Programmable Design of

Protein Interaction Specificity and Logic Gates

Zibo Chen

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

Frank DiMaio

Phil Bradley

Program Authorized to Offer Degree:

Biochemistry

University of Washington

Abstract

Programmable Design of Protein Interaction Specificity and Logic Gates

Zibo Chen

Chair of the Supervisory Committee:

David Baker

Department of Biochemistry

The binding specificity of DNA molecules is straightforward: adenine binds thymine, and cytosine binds guanine. This simple encoding of specificity enables the binding between two DNA molecules to be modularly and programmatically designed. Proteins do not have such straightforward encoding of binding specificity, but could one design for it? This thesis argues such feature can indeed be designed into proteins. I first show orthogonal heterodimers can be designed modularly and programmably to have DNA-like specificity, followed by showing such principles open the door to algorithmic self-assembly of proteins into 2D lattices.

Table of Contents

Acknowledgements

No one finishes a PhD alone, and I was fortunate enough to have great mentors and friends during my 5-year study at UW. My debt is great to those who helped me along the way, without which I will be stuck in an energy landscape minima.

There is no way to pay enough appreciation to my thesis advisors, David Baker and Frank DiMaio. With more than 80 people to manage, David amazingly puts enough time to guide each advisee. David not only taught me how to be a good scientist, but also showed me vividly, on a daily basis, how to perform a form of comedy called deadpan. My understanding of American humor could not have improved so much without David. Frank knew everything about Rosetta, and was always there when I needed help. In fact, with his special ability which we now call the “Frank Magic”, Frank helped solve most of my protein crystal structures. I will forever be in debt to David and Frank, and the best way to repay is to pay it forward to younger generations.

People often stress the importance of a good postdoc mentor during one’s graduate school career, and I cannot be fortunate enough to have Scott Boyken as my mentor. My research projects can be seen as spin-offs of Scott’s, and he was generous enough to let me take a lead of my research. Scott’s scientific excellency, generosity, and kindness will forever be my role model.

I am grateful to have wonderful colleagues and friends in the Baker lab: Gabe Rocklin, Peilong Lu, Chunfu Xu, Hua Bai, Bobby Langan, Sherry Bermeo, Ryan Kibler, Tim Huddy, Nick Woodall, Marc Lajoie, Alfredo Quijano Rubio, among others. Thank you for the great memories.

Science is a collaborative activity, and I had collaborators who are more awesome than the projects we are working together on. Mengxuan Jia, Florian Busch, Zachary Van Aernum, Aniruddha Sahasrabuddhe, and Vicki Wysocki are one of the best collaborators I have seen, and great thanks to all of you for lending your hand on the native mass spectrometry experiments. I would like to additionally thank David Flores-Solis and Nikolaos Sgourakis for solving my first heterodimer structure; Matt Johnson and Justin Kollman for their help with electron microscopy; Andrew Hunt and Michael Jewett for doing fantastic work on in vitro translation; Jocelynn Pearl and John Stamatoyannopoulos for testing my logic gates in their T-cell systems; Andrew Ng and Hana El-Samad for testing my logic gates in their yeast systems.

I would like to additionally thank my thesis committee members, Phil Bradley, Eric Klavins, and Georg Seelig for guidance and stimulating discussions.

I am grateful to be in a wonderful program, BPSD, during my graduate school career. Erin Kirschner has been fantastic keeping tabs on my progress. I would like to thank Charles Asbury and Wendy Thomas for being great program directors.

A special thanks to Shuting Shao, for her love and for her accompany.

And last but not least, thank you Mom and Dad for creating me, surrounding me with love, and giving me the freedom to explore.

Chapter 1: Programmable design of orthogonal protein heterodimers

Abstract

Heterodimeric interaction specificity between two DNA strands, and between protein and DNA, is often achieved by varying side chains or bases coming off the protein or DNA backbone -- for example, the bases participating in Watson-Crick base pairing in the double helix, or the side chains of protein contacting DNA in TALEN-DNA complexes. This modularity enables the generation of an essentially unlimited number of orthogonal DNA-DNA and protein-DNA heterodimers. In contrast, protein-protein interaction specificity is often achieved through backbone shape complementarity (1, 2), which is less modular and hence harder to generalize. Coiled coil heterodimers are an exception, but the restricted geometry of interactions across the heterodimer interface (primarily at the heptad a and d positions (3)) limits the number of orthogonal pairs that can be created simply by varying sidechain interactions (4–7). Here we demonstrate that heterodimeric interaction specificity can be achieved using extensive and modular buried hydrogen bond networks. We used the Crick generating equations (8) to produce millions of four helix backbones with varying degrees of supercoiling around a central axis, identified those accommodating extensive hydrogen bond networks, and used Rosetta to connect pairs of helices with short loops and optimize the remainder of the sequence. 65 of 97 such designs expressed in *E. coli* formed constitutive heterodimers, and crystal structures of four designs were in close agreement with the computational models and confirmed the designed hydrogen bond networks. Following mixing of 32 chains from sixteen heterodimer designs, denaturation in 5M GdnHCl and reannealing, the vast majority of the interactions observed by native mass spectrometry were between the designed cognate pairs. The ability to design large sets of orthogonal protein heterodimers should enable sophisticated protein based control logic for synthetic biology, and illustrates that nature has not fully explored the possibilities for programmable biomolecular interaction modalities.

Introduction

Orthogonal sets of protein-protein and protein-peptide interactions play important roles in biological systems (9, 10). Among the largest such sets are toxin/toxin-antidote pairs arising from the intense evolutionary pressure associated with biological warfare: each toxin is blocked only by its cognate antidote (11). Creation of new specificities based on these protein pairs by sequence redesign has been difficult, often resulting in promiscuous binding (12); the natural specificity results at least in part from complementary variation in backbone conformation (13, 14). More success has been achieved with coiled coils: several orthogonal sets of 2-4 interacting pairs have been created and experimentally validated (15, 16), including the widely used SYNZIPs (17, 18), but interaction promiscuity has again hampered the design of larger orthogonal sets. The Watson-Crick base-pairing analogy suggests that buried hydrogen bonds are a potential route to achieving large numbers of distinct heterodimer interaction specificities. The heterodimeric coiled coils in nature and previous design efforts do not have enough buried positions for extensive hydrogen bond networks (specificity has instead been achieved through hydrophobic packing (3, 19), charge complementarity (20, 21) and individual hydrogen bonds across the interface (22–24)). A solution to the limited coiled coil interaction surface area problem is to use helical hairpins as building blocks rather than single helices, producing 2 helix on 2 helix interfaces rather than 1 helix on 1 helix (as in traditional coiled coils) (25).

Modular and programmable generation of heterodimers

Guided by the example of the DNA double helix, we hypothesized that large sets of designed heterodimers could be generated by incorporating asymmetric buried hydrogen bond networks into regularly repeating backbone structures. We generated helical bundle heterodimers in which each monomer is a helix-turn-helix starting from four-helix backbones produced using a generalization of the Crick coiled coil parameterization (8, 26). For each of the four helices, we exhaustively sampled the helical phase ($\Delta\phi_1$), supercoil radius (R) and offset along the Z-axis (Z offset) (Fig. 1a, a total of 11 free parameters since there is no z offset for the first helix). The supercoil phases of the helices were fixed at 0, 90, 180 and 270 degrees, and the supercoil twist (ω_0) and helical twist (ω_1) held constant at the ideal values for either a two layer left handed super coil (ω_0=-2.85 and ω_1=102.85), or a 5 layer untwisted bundle (ω_0=0 and ω_1=100) (27).

This systematic sampling yielded 27 million untwisted and 60 million left-handed supercoiled backbones for both parallel and antiparallel orientations of opposing helices (Fig. 1b, Extended Data Fig. 1).

To identify the modular hydrogen bond network equivalents to DNA base pairs, we used the Rosetta HBNet algorithm (25) to design buried hydrogen bond networks in the central repeat units of each backbone. This resulted in 2251 unique sidechain hydrogen bond networks involving at least 4 side chain residues with all heavy-atom donors and acceptors participating in hydrogen bonds, and connecting all 4 helices (Fig. 1c; Extended Data Fig 2, Supplementary Table 1). For each of the full length backbones, we then used TMalign (28) to identify all of the geometrically compatible placements of these sidechain hydrogen bond networks (Fig. 1d), selected the backbones accommodating at least two networks, and connected pairs of helices with short loops (Fig. 1e). Low energy sequences were identified using RosettaDesign (29) calculations in which the hydrogen bond networks were held fixed (Fig. 1f). Designs were selected for experimental testing based on the quality of the hydrophobic packing, excluding those with C2 symmetry in the polar networks to disfavor homodimerization of monomers. Requiring the designs to have all buried hydrogen bond donor and acceptor groups participating in hydrogen bonds ensures donor/acceptor matching at the binding interface, as in DNA base pairing (Extended Data Fig. 3). Throughout the text, the designed heterodimers (DHDs) are referred to by numbers with monomers labeled a or b; for example, DHD15_a refers to monomer “a” of design DHD15.

Biophysical characterization

Bicistronic synthetic genes encoding the selected heterodimer designs (Supplementary Table 2) were expressed in *E. coli.* An N-terminal hexahistidine tag on one of the monomers allowed assay of heterodimer formation by co-purification with Ni-NTA affinity chromatography. 94 of the 97 designs were well-expressed in *E. coli* with both monomers present in the Ni-affinity chromatography eluate, as assessed by gel electrophoresis (Supplementary Table 3). Of these, 75.5% (71/94) eluted as a single symmetric peak by size exclusion chromatography (SEC) at the elution volume corresponding to the correct size (Fig. 1g), and 14 eluted as multiple peaks with the major peak at the correct elution volume. Three designs characterized by CD

spectroscopy (at 0.25 mg/ml) were found to be all alpha helical and stable at 95°C (Fig. 1h, Extended Data Fig. 4).

We assessed specificity of heterodimer formation using native mass spectrometry (MS) (30–32). Of the 85 designs that eluted at the correct elution volume in SEC, 39 were exclusive heterodimers at 15uM (Fig. 1i), 13 were heterodimers with a minor population of hetero-tetramers, and 13 formed heterodimers with one (but never both) monomer also present as a homodimer (Extended Data Figs. 5-6). A second affinity purification step showed that the homodimers in the latter case likely arise from unbalanced expression in *E. coli* (Extended Data Fig. 7). Native MS experiments with serially diluted samples suggest the DHDs have affinities in the nanomolar range (Ext. Fig. 18).

We explored the extent to which the heterodimer set could be expanded by permuting the hydrogen bond networks in the different helical repeat units, and by permuting the backbone connectivity. Assigning each unique network a letter, DHD37_XBBA indicates a variant where the second, third and fourth repeat units have hydrogen bond networks B, B, and A, and the first heptad has exclusively hydrophobic residues in the core, while DHD103_1:423 indicates a heterodimer where one monomer consists of the first helix of DHD103 and the other monomer consists of helices 2 through 4 (Extended Data Fig. 8). 13 of 14 hydrogen bond network permuted variants and 9 of 10 “3+1” backbone-permuted heterodimers (generated from five starting “2+2” heterodimers) ran as single peaks on SEC and were constitutive heterodimers by native MS (Extended Data Figs. 9-10). Using the hydrogen bond network permutation approach we were also able to generate a set of partially orthogonal homodimers (Fig. 3g-i, Extended Data Fig. 11).

Structural characterization

To evaluate the accuracy of the designs, we used SAXS (33, 34), NMR and x-ray crystallography. SAXS spectra collected for 44 designs that were constitutive heterodimers by native MS were in close agreement to the expected radius of gyration (R_g), maximum distance (d_{max}), and theoretical spectra computed from the design models (Fig. 1j, Fig. 2f-g, Extended Data Fig. 12, Supplementary Table 4). The NMR structure of DHD13_XAAA closely matches

the design model: the RMSD over all mainchain α carbon (Cα) atoms is 2 Å between the designed structure and the lowest-energy NMR model (Fig. 2e, Extended Data Figs. 13-16, Supplementary Tables 5-8).

We solved the X-ray crystal structures of DHD131, DHD37_1:234, DHD127 and DHD15. The backbones in all four cases were nearly identical to the design models, with backbone Cα atom RMSDs ranging from 0.95 to 1.7 Å. The extensive five-residue buried hydrogen bond network of DHD131 (involving two serines, an asparagine, a tyrosine, and a tryptophan) is nearly identical in the crystal structure, with an additional water molecule bridging the interactions (Fig. 2a). The two designed hydrogen bond networks in DHD37_1:234 are also in close agreement with the crystal structure (Fig. 2b). These networks contain buried histidine and tyrosine aromatic side chains, which likely sterically disfavor self-association of the monomers into homodimers. In DHD127, the histidines in the two hydrogen bond networks adopt a rotamer different from the design model (Fig. 2c), making a hydrogen bond with a water molecule. A crystal structure of DHD15 at pH 7.0 is similar to the design model (Fig. 2d), while a structure at pH 6.5 is of a domain-swapped, hetero-tetramer conformation (Extended Data Fig. 17a-b); native MS at pH 6.5 suggests that the designed heterodimer, rather than the heterotetramer, is dominant in solution (Extended Data Fig. 17c).

Induced dimerization and protein scaffolding using the DHDs

We sought to design protein-induced dimerization systems by covalently linking one monomer each from two different heterodimer pairs via a flexible linker, and testing whether the remaining two monomers from each pair could be brought together by the heterodimerizer fusion (Fig. 3a). We built three such induced dimerization systems in *E. coli*, each with one of the monomeric components hexahistidine tagged. In all of the cases, the three components were expressed at similar levels and copurified by Ni-NTA chromatography, and native MS showed that they formed constitutive heterotrimers at 10 μM (Fig. 3b, Extended Data Fig. 19). No partial complexes (heterodimers) were observed in any of the systems; only monomers and the designed heterotrimers were detected. Surface induced dissociation (SID) mass spectrometry (35, 36) resulted in binary complexes consisting of the heterodimerizer bound to either one of the monomers (Extended Data Fig. 19), with no observed monomer-monomer subcomplexes,

indicating that the interaction between monomers is exclusively mediated through the dimerizer fusion. In yeast two-hybrid assays (Y2H) with monomers from two different heterodimers fused to the DNA binding domain (DBD) and transcriptional activation domain (AD) respectively, expression of the heterodimerizer fusion as a separate polypeptide chain increased signal many fold over background (Fig. 3c).

We also explored whether the DHDs could be used to construct stable synthetic scaffolding proteins (37). Starting from 3 pairs of DHDs (9, 13_XAAA and 37_ABXB), we constructed a scaffold by covalently linking the monomer chain “a” subunits via flexible linkers (Fig. 3d). The scaffold and the 3 separate chain “b” monomers, one with a His tag, were co-expressed in *E. coli*. Native MS of purified sample at 10 μM revealed a heterotetramer of all 4 proteins (Fig. 3e), with surface induced dissociation producing subcomplexes consisting of the scaffold with 1 or 2 of the 3 chain “b” monomers bound (Fig. 3f); no association between chain “b” monomers was detected, consistent with the design model. The scaffold plus monomer assembly is stable up to 95°C and has a guanidine denaturation midpoint of 4 M (Extended Data Figure 19).

All-against-all orthogonality assessment

By generating interfaces with many polar groups that are energetically costly to bury without geometrically matched hydrogen bonding interactions, our design protocol implicitly disfavors non-cognate interactions between monomers from different designed heterodimers (these are unlikely to have a perfect match between hydrogen bond donors and acceptors). Explicit negative design to disfavor non-cognate interactions is computationally intractable given the very large number of possible off-target binding modes. We investigated the orthogonality of the designed heterodimers both in vitro and in cells as described in the following sections, collecting subsets of designs found to be orthogonal into larger sets for all by all interaction assessment.

We investigated the interaction specificity of the designed heterodimers in cells using Y2H experiments (Supplementary Table 10). For 24 designs, strong interactions were observed with the two partners fused to DBD and AD, but not when either partner was fused to both domains, suggesting strong preference for the designed heterodimer over the two possible homodimers (Fig. 4a). To assess orthogonality, the 24 monomers in 12 of these designs were crossed in an

all-by-all Y2H experiment. Interactions were observed for all designed pairs, and 27 of the 552 possible non-cognate interactions (Extended Data Fig. 20). Orthogonality was higher for an 8-pair subset of the designs: of 240 possible non-cognate interactions, only 4 were observed (Fig. 4b; the interacting polar residues in a subset of these designs are depicted schematically in Extended Data Fig. 21). To determine if the designed cognate interactions could outcompete spurius non-cognate interactions, we co-expressed unfused DHD15_b and DHD13_XAAA_a monomers in cells expressing AD and DBD fusions to DHD15_a and DHD13_XAAA_b: this eliminated the off-target interaction (Fig. 4c).

We used native MS to probe all-by-all interaction specificity of the designed monomers when allowed to associate freely in a single pot. To obtain free monomers, purified DHDs were mixed in equimolar amounts, and denatured in 5M GdnHCl at 75°C (Fig. 4d and e). The denaturant was then removed by dialysis to allow refolding and reassociation. The resulting complex mixture was separated and analyzed by online ion exchange chromatography coupled to high-resolution native MS. To evaluate the extent of heterodimer dissociation, an ^{15}N-labeled variant of DHD37_ABXB (DHD37_ABXB_^{15}N) was added as a control for subunit exchange under denaturing conditions; the hybrid labeled/non labeled complexes expected if full exchange was taking place were observed in all cases (Fig. 4f). Sixteen designs (15 unique pairs and one ^{15}N-labeled control) with the highest cognate specificity were pooled together (Extended Data Fig. 22-23), denatured and reannealed in a large scale all by all mixing experiment.

The native MS results on the complex 32-chain mixture are quite striking (Fig. 4f). All 16 designed pairs were recovered, and of the 512 non-cognate binary complexes possible, only 3 were observed. No non-cognate trimers or higher-order oligomers were observed. The 15 orthogonal pairs include several generated using the hydrogen bond network shuffling approach (DHD37_ABXB and DHD37_BBBB) and backbone permutation approaches (DHD13_2:341, DHD13_1:234, DHD37_3:124, DHD94_2:143): the orthogonality of the former is entirely due to positioning of the hydrogen bond networks as the backbone is fixed, and the orthogonality of the latter, to the connectivity of the chain as the sequence is identical. The differences in orthogonality observed in the native MS mixing and Y2H experiments likely arise because the former reports on relative affinity (all monomers are present, and only the lowest energy complexes form) and the latter, on absolute affinity (only two monomers are present at a time). While the orthogonal sets described here are a substantial advance over the current state of the

art, there is clearly room for improvement in the explicit design of orthogonality: the 6x6 set of mutually orthogonal pairs in the Y2H experiments, while larger than sets of designed heterodimers found to be orthogonal in cells previously (2x2 (17)), was selected (based on orthogonality) from a total of 19 pairs tested, and the 15x15 set of mutually orthogonal pairs in the native mass spec experiment, while much larger than previous sets found to be orthogonal in vitro (2x2 (38) by direct binding, 4x4 by CD readout (39)), was selected from a total of 54 pairs tested.

Conclusion

Our results demonstrate that the domain of unbounded sets of orthogonal heterodimeric biomolecules constructed from a single repeating backbone is not limited to nucleic acids. Interaction specificity arises from extensive buried hydrogen bond networks such as the fully connected TYR-SER-TRP-ASN-SER crystallographically confirmed network in Fig. 2a, and also heterogeneity in the size of the residues at the designed interface (Extended Data Fig. 24), analogous to the reinforcement of the hydrogen bond number differences between GC and AT base pairs by steric effects deriving from the different sizes of the bases. The power of native MS for determining interaction specificity in complex mixtures is highlighted by the 32 chain mixing experiment in Fig. 4f; of the effectively unlimited number of possible oligomeric complexes that can be formed from these chains (528 two chain species, 5984 three chain species, etc), the native MS identifies only the 15 designed heterodimers and 3 off target interactions. Evolution has produced toxin-antitoxin pairs with very high orthogonal specificity (11), but the backbones of the two partners are very different (reducing the orthogonality problem since two partners of the same type cannot interact) and attempts to expand the specificity of such sets have generally introduced promiscuity (12). The relatively simple encoding of specificity in DNA gave birth to a broad spectrum of new technology, from DNA origami (40) to artificial circuits (41). The large set of orthogonal interactions, together with the retention of specificity in the fused monomer systems (the induced dimerizer and scaffold of Fig. 3), and the interaction strength hierarchy illustrated by the cognate interaction competition experiment (Fig. 4c), open the door to protein based cellular control circuits providing considerably faster response times and better integration with signalling inputs and outputs than current designed nucleic acid based circuitry.

Main Text Figures

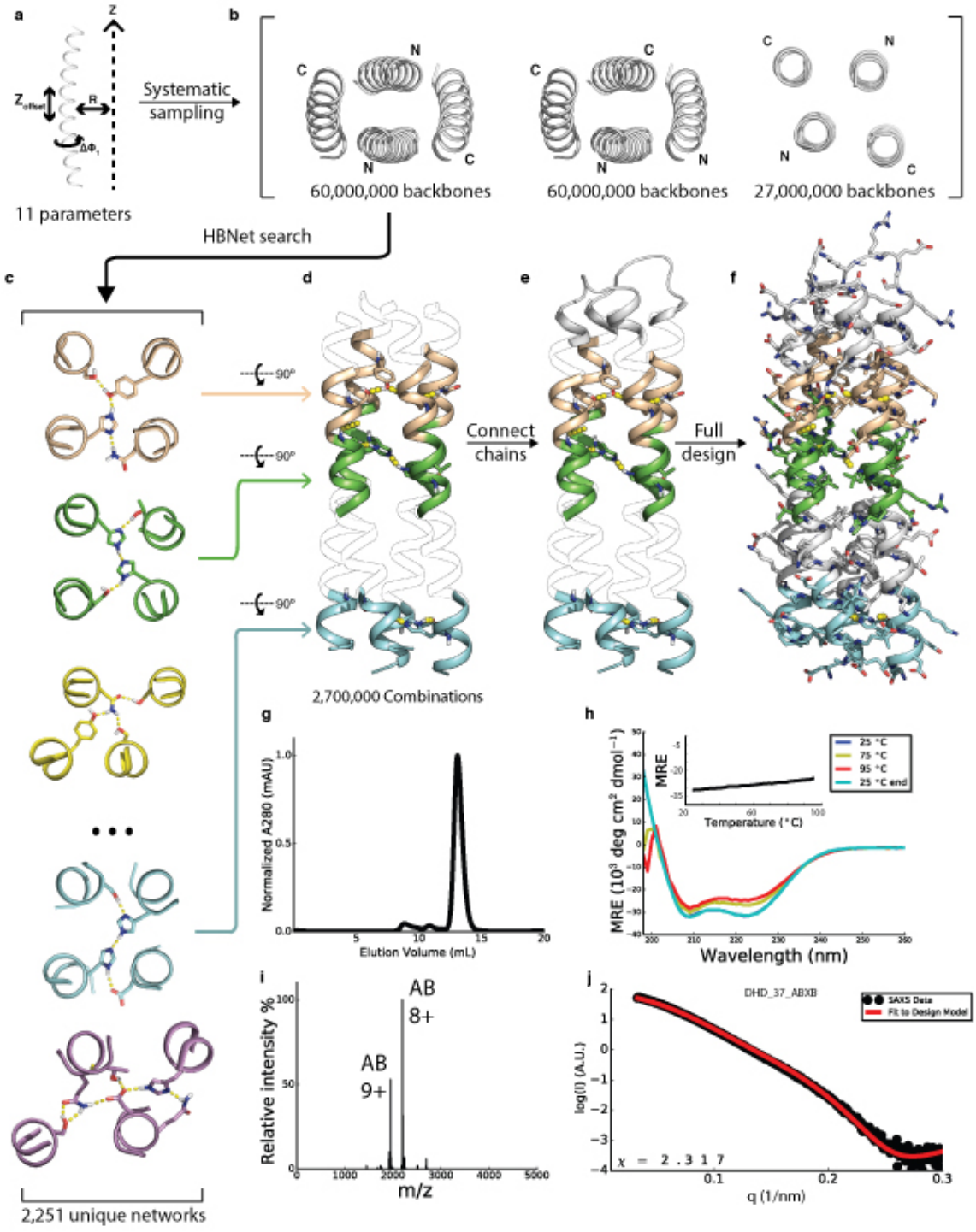


Fig. 1 **Modular heterodimer design. a**, Individual helix generation: the helical phase ($\Delta\phi_1$), supercoil radius (R) and offset along the Z-axis (Z offset) were exhaustively sampled. **b**, Top-down view of parallel twisted (left), antiparallel twisted (middle) and parallel untwisted (right) backbones. **c**, Representative hydrogen bond networks identified using HBNet **d**, Matches of multiple HBnet containing heptads to a single full length backbone. **e**, Addition of loops to connect the 4 helices into two helix hairpins, **f**, Rosetta sequence design. **g-j**, SEC trace, CD spectra and (inset) temperature melt, native MS spectrum and SAXS (black, experimental SAXS data; red, spectra computed from the designed backbones) profile for the design shown in **f**.

