De novo design of obligate ABC heterotrimeric proteins

Sherry Bermeo

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

David Baker, Chair

Frank DiMaio

Dustin Maly

Program Authorized to Offer Degree:

Biochemistry

University of Washington

Abstract

De novo design of obligate ABC heterotrimeric proteins

Sherry Bermeo

Chair of the Supervisory Committee:
David Baker
Department of Biochemistry

The self-association of three distinct proteins, or essentially an ABC heterotrimer, is seen in a few instances in nature but the challenge of designing cooperative 3-sided interfaces has not been tackled yet in the field of protein design. There has been increasing interest towards designing much more diverse higher-order architectures for downstream applications that require the presentation or recruitment of multiple proteins, which could be aided by the specific assembly of a heterotrimer. Therefore, we sought to design cooperative ABC heterotrimeric proteins such that chains A, B, and C are unique and different with respect to amino acid sequence. Inspired by the ease of parametric sampling for coiled coils[1] and previous success of burying hydrogen bonds to mediate homo-oligomer[2] and heterodimer[3] specificity, we strategically searched for buried networks accommodating nonpolar aromatic residues like tyrosine and tryptophan, while

also enhancing for phenylalanine during core packing to create de novo ABC heterotrimeric coiled coils and helical bundles. We hypothesized that the inclusion of these three large aromatic residues, supplemented by tight aliphatic packing, could help drive the desired ABC assembly. These designs were experimentally characterized through size exclusion chromatography, native mass spectrometry, and small angle x-ray scattering. Two of these successful “base” constructs were subsequently parsed through a library of monomeric designed helical repeat (DHR) proteins to ensure that these heterotrimers could be extended while preserving their cooperative specificity. These single stranded and helical bundle heterotrimers DHR extensions or “arms” were then used to design cyclic geometries to highlight the different stoichiometries that can result from the same starting ABC interface, enabling these heterotrimers to serve as potential building blocks for higher-order assembly or as scaffolds for multivalent display.

TABLE OF CONTENTS

List of Figures

ACKNOWLEDGEMENTS

A big thank you to all past and present members of the Baker Lab & Institute for Protein Design, the BPSD program, the biochemistry department, and the Foster School of Business. An extra special thank you to my two postdoc mentors, Tom Linsky and Scott Boyken, for always being super supportive; to Rubul Mout, Danny Sahtoe, and Yang Hsia for kicking some sense into me at a time when I really needed it and for sharing their wisdom throughout the years; to our pre-covid lab volleyball team and all our fun games; to Parisa Hosseinzadeh, Una Nattermann, Gabe Rocklin, Zibo Chen, Hua Bai, Sergey Ovchinnikov, Chris Bahl, and others who made me feel included when I first joined the IPD; and to Andrew Favor, Andrew Norris, and Atty Chang who have all been a joy to work with and have been instrumental in getting this project done. And thank you to all my committee members (especially my advisor) for their feedback, kindness, and encouragement, this degree really could not have been completed without them. Thank you to my friends in Seattle and NYC who have kept me sane and made me laugh all these years. And an extra special shout out to Julia, Alejandra, and Manal for listening with a patient ear and keeping me level headed. Thank you to this pandemic for teaching me the importance of living in the present and changing my perspective on what it means to succeed. Thank you to all my K-12 teachers for instilling an appetite for learning and to my athletic coaches who taught me the importance of accountability and perseverance, they are the reason why I've made it so far in my education. Thank you to the beautiful mountains of Washington, my hometown, and my faith for keeping me grounded. And of course, thank you to my parents and my two younger brothers for all their hugs and for our continuing family dinners. Who I am today and what I have accomplished so far is a reflection of their hard work, love, and support.

Chapter 1. Introduction

Proteins are the building blocks of life and must fold up into specific structures in order to carry out their respective biological functions. This seemingly spontaneous form of self-assembly can actually be classified as a hierarchy with increasing complexity: primary, secondary, tertiary, and quaternary structure. Over half of the proteins found in the Protein Data Bank (PDB) fall at the top of the hierarchy (Figure 1)[4], meaning that two or more subunits assemble to form homo-oligomers or hetero-oligomers, with the former actually being much more commonly found in nature[5]. The smallest oligomer, dimers, are frequently found in nature with homodimers being more abundant than heterodimers[6]. Higher-order stoichiometries seem to be less populated overall, which may potentially be in part due to difficulties in structure determination, but is more likely a result of their unnecessary need for nature's desired functions.

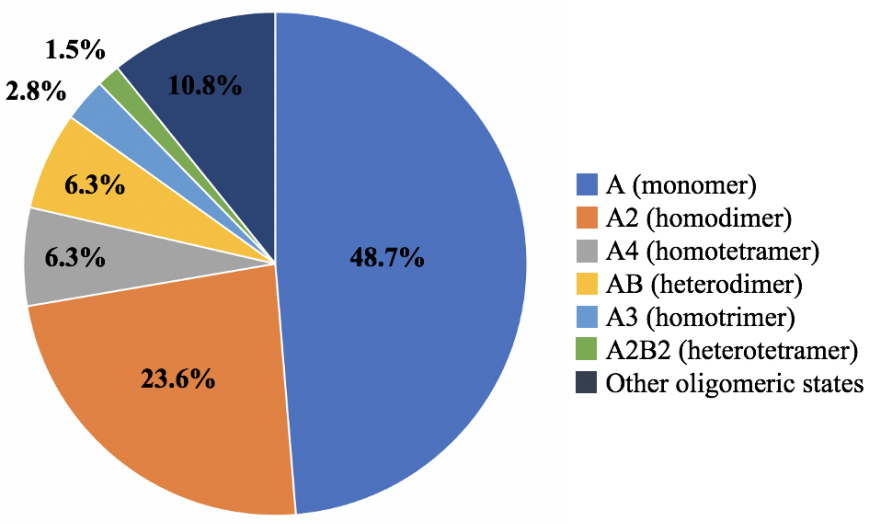


Figure 1. Breakdown of protein oligomeric states found in the PDB. The seven most common stoichiometries are shown, with more than half of the proteins forming heteromers with ranging stoichiometries. Made using statistics presented by Xu and Dunbrack, 2019[4].

1.1 Heterotrimeric Proteins In Nature

Heterotrimeric proteins can be classified by two stoichiometries, AAB/ABB and ABC, but these three component proteins are not as abundant as their oligomeric counterparts. Yet two notable examples in nature that rely on the specific interaction of three components include the G-proteins and laminins, shown in Figure 1.1. Both proteins are ABC heterotrimers since their alpha, beta, and gamma subunits are all different.

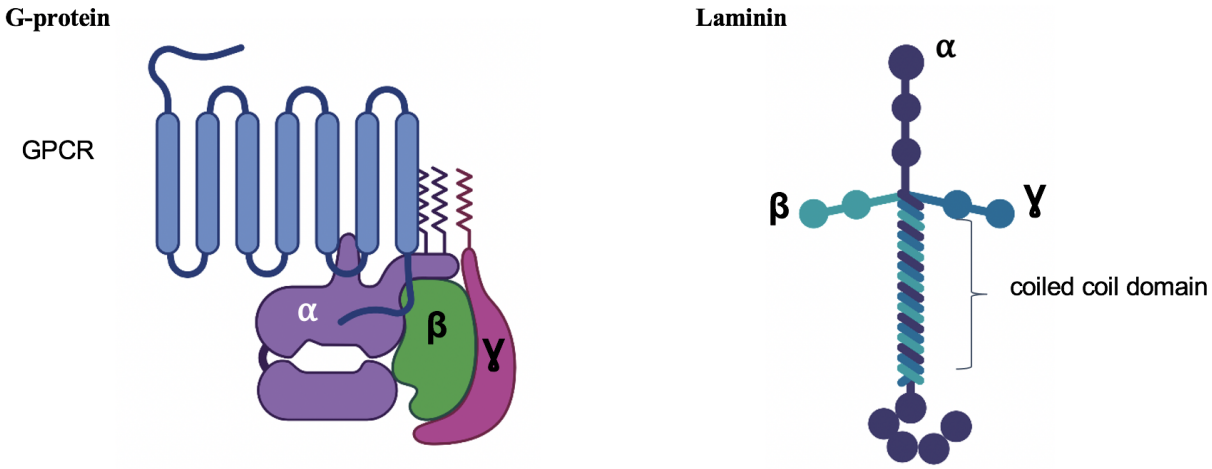


Figure 1.1. G-proteins and laminins. There are two examples of ABC heterotrimeric (three component) proteins in nature, found in different environments and which carry out their own respective functions. Created with BioRender.com.

G-proteins are a family of proteins made up of alpha, beta, and gamma subunits that interact with G protein-coupled receptors (GPCRs) to initiate a signaling cascade in the cell. When GPCR is activated, the alpha unit undergoes a conformational change and dissociates from the beta-gamma heterodimer and then continues to carry out its function[7]. Meanwhile, laminins are heterotrimeric coiled coil glycoproteins that play an integral role in the basement membrane, with a total of 5 different alpha, 3 different beta, and 3 different gamma subunits found in

mammals that combine specifically to carry out the needed function. Although many different combinations between all three chains are possible, only 16 isoforms have been characterized[8]. This could either imply that biochemical characterization of these laminin heterotrimers is quite difficult or that not all combinations actually form in nature, resulting in very specific assemblies that likely arose out of the need for very specific functions.

1.2 Previous attempts at de novo ABC heterotrimer design

In addition, two previous attempts have been made to design fully de novo ABC heterotrimers in an effort to gain a better understanding of the packing requirements needed to achieve this by using coiled coils as a starting scaffold.

The first attempt resulted in only 1 working ABC heterotrimeric coiled coil, which was composed of three peptides with 4.5 heptad repeats each[9]. These peptides were synthesized and mixed together, with a crystal structure indicating the importance of ion pairs at the 'e' and 'g' positions for determining ABC specificity[10]. The second attempt using de novo design by another group resulted in three ABC antiparallel heterotrimeric coiled coils, composed of peptides about 30 residues in length. In this case, the overall goal was to use these coiled coils as asymmetric metal binding sites (which was not observed), but they concluded that specificity was again mediated by favorable electrostatic interactions between the helices[11]. Although the success rate was quite low for these two attempts, it showcased the ability to use these coiled coils as a starting scaffold for heterotrimer design.

1.3 Advantages of designing cooperative ABC heterotrimers

Notably ABC heterotrimers are an interesting design challenge because all three unique components must interact precisely for them to assemble to the desired unit. Although an ABC heterotrimer only has one extra component compared to an AB heterodimer, this design problem is actually much more complex due to possible assembly combinations: an AB heterodimer only has 4 other species it could form (A, B, AA, BB), while an ABC heterotrimer has 15 alternate species (A, B, C, AB, AC, BC, AAA, BBB, CCC, AAB, ABB, AAC, ACC, BBC, BCC). Because an ABC heterotrimer devoid of heterogeneity would be easier to work with for downstream building applications, we sought to design heterotrimers that have cooperative binding properties like that of homo-oligomers[12], such that only the ABC species would form. This would be particularly useful for higher-order design applications downstream because cooperativity can help drive that needed ABC specificity over undesired and alternative assemblies and thus open up new avenues towards the design of much more diverse multi-component architectures.

Using a small scaffold like coiled coils and helical bundles also allows for versatility in applications, facilitating their use for biological purposes without concern over genome size/limitations and enabling them to serve as vectors for displaying multiple proteins of interest by capitalizing on available termini and placement.

Chapter 2. Computational Design of ABC Heterotrimeric bases and arms

By combining the ease of parametric sampling[1], the burial of hydrogen bond networks (which have been previously successful for oligomer design)[2,3], and lessons learned from previous heterotrimer design efforts, we set up a Rosetta pipeline to design cooperative ABC heterotrimeric coiled coils and helical bundles.

We hypothesized that the inclusion of large aromatic nonpolar residues, such as tyrosine (TYR), tryptophan (TRP), and phenylalanine (PHE) in the core would help specific assembly if these amino acids were packed tightly and properly buried. We strategically searched for hydrogen bond networks that would include TYR and TRP, packed around these residues with typical aliphatic nonpolar residues, and then allowed a redesign that required PHE while finding the best adjacent residues to help create this tight packing. Specifically, the use of TYR and TRP was of interest because of its potentially useful "two-fold" entropic cost: its polar groups would want to remain close to buried polars to form hydrogen bonds and their large size would likely help contribute to either steric clashing or cavities depending on what other alternate assemblies could arise. Acidic charged residues, such as aspartate (ASP) and glutamate (GLU), were also considered during the hydrogen bond search because their small size would make them easy to pack around and the burial of a charged residue without its respective hydrogen bonding partners at the core could lead to a destabilization in assembly.

We then built upon these ABC heterotrimers with rigid fusions to other de novo monomeric and heterodimeric proteins to showcase their ability to handle modifications and

contribute as building blocks for higher-order design purposes while still preserving their specificity.

