTSPTYIEKAIEAIEKIARKAIKAIEMLAKNITTEEYKEKAKSAIDEIREKAKEA IKRLEDKRE**
I53_dn10A (2OBX)	**MGHHHHHHHHKDYETVRIAVVRARWHADIVDLCVIAFELEMLLIGGRRFAVD VFDVPGAYEIPLHARTLAETGRYGAVLGTAFVVNGGIYRHEFVASAVIDGMM NVQLSTGVPVLSAVLTPHNYHDSKRHHRFFAMHFIKKGKEAARACVEILAAR EKI**
I53_dn10B (HR00C3_2)	**MIEEVVAEMIDILAESSKKSIEELAKAADNKTTEKAVAEAIEEIARLATAAIQLI EALAKNLASEEFMARAISAIAELAKKAIEAIYRLADNHTTDTFMARAIAAIANL AVTAILAIAALASNHTTEEFMARAISAIAELAKKAIEAILELALEHETDKFMAA AIEAIALLATLAILAIALLASNHTTEEFMAKAIEAIAQLAKLAIIAIYLLALLHES PTYIEKAIEAIEKIARKAIKAIEMLAKNITTEEYKEKAKSAIDEIREKAKEAIKR LEDKRE**
I53_dn11A (2OBX)	MGHHHHHHHHKDYETVRIAVVRARWHADIVDQCVSAFEREMAKIGGDRFAVDVF DVPGAYEIPLHARTLAETGRYGAVLGTAFVVNGGIYRHEFVASAVIDGMMNVQL STGVPVLSAVLTPHEYHDSEIHHKIFFLLFTEKGKEAARACVEILAAREKI
I53_dn11B (HR00C3_2)	MIEEVVAEMIDILAESSKKSIEELARAADNKTTEKAVAEAIEEIARLATAAIQLIEAL AKNLASEEFMARAISAIAELAKKAIEAIYRLADNHTTDTFMARAIAAIANLAVTAI LAIAALASNHTTEEFMARAISAIAELAKKAIEAIYRLADNHTTDKFMAAAIEAIALL ATLAILAIALLASNHTTEEFMAKAISAIAELAKKAIEAIYRLADDHTSPTYIEKAIEA IEKIAKKAIKAIEMLAKNITTEEYQEKARKAILEILEKALEAIRRLEDNRR
T33_dn2A-SOSIP	**MKRGLCCVLLLCGAVFVSPSQEIHARFRRGARAENLWVTVYYGVPVWKDAE TTLFCASDAKAYETKKHNVWATHCCVPTDPNPQEIHLENVTEEFNMWKNN MVEQMHTDIISLWDQSLKPCVKLTPLCVTLQCTNVTNNITDDMRGELKNCSF NMTTELRDKKQKVYSLFYRLDVVQINENQGNRSNNSNKEYRLINCNTSAITQ ACPKVSFEPIPIHYCAPAGFAILKCKDKKFNGTGPCTNVSTVQCTHGIKPVVS TQLLLNGSLAEEEVIIRSENITNNAKNILVQLNESVQINCTRPNNNTVKSIRIGP GQWFYYTGDIIGDIRQAHCNVSKATWNETLGKVVKQLRKHFGNNTIIRFANS**