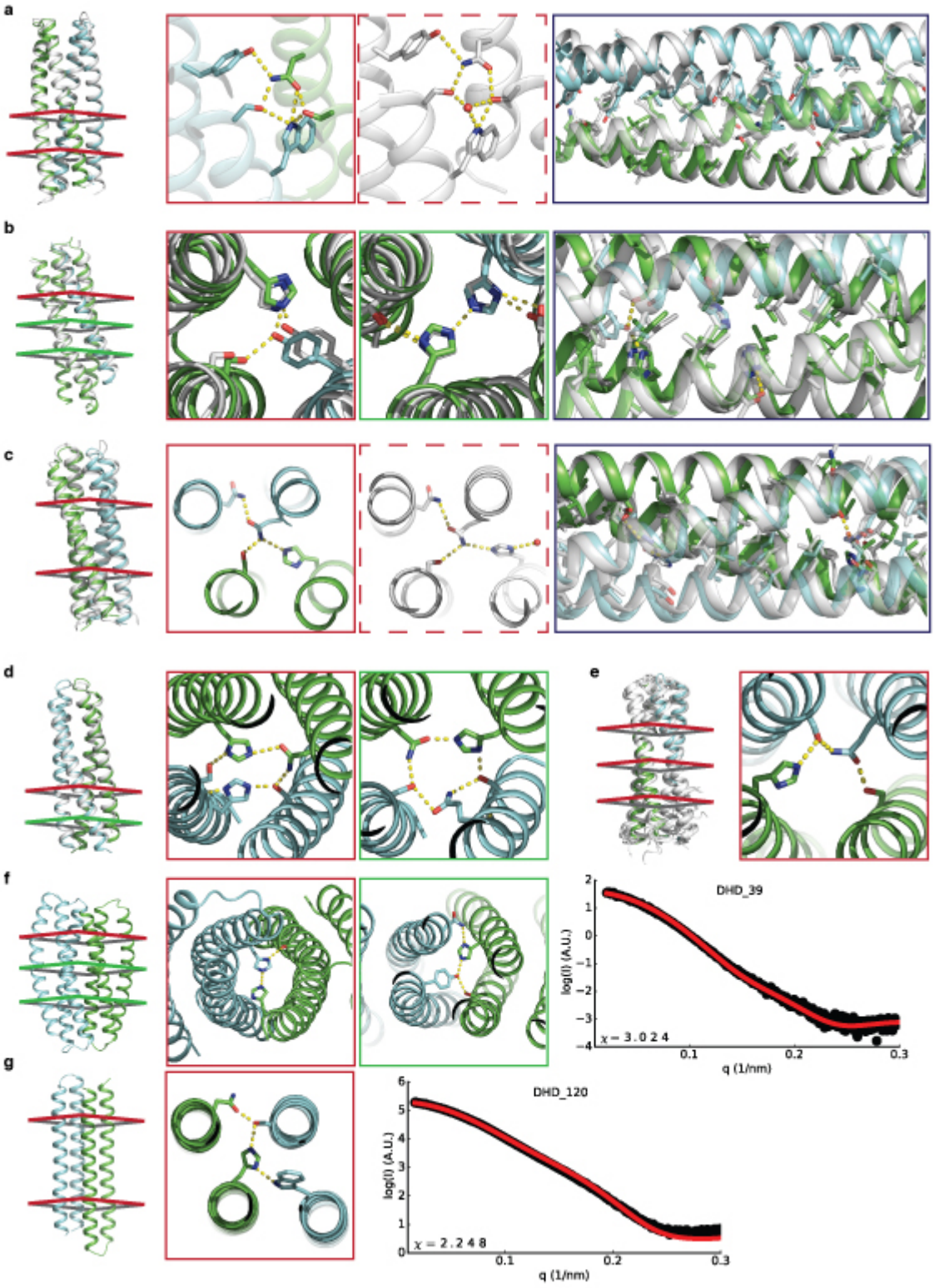


Fig. 2 **Structural characterization of designed heterodimers.** (**a-e**) Crystal and NMR structures (white) superimposed onto the design models with monomer subunits colored green

and cyan; the full backbones are shown with colored cross-sections corresponding to the designed hydrogen-bond networks (yellow dashed lines); outline color corresponds to cross-section color on the left. Solid and dashed red boxes show the designed and crystallographic hydrogen bond networks. **a**, DHD_131, 2.38 Å resolution with 1.03 Å RMSD over all Cα atoms. Black box shows overall hydrophobic packing. **b**, DHD37_1:234, 3.26 Å resolution with 1.37 Å RMSD over all Cα atoms. **c**, DHD_127, 1.75 Å resolution with 1.7 Å RMSD over all Cα atoms. **d**, DHD_15, 3.36 Å resolution with 0.95 Å RMSD over all Cα atoms. The resolution was not sufficient to resolve side chains for hydrogen bond networks. **e**, NMR ensemble (white) of DHD13_XAAA superimposed onto the design model. There were not sufficient assigned sidechain-sidechain NOEs to confirm the presence of the designed hydrogen bond networks. **f-g**, DHD_39 and DHD_120 backbones and designed hydrogen bond networks. Experimental SAXS data (black) are similar to spectra computed from the designed backbones (red).

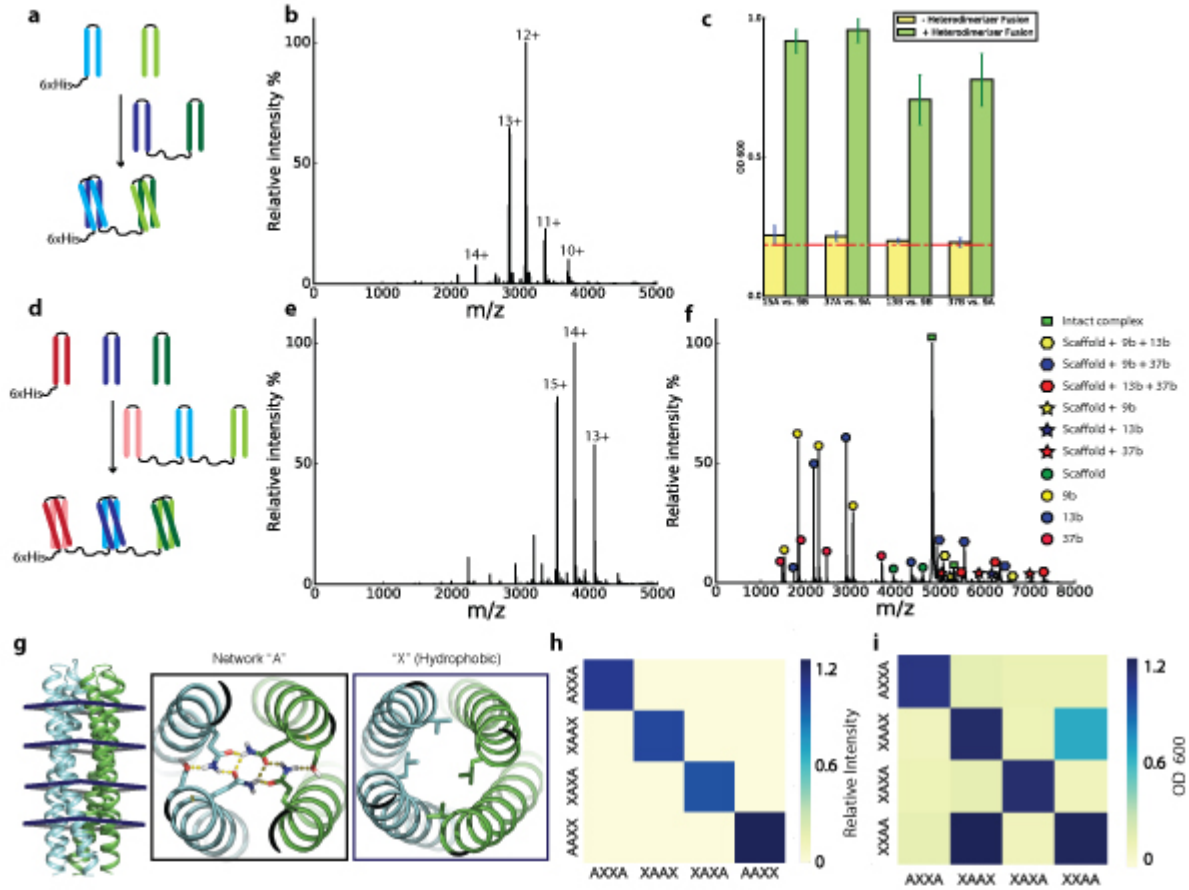


Fig. 3 **New functionality from DHD combinations**. **a**, Induced dimerizer formed from the "b" component of DHD13_XAAA (dark blue) fused to the "b" component of DHD37_ABXB (dark green) with an intervening flexible linker. The "a" components of the two heterodimers (light blue and light green) are brought into close proximity by the heterodimerizer. All three components are co-expressed in *E.coli* with one of the hairpin monomers being hexahistidine tagged. **b**, Native MS spectrum of purified DHD13_XAAA:DHD37_ABXB heterotrimer complex, no heterodimers or monomers were observed. **c**, Y2H data of 4 induced dimerization systems. Yellow, without heterodimerizer fusion; green, with heterodimerizer fusion. Red dashed line indicates background growth when no proteins were attached to AD and DBD. **d**, Synthetic scaffolding proteins based on DHDs 9, 13_XAAA and 37_ABXB. 9_a (pink), 13_XAAA_a (light blue) and 37_ABXB_a (light green) were covalently linked to form the scaffold, recruiting 9_b (red, hexahistidine tagged), 13_XAAA_b (dark blue) and 37_ABXB_b (dark green). **e**, Native MS spectrum of the purified scaffold complex, no heterotrimers, heterodimers or monomers were observed. **f**, SID of the +11 peak in **e**, showing no cross binding among the monomers 9_b,

13_XAAA_b and 37_ABXB_b. **g**, The backbone of 2L4HC2_23 can accommodate hydrogen bond networks at 4 heptad positions. **h**, Native MS mixing data of 4 variants generated using the hydrogen bond shuffling approach, showing a high degree of orthogonality. **i**, Y2H data of 4 variants generated using the hydrogen bond shuffling approach.

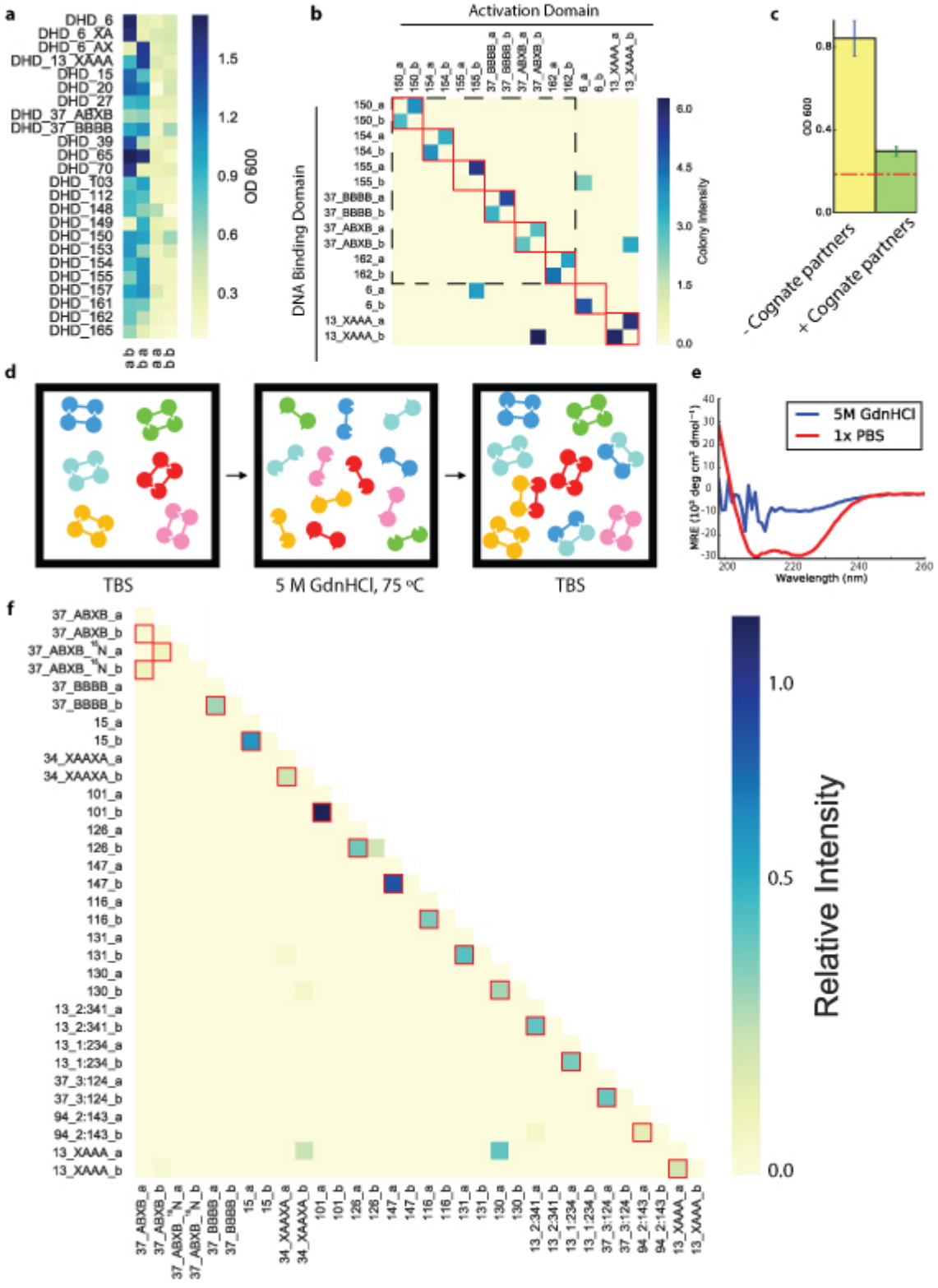


Fig. 4 **All-against-all orthogonality assessment**. **a**, Yeast two-hybrid for 21 heterodimers show constitutive heterodimer formation with little homodimer formation in yeast cells. ab, “a”

monomer fused to Activation domain (AD) and “b” monomer fused to DNA binding domain (DBD). ba, “b” monomer fused to AD and “a” monomer fused to DBD. aa, “a” monomer fused to AD and DBD. bb, “b” monomer fused to AD and DBD. **b**, Y2H all by all testing of 9 pairs of heterodimers, colors indicate colony intensity. Dashed black box indicates a set of 6 orthogonal heterodimers. Red boxes indicate individual heterodimer pairs. **c**, Off-target binding of DHD15_a and DHD13_XAAA_b, in the absence (yellow) or presence (green) of DHD15_b and DHD13_XAAA_a. Red dashed line indicates background growth when no proteins were attached to AD and DBD. **d**, Native MS mixing assay. pairs of heterodimers were mixed in a single test tube and denatured at 5M GdnHCl and 75°C, after which refolding was allowed by dialysis into TBS buffer. In the example shown, orange exchanged with red, and cyan exchanged with blue to form non-cognate pairs. **e**, CD data for a mixture of purified DHDs in PBS (red) or 5M GdnHCl and 75°C (blue). **f**, All-against-all orthogonality test of 16 pairs of heterodimers assessed by the native MS mixing assay. Red boxes indicate cognate interactions. Relative quantifications are based on signal intensities. Exchange of non- and partially ^{15}N-labeled DHD37_ABXB results in a distribution of overlapped species and accounts for the overall low signal intensities depicted.

Computational Methods

Identification of heterodimeric coiled coils in nature

A search on the CC+ database (42, 43) was performed with the following criteria (database last updated on August 10, 2016): less than or equal to 50% structural redundancy, 2 helix, hetero-oligomeric, many chains. The search returned 211 coiled coil heterodimers that fit these criteria.

Computational Design

1. Systematic sampling of parametric helical backbones

We used a generalization of the Crick coiled-coil parameters (44) to independently sample all four helices of the heterodimers supercoiled around the same axis, as described previously (25–27). The supercoil twist (ω_0) and helical twist (ω_1) were coupled and ideal values were used (27) with ω_0 and ω_1 held constant among the helices. A left-handed supercoil results from ω_0=-2.85 and ω_1=102.85, and a straight bundle with no supercoiling from ω_0=0 and ω_1=100. The supercoil phases ($\Delta\phi_0$) for the helices were fixed at 0°, 90°, 180° and 270°, respectively. The offset along the Z-axis (Z offset) for the first helix was fixed to 0 as a reference point, with the rest of the helices independently sampling from -1.51 Å to 1.51 Å, with a step size of 1.51 Å. All helices sampled helical phases ($\Delta\phi_1$) independently, from 0° to 90°, with a step size of 10°. Two of the helices with a $\Delta\phi_0$ separation of 180° sampled the radius from Z-axis (R) from 5 Å to 8 Å, while the other two sampled from 7 Å to 10 Å, all with a step size of 1 Å. Each helix is set to have 35 residues to accomodate 5 heptad repeats. After removing redundant sample points from the overlapping regions of radii sampling, the supercoiled helical bundles contained more than 60 million unique backbones, and the straight helical bundles contained more than 27 million unique backbones.

2. HBNet Search

For each parametrically generated backbone, HBNet (25) was used to search the middle heptad for hydrogen bond networks that connect all four helices, contain at least four side chains contributing hydrogen bonds, have all heavy atom donors and acceptors satisfied, and span the intermolecular interface. Symmetry was not enforced during the HBNet search. For buried interface positions, only non-charged polar amino acids were considered; for residues that were at the boundary between protein core and surface, all polar amino acids were considered. A subsequent Rosetta design calculation was performed to optimize hydrophobic packing, with atom pair restraints from HBNet being put on the newly identified hydrogen bond networks. Finally, a minimization step and

side chain repacking step was performed without atom pair restraints on hydrogen bonding residues to evaluate how well the networks remained intact in the absence of the constraints. Designs with at most 5 alanines in the middle heptad and no buried unsatisfied polar heavy atoms were selected for downstream design.

3. Generating combinations of HBNets with heptad stacking

The purpose of this step is to identify five-heptad backbones (full backbones) that can accommodate at least 3 HBNets. Instead of generating one-heptad backbones and full backbones separately, searching for HBNets in the one-heptad backbones and aligning them to all full backbones, we reasoned the heptad stacking method remains the same if one simply searches for HBNets in the middle heptad on all full backbones, extracts the middle heptads, and aligns them to all full backbones. We therefore extracted the middle heptads containing HBNets, generated all variants of chain ordering, and did pairwise alignment of middle heptads to full backbones using TMalign (28). All alignments with root mean square deviation (RMSD) less than 0.3 were identified and full backbones that can accommodate at least three middle heptads were selected for final design.

4. Connecting parametric helical backbones

Helical backbones are connected with short 2-5 residue loops such that the RMSD of each loop is less than 0.4 RMSD to a nine residues stretch in a native protein. Distance and directionality between helices limit what loops can connect, as such, our closure extends and shrinks helices by up to 3 residues. We then superimpose all short loops from the PDB onto the first and last two helical residues. The loops with the lowest stub-RMSD are minimized using the Rosetta score function onto the helical endpoints to ensure a near perfect closure. Loop quality is assessed by measuring the distance in RMSD to the closest nine stretch in the PDB. The loop with the lowest RMSD is returned as the solution. We repeat this procedure to connect all helices and report the solution with the lowest RMSD.

5. Design calculations

Backbones were regularized using Cartesian space minimization in Rosetta to alleviate any torsional strain introduced by heptad stacking. Two consecutive Rosetta packing rounds were performed with increasing weight on the repulsive energy to optimize hydrophobic packing, while constraining the hydrogen bond network residues. A FastDesign step was subsequently used within a generic Monte Carlo mover to optimize secondary structure shape complementarity, while allowing at most 8% alanine, 3 methionine and 3 phenylalanine in the protein core. The last step of minimization and side chain repacking to identify the movement of HBNets without atom pair constraints is the same as what was described in Step 2.

6. Selection criteria and metrics used to evaluate designs

Designs were selected based on the following criteria: change in polar surface area upon binding (dSASA_polar) greater than 800 Å; secondary structure shape complementarity (ss_sc) score greater than 0.65; holes score around HBNets less than -1.4; no buried unsatisfied heavy atoms; at least one buried bulky polar side chains per monomer. Selected designs were then visually inspected for good packing of hydrophobic side chains, especially the interdigitation of isoleucine, leucine and valine. Surface tyrosines were added at non-interfering positions to aid protein concentration measurement by recording OD280. Surface charge residues for a few of the designs were redesigned to shift the theoretical isoelectric point away from buffer pH.

RMSD calculations

Crystal structures and the corresponding design models were superimposed with TMalign using all heavy atoms. From this alignment, RMSD was calculated across all alpha-carbon atoms, and also across heavy atoms of the hydrogen bond network residues.

Logistic Regression

Designs were first scored with various filters in Rosetta with the filter values reported. Experimental results and Rosetta filter values were used as input to a logistic regression method (45) to find correlations between computational metrics and experimental observations.

Visualization and figures

All structural images for figures were generated using PyMOL (46).

Program code

All program code used in this study can be downloaded from Github repository https://github.com/uagaug/DeNovoHeterodimers

Experimental Methods

Buffer and media recipe

TBM-5052
1.2% [wt/vol] tryptone, 2.4% [wt/vol] yeast extract, 0.5% [wt/vol] glycerol, 0.05% [wt/vol] D-glucose, 0.2% [wt/vol] D-lactose, 25 mM Na_2HPO_4, 25 mM KH_2PO_4, 50 mM NH_4Cl, 5 mM Na_2SO_4, 2 mM $MgSO_4$, 10 µM $FeCl_3$, 4 µM $CaCl_2$, 2 µM $MnCl_2$, 2 µM $ZnSO_4$, 400 nM $CoCl_2$, 400 nM $NiCl_2$, 400 nM $CuCl_2$, 400 nM Na_2MoO_4, 400 nM Na_2SeO_3, 400 nM H_3BO_3

Lysis buffer
20 mM Tris, 300 mM NaCl, 20 mM Imidazole, pH 8.0 at room temperature

Wash buffer
20 mM Tris, 300mM NaCl, 30 mM Imidazole, pH 8.0 at room temperature

Elution buffer
20 mM Tris, 300 mM NaCl, 250 mM Imidazole, pH 8.0 at room temperature

Buffer W
100 mM Tris-HCl pH 8.0, 150 mM NaCl and 1 mM EDTA

Buffer E
Buffer W containing 2.5 mM D-desthiobiotin

TBS buffer
20 mM Tris pH 8.0, 100 mM NaCl

Construction of synthetic genes

For the expression of heterodimers, both monomers were encoded in the same plasmid, separated by a ribosome binding sequence (GAAGGAGATATCATC). Synthetic genes were ordered from Genscript Inc. (Piscataway, N.J., USA) and delivered in pET21-NESG *E. coli* expression vector, inserted between the NdeI and XhoI sites. For the pET21-NESG constructs, a hexahistidine tag and a tobacco etch virus (TEV) protease cleavage site (GSSHHHHHHSSGENLYFQGS) were added in frame at the N-terminus of the second monomer. A stop codon was introduced at the 3' end of the second monomer to stop expression of the C-terminal hexahistidine tag in the vector. For purification with Strep-tactin resin, a streptavidin tag (SAWSHPQFEKGGGSGGGSGGSAWSHPQFEKSGENLYFQGS) coding sequence was cloned in frame 5' of the first monomer sequence.