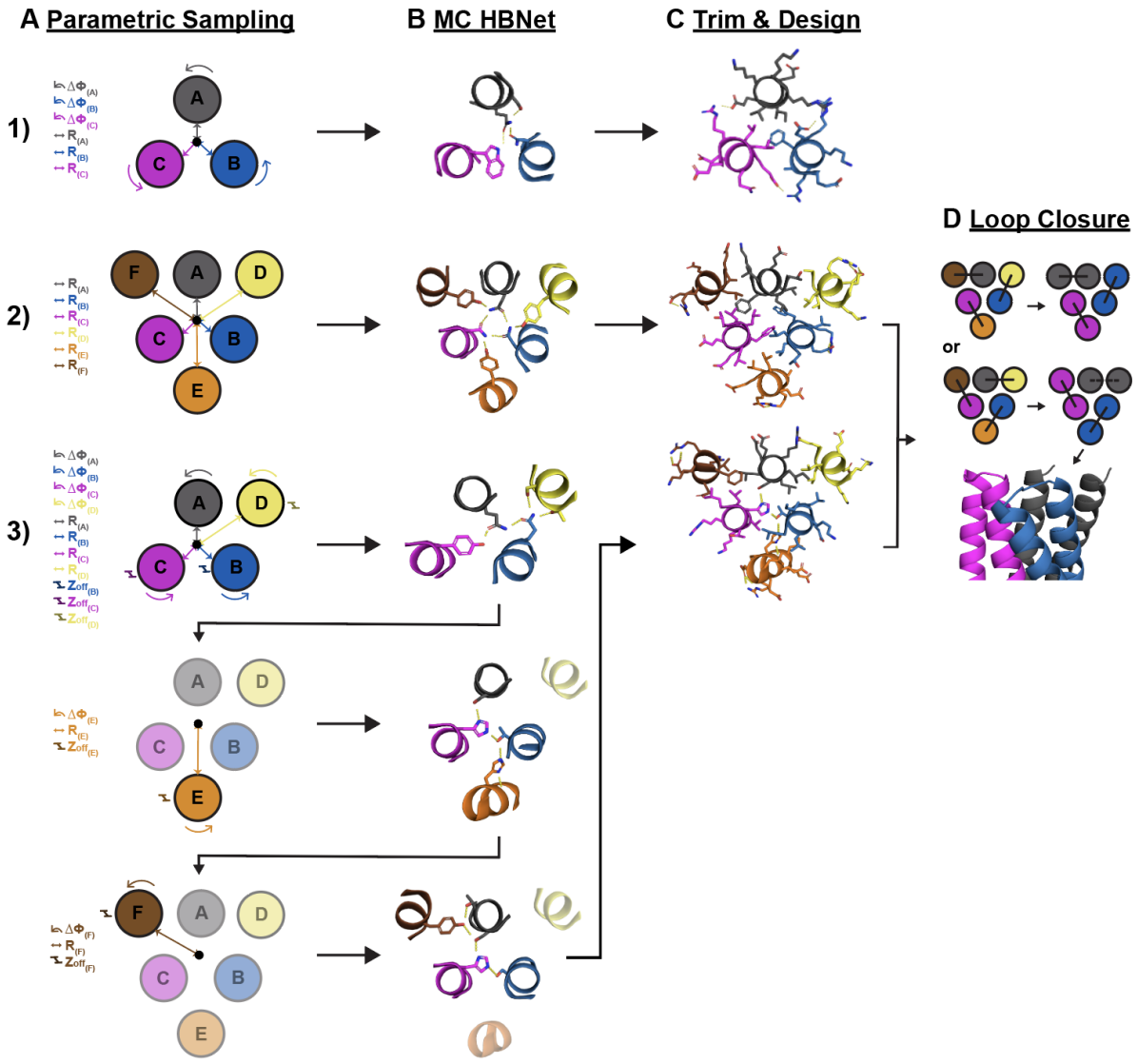


Figure 2. Computational design methodology for ABC heterotrimer bases. (A) Poly-alanine helical backbones are sampled three ways across a number of parameters, as listed on the left. These are then passed through (B) a monte carlo hydrogen bond network search (MC HBNet). Buried networks that have at least two large aromatic residues can be (C) optionally trimmed, packed with required PHE and other nonpolar residues at the core, and decorated with charged/polar residues at the surface to enhance electrostatic interactions across the chains. (D) Chains are then closed for routes 2 and 3 to form three helical hairpin units, which would make the base heterotrimer. Two possible loop combinations are shown for brevity, with loops all facing the same direction (solid lines) or loops facing different directions (dashed lines).

2.1. DESIGN OF SINGLE STRANDED COILED COIL HETEROTRIMERS

The simplest case of an ABC heterotrimer would be a three stranded unit (or also known as a triple coil), in which each chain is a single helix (Fig. 2a, Fig 2.1). A generalized Crick coiled coil parameterization approach[14] was used to sample each individual helix across two parameters, supercoil radius (R) and helical phase ($\Delta\phi_1$). The supercoil phases ($\Delta\phi_0$) were restricted to 0º, 120º, 240º, while the supercoil (ω_0) and helical twist (ω_1) were kept at ideal values -2.85 and 102.85, respectively. This would generate left-handed supercoils with 3.5 residues per turn or a 7 residue (heptad) periodicity across two turns[15]. Helix termini were kept in the same direction for a parallel orientation, while the third helix at supercoil phase 240º was inverted for an antiparallel orientation.

These poly-alanine backbones were then coupled to Rosetta Monte Carlo HBNet[16], an algorithm used to place residues with polar groups across the interface such that all heavy atom donors and acceptors form hydrogen bond networks. A total of three networks were simultaneously searched across the backbone (Fig. 2b), such that the final designs would have a core N-M-N-M-N-M-N stacking pattern (N = nonpolar, M = mixed or polar/nonpolar), and a final count of at least two tyrosine (TYR) and/or tryptophan (TRP) residues. Following this search, the helices for chains B and A were trimmed by one and two heptads, respectively, to make it easier to keep track of each chain during downstream characterization and to also allow for additional electrostatic interactions between each chain. The hydrogen bond networks were then constrained as RosettaDesign[17] was used to find optimal packing around these residues (Fig. 2.1), with the hypothesis that fully hydrophobic heptads above and below the networks would help pad and keep these hydrogen bonding residues in place. A scoring term was used to enforce

phenylalanine (PHE) at the core by penalizing deviations from the specified PHE count during interface design and allow for nonpolar aliphatic redesign around PHE to accommodate this residue and enable sufficient packing.

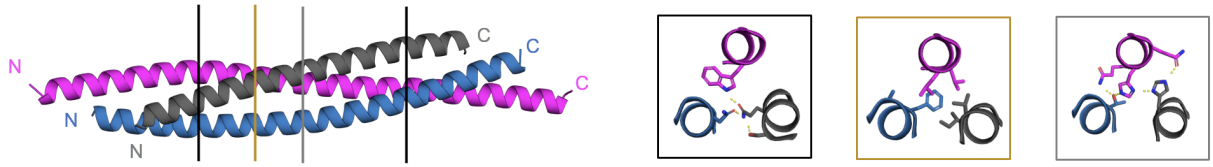


Figure 2.1. An example of a parallel single stranded heterotrimer coiled coil. Colored cross-sections across the single stranded heterotrimer show hydrogen bond networks touching all helices and the presence of aromatic residues across 4 of the 7 heptads that make up the shared ABC interface.

2.2 DESIGN OF HELICAL BUNDLE HETEROTRIMERS

The computational approach for designing heterotrimer coiled coils was modified to include more helices, meaning that the starting unit for each chain would result in a helical hairpin (or a helix-loop-helix motif). Two sampling approaches were taken, the first being a limited sampling approach with 5 out of 6 helices sampled at once and the second was a more in-depth sampling approach taken by breaking up the search problem, as seen in routes 2 and 3 in Figure 2.

Much like the search for coiled coil backbones, the first limited sampling approach only sampled across 2 parameters, supercoil radius (R) and helical phase ($\Delta\phi_1$), for each helix[14]. The supercoil phases ($\Delta\phi_0$) were restricted to 0º, 60º, 120º, 180º, 240º, 300º, while the supercoil (ω_0) and helical twist (ω_1) were kept at ideal values as before to make left-handed coiled coils. All poly-alanine helices were sampled in a parallel orientation and then coupled to Monte Carlo HBNet[16] to find hydrogen bond networks that would satisfy a core N-M-N-M-N stacking pattern.

Each of the 6 helices were required to contribute at least residue with a polar group and the final scaffold also needed a final count of at least two tyrosine and/or tryptophan residues (Fig. 2).

Meanwhile for the more in-depth sampling approach, the first four helices were set to the same supercoil phases ($\Delta\phi_0$) but all four helices were sampled more extensively across 3 parameters, supercoil radius (R), helical phase ($\Delta\phi_1$), and z-offset (Z-off). These four helices were then coupled to Monte Carlo HBNet, requiring a 4-helix network with at least one tyrosine or tryptophan. Backbones that passed this criteria were again passed through parametric sampling to add a fifth helix at supercoil phase 180º and HBNet to find a hydrogen bond network that spanned the first 3 inner helices previously sampled and the new helix added, followed by one more parametric search at supercoil phase 300º coupled to HBNet to specify a hydrogen bond network spanning the first 3 inner helices and the new 6th helix (Fig. 2). Thus the final backbone and HBNet search would yield a core N-M-M-M-N stacking pattern, meaning that 3 of the 5 heptads had residues contributing polar groups for hydrogen bonding.

Three inner helices in the same orientation resulted in a parallel heterotrimer, while an inversion of the 3rd inner helix at supercoil phase 240º resulted in an antiparallel heterotrimer. Backbones were then trimmed and passed through the same downstream packing protocol as the coiled coil heterotrimer, while keeping the hydrogen bond networks constrained. Because of the wider hydrophobic interface, a Developability Index[18] metric added to Rosetta was used to monitor the overall hydrophobic nature of these designs in an effort to prevent sticky mis-assemblies and aggregation. To create the final heterotrimer components, an inner and outer helix were connected with a loop in either a clockwise or counterclockwise orientation (Fig. 2d). The output was filtered by Rosetta energy, secondary structure shape complementarity, tight

packing around the networks, hydrogen bond network satisfaction[19], and high lddt value according to Deep Accuracy Net[20].

2.3 Design of heterotrimer arms using helical repeat proteins

To enable the use of heterotrimers for further downstream design, while also increasing diversity in shape and geometry of each chain, monomeric de novo helical repeat proteins (DHRs)[21] were rigidly fused onto each chain, thereby resulting in heterotrimer constructs being referred to as 1-arm (1 DHR fusion), 2-arm (2 DHR fusions), 3-arm (3 DHR fusions), and 4-arm (4 DHR fusions) heterotrimers. Repeat proteins are an attractive choice for fusion because their tandem repeat units help them serve as protein connectors. They have also been successfully used in conjunction with other de novo helical bundle homo-oligomers and heterodimers previously.[22,23,24]

The number of DHRs that can be added to the heterotrimers is dependent on the number of termini available for fusion and the spacing between each termini to avoid steric clashing. The single stranded ABC heterotrimer has six distinctly placed N and C termini, all of which could be used to flexibly or rigidly fuse proteins. A computational protocol[22] was used to rigidly fuse a library of designed helical repeat proteins onto the available termini of the single stranded heterotrimer. An overlap region of up to 7 residues (a heptad) was searched between the heterotrimer/DHR and those with a low RMSD were accepted (Fig. 2); fusions were found at 5 of the 6 available termini for this construct. We hypothesized that a higher number of repeat protein fusions, especially those making contacts with the heterotrimer backbone or with each other, could help enhance ABC specificity. The best scoring single fusions according to lddt[19] value were then combined to form multiple arms.

While the single stranded coiled coil heterotrimer could theoretically sustain up to six fusions or protein additions due to available termini, helical bundle heterotrimers either have four or all six termini clustered in the same orientation and in close proximity, thereby usually limiting it to a maximum of only 3 potential fusions.

Chapter 3. Experimental validation of ABC Heterotrimeric bases and arms

3.1 Experimental validation of single stranded heterotrimers and arms

3.1.1 *Structural characterization of single stranded heterotrimers and arms*

A total of 20 single stranded coiled coil heterotrimer designs were ordered for tricistronic expression (co-expression) in *Escherichia coli* (*E.coli*) and purified via immobilized metal affinity chromatography (IMAC) pulldown with only one chain having a 6xHis-tag. From this set, 19/20 designs were soluble, 6 had the presence of all three components by liquid chromatography-mass spectrometry (LC-MS), 5 eluted as monodisperse peaks by SEC, but only 1 was an exclusive ABC heterotrimer by native mass spectrometry (Fig. 3.1.1 c) and had good agreement with the design model via small angle X-ray scattering (SAXS) (Fig. 3.1.1 d). Although the heterotrimer was only a three stranded coiled coil, it was determined to be helical and thermostable up to 95ºC by circular dichroism (Fig. 3.1.1 e). Aside from being able to obtain a successful ABC interface, the stability of this coiled coil is important as it demonstrates that an outer helix per unit (resulting in a helix-loop-helix motif) was not necessary for stability. The loss of helical structure was previously noted when helical bundle homo-oligomer designs were reduced to single stranded coils, which perhaps is unsurprising given that the interface was

optimized for a larger bundle, meaning that this much more basic starting scaffold was not considered when designing helical bundle homo-oligomers[2] or heterodimers[3].