	SGGDLEVTTHSFNCGGEFFYCNTSGLFNSTWISNTSVQGSNSTGSNDSITLPCR IKQIINMWQRIGQAMYAPPIQGVIRCVSNITGLILTRDGGSTNSTTETFRPGGG DMRDNWRSELYKYKVVKIEPLGVAPTRCKRRVVGRRRRRRAVGIGAVSLGF LGAAGSTMGAASMTLTVQARNLLSGIVQQQSNLLRAPECQQHLLKDTHWGI KQLQARVLAVEHYLRDQQLLGIWGCSGKLICCTNVPWNSSWSNRNLSEIWD NMTWLQWDKEISNYTQIIYGLLEESQNQQEKNEQDLLELDKWASLWGSMG NLAEKMYKAGNAMYRKGQYTIAIIAYTLALLKDPNNAEAWYNLGNAAYKK GEYDEAIEAYQKALELDPNNAEAWYNLGNAYYKQGDYDEAIEYYKKALRLD PRNVDAIENLIEAEEKQGAS
T33_dn10A-SOSIP	**MKRGLCCVLLLCGAVFVSPSQEIHARFRRGARAENLWVTVYYGVPVWKDAE TTLFCASDAKAYETKKHNVWATHCCVPTDPNPQEIHLENVTEEFNMWKNN MVEQMHTDIISLWDQSLKPCVKLTPLCVTLQCTNVTNNITDDMRGELKNCSF NMTTELRDKKQKVYSLFYRLDVVQINENQGNRSNNSNKEYRLINCNTSAITQ ACPKVSFEPIPIHYCAPAGFAILKCKDKKFNGTGPCTNVSTVQCTHGIKPVVS TQLLLNGSLAEEEVIIRSENITNNAKNILVQLNESVQINCTRPNNNTVKSIRIGP GQWFYYTGDIIGDIRQAHCNVSKATWNETLGKVVKQLRKHFGNNTIIRFANS SGGDLEVTTHSFNCGGEFFYCNTSGLFNSTWISNTSVQGSNSTGSNDSITLPCR IKQIINMWQRIGQAMYAPPIQGVIRCVSNITGLILTRDGGSTNSTTETFRPGGG DMRDNWRSELYKYKVVKIEPLGVAPTRCKRRVVGRRRRRRAVGIGAVSLGF LGAAGSTMGAASMTLTVQARNLLSGIVQQQSNLLRAPECQQHLLKDTHWGI KQLQARVLAVEHYLRDQQLLGIWGCSGKLICCTNVPWNSSWSNRNLSEIWD NMTWLQWDKEISNYTQIIYGLLEESQNQQEKNEQGSGSGSGSGGEEAELAYL LGELAYKLGEYRIAIRAYRIALKRDPNNAEAWYNLGNAYYKQGDYDEAIEYY QKALELDPNNAEAWYNLGNAYYKQGDYDEAIEYYEKALELDPENLEALQNL LNAMDKQG**
I53_dn5B-SOSIP	**MKRGLCCVLLLCGAVFVSPSQEIHARFRRGARAENLWVTVYYGVPVWKDAE TTLFCASDAKAYETKKHNVWATHCCVPTDPNPQEIHLENVTEEFNMWKNN MVEQMHTDIISLWDQSLKPCVKLTPLCVTLQCTNVTNNITDDMRGELKNCSF NMTTELRDKKQKVYSLFYRLDVVQINENQGNRSNNSNKEYRLINCNTSAITQ ACPKVSFEPIPIHYCAPAGFAILKCKDKKFNGTGPCTNVSTVQCTHGIKPVVS TQLLLNGSLAEEEVIIRSENITNNAKNILVQLNESVQINCTRPNNNTVKSIRIGP GQWFYYTGDIIGDIRQAHCNVSKATWNETLGKVVKQLRKHFGNNTIIRFANS SGGDLEVTTHSFNCGGEFFYCNTSGLFNSTWISNTSVQGSNSTGSNDSITLPCR IKQIINMWQRIGQAMYAPPIQGVIRCVSNITGLILTRDGGSTNSTTETFRPGGG DMRDNWRSELYKYKVVKIEPLGVAPTRCKRRVVGRRRRRRAVGIGAVSLGF LGAAGSTMGAASMTLTVQARNLLSGIVQQQSNLLRAPECQQHLLKDTHWGI KQLQARVLAVEHYLRDQQLLGIWGCSGKLICCTNVPWNSSWSNRNLSEIWD NMTWLQWDKEISNYTQIIYGLLEESQNQQEKNEQGSGSGSGSGGEEAELAYL LGELAYKLGEYRIAIRAYRIALKRDPNNAEAWYNLGNAYYKQGRYREAIEYY QKALELDPNNAEAWYNLGNAYYERGEYEEAIEYYRKALRLDPNNADAMQNL LNAKMREELEAS**
I53_dn5B-HA	**MKAILVVLLYTFTTANADTLCIGYHANNSTDTVDTVLEKNVTVTHSVNLLED KHNGKLCKLRGVAPLHLGKCNIAGWILGNPECESLSTASSWSYIVETSNSDN GTCFPGDFINYEELREQLSSVSSFERFEIFPKTSSWPNHDSNKGVTAACPHAGA KSFYKNLIWLVKKGNSYPKLNQSYINDKGKEVLVLWGIHHPSTTADQQSLYQ NADAYVFVGTSRYSKKFKPEIATRPKVRDQEGRMNYYWTLVEPGDKITFEAT GNLVVPRYAFTMERNAGSGIIISDTPVHDCNTTCQTPEGAINTSLPFQNIHPITI GKCPKYVKSTKLRLATGLRNVPSIQSRGLFGAIAGFIEGGWTGMVDGWYGY HHQNEQGSGYAADLKSTQNAIDKITNKVNSVIEKMNTQFTAVGKEFNHLEKR IENLNKKVDDGFLDIWTYNAELLVLLENERTLDYHDSNVKNLYEKVRNQLKN**