For the co-expression of 3 and 4 proteins from the same plasmid (induced dimerization and synthetic scaffold designs), synthetic genes were cloned in the pRSFDuet-1 expression vector. The first (in the case of 3 proteins) or first two (in the case of 4 proteins) genes were cloned between NcoI and HindIII sites, with a ribosome binding site separating the 2 proteins in the latter case.The last two genes were cloned between NdeI and XhoI sites, separated by a ribosome binding site. A hexahistidine tag and a TEV protease cleavage site coding sequence were cloned in frame 5' of the last gene.

Genes for yeast-two-hybrid (Y2H) studies were cloned into plasmids bearing the GAL4 transcription activation domain (poAD) and the GAL4 DNA-binding domain (poDBD).

Protein expression

Plasmids were transformed into chemically competent *E. coli* expression strains BL21(DE3)Star (Invitrogen) or Lemo21(DE3) (New England Biolabs) for protein expression. Single colonies were picked from agar plates following transformation and growth overnight, and 5 ml starter cultures were grown at 37°C in Luria-Bertani (LB) medium containing 100 µg/mL carbenicillin (for pET21-NESG vectors) or kanamycin (for pRSFDuet-1 vectors) with shaking at 225 rpm for 18 hours at 37°C. Starter cultures were diluted into 500 ml TBM-5052 containing 100 µg/mL carbenicillin or kanamycin, and incubated with shaking at 225 rpm for 24 hours at 37°C.

For expression of $^{13}C^{15}N$- or ^{15}N-labeled protein, the plasmids were transformed into the Lemo21(DE3) *E. coli* expression strain and plated on M9/glucose plates containing 50 µg/mL carbenicillin. For the starter culture, a single colony was used for inoculation of 50 mL LB medium with 50 µg/mL carbenicillin in a 250 mL baffled flask , and incubated with shaking at 225 rpm for 18 hours at 37°C. 10 mL of the starter culture was then transferred to a 2 L baffled flask containing 500 mL of Terrific Broth (Difco), with 25 mM Na2HPO4, 25 mM KH2PO4, 50 mM NH4Cl, 5 mM Na2SO4, and 100 µg/mL carbenicillin. The culture was grown at 37°C to an OD600 of approximately 1.0, then centrifuged at 5000 rcf for 15 minutes to pellet the cells. The Terrific Broth medium was removed, and the cells were washed briefly with 30 mL of phosphate buffered saline (PBS). The cells were then transferred to a fresh 2 L baffled flask containing 500 mL of labeled media (25 mM Na2HPO4, 25 mM KH2PO4, 50 mM 15NH4Cl, 5 mM Na2SO4, 0.2% (w/v) ^{13}C glucose), and 100 µg/mL carbenicillin. The cells were allowed to grow at 37°C for 2 hours, before IPTG (Carbosynth) was added to 1mM and the temperature was reduced to 18°C. The labeled glucose and NH4Cl were obtained from Cambridge Isotopes.

Affinity purification

Cells were harvested by centrifugation for 15 minutes at 5000 rcf at 4°C and resuspended in 20 ml lysis buffer. Lysozyme, DNAse, and EDTA-free cocktail protease inhibitor (Roche) were added to the resuspended cell pellet before sonication at 70% power for 5 minutes. For Immobilized metal affinity chromatography (IMAC), lysates were clarified by centrifugation at

4°C and 18,000 rpm for at least 30 minutes and applied to Ni-NTA (Qiagen) columns pre-equilibrated with lysis buffer. The column was washed two times with 5 column volumes (CV) of wash buffer, followed by 5 CV of elution buffer buffer. For Strep tag purification, elution fractions from IMAC were applied to Strep-Tactin® Superflow resin (IBA) pre-equilibrated in Buffer W. The column was washed with 5 CV Buffer W, before applying 3 CV Buffer E to elute proteins off the column. Mass and purity of eluted proteins were confirmed using electrospray ionization mass spectrometry (ESI-MS) on a Thermo Scientific TSQ Quantum Access mass spectrometer.

Size-exclusion chromatography (SEC)

N-terminal hexahistidine tags and streptavidin tags were cleaved with TEV protease overnight at room temperature, at a ratio of 1 mg TEV for 100 mg of protein. Prior to addition of TEV, buffer was exchanged into lysis buffer. After TEV cleavage, sample was passed over an additional Ni-NTA column and washed with 1.5 CV of lysis buffer, flow through were collected and further purified by SEC using a Superdex 75 10/300 increase column (GE Healthcare) in TBS buffer.

Circular dichroism (CD) measurements

CD wavelength scans (260 to 195 nm) and temperature melts (25 to 95°C) were performed using an AVIV model 420 CD spectrometer. Temperature melts were carried out at a heating rate of 4°C/min and monitored by the change in ellipticity at 222 nm; protein samples were diluted to 0.25 mg/mL in PBS pH 7.4 in a 0.1 cm cuvette. Guanidinium chloride (GdmCl) titrations were performed on the same spectrometer with automated titration apparatus in PBS pH 7.4 at 25°C, with a protein concentration of 0.025 mg/mL in a 1 cm cuvette with stir bar. Each titration consisted of at least 40 evenly distributed GdmCl concentration points with one minute mixing time for each step. Titrant solution consisted of the same concentration of protein in PBS + GdmCl.

Nuclear Magnetic Resonance (NMR)

SEC purified $^{13}C^{15}N$-labeled protein was concentrated to >1 mM and buffer exchanged into 50 mM NaCl, 20 mM sodium phosphate, 10% D_2O, 0.01% NaN3 at pH 6.3. Sample was loaded into a 5.0 mm Shigemi tube and 4 NMR experiments were recorded and analyzed: 2D-TROSY-HSQC, 4D HNCH NOESY, 4D HNCH TOCSY and 4D HCCH NOESY. The data were acquired with a non-uniform sampling scheme (NUS) and subsequently reconstructed with the SMILE program in nmrPipe. For the NOESY experiments, the mixing time was 120 ms and for the NUS protocol the data was recorded with 0.3% and 3.0% of sparsity for the HNCH and HCCH experiments, respectively. The final spectra were loaded and analyzed on Sparky 3.115.

The spin systems were identified based on supervised NMR data analysis and 148 residues were successfully assigned (93.67%). The completeness in terms of protons assigned was 87.4% (Table S4). For the structural determination, 3423 peaks were extracted from the NOESY data. 24 long-range contacts (|i-j|>4) were manually assigned with the 4D HNCH NOESY experiment and 29 in 4D HCCH. Due to the lack of stereospecific assignments (ambiguous data) the NOE contacts were considered as non-stereospecific assignments for the methyl groups of Leu and Val residues. Those contacts were principally located at the beginning, center and end of both sequences. The assignments, chemical shifts and proton-proton constraints were used for the RASREC or AutoNoe structural calculations in ROSETTA 3.

Crystallization of protein samples

Purified protein samples were concentrated to approximately 20 mg/ml in 25 mM Tris pH 8.0 and 150 mM NaCl. Samples were screened with a 5-position deck Mosquito crystal (ttplabtech) with an active humidity chamber, utilizing the following crystallization screens: JCSG+ (Qiagen), Crystal Screen (Hampton Research), PEG/Ion (Hampton Research), PEGRx HT (Hampton Research), Index (Hampton Research) and Morpheus (Molecular Dimensions). The optimal conditions for crystallization of the different designs were found as follows: OPHD_37_N3C1, 0.15 M potassium bromide and 30% w/v polyethylene glycol monomethyl ether 2000; OPHD_127, 0.12 M ethylene glycols, 0.1 M buffer system 3 pH 8,5, and 50% v/v precipitate mix 1 from the Morpheus screen; OPHD_15, 0.2 M Ammonium sulfate, 0.1 M BIS-TRIS pH 6.5, 18% v/v Polyethylene glycol 400; OPHD_15, 0.1 M Imidazole pH 7.0, and 25% v/v Polyethylene glycol monomethyl ether 550; OPHD_131, 0.2 M Ammonium acetate, 0.1 M HEPES pH 7.5, 25% w/v Polyethylene glycol 3,350. Crystals were obtained after 1 to 14 days by the hanging drop vapor diffusion method with the drops consisting of a 1:1, 2:1 and 1:2 mixture of protein solution and reservoir solution.

X-ray data collection and structure determination

The crystals of the designed proteins were looped and placed in the corresponding reservoir solution, containing 20% (v/v) glycerol if the reservoir solution did not contain cryoprotectant, and flash-frozen in liquid nitrogen. The X-ray data sets were collected at the Advanced Light Source at Lawrence Berkeley National Laboratory with beamlines 8.2.1 and 8.2.2. Data sets were indexed and scaled using either XDS (47) or HKL2000 (48). Initial models were generated by the molecular-replacement method with the program PHASER (49) within the Phenix software suite (50), using the design models as the initial search models. Efforts were made to reduce model bias through refinement with simulated annealing using Phenix.refine (51), or, if the resolution was sufficient, by using Phenix.autobuild (52) with rebuild-in-place set to false, simulated annealing and prime-and-switch phasing. Iterative rounds of manual building in COOT (53) and refinement in Phenix were used to produce the final models. Due to the high degree of self-similarity inherit in coiled-coil-like proteins, datasets for the reported structures suffered from a high degree of pseudo translational non-crystallographic symmetry, as report by

Phenix.Xtriage, which complicated structure refinement and may explain the higher than expected R values reported. RMSDs of bond lengths, angles and dihedrals from ideal geometries were calculated with Phenix (50). The overall quality of all final models was assessed using the program MOLPROBITY (54). Summaries of diffraction data and refinement statistics are provided in Supplementary Table 9.

Small Angle X-ray Scattering (SAXS)

Samples were purified by SEC in 25 mM Tris pH 8.0, 150 mM NaCl and 2% glycerol; fractions preceding the void volume of the column were used as blanks for buffer subtraction. Scattering measurements were performed at the SIBYLS 12.3.1 beamline at the Advanced Light Source. The X-ray wavelength (λ) was 1.27 Å, and the sample-to-detector distance was 1.5 m, corresponding to a scattering vector q ($q = 4\pi \sin\theta/\lambda$, where 2θ is the scattering angle) range of 0.01 to 0.3 $Å^{-1}$. A series of exposures, in equal sub-second time slices, were taken of each well: 0.3 second exposures for 10 seconds resulting in 32 frames per sample. For each sample, data was collected for two different concentrations to test for concentration-dependent effects; “low” concentration samples ranged from 2-3 mg/mL and “high” concentration samples ranged from 5-7 mg/mL. Data was processed using the SAXS FrameSlice online serve and analyzed using the ScÅtter software package (33, 55). FoXS (56, 57) was used to compare design models to experimental scattering profiles and calculate quality of fit (χ) values.

Yeast two-hybrid assay

For each pair of binders tested, chemically competent cells of yeast strain PJ69-4a (MATa trp1-901 leu2-3,112 ura3-52 his3-200 gal4(deleted) gal80(deleted) LYS2::GAL1-HIS3 GAL2-ADE2 met2::GAL7-lacZ) were transformed with the appropriate pair of plasmids containing DNA binding domain or activation domains, using the LiAc/SS carrier DNA/PEG method (58). In the case of induced dimerization, the heterodimerizer was cloned downstream of one of the “monomer proteins”, separated by a p2a and nuclear locolization sequence (GSGATNFSLLKQAGDVEENPGPGDKAELIPEPPKKKRKVELGTA). The p2a sequence ensures translational cleavage to make the heterodimerizer a separate protein from the “monomer protein”. The selection of transformed yeast cells was performed in synthetic dropout (SDO) media lacking tryptophan and leucine for 48 hours with shaking at 1000 rpm at 30°C. The resulting culture was diluted 1:100 and grown for 16 hours in fresh SDO media lacking tryptophan and leucine, before transferring to a 96 well plate and diluted 1:100 into SDO media containing 100 mM 3-Amino-1,2,4-triazole (3-AT), lacking tryptophan, leucine and histidine. This process was repeated twice, resulting in duplicates. The culture was incubated with shaking at 1000 rpm at 30°C. Since bringing the DNA binding domain and the transcription activation domain into proximity is necessary for the growth of yeast cells in media lacking histidine, binding of two proteins was indicated by the growth of yeast cells (59, 60). The optical density of yeast cells was recorded after 48 hours. For Y2H assay on agar plates, the 1:100 diluted overnight culture was transferred onto Nunc™ OmniTray™ (Thermo Fisher) using a 96 Solid

Pin Multi-Blot Replicator (V&P Scientific), with the agar lacking tryptophan, leucine and histidine, and containing 100 mM 3-AT. The plates were imaged daily until Day 5 to monitor the sizes of colonies. Images were analyzed by the ColonyArea (61) package on ImageJ.

Native MS assessment of heterodimer affinity

Samples were buffer exchanged into 200 mM ammonium acetate using Micro Bio-Spin 6 columns (Bio-Rad). Protein concentrations were determined spectroscopically by a NanoDrop 2000c (Thermo Fisher Scientific). After dilution with 200 mM ammonium acetate, proteins were allowed to equilibrate for 24h at 4 ^{0}C. Mass spectra were subsequently recorded by nanoESI-MS using an Exactive Plus EMR Orbitrap instrument (Thermo Fisher Scientific) modified to incorporate a quadrupole mass filter and allow surface-induced dissociation (35, 36, 62).

Native MS of individual heterodimers

Sample purity and integrity were first analyzed using a self-packed buffer exchange column (63) (P6 polyacrylamide gel, BioRad, Hercules CA), coupled online to an Exactive Plus EMR Orbitrap instrument (Thermo Fisher Scientific) modified to incorporate a quadrupole mass filter and allow surface-induced dissociation. For online buffer-exchange, 200 mM ammonium acetate, pH 6.8 (AmAc) was used as a mobile phase. Samples that showed specific dimer formation and a good correlation with the theoretical monomer/dimer masses (Extended Data Fig. 5) were selected for mixing experiments.

Native MS mixing assay and data analysis

In the mixing experiment, heterodimers were mixed in equimolar ratio of 10 uM. Guanidine hydrochloride (GdnHCl) was added to a final concentration of 5M and the mixture was incubated at 75 ^{0}C for 30 min to ensure complete denaturation. To allow for the relative quantification of exchanged species, a control mixing experiment was performed where the denaturation and refolding steps were omitted. The mixtures were then dialyzed against 150 mM AmAc solution for refolding and subsequent formation of protein protein interactions. 8 µL of sample was injected on a ProPac WCX-10 column and separately, a ProPac WAX-10 column (Thermo Scientific) and separated using a Dionex UltiMate 3000 HPLC (Thermo Scientific) by a salt gradient elution from 20 mM AmAc to 1000 mM AmAc over a period of 55 min. The eluting proteins were detected online by a modified Exactive Plus EMR Orbitrap mass spectrometer. LC-MS analysis was performed for mixtures in full MS mode (no collision voltage applied) and all-ion fragmentation (MSMS) mode with high energy collision-induced dissociation (HCD) 100 V and surface-induced dissociation (SID) 85 V, respectively (manuscript in preparation). Details on instrument settings are in Supplementary Table 11. Data were deconvoluted using Xcalibur™ (Thermo Scientific), UniDec (64) and Intact Mass™ (Protein Metrics (65)). The detailed deconvolution parameters are listed in Supplementary Table 12. The deconvoluted mass lists from Intact Mass™ were searched against a theoretical mass list of all possible monomers to tetramers combinations. Only one trimeric species was found, which corresponds to the cognate

13_XAAA_b + 13_2:341_a + 13_1:234_a hetero-trimer formation, which constitutes the 13_XAAA design with the two helices of its “a” monomer coming from 13_2:341_a and 13_1:234_a. Dimers were identified based on the full MS runs and MSMS runs with both subunits being detected at the same retention time. The mass tolerance was set to 2 Da and intensity tolerance was set to 1% of the highest intensity. The relative intensity was calculated using the equation

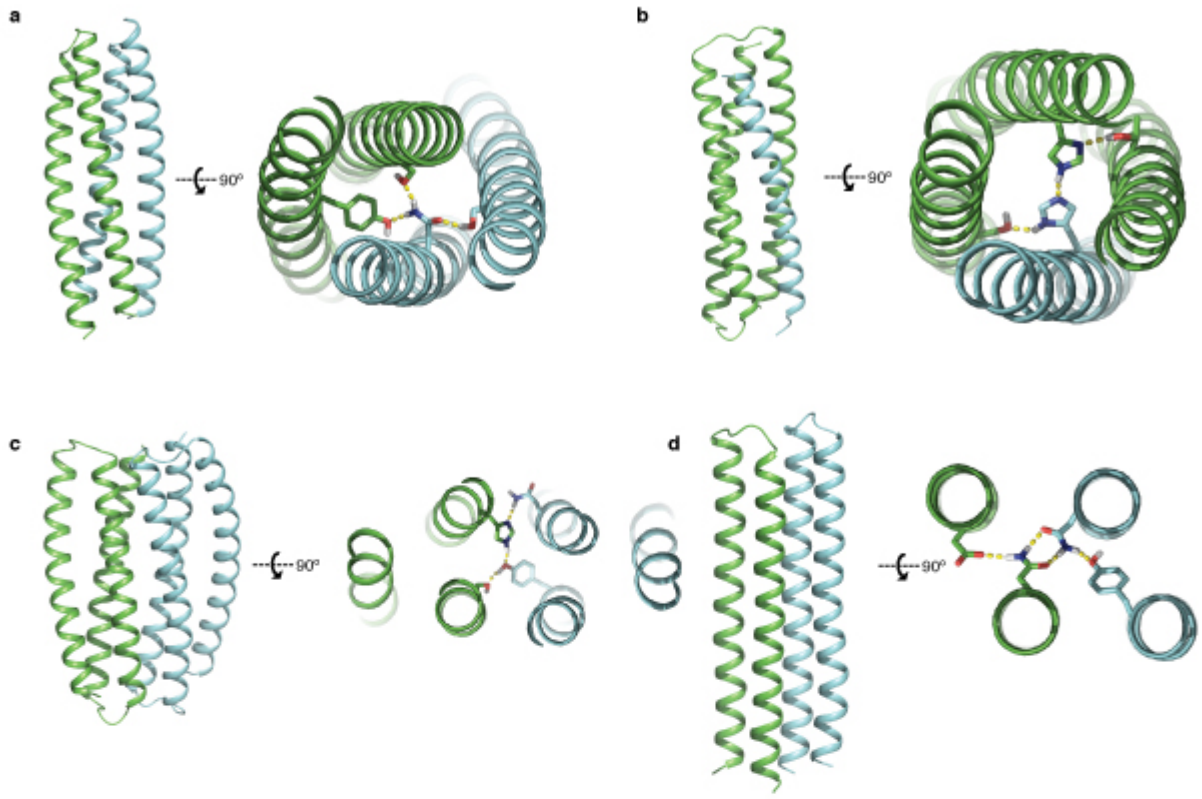
$$I_r(N_aM_b) = \frac{I_{DN}(N_aM_b)}{2I_N(N_aN_b)} + \frac{I_{DN}(N_aM_b)}{2I_N(M_aM_b)}$$

, where $I_r(N_aM_b)$ is the relative intensity of a dimer N_aM_b identified in the mixing experiment. $I_{DN}(N_aM_b)$ is the intensity of the N_aM_b species in the run involving denaturation and refolding. $I_N(N_aN_b)$ and $I_N(M_aM_b)$ are the intensity of the cognate pairs N_aN_b and M_aM_b in the run that skipped denaturation and refolding. The native MS mixing workflow is shown in Extended Data Fig. 21.

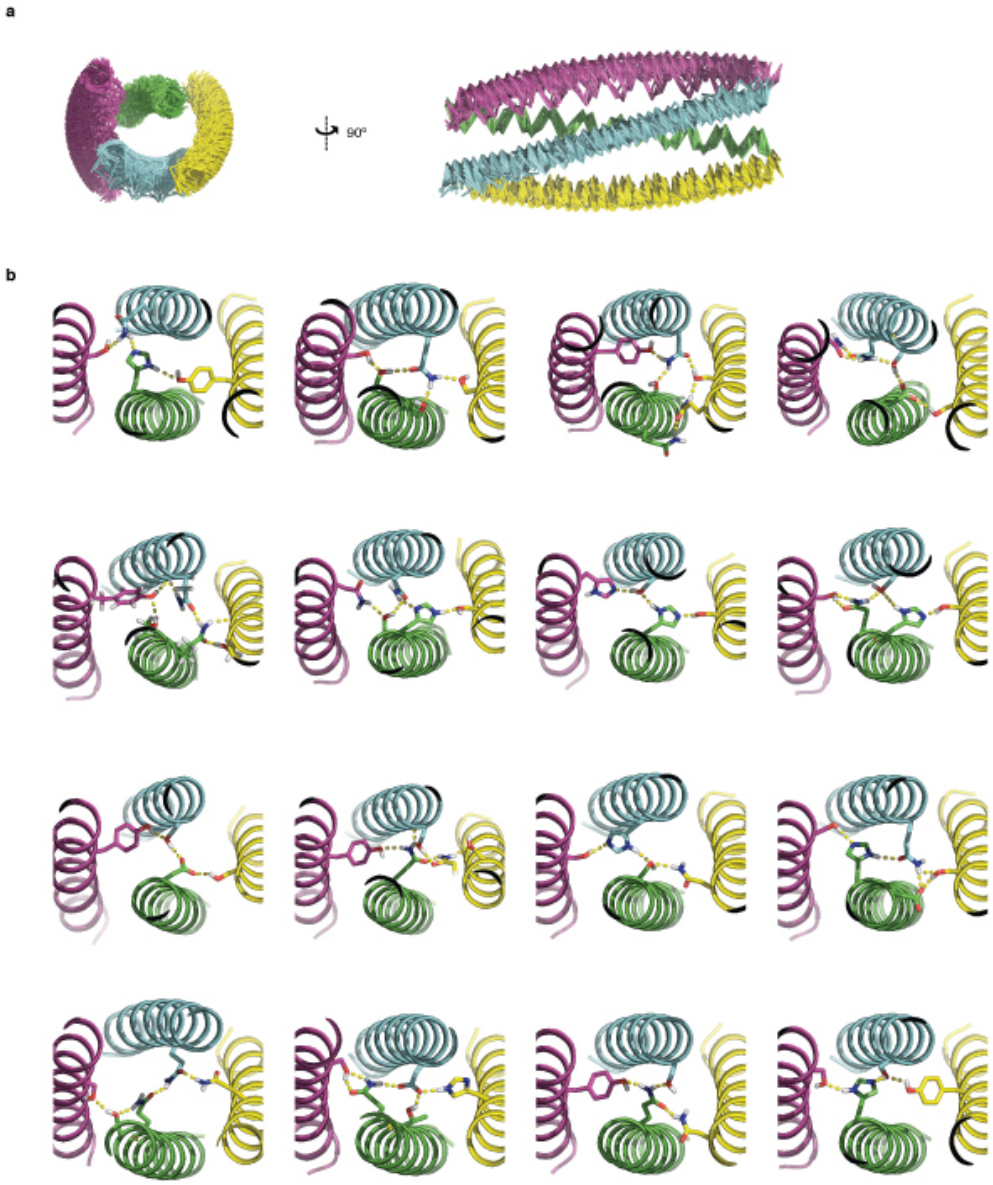
Native MS of higher-order hetero-oligomers

Samples were buffer exchanged into 200 mM ammonium acetate using Micro Bio-Spin 6 columns (Bio-Rad). 20% (v/v) 200 mM triethylammonium acetate (Sigma) was added for charge-reduction. Surface-induced dissociation (SID) was performed on an in-house modified SYNAPT G2 HDMS (Waters Corporation) with an SID device incorporated between a truncated trap travelling wave ion guide and the ion mobility cell (35). The following instrument parameters were used: sampling cone, 20 V; extraction cone, 2 V; source temperature, 20 °C; trap gas flow, 2 mL/min; trap bias, 45V. The SID settings are listed in Supplementary Table 13.