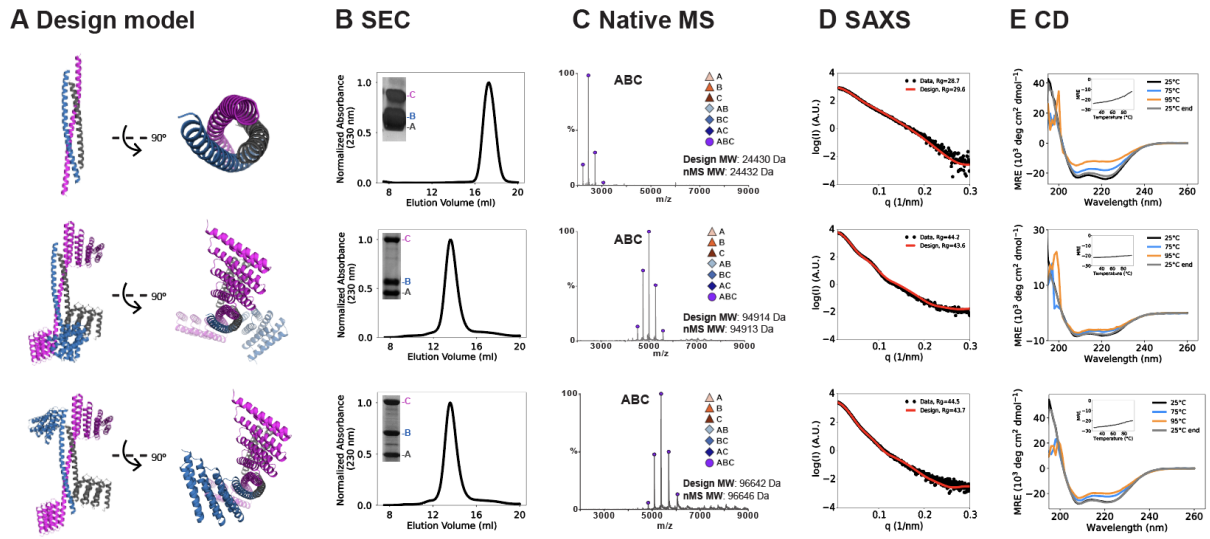


Figure 3.1.1. Experimental characterization of a co-expressed parallel ABC heterotrimeric single stranded coiled coil and subsequent 4-arm fusions. (A) Design models of a single stranded heterotrimer and its two 4-arms fusions are colored by chain and shown in cartoon representation. (B) SEC traces and SDS-page gel (inset) show 3 chains eluting at a monodisperse peak, (C) native mass spec shows only the ABC heterotrimer forming, (D) SAXS profiles indicate a good fit between the design model and experimental data collected, and (E) CD spectra shows designs are helical and thermal melt (inset) indicates they are stable up to 95ºC.

Following successful low-resolution characterization of this ABC coiled coil, the construct was parsed through HelixFuse[22] to add repeat proteins as "arms." A small set of 7 designs were ordered, two 3-arm constructs, four 4-arm constructs, and one 5-arm construct, with only one chain having the 6xHis-tag for IMAC pull-down. A Strep tag was added to a second chain in case another pull-down was needed to alleviate potential expression imbalance and to ensure ABC affinity regardless of which chain had a tag. Both pulldown experiments resulted in the presence of 3 components by SDS-page gel (not shown). Two 4-arm heterotrimers retained exclusive ABC heterotrimer assembly via SEC, Native MS, SAXS, and CD, with an

inset temperature scan indicating both arms are more stable than the original coiled coil heterotrimer when heated up to 95ºC (Fig 3.1.1 a-e). It is not quite clear why the varying number of 3, 4, or 5-arm fusions results in differences in ABC affinity or other non-ABC states for this scaffold, but this could perhaps be due to the general sensitivity of coiled coils to changes in amino acid sequence or issues with individual single fusions (not tested). Nonetheless it is promising that this heterotrimer can sustain rigid repeat protein fusions as it suggests that the base coiled coil construct is likely folding up as designed.

3.1.2 *Using a helical wheel diagram to understand heterotrimer packing*

A helical wheel[25] (Fig. 3.1.2 a,b) was made to compare the packing of this de novo single stranded heterotrimer to that of a classical coiled coil[26], which shows hydrophobic residues at positions ‘a’ and ‘d’ and charged residues mediating electrostatic interactions at positions ‘e’ and ‘g.’ Because the chains are all different sizes in this design, the helical wheel only took into account the shared interface between all three chains across 7 heptads. The residues contributing to the designed hydrogen bond networks were found across positions ‘a,’ ‘d,’ ‘e,’ and ‘g’ and are highlighted in bold as shown in Figure 3.1.2 a & b. Position ‘g’ had the most variability among the chains, being fully hydrophobic on chain B, mixed (nonpolar and polar) on chain A, and fully polar on chain C.

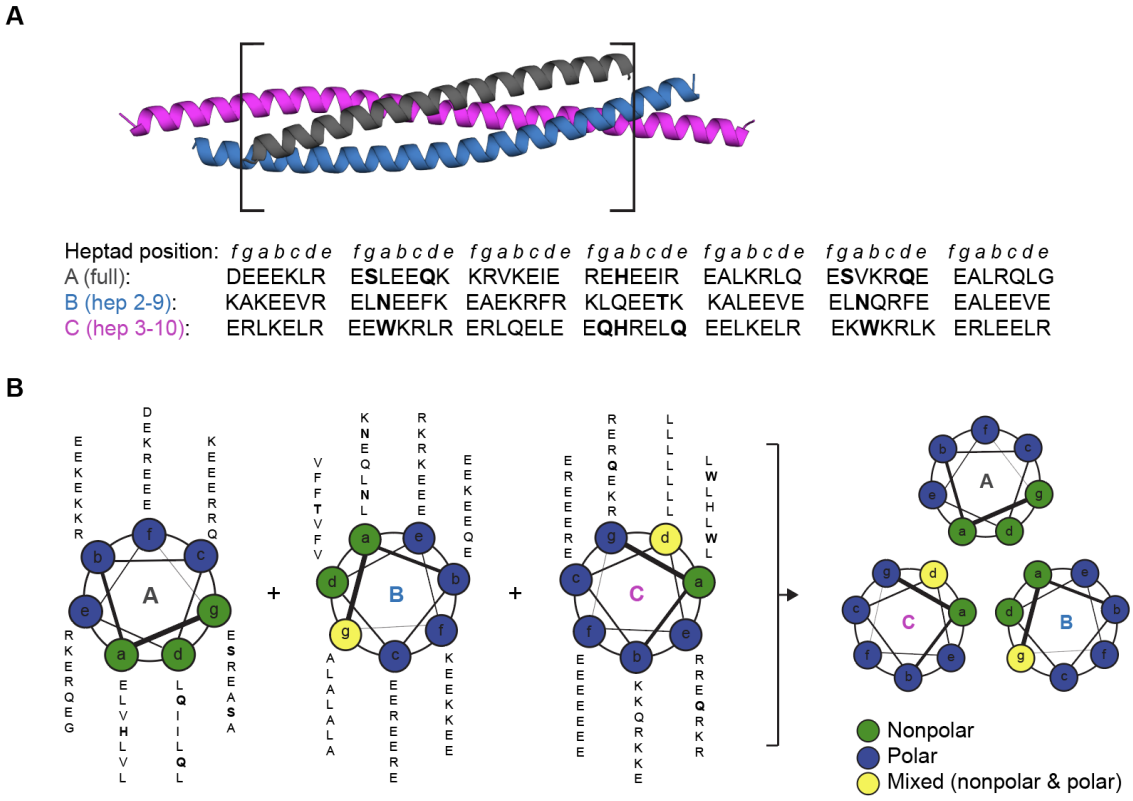


Figure 3.1.2. Helical wheel for ABC single stranded heterotrimer. (A) Shared interfaces of all three chains of the heterotrimer are broken up by heptad and helical wheel position, with the amino acid letters in bold indicating the residues found by Monte Carlo HBNet. (B) A helical wheel was drawn for each chain and then combined to form a heterotrimer wheel, with positions color-coded to indicate the presence of nonpolar, polar, or mixed (nonpolar and polar) residues.

To determine the importance of amino acid composition on position 'g' and whether the conversion from nonpolar to polar residues would affect ABC binding, two variants of the heterotrimer were tested (Fig. 3.1.2 c & d). In the first variant, alanine residues at position 'g' on chain A were changed to glutamate (making this position now polar instead of mixed) while alanine residues at position 'g' on chain B were changed to serine, threonine, or glutamate (making 'g' now mixed instead of nonpolar). In the second variant, the mutations from the first variant were kept as is but leucine residues at position 'g' on chain B were also changed to

glutamate in an effort to enhance electrostatic interactions and preserve the ABC assembly per the conclusions of the two previous published attempts at heterotrimer design by other groups[4,6].

When the same pulldown approach as the parent construct was carried out experimentally, both variants had a loss of affinity for chain B, indicating that the presence of nonpolar residues on 'g' was essential. This finding may also suggest that ABC heterotrimer specificity could benefit from a wider hydrophobic interface, meaning that more helices would need to be sampled parametrically and thus resulted in using helical bundles as a starting base scaffold.

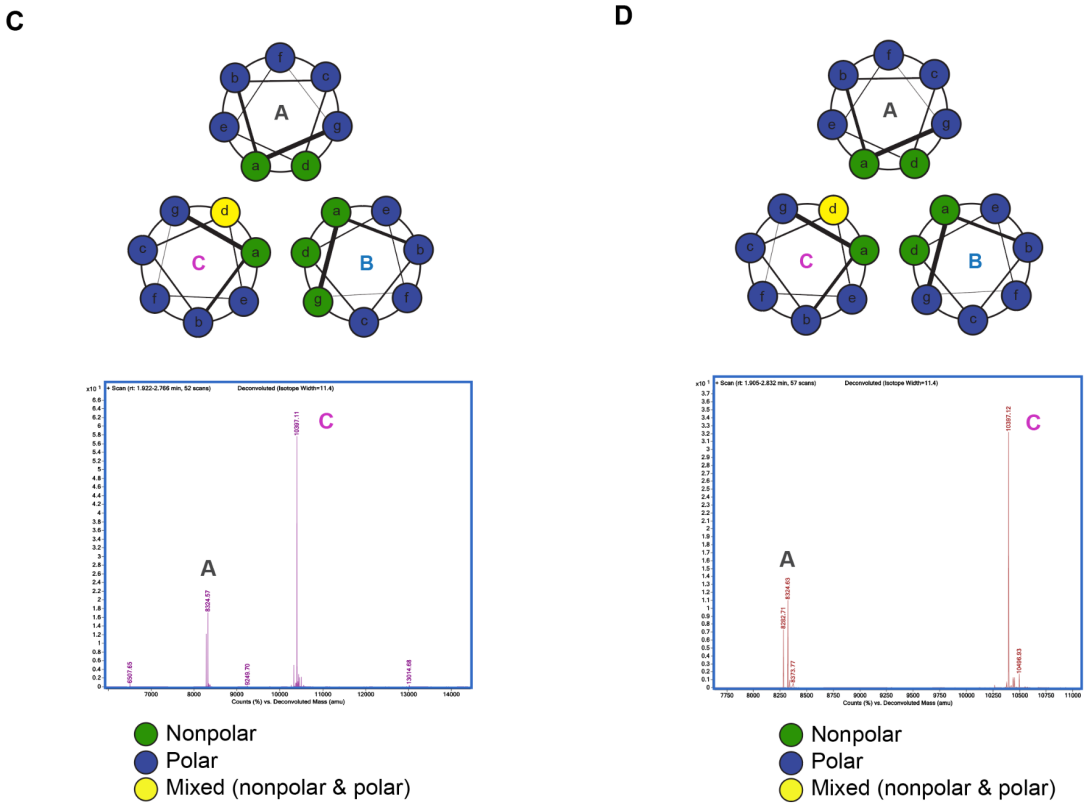


Figure 3.1.2. Point mutations at position 'g' affect ABC affinity. (C) A helical wheel for the first heterotrimer variant (left) shows the conversion of 'g' on chain A from mixed to polar and 'g' on chain B from nonpolar to mixed. Experimental validation of this design by LC-MS (right) indicates the absence of chain B in the eluate. (D) A helical wheel for the second heterotrimer variant (left) shows the conversion of 'g' on chain B to fully polar due to