NAKEIGNGCFEFYHKCDNTCMESVKNGTYDYPKYSEEAKLNREKIDGVSAEE AELAYLLGELAYKLGEYRIAIRAYRIALKRDPNNAEAWYNLGNAYYKQGRY REAIEYYQKALELDPNNAEAWYNLGNAYYERGEYEEAIEYYRKALRLDPNNA DAMQNLLNAKMREEGGWELQHHHHHH

Appendix Table 2.3. Biophysical attributes of designed trimers (top) and two-component nanoparticles (bottom) compared to experimentally-measured data (exp). Molecular weights (MW) were obtained from the ASTRA software. R_g and D_{max} calculations obtained from Scatter3 SAXS analysis software with the determined q_{max} values. *X* values computed from the FoXS online SAXS web server between the designed model and the experimental scattering data.

Design	MW kDa (A)	MW kDa (B)	MW kDa (model)	MW kDa (exp)	Rg Å (Model)	Rg Å (Exp)	Dmax Å (Model)	Dmax Å (exp)	*X*	qmax 1/nm
1na0C3_2	14.99	-	44.97	48	26.4	29.5	84	86	1.4	0.23
3ltjC3_1v2	20.90	-	62.69	56	27.7	31.3	88	94	1.1	0.18
3ltjC3_11	22.21	-	66.62	50	28.3	30.3	87	92	1.6	0.20
HR04C3_5v2	23.05	-	69.14	71	25.9	28.6	82	86	1.5	0.25
T33_dn2	13.82	14.88	344.45	397	61.4	64.7	169	169	4.8	0.17
T33_dn5	13.72	21.49	422.42	422	66.7	69.4	177	193	1.7	0.16
T33_dn10	14.07	31.42	545.88	556	62.3	60.1	169	170	2.3	0.20
I53_dn5	17.19	15.33	1951.57	2000	95.9	97.1	241	243	1.2	0.21

Appendix Table 2.4. Crystallography table for data collected and refinement statistics for designs 1na0C3_2 and 3ltjC3_1v2. Statistics for the highest-resolution shell are shown in parentheses.

	1na0C3_2 (PDB ID)	**3ltjC3_1 (PDB ID)**
Data Acquisition		
Space group	P 1 21 1	R 3 :H
Cell dimensions a, b, c (Å) α , β , γ (°)	 69.47, 64.91, 99.03 90, 106.33, 90	 88.182, 88.182, 65.244 90, 90, 120
R_{merge}	0.1427 (0.8023)	0.1189 (1.041)
$CC_{1/2}$	0.994	0.997 (0.623)
$\langle I/\sigma_I \rangle$	9.5 (2.16)	9.35 (1.30)

Completeness (%)	99.86	91.51 (77.88)
Multiplicity	4.6 (4.6)	5.6 (5.1)
Wilson B-factor ($Å^2$)	37.11	40.43
Refinement		
Resolution range (Å)	38.34 - 2.53 (2.62 - 2.53)	44.09 - 2.303 (2.386 - 2.303)
No. of reflections	28506 (2837)	46901 (4290)
R_{work}(%)/ R_{free}(%)	0.2097 (0.2841) / 0.2556 (0.3460)	0.2021 (0.2555) / 0.2261 (0.2240)
Average B-factors ($Å^2$) Protein Water	 52.08 43.30	 49.36 50.80
r.m.s.d. deviations Bond length (Å) Bond angles (°)	 0.005 0.95	 0.002 0.44
Ramachandran favored (%)	99.00	100.00
Ramachandran outliers (%)	0.00	0.00

Appendix Table 3.1. Data collection and refinement statistics for E2/EpoR and dimer C_R3. Statistics for the highest-resolution shell are shown in parentheses.

[a] These statistics are for data that were truncated by STARANISO to remove poorly measured reflections affected by anisotropy.

[b] The resolution limits for three directions in reciprocal space are indicated here. To accomplish this, STARANISO computed an ellipsoid postfitted by least squares to the cutoff surface, removing points where the fit was poor. Note that the cutoff surface is unlikely to be perfectly ellipsoidal, so this is only an estimate.