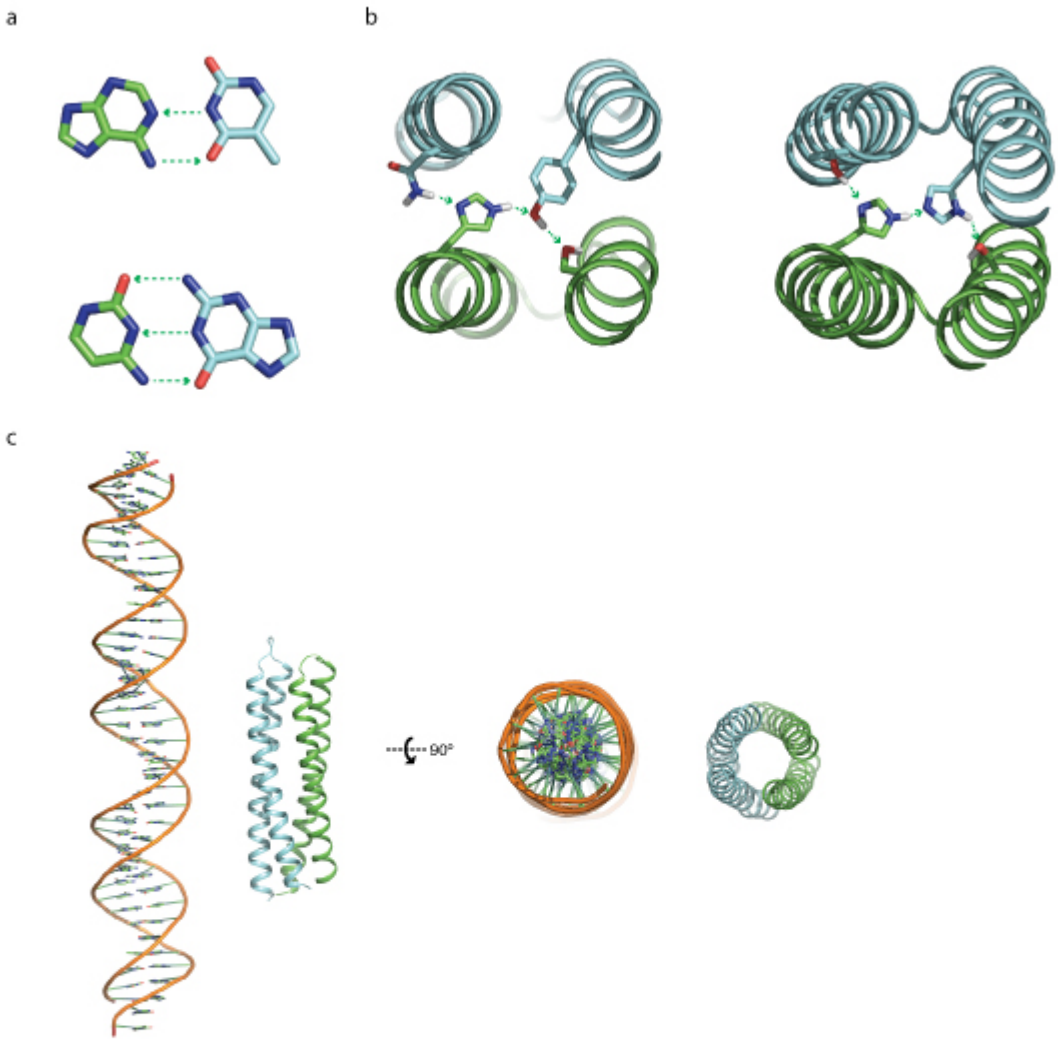
Additional Figures



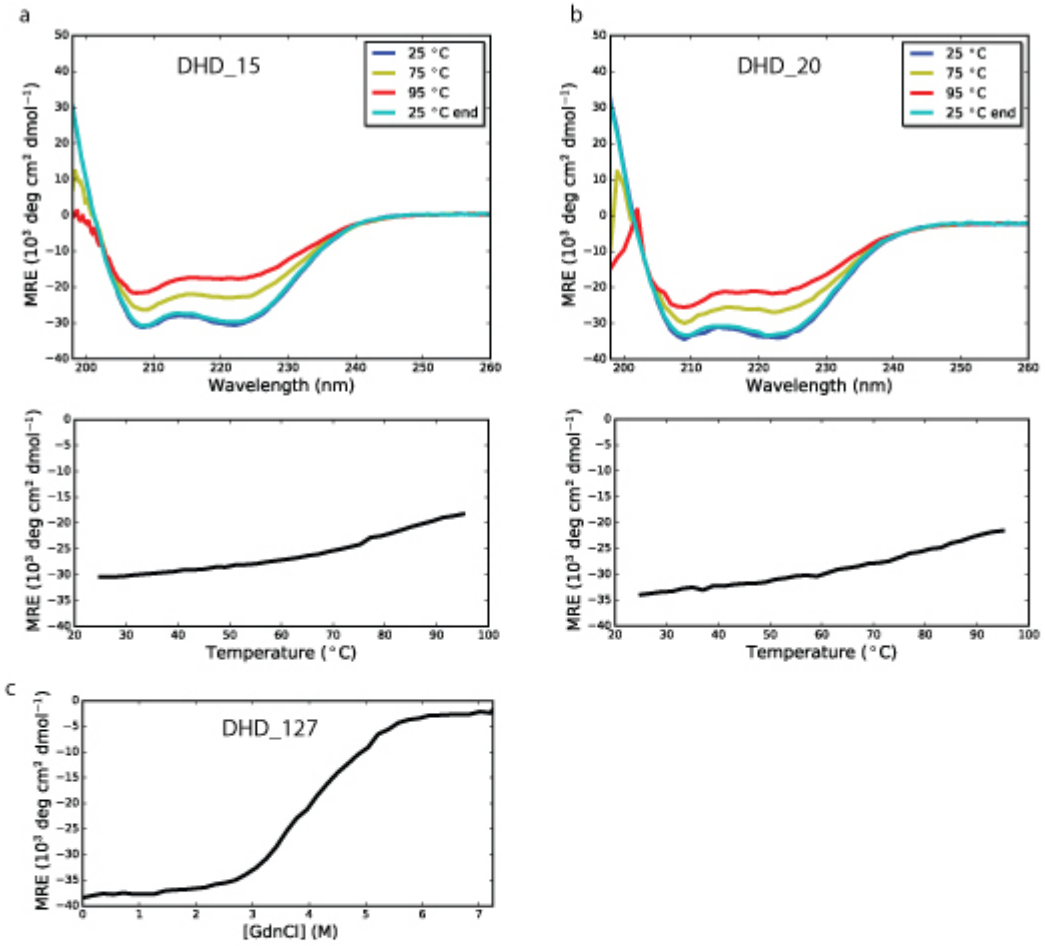
Extended Data Fig. 1 Overview of different topologies designed in this study. a-d, Overall topology on the left and example HBNets on the right. **a,** A left-handed supercoil backbone, with each monomer being helix hairpins. **b,** A backbone permuted “3+1” design, one monomer is a single helix and the other is a three helix bundle. **c,** A left-handed supercoil backbone, with each monomer being a three helix bundle. **d,** A straight, untwisted backbone, with each monomer being helix hairpins.



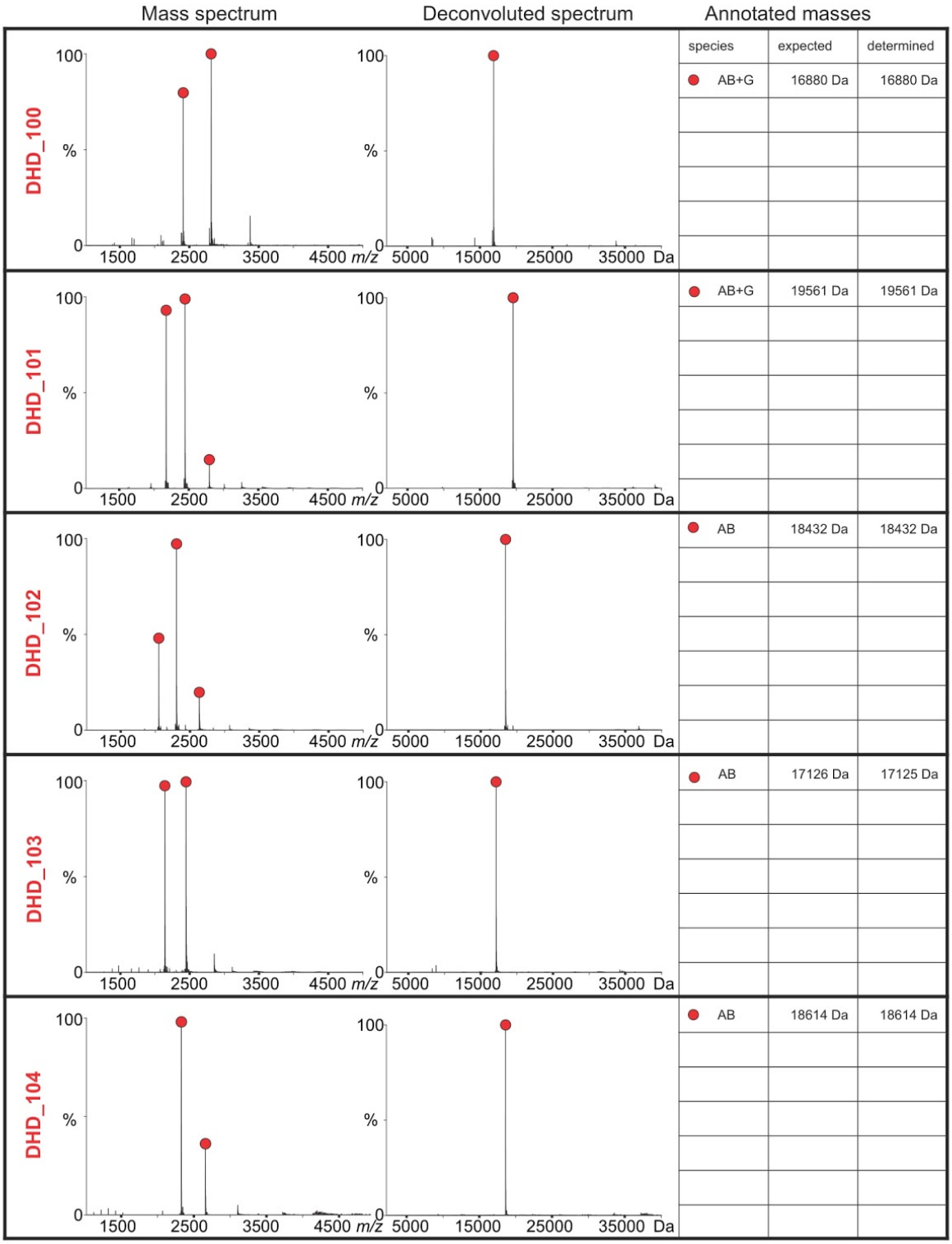
Extended Data Fig. 2 Example HBNets resulting from the systematic search. a, Overlay of 50 backbones with different Crick parameters for each helix. **b,** Example hydrogen bond networks from the systematic search, each involving at least 4 residues and contacting all 4 helices.



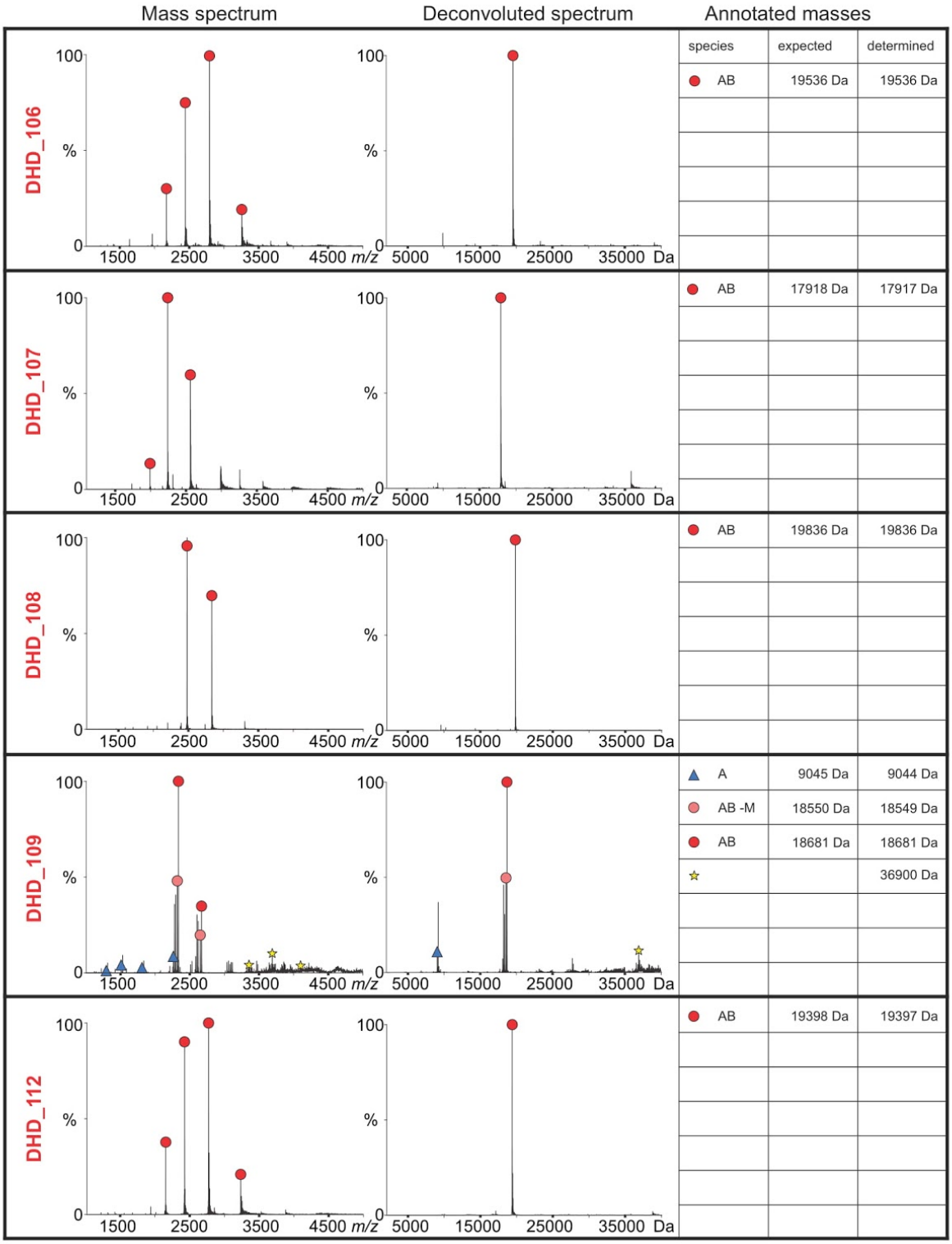
Extended Data Fig. 3 Comparing designed protein heterodimers to DNA. a, Hydrogen bond pairing in DNA bases. Top, A-T pairing. Bottom, C-G base pairing. Green arrows point from hydrogen bond donors to acceptors. **b,** Hydrogen bond pairing in designed protein hydrogen bond networks. **c,** Comparing a designed protein heterodimer with B-form DNA on the same scale.



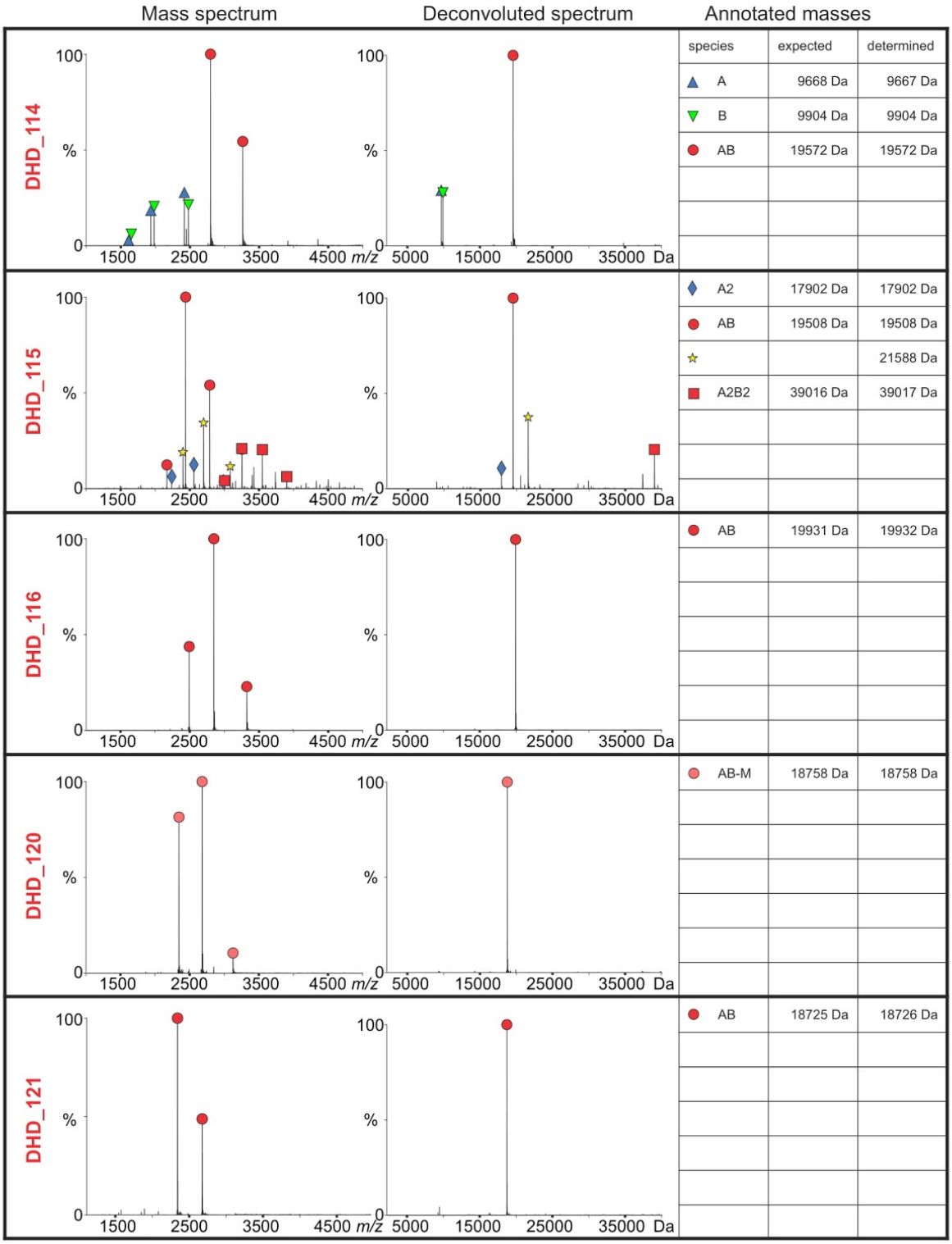
Extended Data Fig. 4 Thermal and chemical denaturation of DHDs. a-b, CD spectra for thermal denaturation of DHD_15 and DHD_20, respectively. Top, wavelength scan at 25°C, 75°C, 95°C, and final 25°C. Designs were alpha helical and stable up to 95°C. Bottom, CD temperature melts, monitoring absorption at 222 nm as temperature was increased from 25°C to 95°C. **c,** GdnHCl denaturation of DHD_127 by CD monitoring absorption at 222 nm.

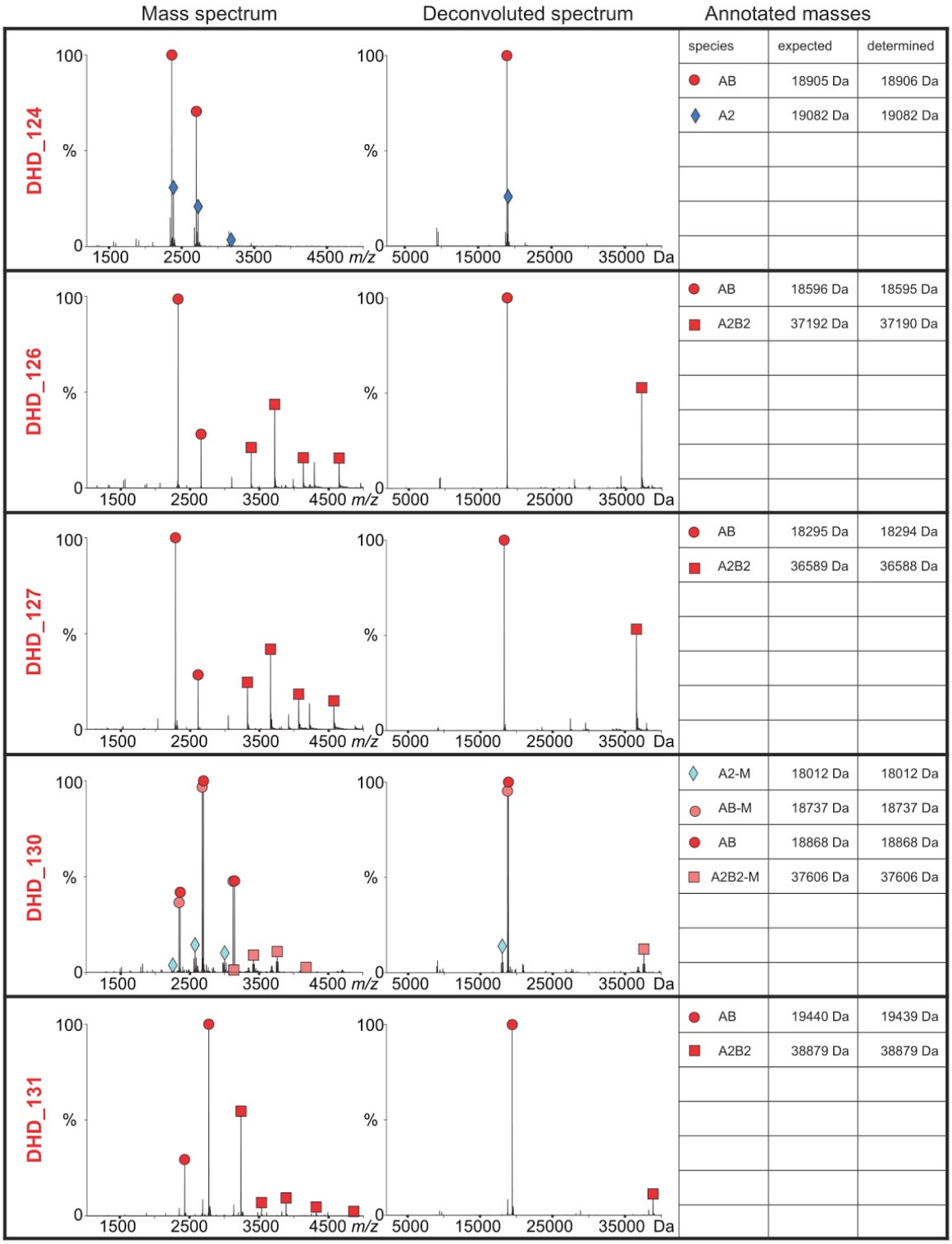


species	expected	determined
AB+G	16880 Da	16880 Da
AB+G	19561 Da	19561 Da
AB	18432 Da	18432 Da
AB	17126 Da	17125 Da
AB	18614 Da	18614 Da