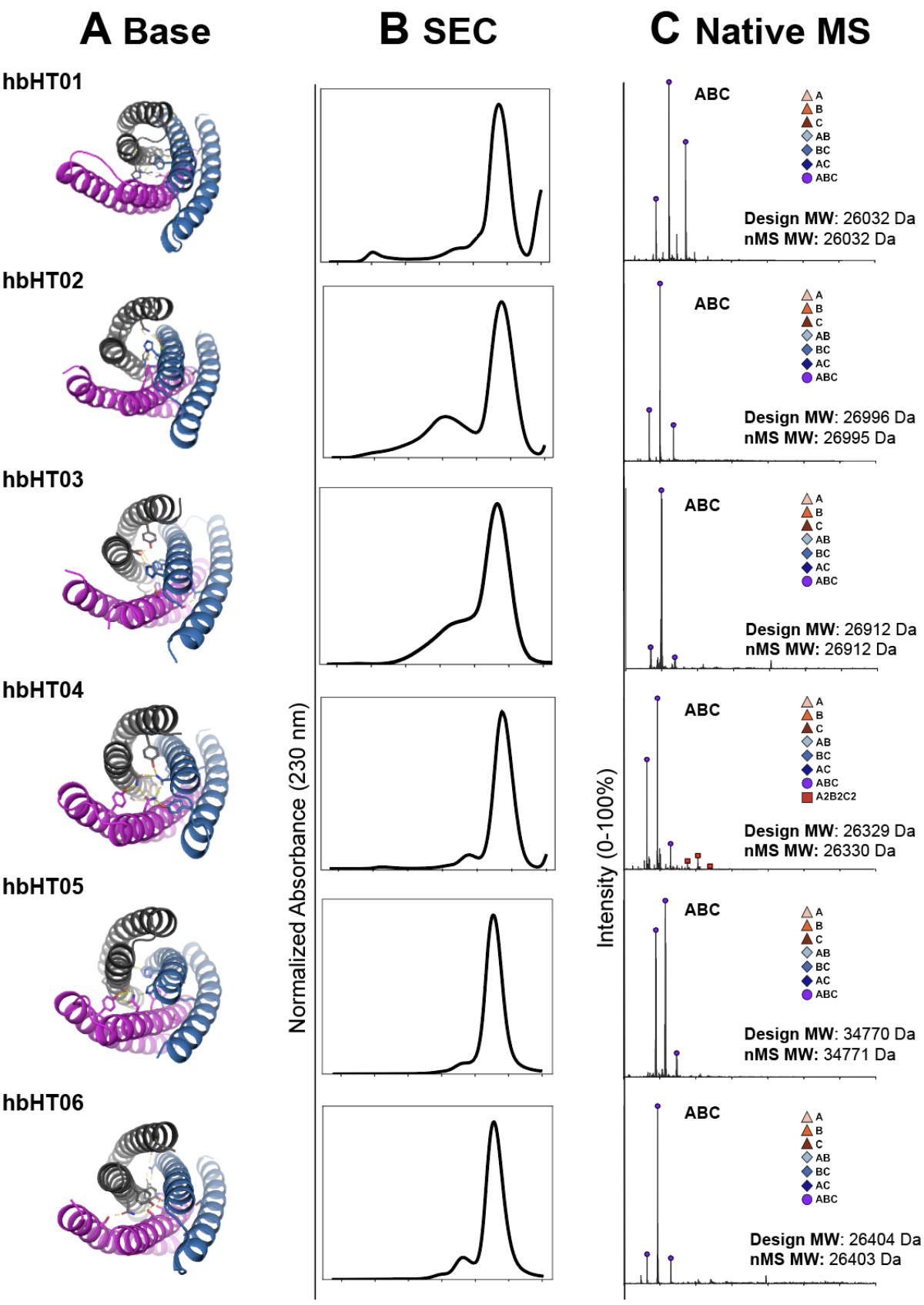
glutamate substitutions made to enhance electrostatics and LC-MS (right) still shows a loss of affinity for chain B in the eluate.

3.2 EXPERIMENTAL VALIDATION OF HELICAL BUNDLE HETEROTRIMERS AND ARMS

3.2.1 *Structural characterization of helical bundle heterotrimers*

These helical bundle designs were ordered for tricistronic expression (co-expression) in E.coli and purified via IMAC with only one chain having the 6xHis-tag. A total of 8 heterotrimers showed monodisperse SEC peaks (Fig 3.2.1 b), ABC heterotrimer formation by native mass spec (Fig. 3.2.1 c), and good alignment to the design models using SAXS (data not shown). From this set of validated heterotrimers, only 1 of them was an antiparallel heterotrimer.

While characterizing these constructs, it was noted that other helical bundle heterotrimer designs showed the presence of 3 components by SDS-page gel and LC-MS, but these designs showed the formation of the ABC species and other non-specific mixtures by native mass spec and were thus not counted as successes as it would be difficult to build with these non-cooperative constructs. Because native mass spec is also not quantitative, it was difficult to approximate ratios of ABC and other states present in solution. In a few cases, the ABC heterotrimer was determined to only form to the desired state by native mass spec (when the respective SEC fraction was analyzed) but the overall SEC peak was quite broad or was joined with other soluble aggregates that would make continuing to build with these constructs challenging.


A Base
B SEC
C Native MS
hbHT01
hbHT02
hbHT03
hbHT04
hbHT05
hbHT06
Normalized Absorbance (230 nm)
Intensity (0-100%)
ABC
A
B
C
AB
BC
AC
ABC
A2B2C2
Design MW: 26032 Da
nMS MW: 26032 Da
Design MW: 26996 Da
nMS MW: 26995 Da
Design MW: 26912 Da
nMS MW: 26912 Da
Design MW: 26329 Da
nMS MW: 26330 Da
Design MW: 34770 Da
nMS MW: 34771 Da
Design MW: 26404 Da
nMS MW: 26403 Da

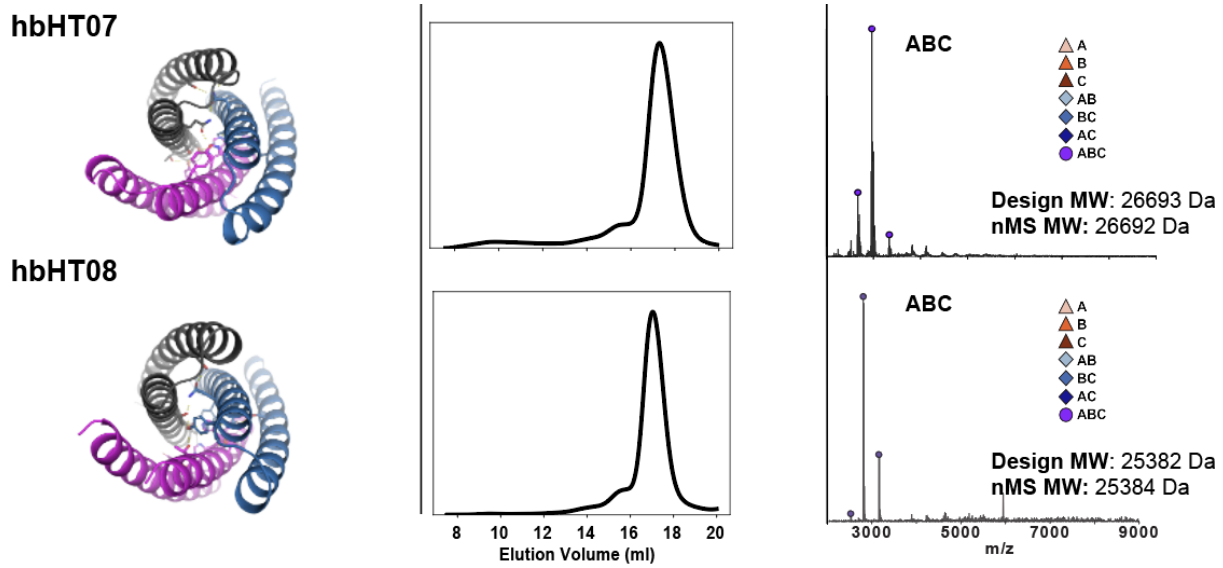
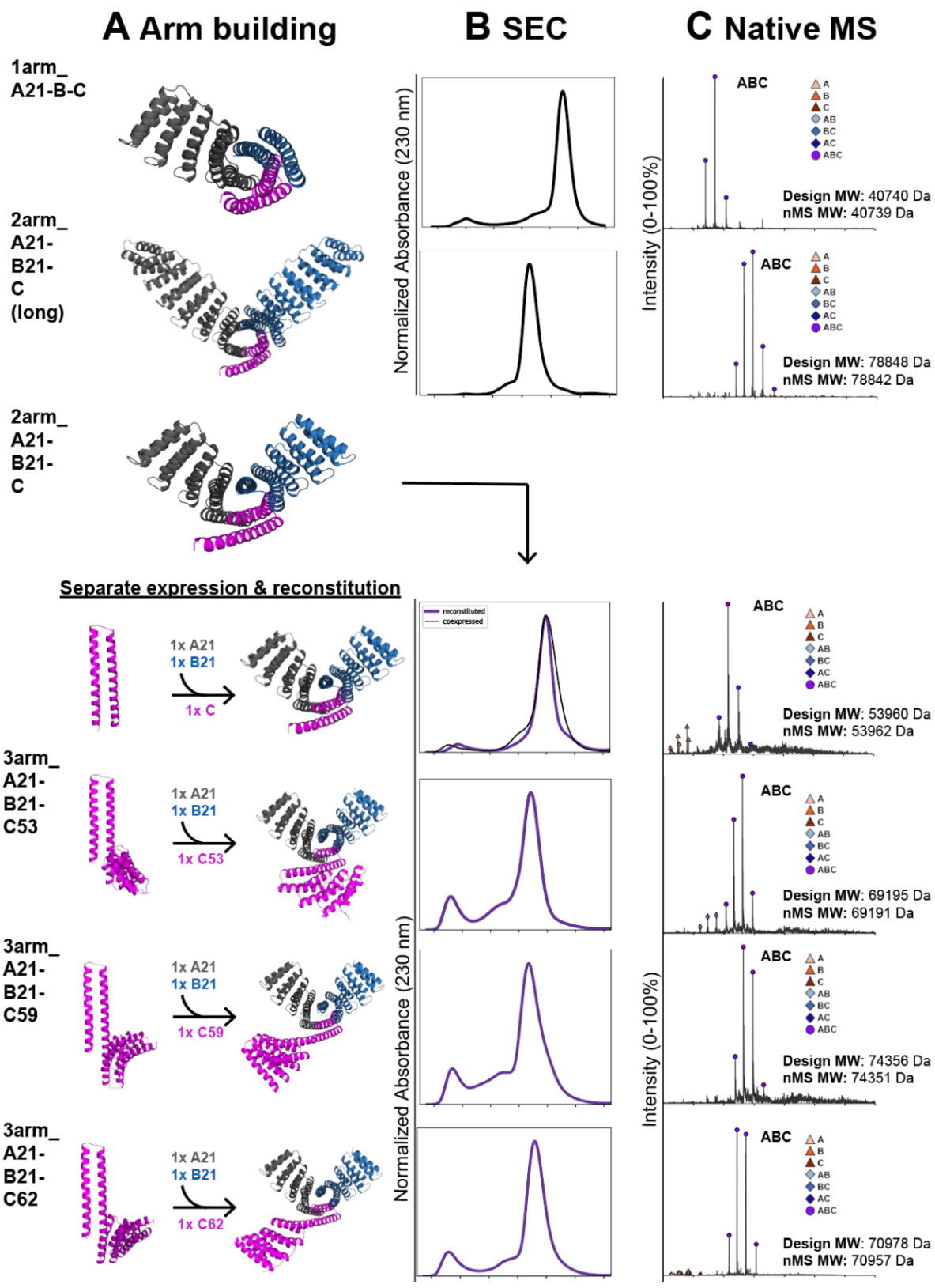


Figure 3.2.1. Experimental characterization of co-expressed ABC heterotrimeric helical bundles. (A) Design models for helical bundle heterotrimers are colored by chain and shown from a top-down view with a cartoon representation and with hydrogen bond networks in stick representation. (B) Monodisperse SEC traces and (C) native mass spec shows only the ABC heterotrimer forming in almost all cases.

3.2.2 Hierarchical building of heterotrimer arms using co-expression and separate expression

To create 3-arm heterotrimers, a hierarchical building approach was carried out to ensure that each new fusion did not disrupt ABC formation. This was applied to one the helical bundle heterotrimers (hbHT02) as highlighted in Figure 3.2.1. A DHR fusion was made to chain A to create a 1-arm construct, followed by subsequent 2-arm fusions (with only one shown as an example) by keeping chain A constant and varying chain B DHR fusions. These were also tricistronically expressed and determined to be successful by SEC (Fig 3.2.2 b), native mass spec (Fig. 3.2.2 c), and SAXS (data not shown).