[c] The anisotropic completeness was obtained by least squares fitting an ellipsoid to the reciprocal lattice points at the cutoff surface defined by a local mean $I/\sigma I$ threshold of 1.0, rejecting outliers in the fit due to spurious deviations (including any cusp), and calculating the fraction of observed data lying inside the ellipsoid so defined. Note that the cutoff surface is unlikely to be perfectly ellipsoidal, so this is only an estimate.

	E2/EpoR	C_R3
Wavelength (Å)	1.033202	0.97946
Number of crystals	1	1
Resolution range (Å)	29.76 - 2.091 (2.166 - 2.091)	33.25 - 1.21 (1.253 - 1.21)
Ellipsoidal[a] resolution limit (Å) (direction)[b]	-	-

Space group	P 21 21 21	P 1 21 1
Unit cell (a, b, c in Å; α, β, γ in °)	69.25, 122.16, 132.57; 90, 90, 90	36.25, 89.81, 44.27; 90, 99.132, 90
Total reflections	913420 (90406)	564020 (55265)
Unique reflections	67078 (6559)	83274 (8248)
Multiplicity	13.6 (13.8)	6.8 (6.7)
Completeness (%)	99.40 (96.63)	88.20 (72.00)
Completeness (ellipsoidal)[c] (%)	-	-
Mean I/sigma(I)	12.15 (0.50)	14.80 (1.34)
Mean I/sigma(I) (ellipsoidal)[a]	-	-
Wilson B-factor	52.59	14.14
R-merge	0.1322 (2.671)	0.05076 (1.564)
R-pim	0.03702 (0.7394)	0.0208 (0.6433)
CC1/2	0.999 (0.693)	0.999 (0.669)
Resolution range of final refinement (Å)	29.76 – 2.091	33.25 - 1.21
Reflections used in refinement	66813 (6428)	74921 (6106)
Reflections used for R-free	1997 (208)	1670 (144)
R-work/R-free	0.2036/0.2314	0.1414/0.1710
Number of non-hydrogen atoms	6295	2644
macromolecules	5751	2411
ligands	51	50
solvent	493	183
RMS(bonds) (Å)	0.003	0.009
RMS(angles) (°)	0.61	1.03
Ramachandran favored (%)	97.59	99.36
Ramachandran allowed (%)	2.28	0.64
Ramachandran outliers (%)	0.13	0.00
Rotamer outliers (%)	1.84	1.18
Clashscore	2.27	4.35
Number of TLS groups	22	
PDB ID	6MOE	6MOG

Appendix Table 3.2. Data collection and refinement statistics for C_R3/EpoR, C_angle_R5/EpoR, A_angle_R5/EpoR, A_dist_R5/EpoR, and M_12/EpoR. Statistics for the highest-resolution shell are shown in parentheses.

[a] These statistics are for data that were truncated by STARANISO to remove poorly measured reflections affected by anisotropy.

[b] The resolution limits for three directions in reciprocal space are indicated here. To accomplish this, STARANISO computed an ellipsoid postfitted by least squares to the cutoff surface, removing points where the fit was poor. Note that the cutoff surface is unlikely to be perfectly ellipsoidal, so this is only an estimate.
[c] The anisotropic completeness was obtained by least squares fitting an ellipsoid to the reciprocal lattice points at the cutoff surface defined by a local mean $I/\sigma I$ threshold of 1.0, rejecting outliers in the fit due to spurious deviations (including any cusp), and calculating the fraction of observed data lying inside the ellipsoid so defined. Note that the cutoff surface is unlikely to be perfectly ellipsoidal, so this is only an estimate.