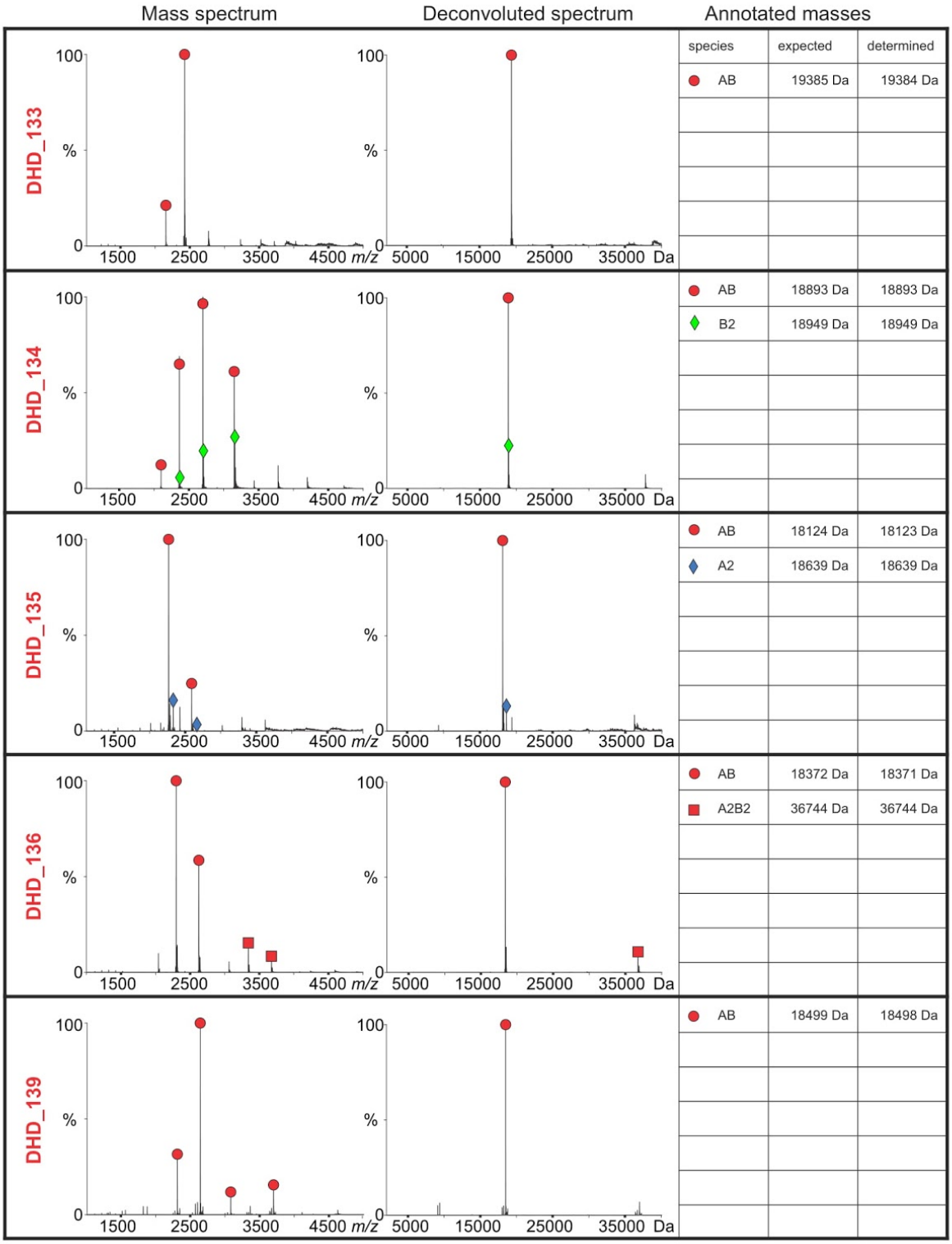


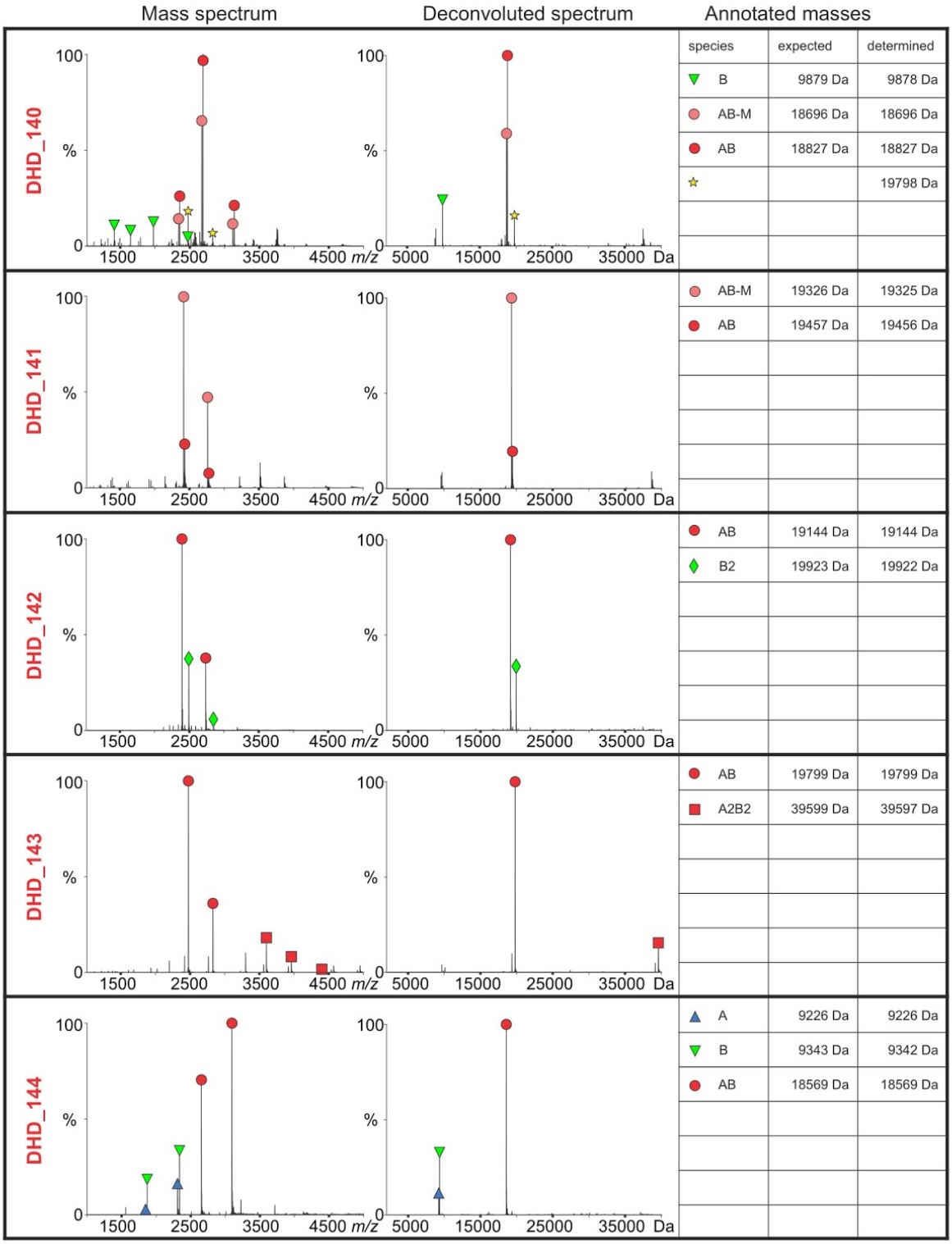
species	expected	determined
AB	19536 Da	19536 Da
AB	17918 Da	17917 Da
AB	19836 Da	19836 Da
A	9045 Da	9044 Da
AB -M	18550 Da	18549 Da
AB	18681 Da	18681 Da
☆		36900 Da
AB	19398 Da	19397 Da


Mass spectrum
Deconvoluted spectrum
Annotated masses
DHD_114
DHD_115
DHD_116
DHD_120
DHD_121
100
%
0
1500
2500
3500
4500 m/z
5000
15000
25000
35000 Da
species
expected
determined
A
9668 Da
9667 Da
B
9904 Da
9904 Da
AB
19572 Da
19572 Da
A2
17902 Da
17902 Da
AB
19508 Da
19508 Da
21588 Da
A2B2
39016 Da
39017 Da
AB
19931 Da
19932 Da
AB-M
18758 Da
18758 Da
AB
18725 Da
18726 Da



species	expected	determined
AB	18905 Da	18906 Da
A2	19082 Da	19082 Da
AB	18596 Da	18595 Da
A2B2	37192 Da	37190 Da
AB	18295 Da	18294 Da
A2B2	36589 Da	36588 Da
A2-M	18012 Da	18012 Da
AB-M	18737 Da	18737 Da
AB	18868 Da	18868 Da
A2B2-M	37606 Da	37606 Da
AB	19440 Da	19439 Da
A2B2	38879 Da	38879 Da


Mass spectrum
Deconvoluted spectrum
Annotated masses
DHD_133
DHD_134
DHD_135
DHD_136
DHD_139
100
%
0
1500
2500
3500
4500 m/z
5000
15000
25000
35000 Da
species
expected
determined
AB
19385 Da
19384 Da
AB
18893 Da
18893 Da
B2
18949 Da
18949 Da
AB
18124 Da
18123 Da
A2
18639 Da
18639 Da
AB
18372 Da
18371 Da
A2B2
36744 Da
36744 Da
AB
18499 Da
18498 Da



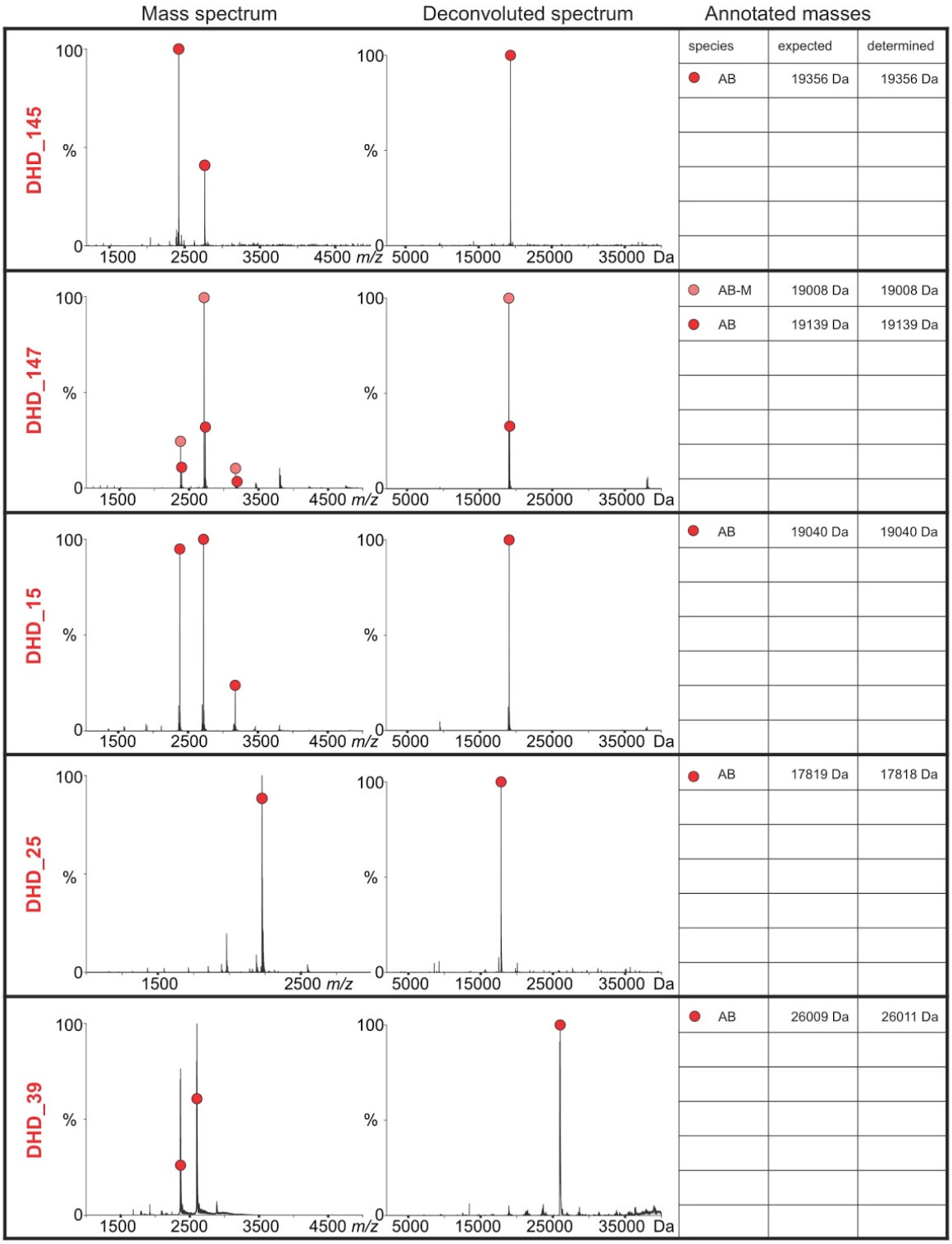
species	expected	determined
B	9879 Da	9878 Da
AB-M	18696 Da	18696 Da
AB	18827 Da	18827 Da
☆		19798 Da

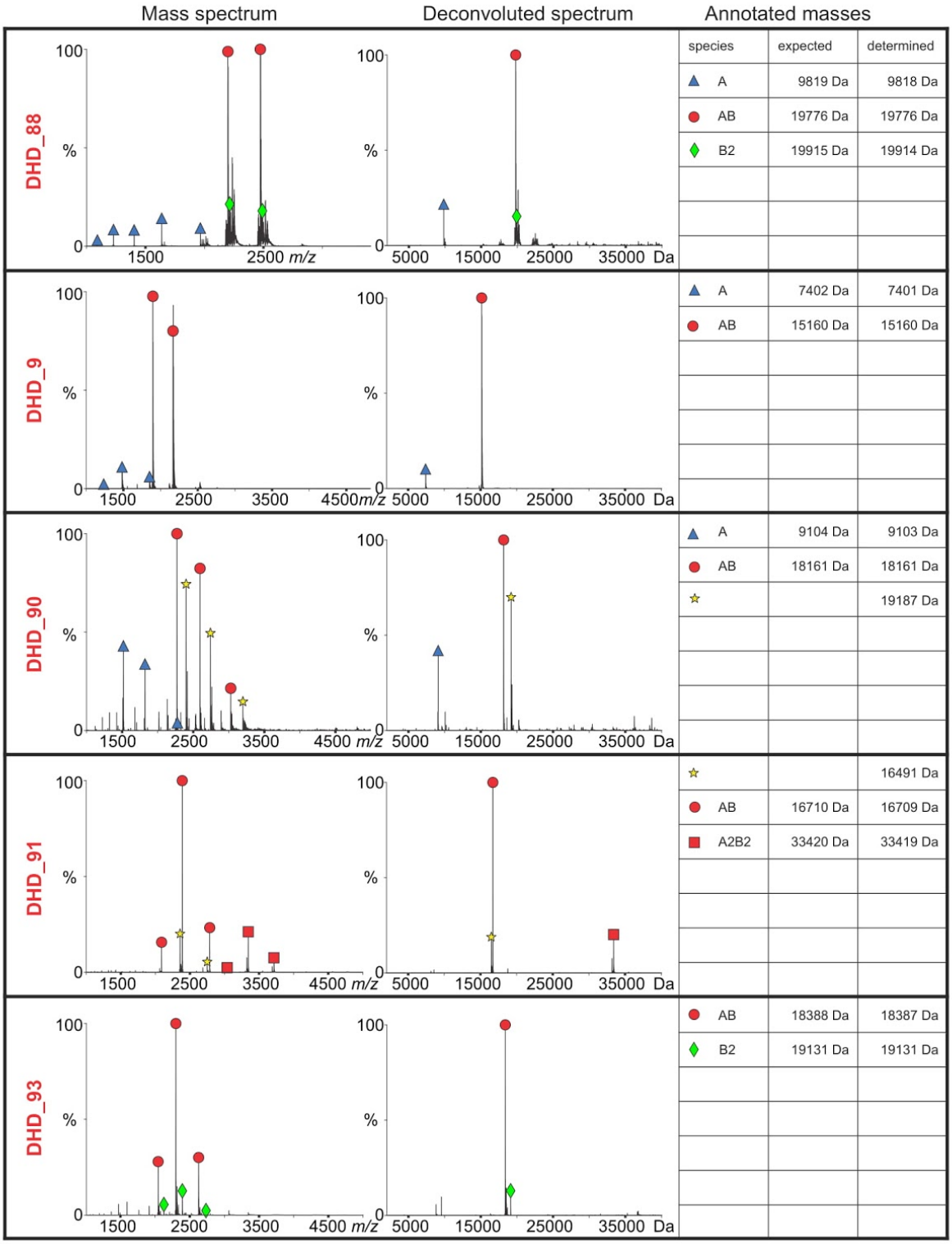
species	expected	determined
AB-M	19326 Da	19325 Da
AB	19457 Da	19456 Da

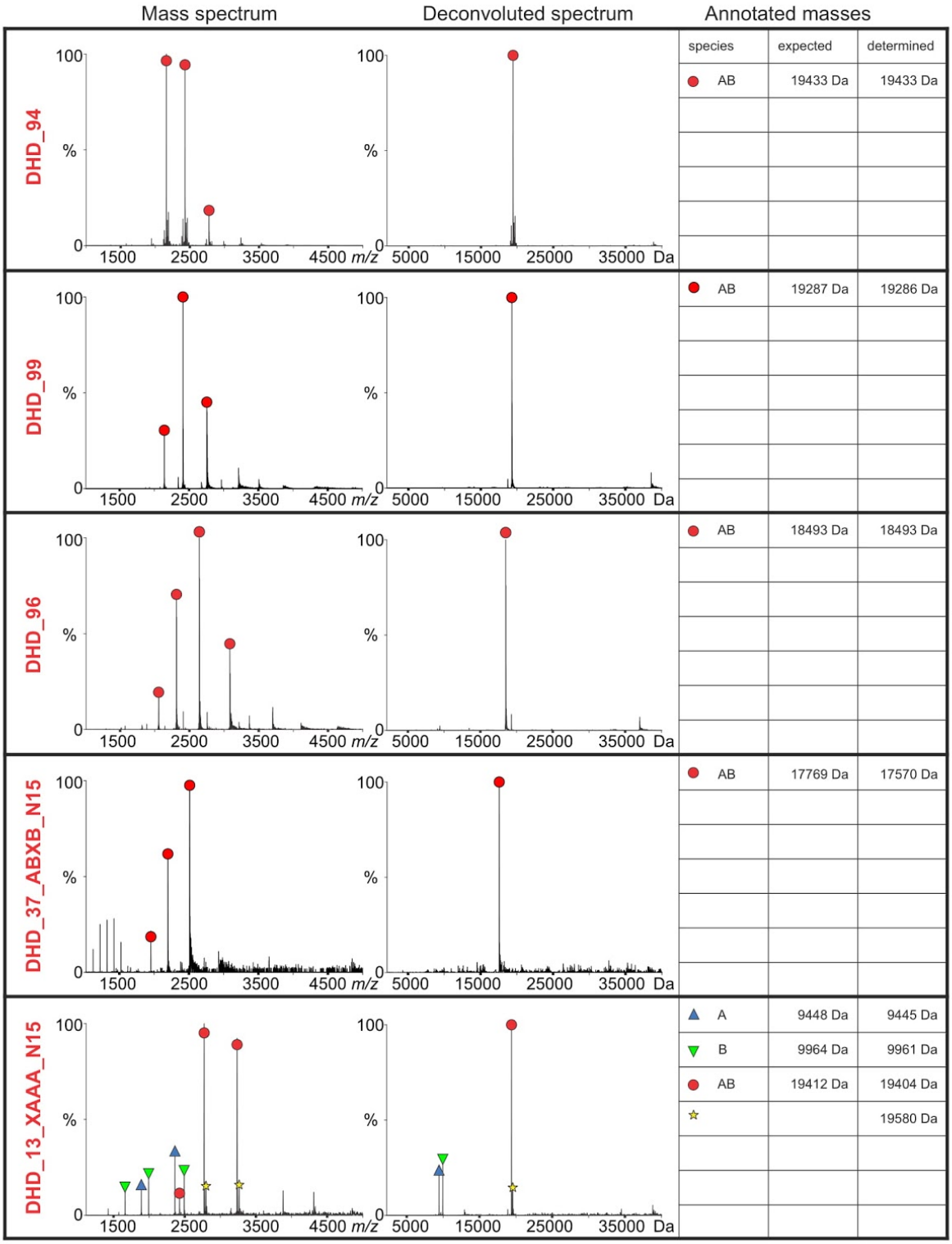
species	expected	determined
AB	19144 Da	19144 Da
B2	19923 Da	19922 Da

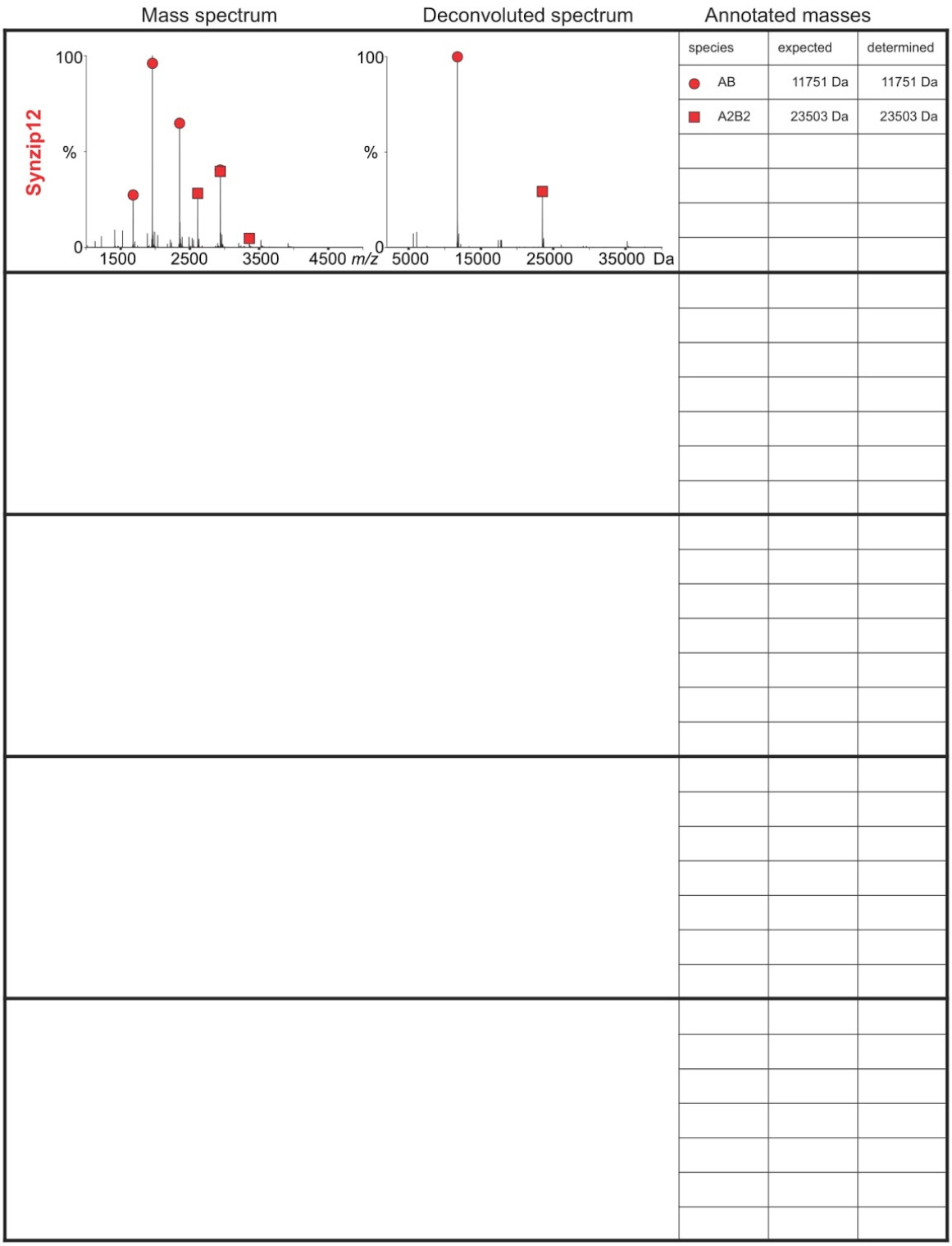
species	expected	determined
AB	19799 Da	19799 Da
A2B2	39599 Da	39597 Da

species	expected	determined
A	9226 Da	9226 Da
B	9343 Da	9342 Da
AB	18569 Da	18569 Da


Mass spectrum
Deconvoluted spectrum
Annotated masses
DHD_145
DHD_147
DHD_15
DHD_25
DHD_39
100
%
0
1500
2500
3500
4500 m/z
5000
15000
25000
35000 Da
species
expected
determined
AB
19356 Da
19356 Da
AB-M
19008 Da
19008 Da
AB
19139 Da
19139 Da
AB
19040 Da
19040 Da
AB
17819 Da
17818 Da
AB
26009 Da
26011 Da