A Arm building
B SEC
C Native MS
1arm_ A21-B-C
2arm_ A21-B21-C (long)
2arm_ A21-B21-C
Separate expression & reconstitution
1x A21
1x B21
1x C
3arm_ A21-B21-C53
1x C53
3arm_ A21-B21-C59
1x C59
3arm_ A21-B21-C62
1x C62
reconstituted
coexpressed
Normalized Absorbance (230 nm)
Intensity (0-100%)
ABC
A
B
C
AB
BC
AC
ABC
Design MW: 40740 Da
nMS MW: 40739 Da
Design MW: 78848 Da
nMS MW: 78842 Da
Design MW: 53960 Da
nMS MW: 53962 Da
Design MW: 69195 Da
nMS MW: 69191 Da
Design MW: 74356 Da
nMS MW: 74351 Da
Design MW: 70978 Da
nMS MW: 70957 Da

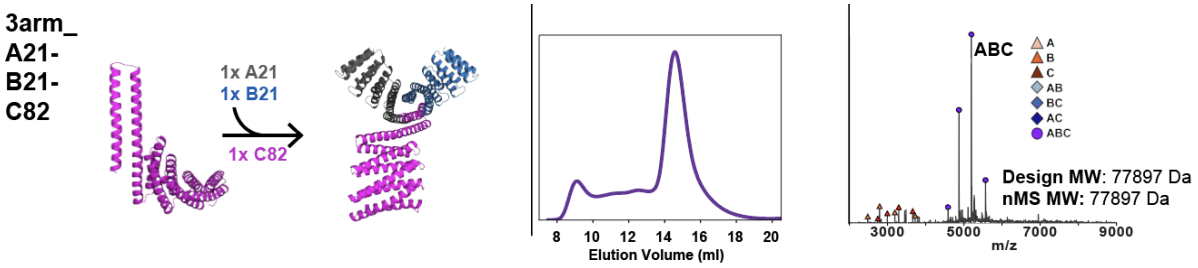


Figure 3.2.2. Experimental characterization of hierarchical arm extensions for a heterotrimer base and successful heterotrimer reconstitution via separate expression. (A) Design models for hierarchical 1, 2, and 3-arms for hbHT02 are colored by chains and shown in a cartoon representation, followed by (B) SEC and (C) native mass spec indicating the ABC heterotrimer is still intact. These were tricistronically expressed. Components for 2arm_A21-B21-C before repeat propagation and subsequent 3-arm fusions were separately expressed, respective components were mixed at a 1:1:1 ratio, and then annealed to reconstitute the ABC heterotrimer.

Because of the hydrophobic nature of the core, each chain of the heterotrimer will have a tendency to self-associate when expressed separately (data not shown). We hypothesized that the independent chain oligomerization is actually a molten globule state[27] along the protein folding pathway that results out of circumstance due to the absence of its binding partners; this oligomerization would be more energetically favorable than a partially or fully unfolded monomeric state but less favorable than the preferred ABC heterotrimer state. If heat could be applied to disrupt the sticky interfaces followed by a controlled cooling protocol, much like the denaturing and annealing phases of standard PCR that allow primers to bind to template DNA via hydrogen bonding, then the desired ABC state could potentially form in a reasonable time frame.

To test this idea, chains A, B, and C of a smaller version of the 2-arm propagated crystal structure (2arm_A21-B21-C) were expressed separately; the 2-arm propagated/extended construct was made because this original and smaller 2-arm version failed to crystallize (this was

also tested and works just as nicely, data not shown for brevity). An equimolar amount of A, B, and C were mixed together and annealed[28] for almost an hour using the thermocycler to allow these interfaces to reassemble in the presence of all components. After running this post-annealed mixture through SEC, a distinct monodisperse peak is observed belonging to the ABC heterotrimer state as confirmed by native mass spec (Fig. 3.2.2). Subsequently, four different repeat protein fusions to chain C were tested and all 3-arm heterotrimers were successfully reconstituted via separate expression. Because the expression levels of each chain cannot be precisely controlled with protein coexpression, proper assembly may be inhibited by this imbalance and so we wanted to find a way to reconstitute these heterotrimers via separate expression and in a relatively quick timeframe. By now having a feasible approach that allows us to explicitly mix each component together and successfully achieve the desired ABC species, we can think about using these heterotrimers in conjunction with other de novo proteins to mediate the specific assembly of complex multi-component protein nanosystems.

Chapter 4. High-resolution structure determination and systematic replacement of networks

4.1 Using x-ray crystallography for high-resolution structure determination

X-ray crystallography is a high-resolution structure determination method that measures the X-ray diffraction of a protein crystalline. This diffraction pattern can then be parsed through software to determine if the solved output matches the intended designed structure. Because hydrogen bonds occur at 2.8-3.1Å, a high resolution structure was needed to conclude their

importance in the overall packing of the designed ABC heterotrimers. A crystal structure for each of the following was solved: hbHT02 heterotrimer, its 1-arm version (1arm_A21-B-C), and its propagated 2-arm version (2arm_A21-B21-C long).

4.1.1 Crystal structure of a helical bundle heterotrimer base

The first crystal structure (Fig. 4.1.1) was solved at a 2.35 Å resolution. The design model, shown in darker colors, had an overall 2.2 Å Cα agreement to the crystal structure, shown in lighter colors.

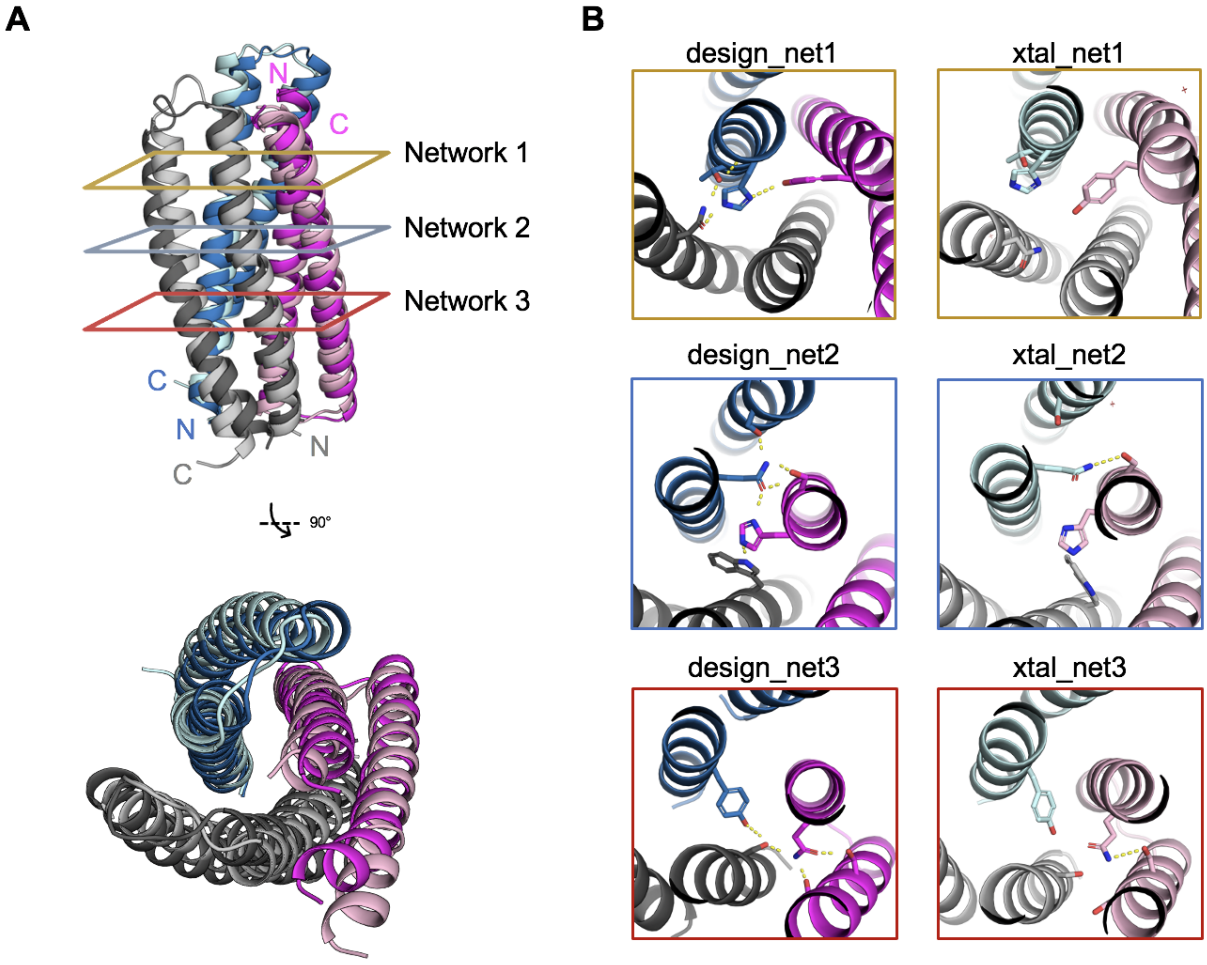


Figure 4.1.1 (A & B). Crystal data shows hydrogen bond networks do not form as designed but heterotrimers still assemble in the correct orientation. (A) Crystal structure of base heterotrimer in lighter colors is aligned to

the design model in darker colors, shown in cartoon representation from a top-down view. (B) Colored cross-sections of the helical bundle base heterotrimer highlight hydrogen bond networks (left), compared to those same residues in the crystal structure (right).

Fig. 4.1.1 shows a colored cross-section of each designed network compared to the corresponding "network" in the crystal structure. It is clear that the hydrogen bond networks designed are not forming as intended, but it is interesting to see that the overall placement of these residues is relatively close in space to the design model and that the ABC heterotrimer still assembles with buried polar groups in the core (water was not detected here). The placement of these residues also appears to be effective in specifying orientation as the chain C inversion in the design model matches that of the crystal structure. Given the distance between residues in the crystal structure, one hydrogen bond could be possible in network 2, between Q57 on chain B and S17 on chain C, and in network 3, between Q26 and S48 on chain C. It is hypothesized that there may be some minor flexibility or movement in the protein core that enables for potential hydrogen bonding since x-ray crystallography really only captures a snapshot of the protein.

For the first network in the crystal structure, N27 on chain A prefers to form hydrogen bonds with neighboring R26 and the chain A backbone, while T55 makes an hbond with its chain B backbone. In the second network, W57 is too far away from the rest of the network residues (~5 Å) to make any potential contacts. In the final third network, Q26 on chain C and Y68 on chain B are ~4.3 Å apart but given their current rotamer placement, they are unlikely to interact. This indicates that polar groups buried in the core of the protein do not need to be fully hydrogen bonding in order to dictate specificity in assembly. Though this bonding may help, it is possible that each unsatisfied polar group is contributing an entropic cost that prevents alternate conformations from forming that would likely result in an even higher entropic folding cost due

to “even worse” satisfaction between the polar group and potential cavities/clashing from the large nonpolar aromatic residues contributing polar groups.

If each chain of the crystal structure is individually aligned to each respective chain of the design model in the absence of the other two binding partners, the backbone alignment is good: chain A has a 1.5 Å Cα agreement, chain B has a 1.2 Å Cα agreement, and chain C has a 2.3 Å Cα agreement (Fig. 4.1.1C).

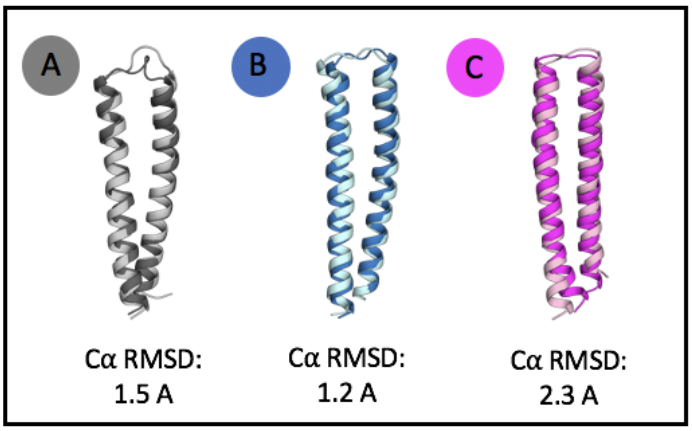


Figure 4.1.1 (C). Good backbone alignment between each individual chain of the heterotrimer base design to its corresponding crystal chain. Each chain was aligned with its matching chain in the absence of the other two binding partners, with chain C having the greatest RMSD deviation.

When all 3 chains are in the presence of one another, there is a bit of movement detected across all chains to accommodate the heterotrimer assembly. It is interesting to note that because these buried polar groups are not interacting with water or other nearby polar groups, they would be deemed “unsatisfied” and the number of these “buried unsatisfied” polar groups would be too high to pass through Rosetta as an acceptable heterotrimer design[19]. This crystal structure will be further analyzed with Rosetta’s energy function to determine what can be learned for future de novo designs containing buried polar hydrogen bond networks.

4.1.2 Crystal structures of helical bundle heterotrimer arms

The second crystal structure obtained at 3.35Å resolution (Fig. 4.1.2) is the 1-arm version of the solved hbHT02 base heterotrimer design. When the original design model (before being informed by the base crystal structure output) is aligned to the 1-arm crystal output, there is deviation in the repeat protein fusion added to chain A. This is due to the actual placement of the chain A outer helix in the crystal structure, as the designed fusion itself is rigid (determined by proper helix formation) and the designed helical repeat protein used here was previously verified by crystallography[21].