	C_R3/EpoR	C_angle_R5/ EpoR	A_angle_R5/ EpoR	A_dist_R7/ EpoR	M_R12/E poR
Wavelength (Å)	0.999924	0.97741	0.97741	1.00001	1.033202
Number of crystals	2	1	1	1	1
Resolution range (Å)	68.02 - 3.2 (3.315 - 3.2)	48.821 – 2.065 (2.421 – 2.065)	48.739 – 2.431 (2.730 – 2.431)	46.793 - 5.101 (5.283 - 5.101)	44.969 – 3.163 (3.400 – 3.163)
Ellipsoidal[a] resolution limit (Å) (direction)[b]	3.469 (**a***) 3.469 (**b***) 3.167 (**c***)	3.026 (0.894 **a*** - 0.447 **b***) 3.026 (**b***) 2.065 (**c***)	3.387 (**a***) 3.387 (**b***) 2.448 (**c***)	-	3.144 (0.894 **a*** - 0.447 **b***) 3.144 (**b***) 4.58 (**c***)
Space group	P 43 21 2	P 65 2 2	I 41 2 2	P 62 2 2	P 63
Unit cell (a, b, c in Å; α, β, γ in °)	96.1931, 96.1931, 286.401; 90, 90, 90	112.748, 112.748, 280.793; 90, 90, 120	130.42, 130.42, 293.454; 90, 90, 90	117.09, 117.09, 233.587; 90, 90, 120	168.109, 168.109, 78.029; 90, 90, 120
Total reflections	205473 (6189)	1101540 (37236)	652937 (32944)	41433 (4116)	126061 (4958)
Unique reflections	18878 (944)	29498 (1475)	25096 (1280)	4209 (395)	14501 (724)
Multiplicity	11.2 (11.3)	37.3 (25.2)	26.0 (25.7)	9.8 (10.4)	8.7 (6.8)
Completeness (%)	79.96 (17.52)	45.01 (1.59)	52.4 (9.2)	99.4 (98.74)	66.8 (17.3)
Completeness (ellipsoidal)[c] (%)	92.1 (71.9)	93.67 (73.00)	95.1 (78.7)	-	92.5 (57.4)
Mean I/sigma(I)	3.07 (0.56)	5.48 (-0.01)	11.04 (0.46)	9.92 (0.90)	6.2 (0.24)
Mean I/sigma(I) (ellipsoidal)[a]	3.7 (1.4)	9.4 (1.6)	14.7 (1.5)	-	7.6 (1.0)
Wilson B-factor	29.34	44.32	44.92	337.23	109.20
R-merge	0.8886 (5.596)	0.4664 (31.43)	0.273 (2.759)	0.116 (2.75)	0.198 (2.046)
R-pim	0.2813 (1.771)	0.08034 (6.339)	0.054 (0.551)	0.03905 (0.884)	0.070 (0.832)
CC1/2	0.965 (0.552)	1 (0.151)	0.999 (0.517)	0.996 (0.585)	0.995 (0.491)

Resolution range of final refinement (Å)	68.019 – 3.200	48.821 – 2.065	48.739 – 2.431	41.35 - 5.101	44.969 – 3.163
Reflections used in refinement	18483 (398)	29491 (102)	25133 (692)	4194 (392)	14455 (122)
Reflections used for R-free	891 (27)	1573 (6)	1278 (27)	420 (39)	1368 (9)
R-work/R-free	0.2390/0.2820	0.2145/0.2507	0.2164/0.2563	0.2527/0.3012	0.2212/0.2637
Number of non-hydrogen atoms	5510	3391	3366	3739	6519
macromolecules	5468	3275	3296	3739	6511
ligands	36	82	48		6
solvent	6	34	22		2
RMS(bonds) (Å)	0.004	0.004	0.004	0.008	0.004
RMS(angles) (°)	0.99	0.92	0.95	1.60	0.90
Ramachandran favored (%)	95.35	97.64	96.80	96.94	94.57
Ramachandran allowed (%)	4.51	2.36	3.20	2.65	5.43
Ramachandran outliers (%)	0.14	0.00	0.00	0.41	0.00
Rotamer outliers (%)	3.37	1.48	1.79	3.95	1.84
Clashscore	4.68	3.88	5.77	21.64	4.05
Number of TLS groups	22	14	13		24
PDB ID	6MOH	6MOI	6MOJ	6MOK	6MOL

VITA

Born in Seattle, WA, George completed his undergraduate degree at the University of Washington (UW) in 2013, where he worked as a tutor and in another laboratory in the Seattle Institute of Biomedical and Clinical Research studying the specifics of glucose-stimulated insulin secretion and implications in diabetes. After entering the Biochemistry graduate program at the UW, he refocused his research interests on computational structural biology to study and engineer protein structures under the supervision of David Baker, Professor of Biochemistry and Director of the Institute for Protein Design. George's primary scientific expertise now lies in computational protein engineering & design with the Rosetta Macromolecular Modeling Suite, as well as experimental preparation and performance. During the summer of 2017, he took leave from graduate school for an internship at Amgen (South San Francisco), obtaining substantial exposure to state-of-the-art tools and methods in the biotech industry. Combining this knowledge with the experience and field exposure acquired throughout graduate school, he is now making efforts to develop a new generation of protein-based biotherapeutics. George also enjoys putting biochemistry to work both in the lab and in kitchen - cooking (and eating) is a favorite hobby, among others such as cycling, gaming, and travel.