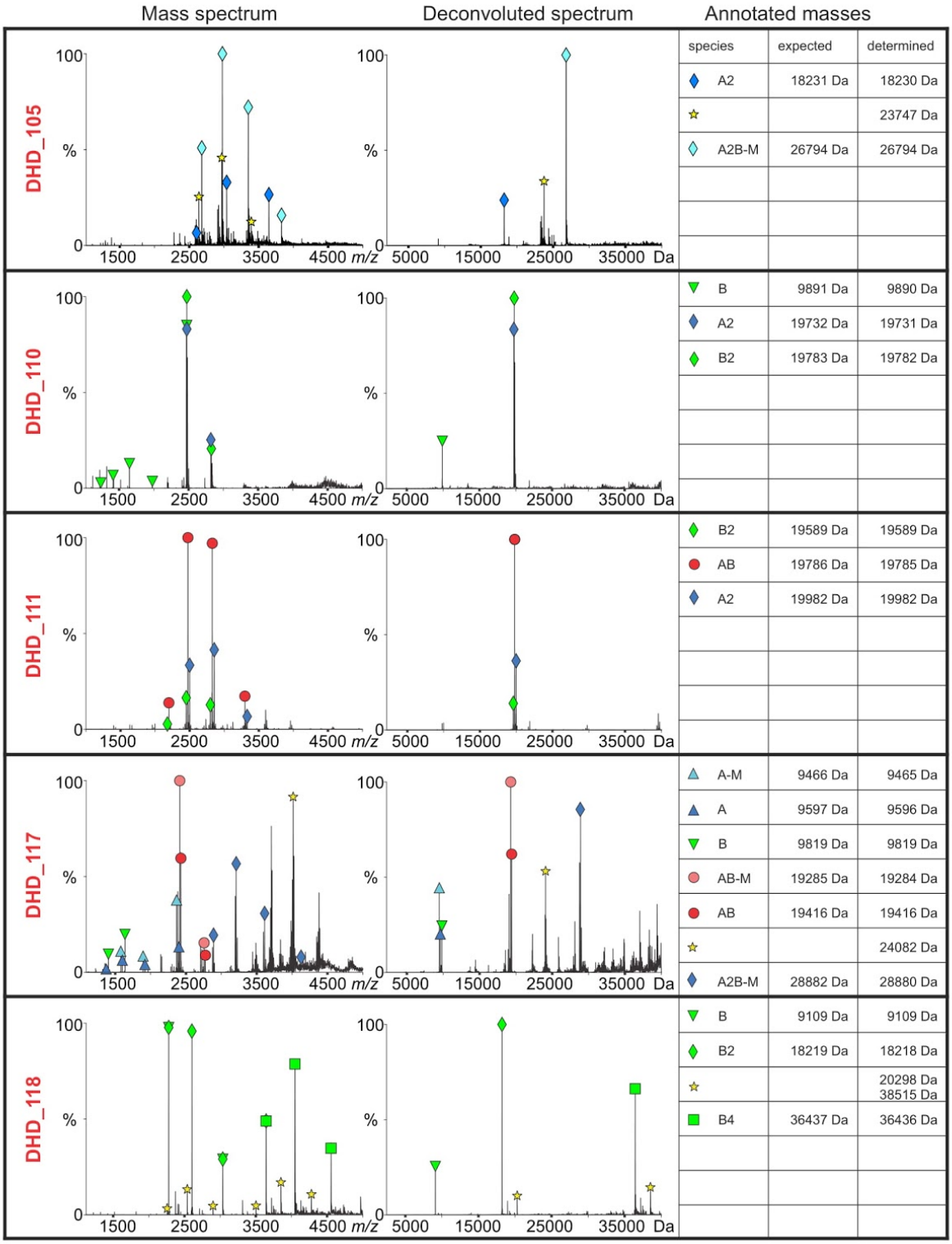

Mass spectrum
Deconvoluted spectrum
Annotated masses
DHD_88
DHD_9
DHD_90
DHD_91
DHD_93
100
%
0
1500
2500 m/z
4500 m/z
5000
15000
25000
35000 Da
species
expected
determined
A 9819 Da 9818 Da
AB 19776 Da 19776 Da
B2 19915 Da 19914 Da
A 7402 Da 7401 Da
AB 15160 Da 15160 Da
A 9104 Da 9103 Da
AB 18161 Da 18161 Da
19187 Da
16491 Da
AB 16710 Da 16709 Da
A2B2 33420 Da 33419 Da
AB 18388 Da 18387 Da
B2 19131 Da 19131 Da


Mass spectrum
Deconvoluted spectrum
Annotated masses
DHD_94
DHD_99
DHD_96
DHD_37_ABXB_N15
DHD_13_XAAA_N15
100
%
0
1500
2500
3500
4500 m/z
5000
15000
25000
35000 Da
species
expected
determined
AB
19433 Da
19433 Da
AB
19287 Da
19286 Da
AB
18493 Da
18493 Da
AB
17769 Da
17570 Da
A
9448 Da
9445 Da
B
9964 Da
9961 Da
AB
19412 Da
19404 Da
19580 Da

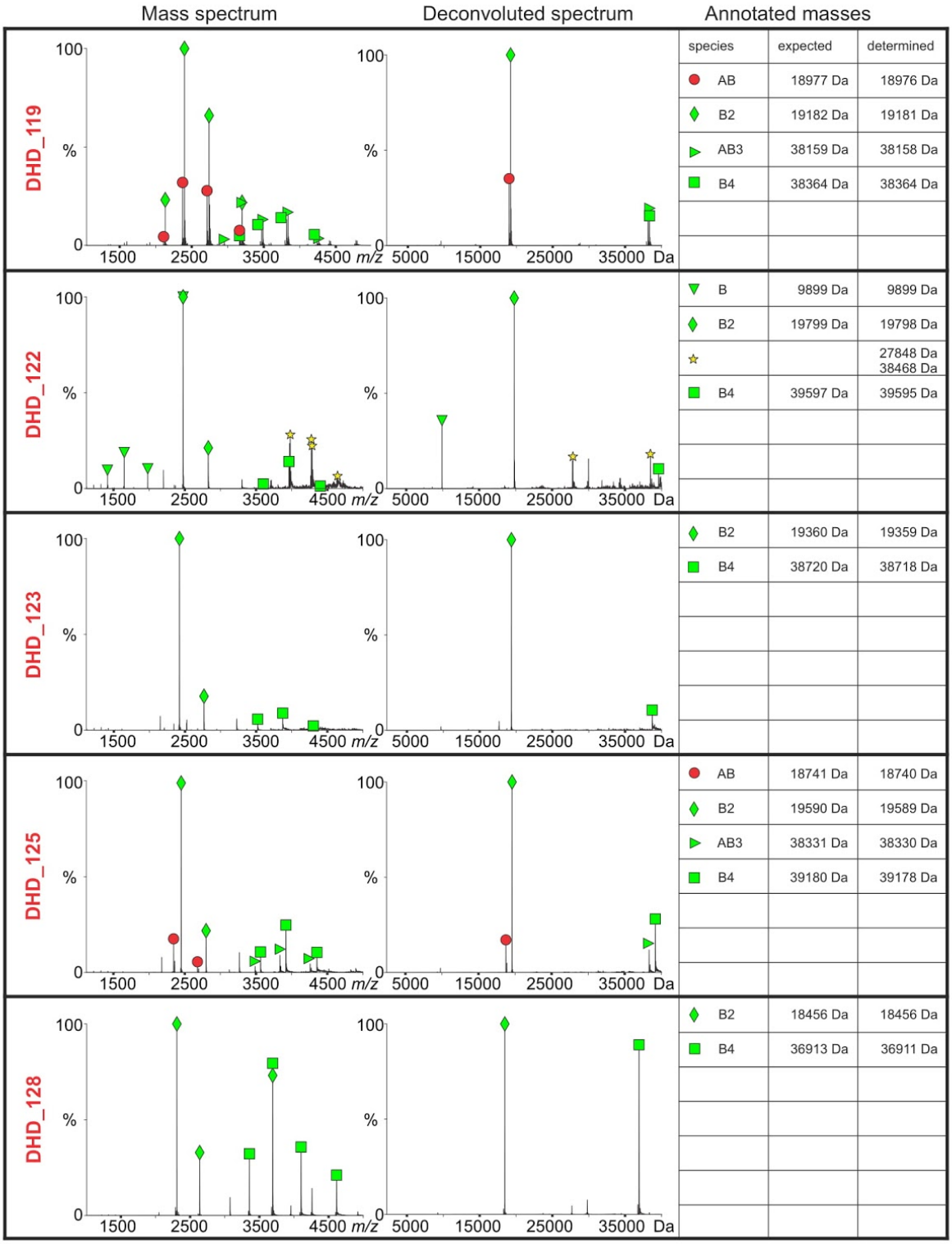


Extended Data Fig. 5 Native MS spectra of designs showing constitutive heterodimers. A, monomer A. B, monomer B. AB+G, a heterodimer with an additional N-terminal Gly. AB-M, a

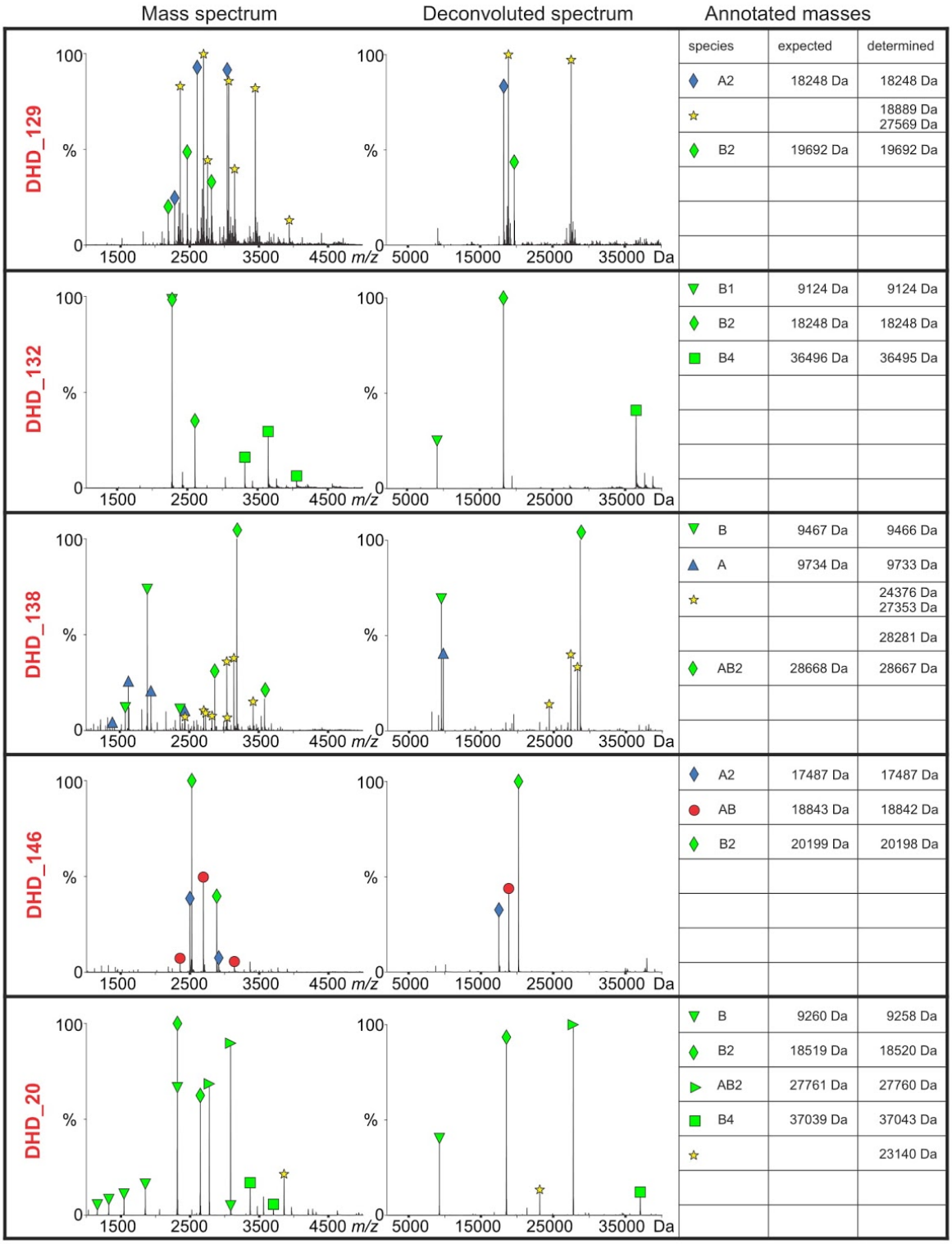
heterodimer with the N-terminal Met cleaved off. Yellow star, contaminant. A2, a homodimer of monomer A. B2, a homodimer of monomer B. A2B2, a hetero-tetramer consisting of 2 copies of monomer A and 2 copies of monomer B. Synzip12, a previously reported *de novo* coiled coil heterodimer (17).

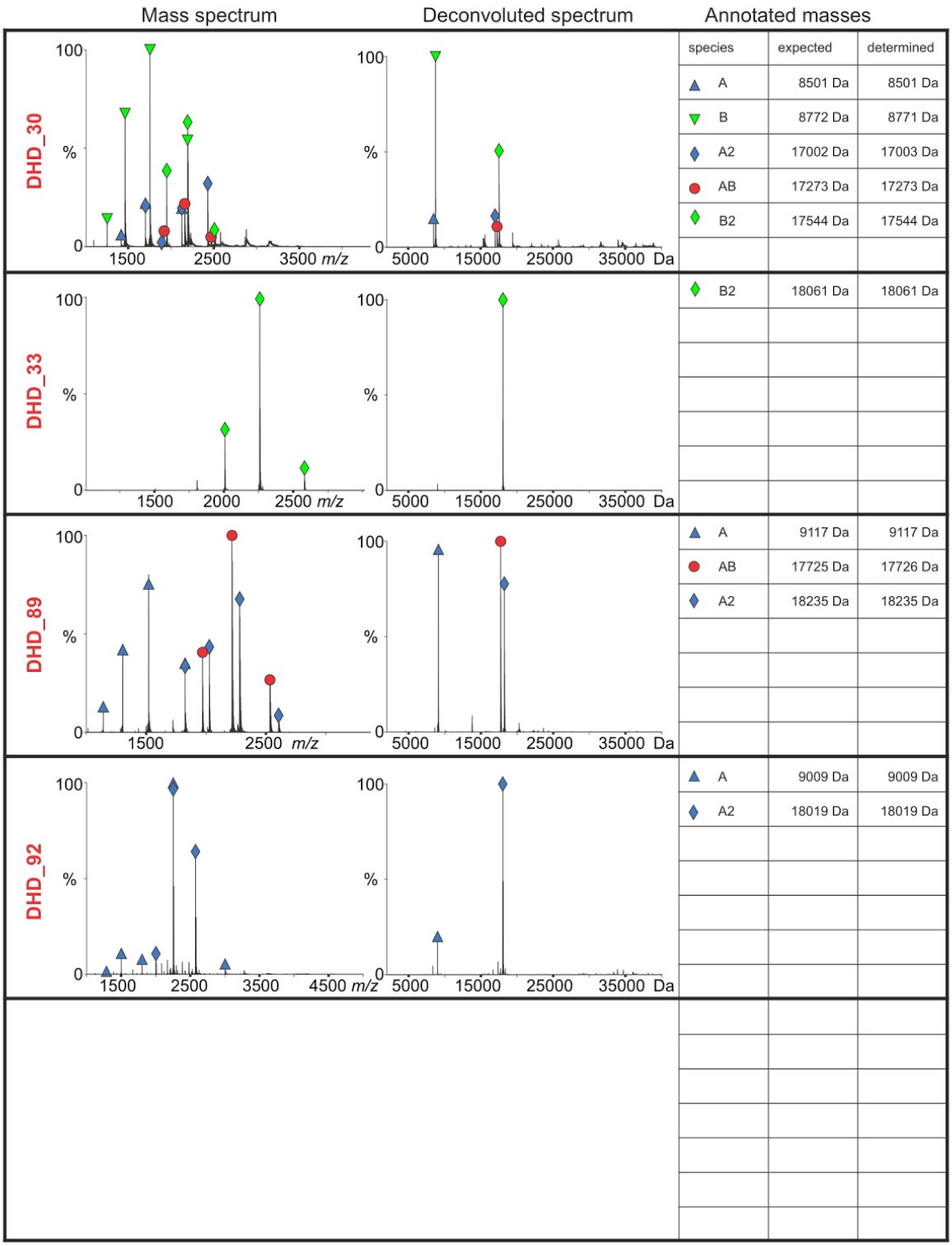


species	expected	determined
A2	18231 Da	18230 Da
		23747 Da
A2B-M	26794 Da	26794 Da
B	9891 Da	9890 Da
A2	19732 Da	19731 Da
B2	19783 Da	19782 Da
B2	19589 Da	19589 Da
AB	19786 Da	19785 Da
A2	19982 Da	19982 Da
A-M	9466 Da	9465 Da
A	9597 Da	9596 Da
B	9819 Da	9819 Da
AB-M	19285 Da	19284 Da
AB	19416 Da	19416 Da
		24082 Da
A2B-M	28882 Da	28880 Da
B	9109 Da	9109 Da
B2	18219 Da	18218 Da
		20298 Da 38515 Da
B4	36437 Da	36436 Da

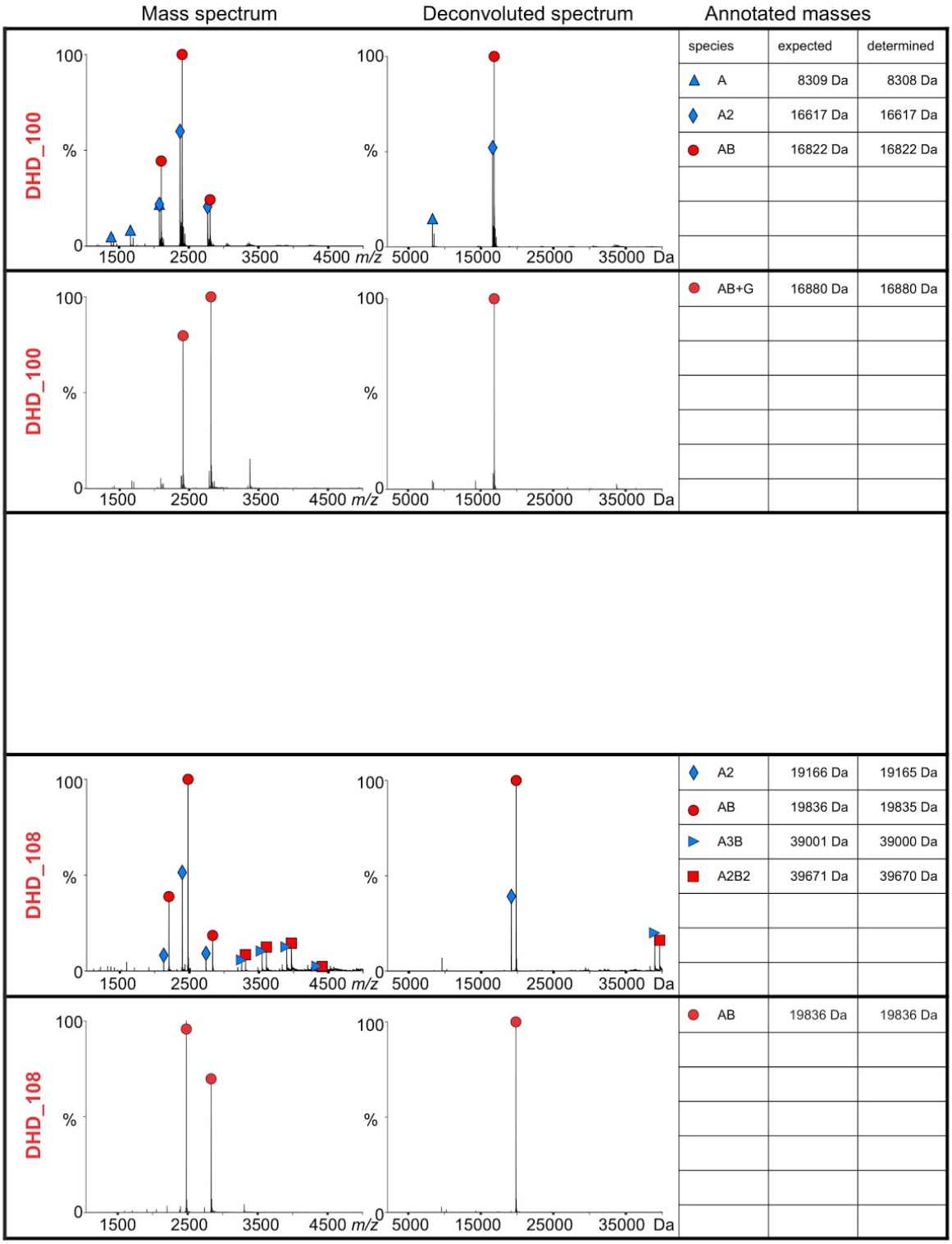


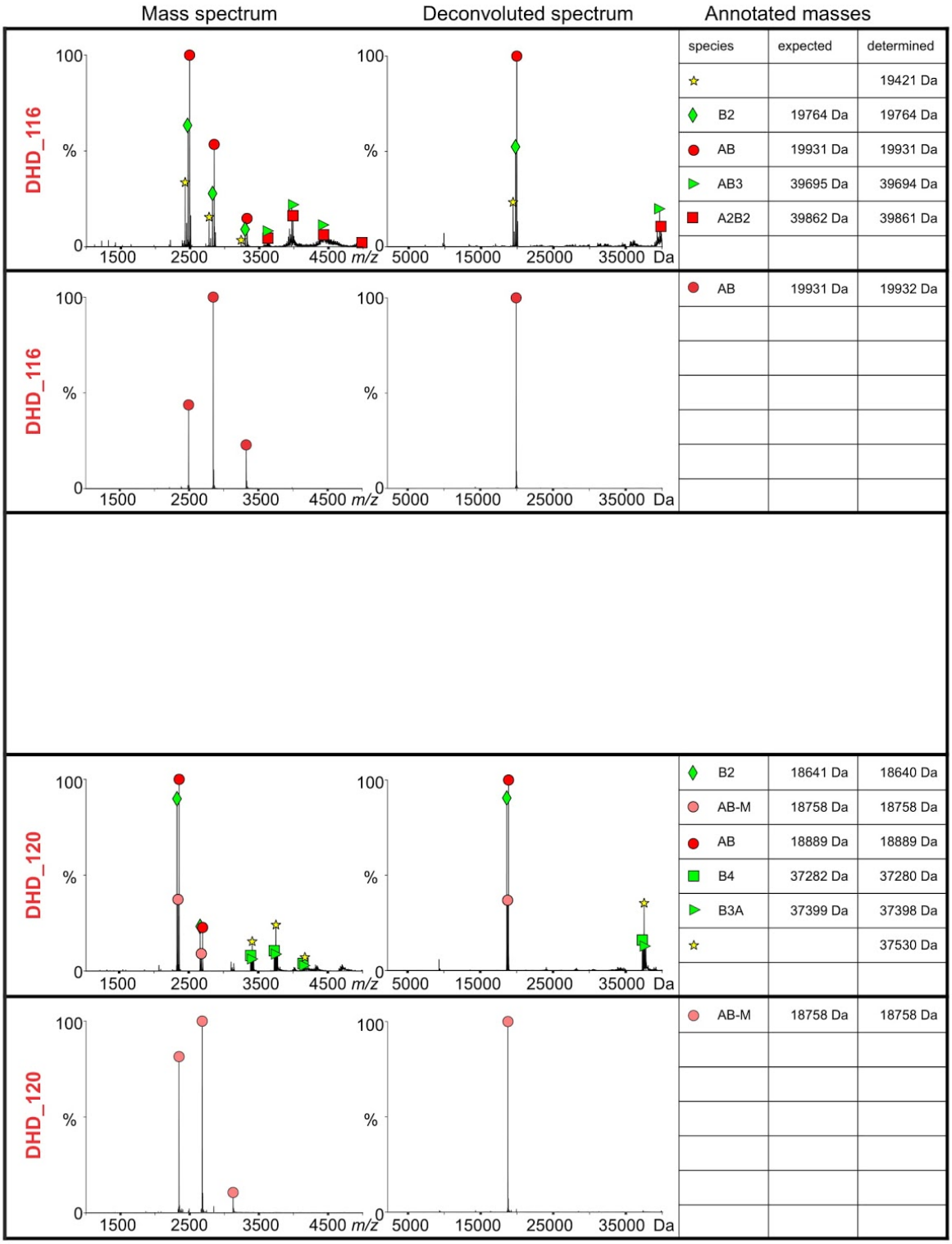
species	expected	determined
AB	18977 Da	18976 Da
B2	19182 Da	19181 Da
AB3	38159 Da	38158 Da
B4	38364 Da	38364 Da
B	9899 Da	9899 Da
B2	19799 Da	19798 Da
☆		27848 Da 38468 Da
B4	39597 Da	39595 Da
B2	19360 Da	19359 Da
B4	38720 Da	38718 Da
AB	18741 Da	18740 Da
B2	19590 Da	19589 Da
AB3	38331 Da	38330 Da
B4	39180 Da	39178 Da
B2	18456 Da	18456 Da
B4	36913 Da	36911 Da