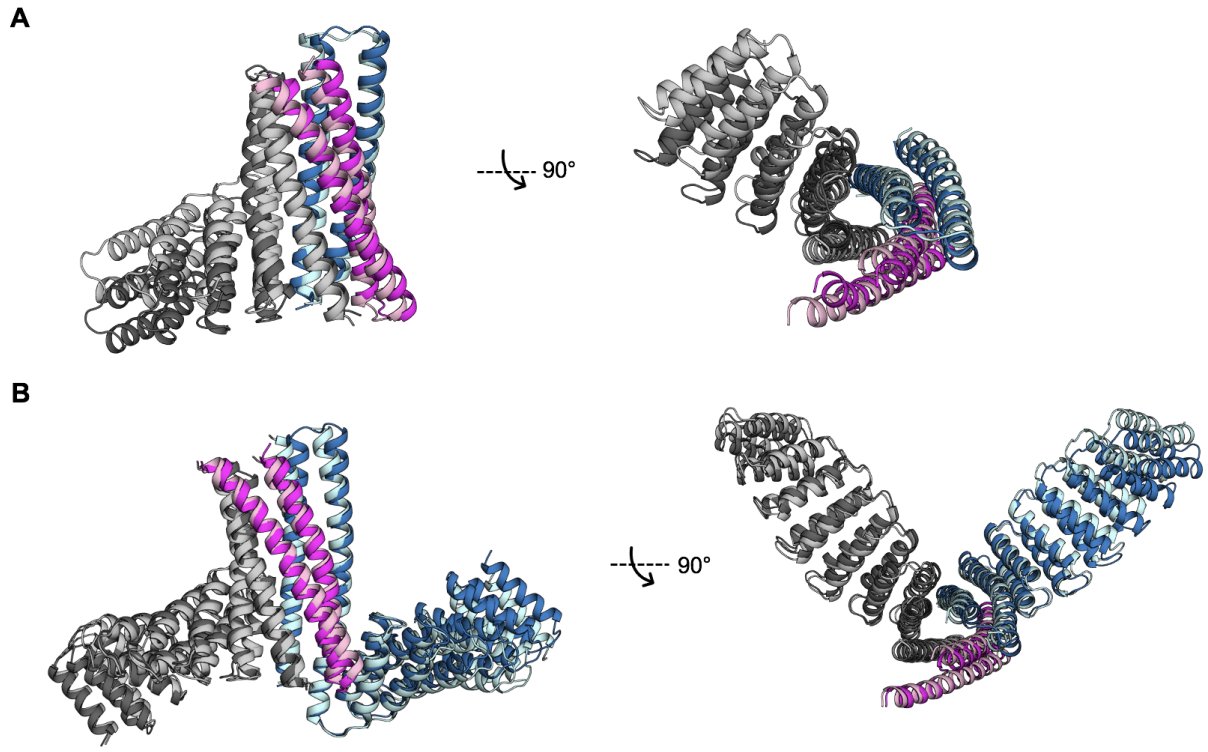


Figure 4.1.2. Crystal structures of a 1-arm and propagated 2-arm heterotrimer. (A) Crystal structure of a 1-arm heterotrimer in lighter colors is aligned to the design model in darker colors, shown in a top-down and side view in cartoon representation. (B) Crystal structure of a propagated version of a 2-arm heterotrimer built from the previous 1-arm trimer is shown in lighter colors, with the design model informed by (C) in darker colors.

The third crystal structure obtained at 3.35Å resolution (Fig. 4.1.2) is the 2-arm version of the 1-arm heterotrimer with a repeat protein fusion now on chain B and with both of the helical repeat proteins propagated. The DHR on chain A and B is the same. Informed by the crystal structure for the 1-arm heterotrimer, chain A in the 2-arm design model has a 1.78 Å alignment between the design model (darker colors) and crystal structure (lighter colors). Only slight deviations are noted along the chain B repeats being propagated. The backbone of the first solved base heterotrimer crystal structure also aligns to each 1-arm and 2-arm crystal structure within 1Å, indicating confidence in the design assembly and crystallographic output.

4.2 Systematic replacement of hydrogen bond networks in crystal structure

Since the designed hydrogen bond networks for crystal structure hbHT02 are not forming as intended, there is a natural curiosity to determine the importance of these residues for exclusive ABC heterotrimeric assembly. A systematic replacement of these "networks" was carried out, in which each "network" in the crystal structure was changed to either Rosetta-preferred nonpolar residues or only alanine while keeping the other residues intact.

4.2.1 Nonpolar substitutions

The residues that made up the "networks" in the crystal structure were first substituted with Rosetta-preferred nonpolar amino acids and these are referred to as "sub_net1," "sub_net2," and "sub_net3." Meanwhile, the combination of all three "network" conversions to nonpolar amino acids is referred to as "sub_netall." These four constructs were purified via the same IMAC pull-down approach as all other heterotrimer designs. All four constructs indicated the presence

of components A, B, and C in the eluate by LC-MS but had broad or messy SEC spectra (Fig. 4.2.1) compared to the original heterotrimer design, indicating increased complex heterogeneity. The specific amino acid substitutions made are shown in each SEC spectra and colored by the respective crystal structure coloring scheme: light gray for chain A, light cyan for chain B, and pink for chain C. These substitutions indicate that every "network" in the core is indeed important for exclusive ABC assembly, and thus perhaps not surprisingly, the messiest SEC trace was observed for the "sub_netall" construct which only had hydrophobic residues across the core.

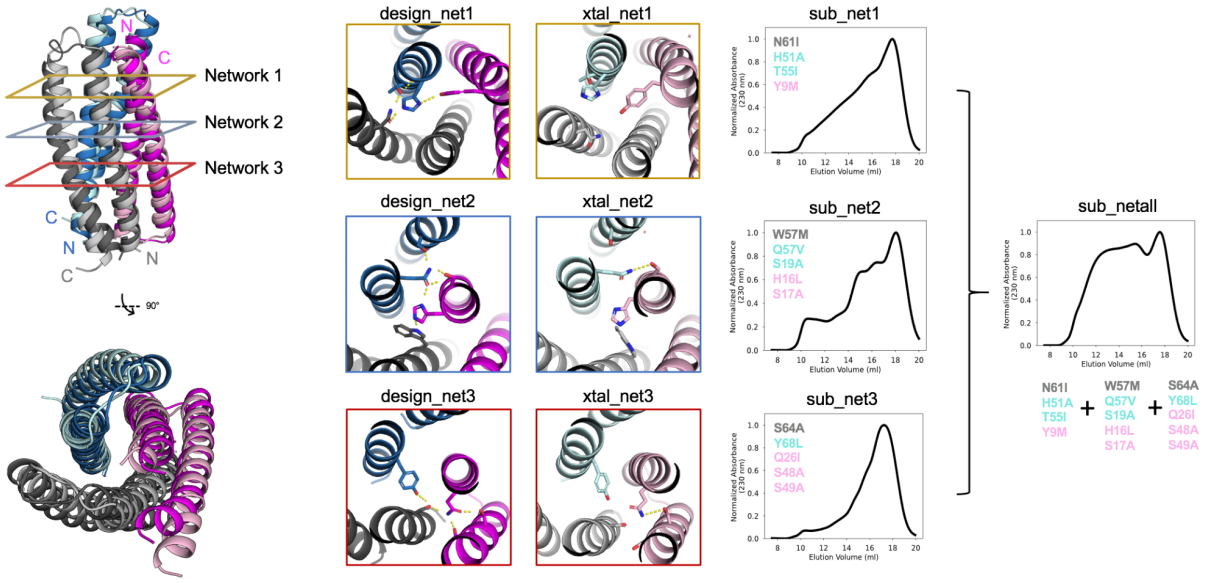


Figure 4.2.1: Systematic replacement of these residues indicates their importance for mediating ABC specificity. Colored cross-sections of the helical bundle base heterotrimer highlight hydrogen bond networks (left), compared to (middle) those same residues in the crystal structure. Replacement of these residues to nonpolar amino acid substitutions (right) shows broad SEC output indicating loss of ABC heterotrimer specificity and a combination of all 3 replacements shows the messiest SEC spectra seen for this design. Mutations are colored by chain for each construct run through SEC.

4.2.2 Alanine substitutions

In parallel, four more systematic mutations were made to hbHT02 crystal structure. In this case, all the residues at the interface that were supposed to be part of the designed hydrogen bond networks were mutated to only alanine on a per "network" basis like before (Fig 4.2.2). Because of alanine's small size, it was expected that this change would create cavities at the interface that would destabilize and disfavor any form of assembly, especially since each chain was contributing at least one large aromatic residue. In the eluate, only chains A and B were detected for "subALA_net1," while all three components were detected for "subALA_net2" and "subALA_net3," and only the His-tagged chain was detected for "subALA_netall" by LC-MS. After running SEC, only chain A was detected in the main peak for "subALA_net2," while only chains A and C were detected in the main peak for "subALA_net3" as determined by LC-MS (Fig. 4.2.2).

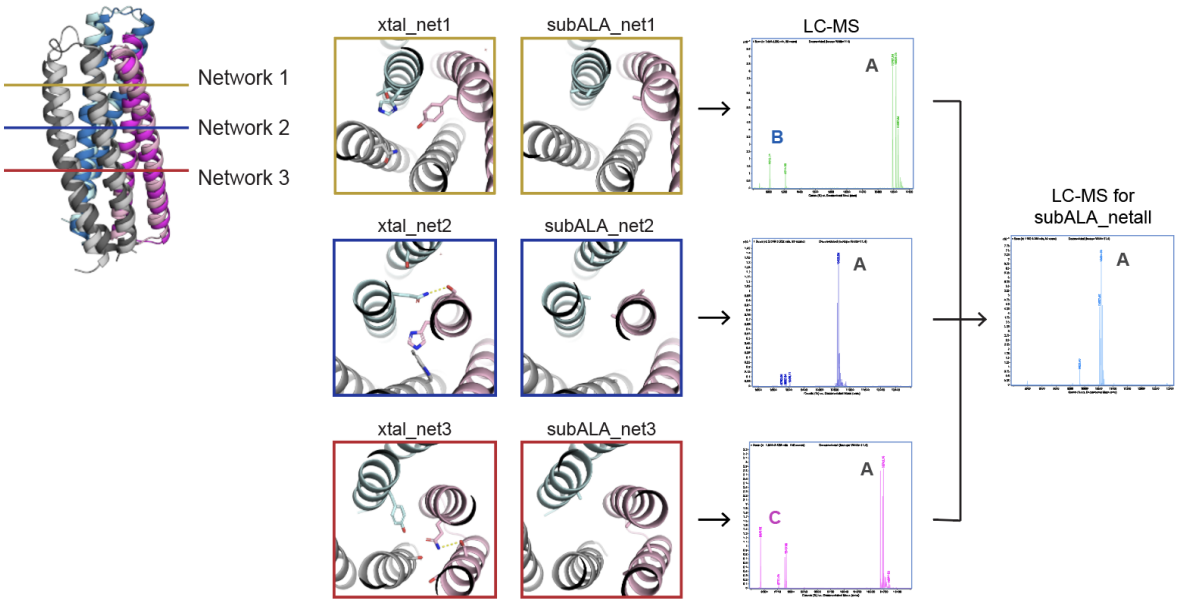


Figure 4.2.2: Systematic replacement to alanine indicates the importance of proper packing. Colored cross-sections of the helical bundle base heterotrimer highlight hydrogen bond networks (left) compared to alanine

conversions (middle), which show that ABC affinity is lost as determined by LC-MS (right). The final combination of alanine substitutions indicates only the presence of the 6xHis-tagged chain.

Perhaps not surprisingly, these aromatic to alanine substitutions correlate with loss of affinity for the chain contributing a large aromatic residue: "subALA_net1" loses affinity for chain C which previously had a tyrosine, "subALA_net2" loses affinity for chains B and C since the His tag is on chain A which previously had a tryptophan, and "subALA_net3" loses affinity for chain B which previously had a tyrosine. The combination of all alanine substitutions in the core, unsurprisingly, resulted in too many cavities for the desired ABC assembly to form or for any chain to bind to the 6xHis-tagged chain for that matter.

4.2.3 What can we learn about replacing these residues?

The proper assembly of the base helical bundle, 1-arm, and 2-arm heterotrimer without the proper formation of the designed hydrogen bond networks indicates that long, extensive networks are likely not necessary across the interface to mediate ABC heterotrimer specificity. Rather, (1) the strategic use of (even unsatisfied) residues with polar groups in the core may be enough to impose an energetic penalty to folding that discourages mis-assemblies and (2) the use of large aromatic nonpolar amino acids, like tyrosine, tryptophan, and phenylalanine can help mediate specific packing in these helical bundles. It may actually be advantageous in the future to incorporate buried residues with polar groups as an explicit negative design tool, in which any other combination or orientation of these residues would result in far more destabilizing alternative assemblies.

Chapter 5. Using ABC heterotrimers as building blocks for higher-order design

5.1 Applying cyclic symmetry to heterotrimers for higher-order building

In an effort to use this specific ABC heterotrimer assembly to build more complex heteromeric constructs, another helical fusion software named WORMS[22] was used to splice working heterotrimer arm designs though the same helical repeat protein library as before to create user-specified cyclic architectures. This would serve as a proof of principle that these heterotrimers can be modified for downstream use in higher-order assembly and potentially result in an array of heteromers with diverse shapes.

5.1.1 Design and validation of cyclic rings using single stranded 4-arm heterotrimers

To simplify the number of working components for cyclic ring design, two (out of three) chains of the working 4-arm single stranded ABC heterotrimers from Figure 3.1.1 were rigidly joined together using WORMS to make one component while the third chain became the second component of the assembly, resulting in cyclic symmetries of an AB dimer unit. A total of 12 rings were ordered to test this cyclization approach via bicistronic expression: six A2B2 heterotetramers, four A3B3 heterohexamers, and two A4B4 heterooctamers. A 6xHis-tag was added to only the non-cyclized chain. All 12 constructs were soluble and had the presence of both components, although only 7 had approximately equimolar ratios of each by SDS page. From these, two designs initially showed monodisperse peaks by SEC. It was noted that the rest of the SEC spectra were either quite broad indicating heterogeneity, eluted later than expected

indicating smaller cyclic assemblies (seen for two A3B3 cases that were likely forming A2B2 instead), or had the expected crown peak but it was not monodisperse due to potential alternate mis-assemblies. Therefore, the annealing protocol previously described to reconstitute the helical bundle heterotrimers was used here in an attempt to give these bicistronic-expressed crowns a second chance at assembly. After running post-annealed samples via SEC, one more A2B2 heterotetramer design showed a monodisperse peak as well (data not shown).

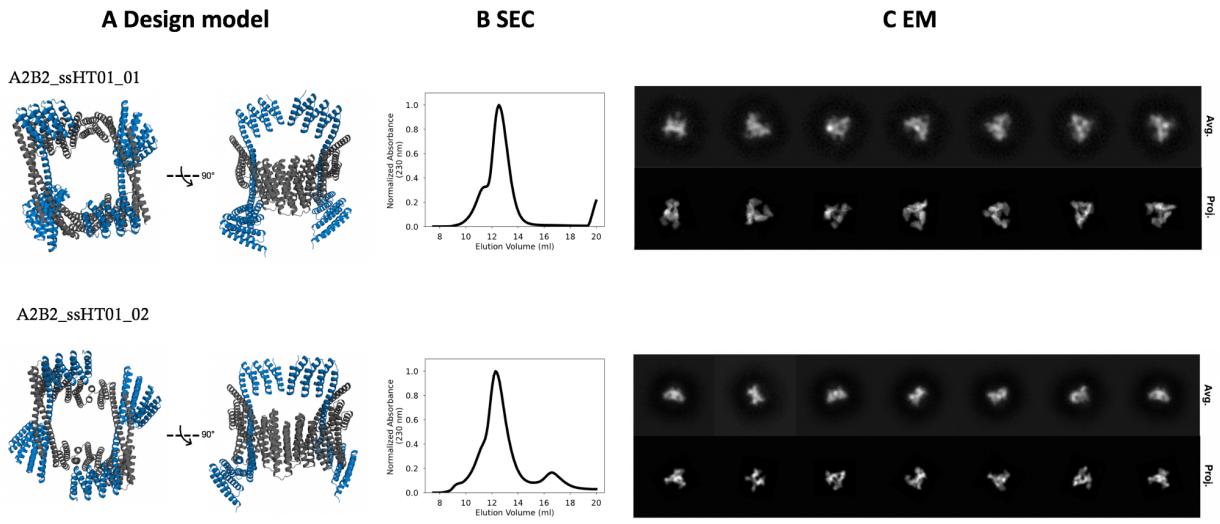


Figure 5.1.1. Single stranded 4-arm heterotrimer can be used to design A2B2 heterotetramers with 4 termini available at each symmetric point for further building downstream. (A) A 4-arm heterotrimer from Figure 3.1.1 was used to make A2B2 heterotetramer design models, shown in a cartoon representation. (B) SEC indicates designs eluting at the right point and a fraction from each representative peak was viewed under electron microscopy. (C) Negative stain electron microscopy micrographs for 2D class averages (experimental) and corresponding projections (design) are shown.

These constructs will be tested via separate expression using the annealing approach proposed and without annealing to gain a better understanding of how well these single stranded ABC heterotrimers can be reconstituted, since this idea had only been previously tested with the helical bundle heterotrimeric interface solved by crystallography.

5.1.2 *Design of cyclic rings using 2-arm crystal structure*

Because the crystal structure of a propagated 2-arm ABC heterotrimer helical bundle was solved, it would be quite easy to use this as WORMS[22] input in conjunction with the repeat protein database mentioned in the previous section and a new alpha/beta (LHD) heterodimer[23] input library to create 3 unique component A3B3C3, A4B4C4, and A5B5C5 complexes. A total of 10 rings were ordered for tricistronic expression and separate expression and reconstitution: five A3B3C3 complexes, two A4B4C4 complexes, and three A5B5C5 complexes (shown in Figure 5.1.2). A 6xHis-tag was added to the chain C hairpin for tricistronic expression. Because these constructs are quite large, negative stain electron microscopy will be the best technique to determine if these 9, 12, and 15 component geometries are assembling as desired.

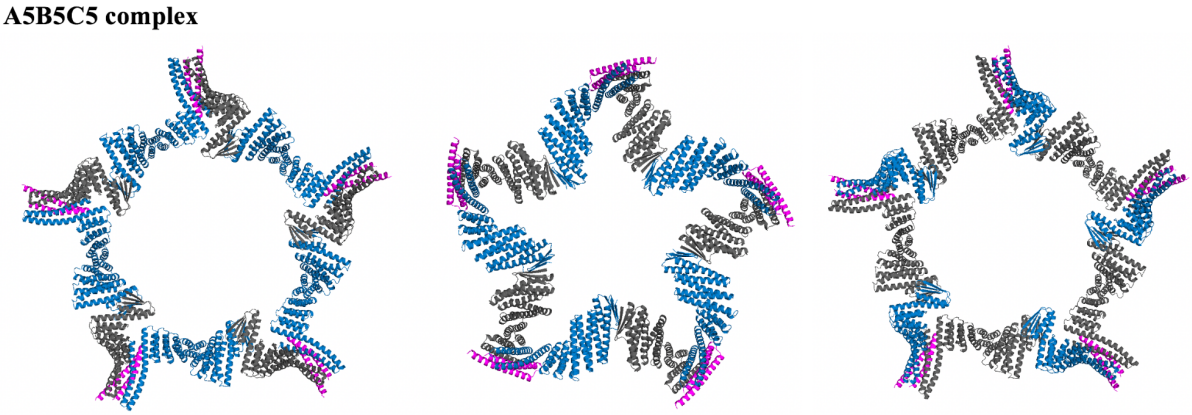


Figure 5.1.2. The propagated 2-arm crystal structure can be used to make large A5B5C5 complexes with 2 termini available at each symmetric point for further building downstream. Three 15-mer A5B5C5 computational designs shown in cartoon representation.

5.1.3 *Availability of termini for building*

Aside from the number of unique chains found in these architectures, the potential for building with these stems from the number of available termini at each chain and its placement. For the

A2B2 complexes made from a single stranded 4-arm heterotrimer, each symmetric point has 4 unique termini available. This means that each complex can sustain or display two copies of up to 4 unique proteins or a total of 8 copies of the same protein via rigid fusions or flexible linkers. This makes it a unique construct given that a typical A2B2 heterotetramer would only be able to sustain or display up to 2 different proteins at each symmetric point.

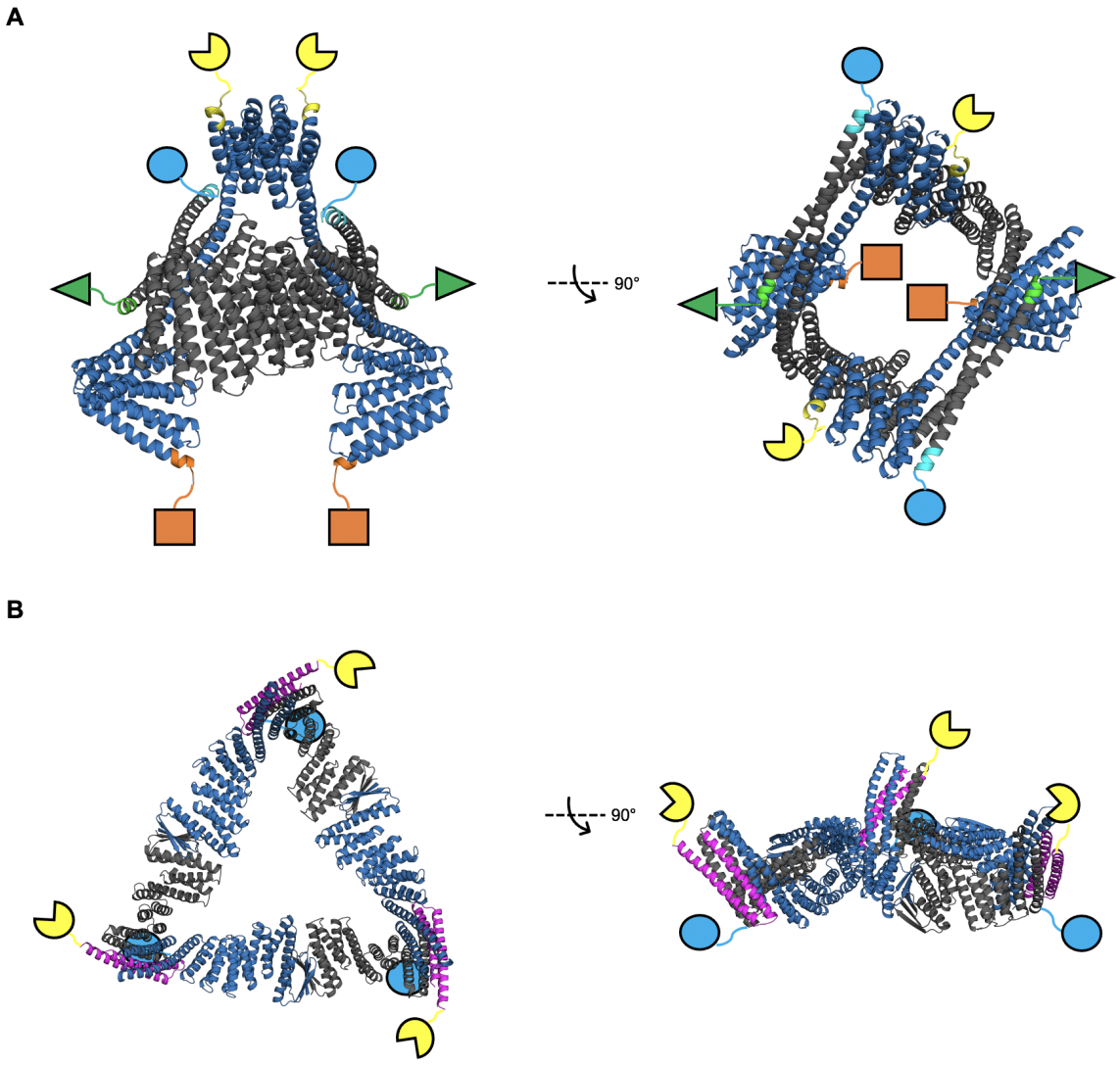


Figure 5.1.3. Availability of cyclic heteromer termini for further building. (A) The A2B2 heterotetramers made from a single stranded 4-arm construct, shown in cartoon representation, are capable of displaying up to 4 unique

proteins at each symmetric point, as shown by the different shapes. (B) Meanwhile a 9-mer A3B3C3 scaffold, shown in cartoon representation, has 2 unique termini at each symmetric point available for rigid or flexible fusions of other proteins.

On the contrary, the A3B3C3, A4B4C4, and A5B5C5 complexes would only have 2 unique termini available at each symmetric point (Fig. 5.1.3), but they would still allow for building outside the plane and within the same plane of the cyclic ring.

These larger multi-component constructs would highlight the ability to successfully mix ABC cooperative heterotrimers with LHD heterodimers to create specific user-desired geometries, indicating that these ABC heterotrimers behave well and would be useful team players for future de novo design efforts and applications.

Chapter 6. Conclusion

We have been able to successfully tackle a 3-sided interface design problem by designing multiple heterotrimeric coiled coil and helical bundle scaffolds that are cooperative in nature. Their ability to also sustain protein additions, like that of rigid helical fusions to monomeric repeat proteins, also indicate their usefulness in downstream higher-order protein design. Overall, this approach was met with a lower success rate than desired as the difficulty lay in getting only the ABC complex designed for, as opposed to achieving ABC and/or other alternate homo and hetero-oligomeric states (implying the design was not specific and cooperative). Still, these heterotrimers can be utilized as a driving force for specific assembly, used to create larger complexes with the capacity to continue building from them using available termini, and ultimately aid in diversifying the geometry of larger architectures.