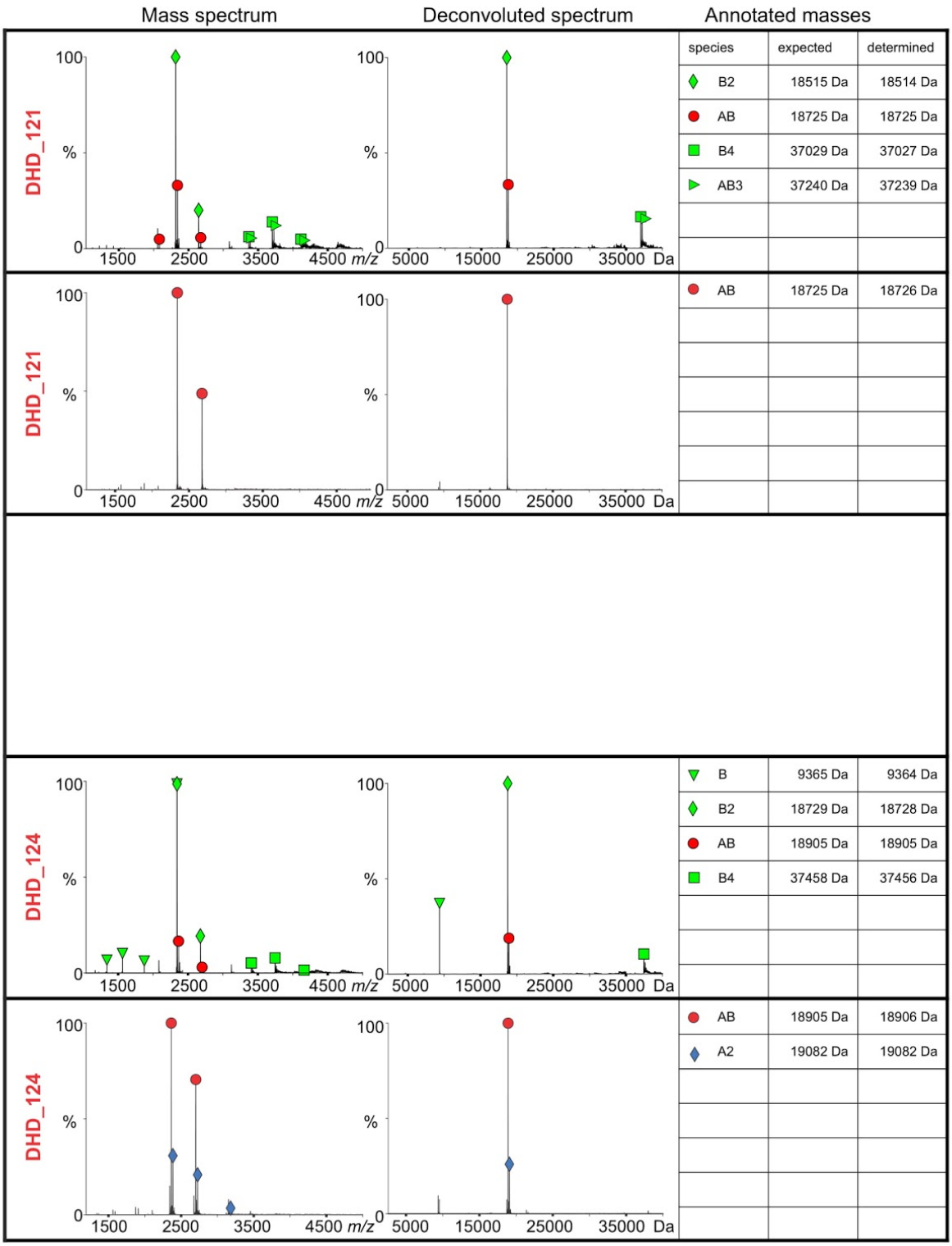

Mass spectrum
Deconvoluted spectrum
Annotated masses
DHD_129
DHD_132
DHD_138
DHD_146
DHD_20
100
%
0
1500
2500
3500
4500 m/z
5000
15000
25000
35000 Da
species
expected
determined
A2 18248 Da 18248 Da
18889 Da
27569 Da
B2 19692 Da 19692 Da
B1 9124 Da 9124 Da
B2 18248 Da 18248 Da
B4 36496 Da 36495 Da
B 9467 Da 9466 Da
A 9734 Da 9733 Da
24376 Da
27353 Da
28281 Da
AB2 28668 Da 28667 Da
A2 17487 Da 17487 Da
AB 18843 Da 18842 Da
B2 20199 Da 20198 Da
B 9260 Da 9258 Da
B2 18519 Da 18520 Da
AB2 27761 Da 27760 Da
B4 37039 Da 37043 Da
23140 Da

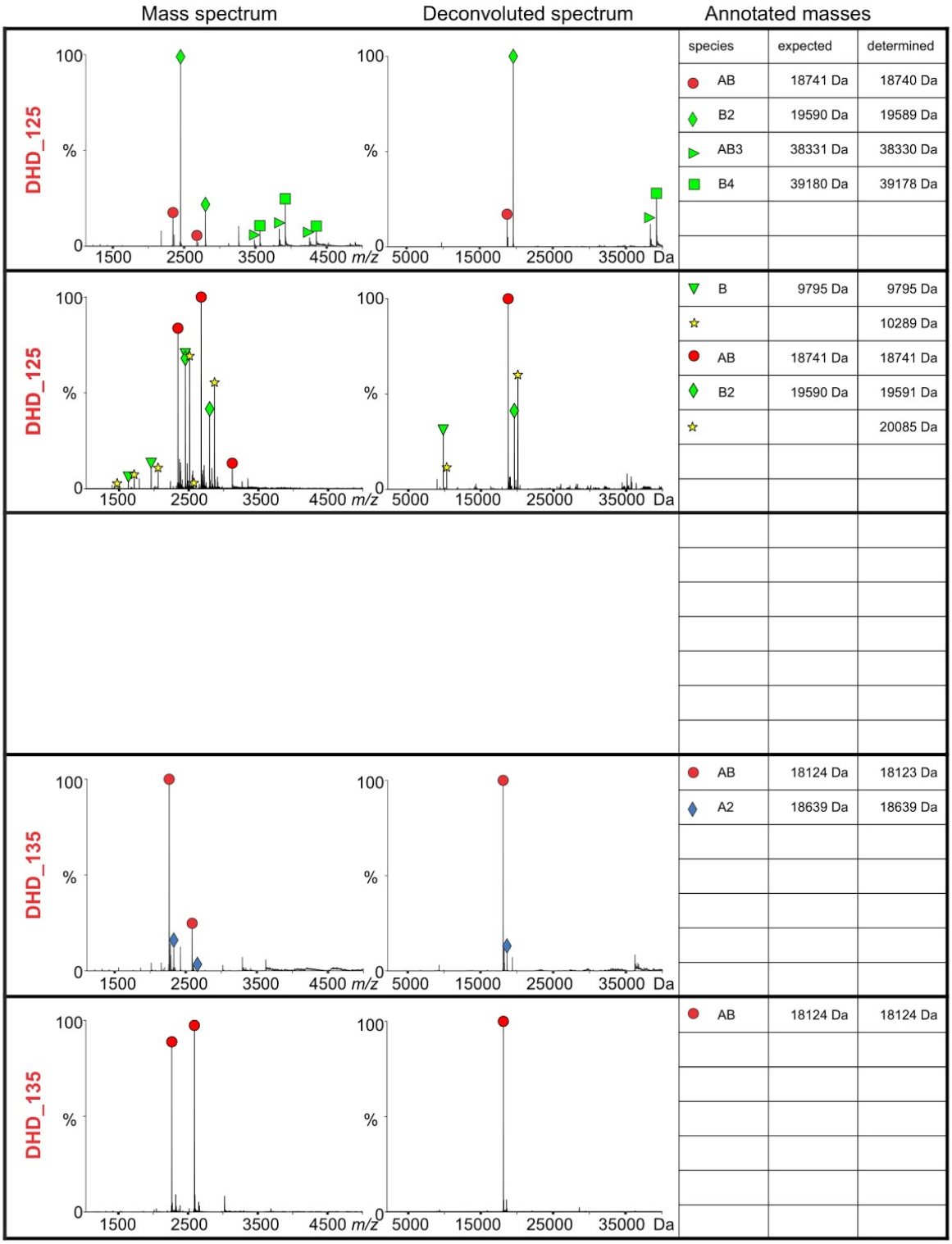


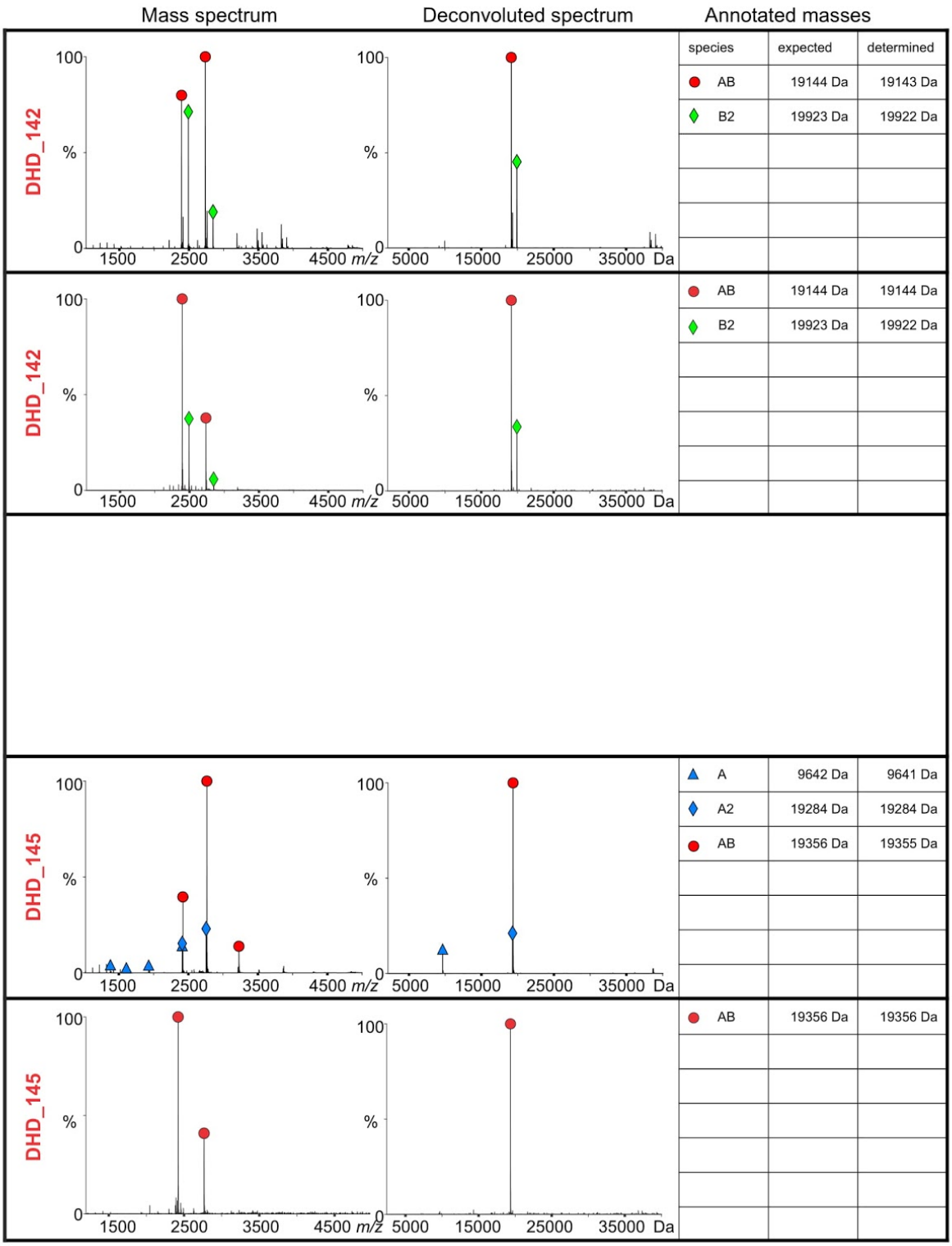
Extended Data Fig. 6 Native MS spectra of designs that assemble to off-target oligomeric states.


Mass spectrum
Deconvoluted spectrum
Annotated masses
DHD_100
DHD_100
DHD_108
DHD_108
100
%
0
1500
2500
3500
4500 m/z
5000
15000
25000
35000 Da
species
expected
determined
A
8309 Da
8308 Da
A2
16617 Da
16617 Da
AB
16822 Da
16822 Da
AB+G
16880 Da
16880 Da
A2
19166 Da
19165 Da
AB
19836 Da
19835 Da
A3B
39001 Da
39000 Da
A2B2
39671 Da
39670 Da
AB
19836 Da
19836 Da


Mass spectrum
Deconvoluted spectrum
Annotated masses
DHD_116
DHD_116
DHD_120
DHD_120
100
%
0
1500
2500
3500
4500 m/z
5000
15000
25000
35000 Da
species
expected
determined
B2 19764 Da 19764 Da
AB 19931 Da 19931 Da
AB3 39695 Da 39694 Da
A2B2 39862 Da 39861 Da
19421 Da
AB 19931 Da 19932 Da
B2 18641 Da 18640 Da
AB-M 18758 Da 18758 Da
AB 18889 Da 18889 Da
B4 37282 Da 37280 Da
B3A 37399 Da 37398 Da
37530 Da
AB-M 18758 Da 18758 Da

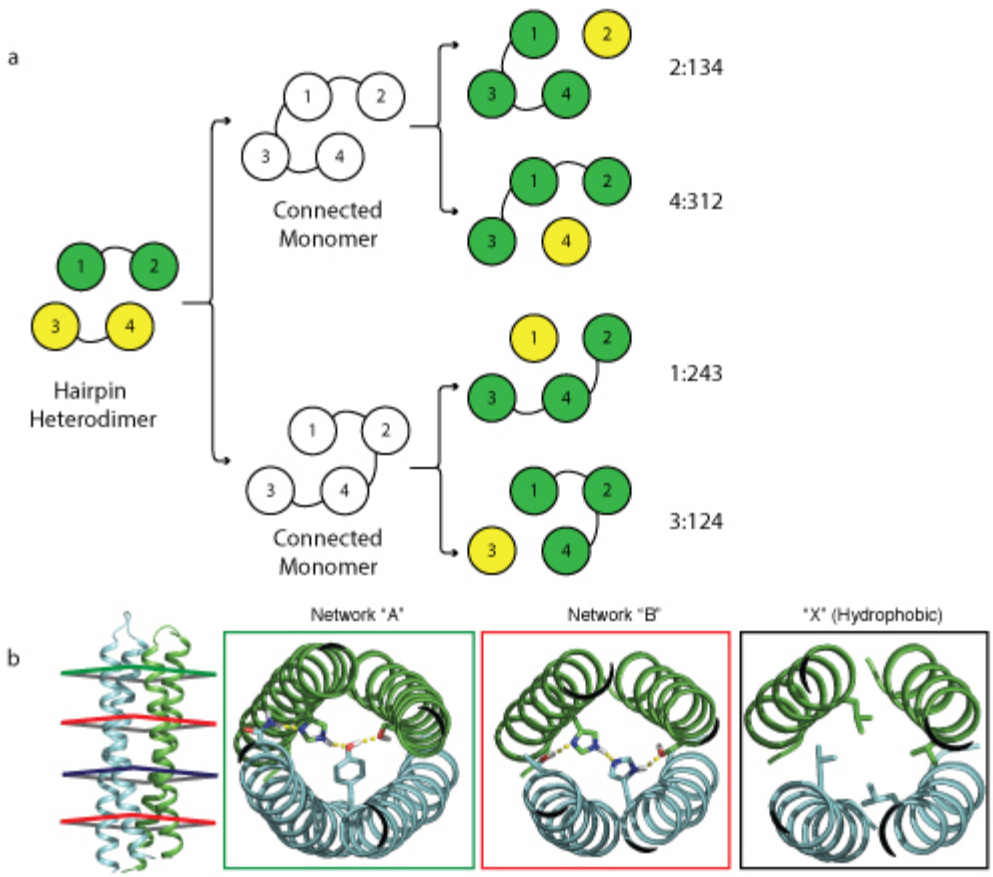

Mass spectrum
Deconvoluted spectrum
Annotated masses
DHD_121
100
%
0
1500
2500
3500
4500 m/z
5000
15000
25000
35000 Da
species
expected
determined
B2
18515 Da
18514 Da
AB
18725 Da
18725 Da
B4
37029 Da
37027 Da
AB3
37240 Da
37239 Da
DHD_121
AB
18725 Da
18726 Da
DHD_124
B
9365 Da
9364 Da
B2
18729 Da
18728 Da
AB
18905 Da
18905 Da
B4
37458 Da
37456 Da
DHD_124
AB
18905 Da
18906 Da
A2
19082 Da
19082 Da


Mass spectrum
Deconvoluted spectrum
Annotated masses
DHD_125
DHD_125
DHD_135
DHD_135
100
%
0
1500
2500
3500
4500 m/z
5000
15000
25000
35000 Da
species
expected
determined
AB 18741 Da 18740 Da
B2 19590 Da 19589 Da
AB3 38331 Da 38330 Da
B4 39180 Da 39178 Da
B 9795 Da 9795 Da
10289 Da
AB 18741 Da 18741 Da
B2 19590 Da 19591 Da
20085 Da
AB 18124 Da 18123 Da
A2 18639 Da 18639 Da
AB 18124 Da 18124 Da

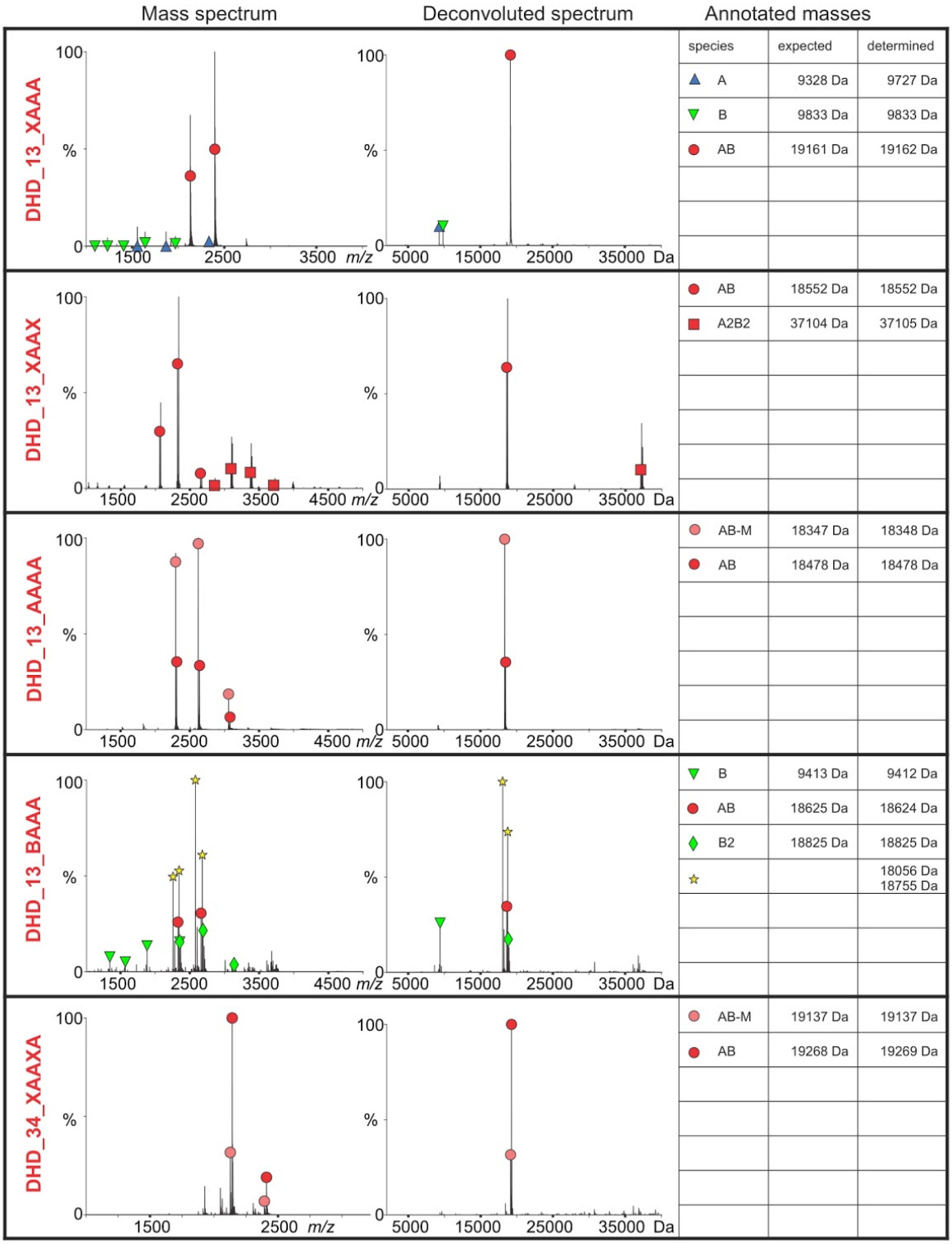


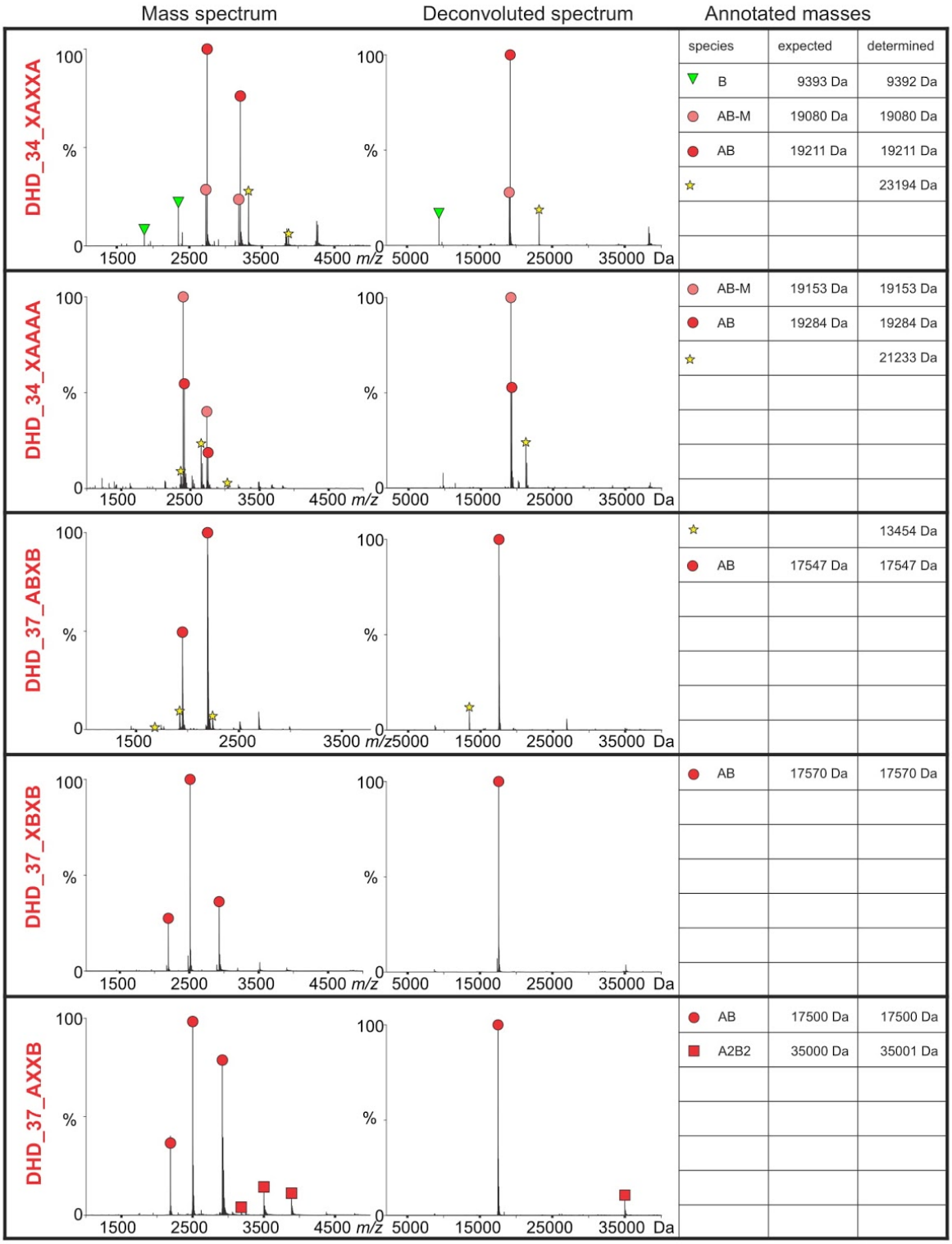
Extended Data Fig. 7 Double purification eliminated homodimers in most cases. For each given design, genes were ordered either with only one monomer having a hexahistidine tag