One of the benefits of building from the ground up is the many different shapes that can potentially result from the design of higher-order oligomers, such that the final user-desired geometry is not limited by the starting topology of the heterotrimer. Furthermore, the orientation of the termini (as seen for the single stranded heterotrimers) and directionality of the chains (as seen in the helical bundle heterotrimers) make it possible to attach multiple unique proteins for displaying or building outside of the respective plane. Thus these scaffolds can quite literally be used to build or display "up" and "down" rather than linearly or all in the same orientation, which is not possible with other current de novo homo and hetero-oligomeric scaffolds.

The cooperative nature and small starting size of these ABC heterotrimer scaffolds make these constructs versatile for a broad range of applications that require three unique components. Future efforts will require approaches that meet the design of cooperative ABC and AAB heterotrimers and other higher-order hetero-oligomers with a better success rate in order to continue exploring the potential complexity that can be achieved in the de novo protein landscape.

BIBLIOGRAPHY

1. Grigoryan G. & Degrado W. F. Probing designability via a generalized model of helical bundle geometry. *J. Mol. Biol*. 405, 1079–1100. (2011).
2. Boyken S.E. et al. De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity. *Science* 352: 680–687. (2016).
3. Chen Z et al. Programmable design of orthogonal protein heterodimers. *Nature* 565: 106-111. (2019).
4. Xu Q & Dunbrack R.L. Principles and characteristics of biological assemblies in experimentally determined protein structures. *Current Opinion in Structural Biology* 55, 34-49. (2019).
5. Hashimoto K & Panchenko A.R. Mechanisms of protein oligomerization, the critical role of insertions and deletions in maintaining different oligomeric states. *PNAS* 107 (47): 20352-20357. (2010).
6. Mei G. et al. The Importance of Being Dimeric. *FEBS J*. 272: 16-27. (2005).
7. Oldham W. & Hamm H. Heterotrimeric G protein activation by G-protein-coupled receptors. *Nat Rev Mol Cell Biol* 9: 60–71. (2008).
8. Hohenester E & Yurchenco P.D. Laminins in basement membrane assembly. *Cell Adh Migr* 7(1): 56-63. (2012).
9. Nautiyal S et al. A designed heterotrimeric coiled coil. *Biochemistry* 34, 11645-11651. (1995).
10. Nautiyal S. & Alber T. Crystal structure of a designed, thermostable, heterotrimeric coiled coil. *Prot. Sci.* 8, 84-90. (1999).
11. Lombardi A., Bryson J. W. & DeGrado, W. F. De novo design of heterotrimeric coiled coils. *Biopolymers* 40, 495-504. (1996).
12. Goodsell D. S. & Olson A. J. Structural symmetry and protein function. *Annu Rev Biophys Biomol Struct* 29, 105-153. (2000).
13. Crick F.H.C. The packing of a-helices: simple coiled-coils. *Acta Cryst* 6:689-697. (1953).
14. Crick F.H.C. The Fourier transform of a coiled-coil. *Acta Cryst* 6: 685–689. (1953).
15. Huang P.S. et al. High thermodynamic stability of parametrically designed helical bundles. *Science* 346, 481-485 (2014).
16. Maguire J.B. et al. Rapid Sampling of Hydrogen Bond Networks for Computational Protein Design. *JCTC* 14(5): 2751-2760. (2018).
17. Alford R.F. et al. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *JCTC* 13(6): 3031-3048.
18. Lauer T.M. et al. Developability index: a rapid in silico tool for the screening of antibody aggregation propensity. *J Pharm Sci* 101(1): 102-115. (2012).
19. Coventry B & Baker D. Protein sequence optimization with a pairwise decomposable penalty for buried unsatisfied hydrogen bonds. *PLoS Comput Biol* 17(3): e1008061. (2021).
20. Hiranuma N., Park H., Baek M. et al. Improved protein structure refinement guided by deep learning based accuracy estimation. *Nature Communications* 12, 1340. (2021).
21. Brunette T.J., Parmeggiani F., Huang P.S. et al. Exploring the repeat protein universe through computational protein design. *Nature* 528(7583): 580-584. (2015).

22. Hsia Y., Mout R., Sheffler W. et al. Design of multi-scale protein complexes by hierarchical building block fusion. *Nature Communications* 12, 2294. (2021).
23. Sahtoe D.D., Praetorius F. et al. Reconfigurable asymmetric protein assemblies through implicit negative design. *Submitted.* (2021).
24. Divine R. et al. Designed proteins assemble antibodies into modular nanocages. Science 372, 6537. (2021).
25. Schiffer, M & Edmundson A.B. Use of helical wheels to represent the structure of proteins and to identify segments with helical potential. *Biophys J* 7(2): 121-135. (1967).
26. Mason J.M., Muller K.M., Arndt K.M. Considerations in the design and optimization of coiled coil structures. *Protein Engineering Protocols* 3: 35-70. (2007).
27. Bychkova V.E., et al. The molten globule concept: 45 years later. *Biochemistry Moscow* 83, S33-S47. (2018).
28. Egelman E.H., et al. Structural plasticity of helical nanotubes based on coiled-coil assemblies. *Structure* 23(2): 280-289. (2015).
29. Studier F.W. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* 41(1): 207-234. (2005).
30. Barth M. & Schmidt C. Native mass spectrometry - a valuable tool in structural biology. *J. Mass Spectrom.* 55(10) e4578. (2020).
31. VanAernum Z.L. et al. Rapid online buffer exchange for screening of proteins, protein complexes and cell lysates by native mass spectrometry. *Nature Protocols* 15, 1132-1157. (2020).
32. Dyer K. N. et al. High-throughput SAXS for the characterization of biomolecules in solution: a practical approach. *Methods Mol. Biol*. 1091, 245–258. (2014).
33. Schneidman-Duhovny D, Hammel M, Tainer JA, and Sali A. Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophysical Journal* 105 (4), 962-974. (2013).

Appendix A: Materials and Methods

Gene preparation

Genes were codon optimized for bacterial expression and most were ordered in pET29b+ vector between the NdeI and XhoI restriction sites, with a T7 promoter and Kanamycin resistance gene. hbHT01and hbHT02 were ordered in pET21b(+) between NdeI and XhoI restriction sites, with a T7 promoter and Ampicillin resistance gene, and with an additional de novo EHEE protein added to the N-term of the first chain to increase molecular weight for SDS-PAGE differentiation. All constructs, unless explicitly specified, were ordered for tricistronic expression using two different ribosome binding site (RBS) sequences, one RBS (TAAGAAGGAGATATCATCATG) was appended after the first chain and the other RBS (TAAAGAAGGAGATATCATATG) was appended after the second chain. The last chain in all sequences had a cleavable N-term 6xHis-tag, with recognition sequences for tobacco etch virus (TEV) or thrombin cleavage. A stop codon was added after the last chain. The cyclic ring designs were ordered in the same manner, with the exception of using the C-terminal His-tag in frame to reduce overall DNA synthesis length. An additional Strep-tag II (WSHPQFEK) was added to the N-term of chain B for the single stranded coiled coil arms and bicistronic crowns made from these arms. Genes were ordered from Integrated DNA Technologies (IDT) or Genscript.

Protein expression and purification

Plasmids were transformed into either BL21(DE3) or Lemo21(DE3) *E.coli* cells using a 30 second heat shock protocol, added to autoinduction media[29], and incubated at 225 rpm for 20-22 hours at 37ºC. Cell pellets were obtained by centrifugation at 4000 x *g* for 15 minutes, resuspended in 30 ml of lysis buffer (25mM Tris-HCl pH 8.0, 300 mM NaCl, 20mM imidazole, with added protease inhibitor PMSF), lysed with a sonicator at 85% amplitude with 15 second on/off cycles for a total of 2 minutes and 30 seconds, and then spun in the centrifuge at 24,000 x *g* for 30 minutes. Cleared lysate was poured over an Ni-NTA column pre-equilibrated with 3 column volume (CV) of lysis buffer, washed with wash buffer (25mM Tris-HCl pH 8.0, 300 mM NaCl, 30mM imidazole) at 2 X 10 CV, and eluted with elution buffer (25mM Tris-HCl pH 8.0, 300 mM NaCl, 250mM imidazole) at 6 CV. For Strep-tag purification, Strep-Tactin XT Superflow high capacity resin (IBA) was equilibrated with 2 CV of Buffer W (100 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA), IMAC eluate was poured over, column was washed with 5 CV of Buffer W, and protein was eluted with 3 CV of Buffer BXT (100 mM Tris-HCl, pH 8.0; 150 mM NaCl; 1 mM EDTA; 50 mM biotin).

For TEV or thrombin cleavage, imidazole was cleared out through a buffer exchange into TBS buffer (25mM Tris-HCl pH 8, 150mM NaCl) and enzyme was applied for overnight cleavage. PMSF was used to stop thrombin cleavage and a second IMAC pulldown was carried out for either cleavage reaction. Flow-through was collected and run through SEC (25mM Tris-HCl pH 8.0, 100mM NaCl).

SDS-PAGE

All protein samples were mixed with Bio-Rad Tricine Sample Buffer or 2x Laemmli Sample Buffer, heated for 10 minutes at 95ºC, and loaded onto either Tris-Tricine or Tris-Glycine gels along with 5 ul of BioRad's Precision Plus Protein Dual Xtra Protein Standard. Gel was run for 30 min at 200V (Tris-Glycine) or for an hour at 120V (Tris-Tricine) and then either stained with Coomassie Brilliant Blue dye or with Genscript's eStain.

Size exclusion chromatography (SEC)

An AKTA PURE FPLC system was used. Heterotrimer bases, arm extensions, and single stranded cyclic rings were passed through a Cytvia Superdex S200 Increase 10/300 GL column, while helical bundle alpha/beta cyclic rings were passed through a Cytvia Superdex S6 Increase 10/300 GL column. Samples ran at a flow rate of 0.75 ml/min and fractions were collected at 0.5 ml.

Mass spectrometry

The fraction corresponding to the SEC peak was concentrated to 1-2 mg/ml and run through Agilent 6230 LC/MS TOF through an AdvanceBio RP desalting column. The mass of the proteins was quantified using intact mass spectrometry in positive mode.

CD

Samples were run over SEC through PBS (Phosphate-Buffered Saline pH 7.4) buffer, concentrated to 0.25 mg/ml, and placed in a 1 mm pathlength cuvette. A JASCO-1500 was used for wavelength scans (190-260 nm) and temperature melts (25-95ºC).

SAXS[32]

Purified protein samples were run through 25mM Tris, 150mM NaCl, and 2% glycerol buffer for SEC. Samples were concentrated using a 10K molecular weight cutoff (MWCO) benchtop spin concentrator and flow-through from the concentrator was used as a buffer blank. A 1.5-2.5 mg/ml low concentration range and a 3-6 mg/ml high concentration range were used for shipping to the SIBYLS High Throughput SAXS Advanced Light Source in Berkeley, California. Scattering output was then fit to the theoretical design mode using the FOXS[33] Server (https://modbase.compbio.ucsf.edu/foxs/). ScÅtter3 was used for analysis (https://bl1231.als.lbl.gov/scatter/).

X-ray crystallography preparation

Protein stored in 25mM Tris, pH 8.0, 100mM NaCl buffer was concentrated to 40 mg/ml (hbHT02), 30 mg/ml (1arm_A21-B-C), and 41 mg/ml (2arm_A21-B21-C long) for setting up the first drop of crystal trays. They were then diluted as follows for setting up the second drop of crystal trays: 9.4 mg/ml (hbHT02), 8.2 mg/ml (1arm_A21-B-C), 12.8 mg/ml (2arm_A21-B21-C long). Samples were set up by using Mosquito Crystal (TPP) as sitting drops and vapor-diffusion for crystallization trials in the JCSG Core Suite I-IV (NeXtal). The crystal trays were incubated/imaged by Rock Imager systems at 4°C and 18°C. Crystals of hbHT02 base, 1arm_A21-B-C, and 2arm_A21-B21-C long were grown and cryoprotected under the following conditions:

hbHT02 base (xRR0072 I-4C, B2.1, RRX0645):

Buffer condition: 0.1M HEPES pH 7.5, 20% PEG8000

Temperature: 4°C

Cryoprotected by supplementing the reservoir with 15% Ethylene glycol

1arm_A21-B-C (xRR0073 II-4C, F11.1, RRX0658):

Buffer condition: 1M LiCl, 0.1M Sodium citrate pH 5, 20% PEG6000

Temperature: 4°C

Cryoprotected by supplementing the reservoir with 15% Ethylene glycol

2arm_A21-B21-C long (xRR0067 II-18C, G10.2, RRX0491):

Buffer condition: 2M $(NH4)_2SO4$, 0.1M sodium acetate pH 4.6

Temperature: 18°C

Cryoprotected by supplementing the reservoir with 2.2M sodium malonate pH 5