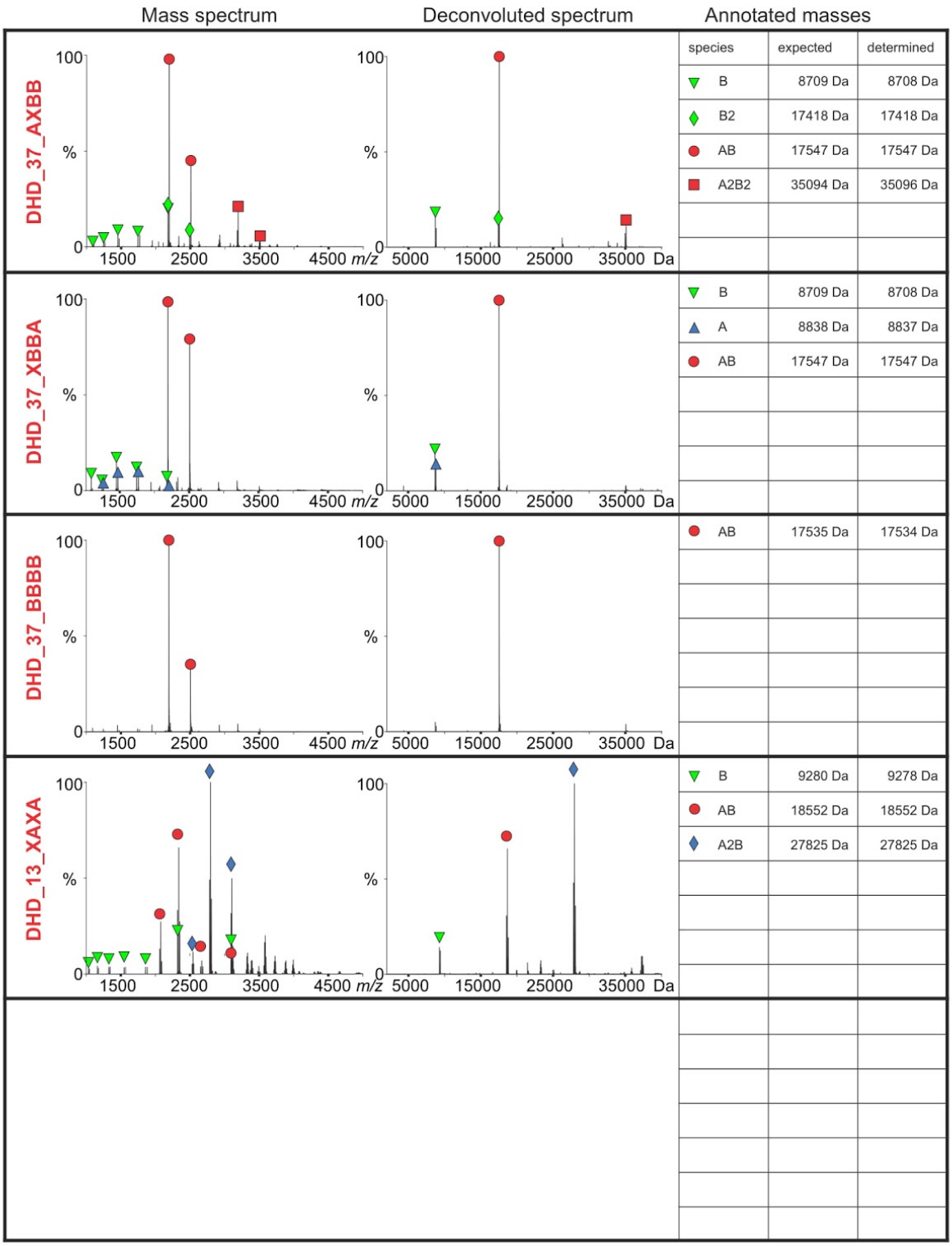
(top), or with an additional streptavidin tag on the other monomer (bottom). A second affinity purification step with Strep-Tactin resin (IBA) showed that the homodimers likely arise from unbalanced expression in *E. coli*.



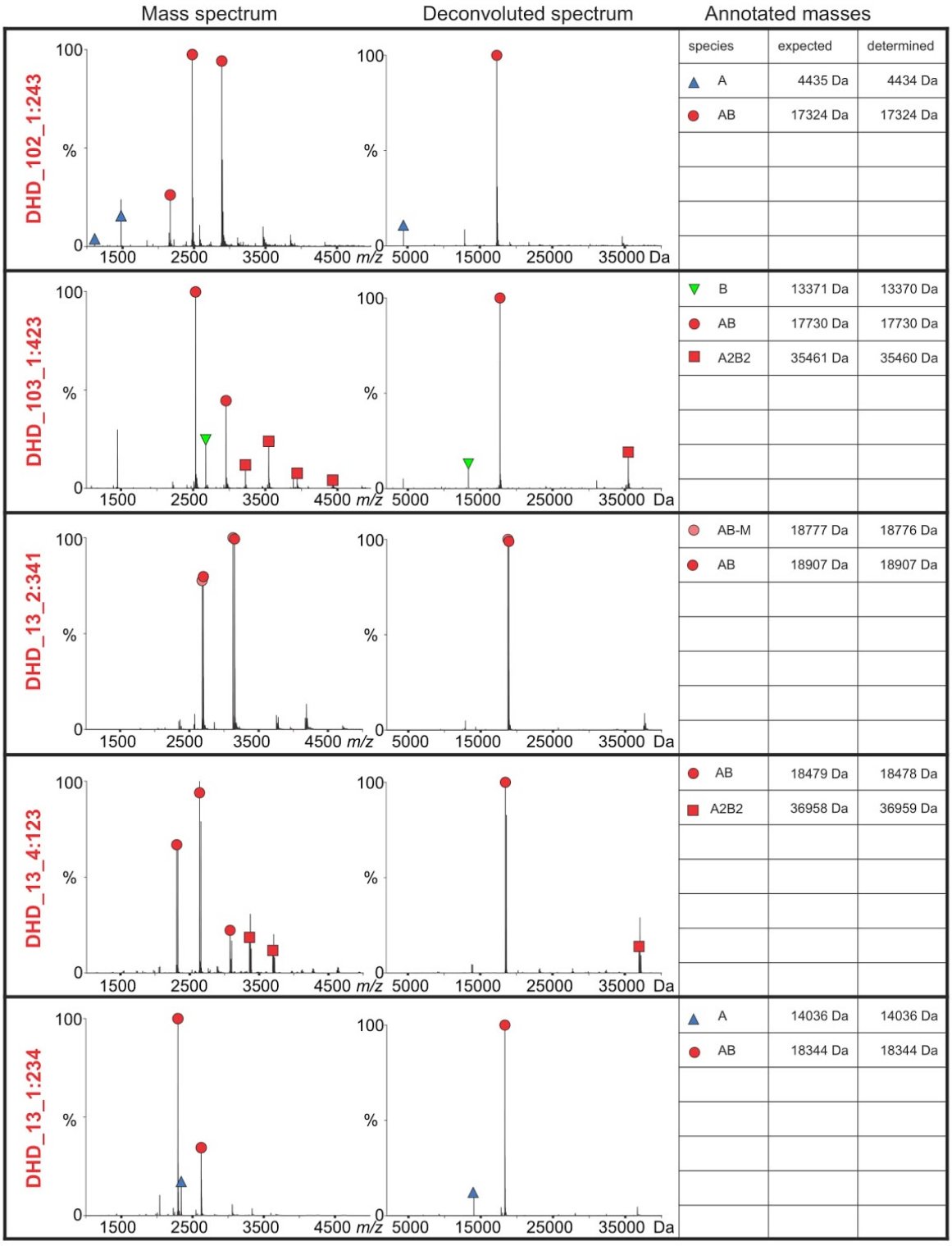
Extended Data Fig. 8 Backbone and hydrogen bond network permutations. a, On a 2+2 backbone (left), two loops were designed to connect the 4 helices into a single monomer in 2 different ways (middle), after which 4 different cut points were introduced to generate 4 possible backbone permuted heterodimers of a single helix and a three helix bundle (3+1 heterodimers, right). For example, 2:134 refers to a heterodimer where the original helix 2 is a single helix, and helices 1, 3, and 4 were connected into a 3 helix bundle. **b,** Hydrogen bond network permutation. Each unique network was assigned a letter (Networks “A” and “B” in this case), and with the hydrophobic packing assigned X, the backbone on the left reads “ABXB”, since its first heptad accommodates network “A”, its second and fourth heptad accommodate network “B”, and its third heptad accommodates hydrophobic packing only.

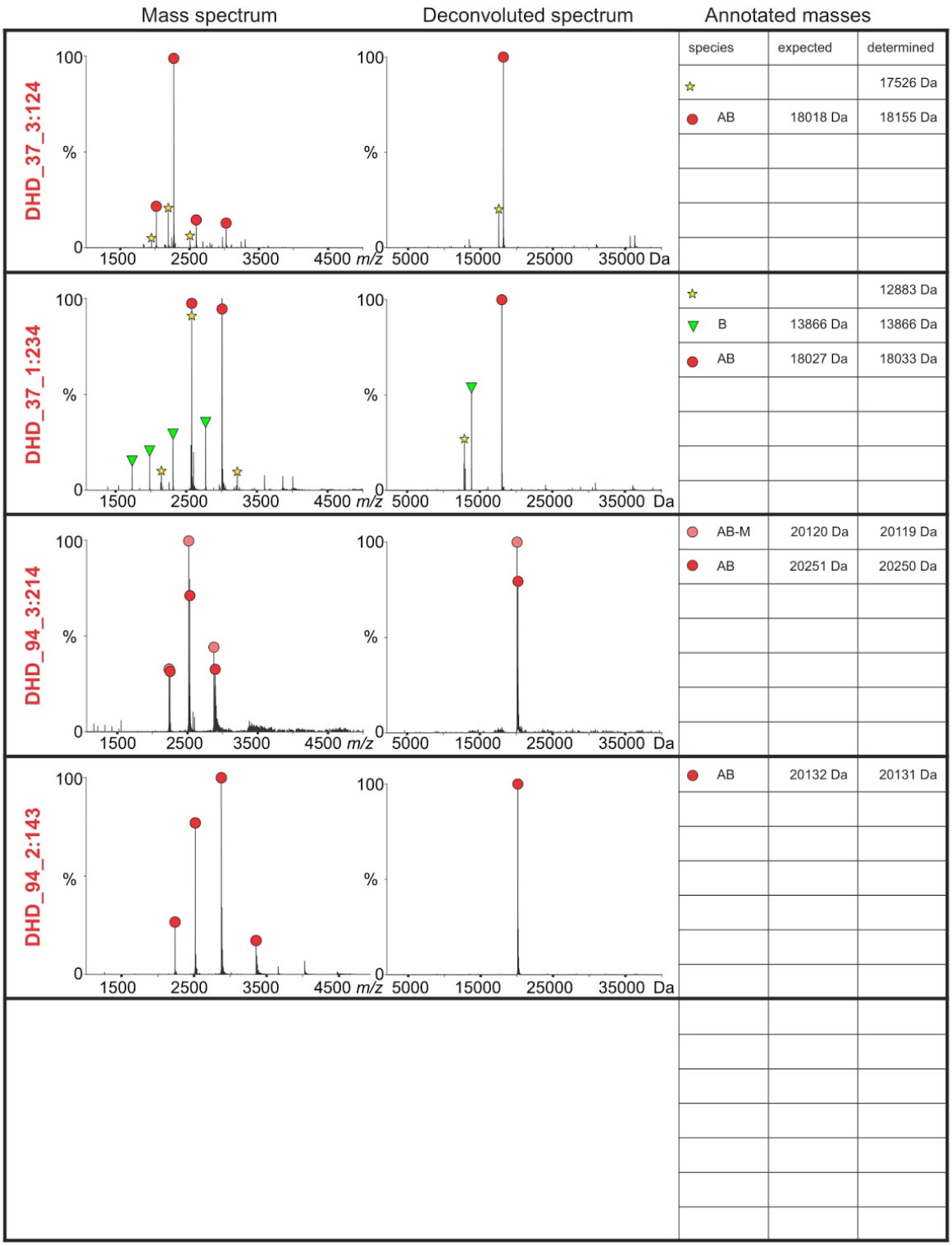

Mass spectrum
Deconvoluted spectrum
Annotated masses
DHD_13_XAAA
DHD_13_XAAX
DHD_13_AAAA
DHD_13_BAAA
DHD_34_XAAXA
100
%
0
1500
2500
3500
4500
m/z
5000
15000
25000
35000
Da
species
expected
determined
A
9328 Da
9727 Da
B
9833 Da
9833 Da
AB
19161 Da
19162 Da
AB
18552 Da
18552 Da
A2B2
37104 Da
37105 Da
AB-M
18347 Da
18348 Da
AB
18478 Da
18478 Da
B
9413 Da
9412 Da
AB
18625 Da
18624 Da
B2
18825 Da
18825 Da
18056 Da
18755 Da
AB-M
19137 Da
19137 Da
AB
19268 Da
19269 Da


Mass spectrum
Deconvoluted spectrum
Annotated masses
DHD_34_XAXXA
DHD_34_XAAAA
DHD_37_ABXB
DHD_37_XBXB
DHD_37_AXXB
100
%
0
1500
2500
3500
4500 m/z
5000
15000
25000
35000 Da
species
expected
determined
B 9393 Da 9392 Da
AB-M 19080 Da 19080 Da
AB 19211 Da 19211 Da
23194 Da
AB-M 19153 Da 19153 Da
AB 19284 Da 19284 Da
21233 Da
13454 Da
AB 17547 Da 17547 Da
AB 17570 Da 17570 Da
AB 17500 Da 17500 Da
A2B2 35000 Da 35001 Da

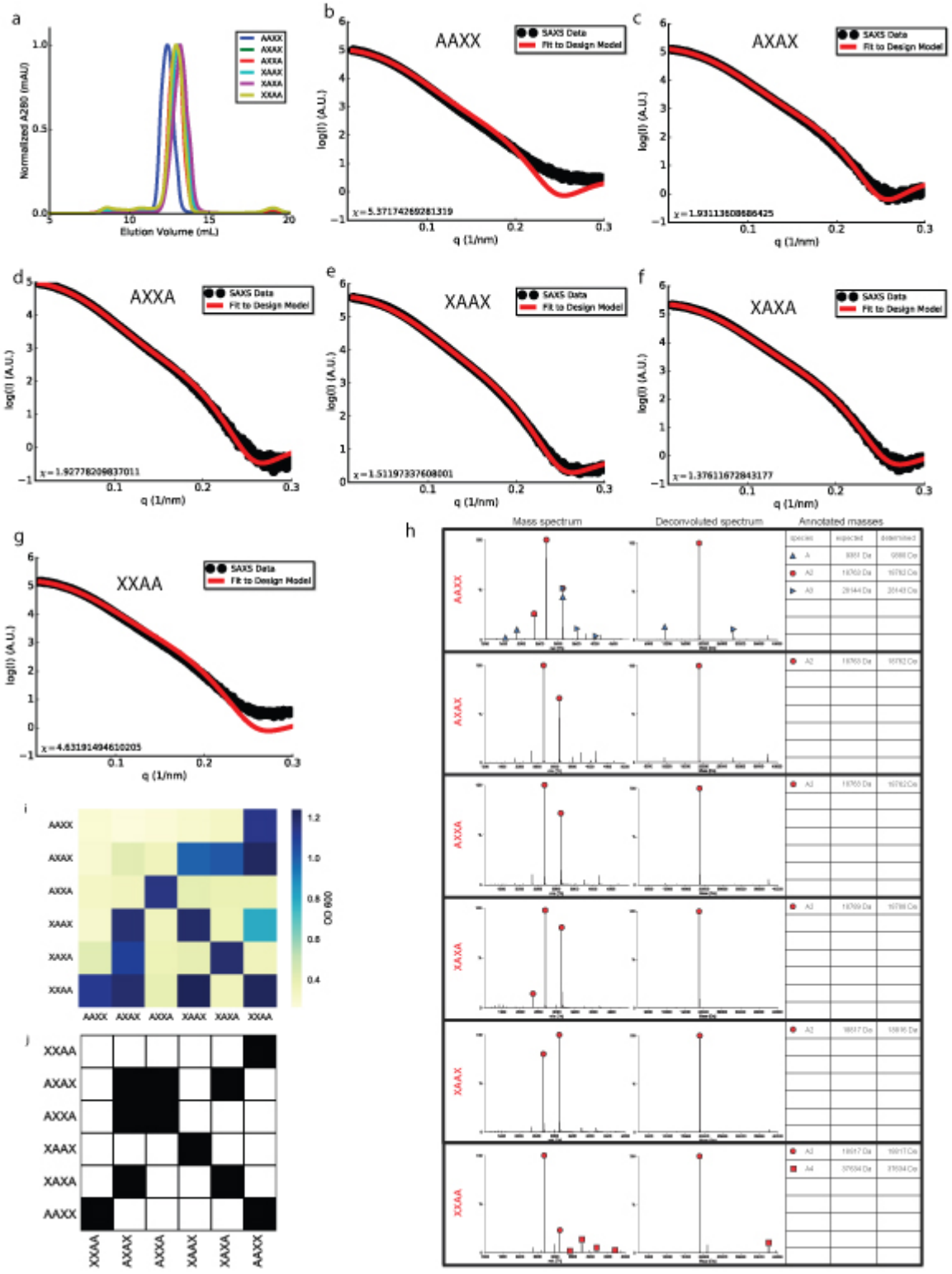


Extended Data Fig. 9 Native MS spectra of hydrogen bond network permuted designs. All but DHD_13_XAXA showed constitutive heterodimer formation with no off-target oligomers.


Mass spectrum
Deconvoluted spectrum
Annotated masses
DHD_102_1:243
DHD_103_1:423
DHD_13_2:341
DHD_13_4:123
DHD_13_1:234
100
%
0
1500
2500
3500
4500 m/z
5000
15000
25000
35000 Da
species
expected
determined
A 4435 Da 4434 Da
AB 17324 Da 17324 Da
B 13371 Da 13370 Da
AB 17730 Da 17730 Da
A2B2 35461 Da 35460 Da
AB-M 18777 Da 18776 Da
AB 18907 Da 18907 Da
AB 18479 Da 18478 Da
A2B2 36958 Da 36959 Da
A 14036 Da 14036 Da
AB 18344 Da 18344 Da

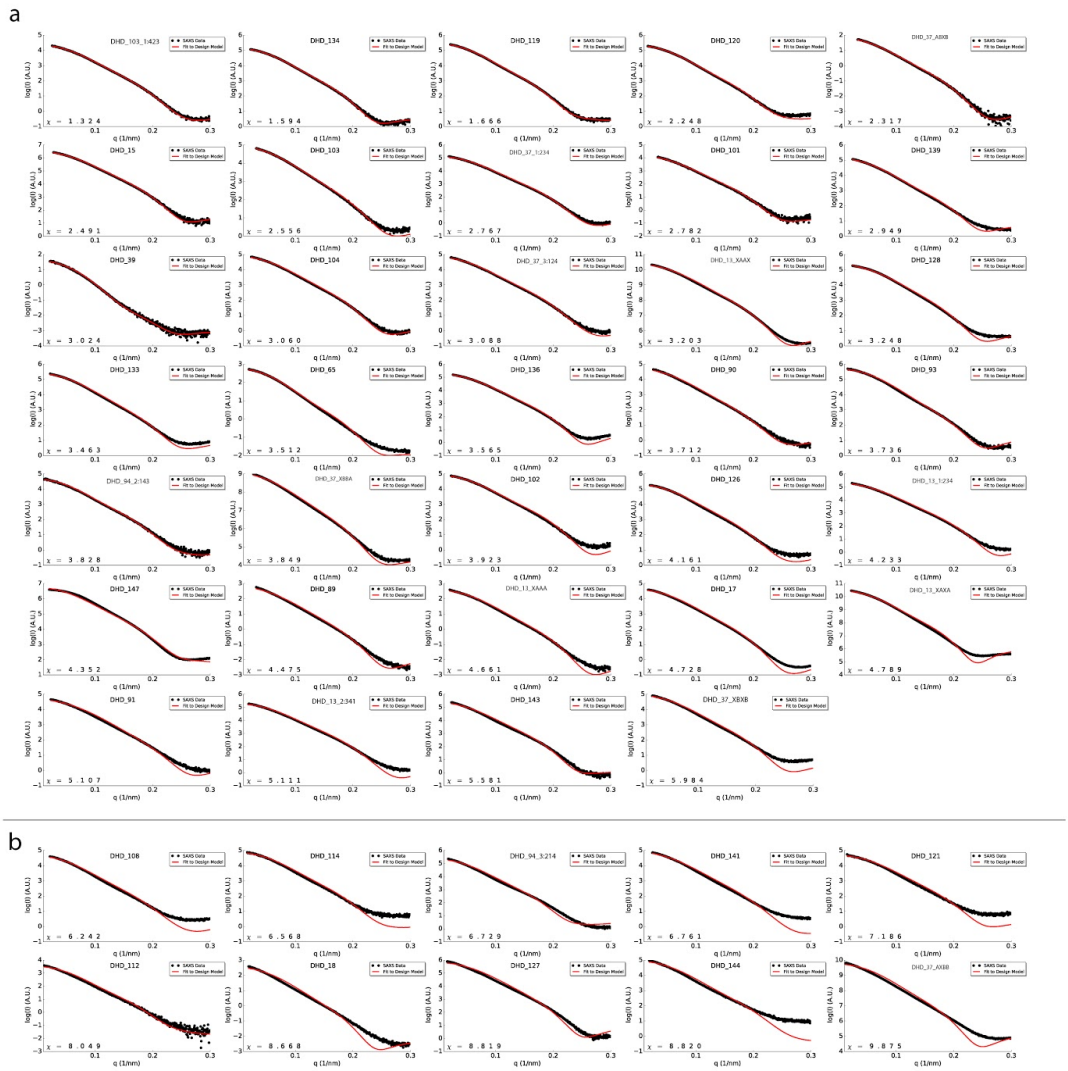


Extended Data Fig. 9 Native MS spectra of backbone permuted designs. All showed constitutive formation of designed heterodimers.

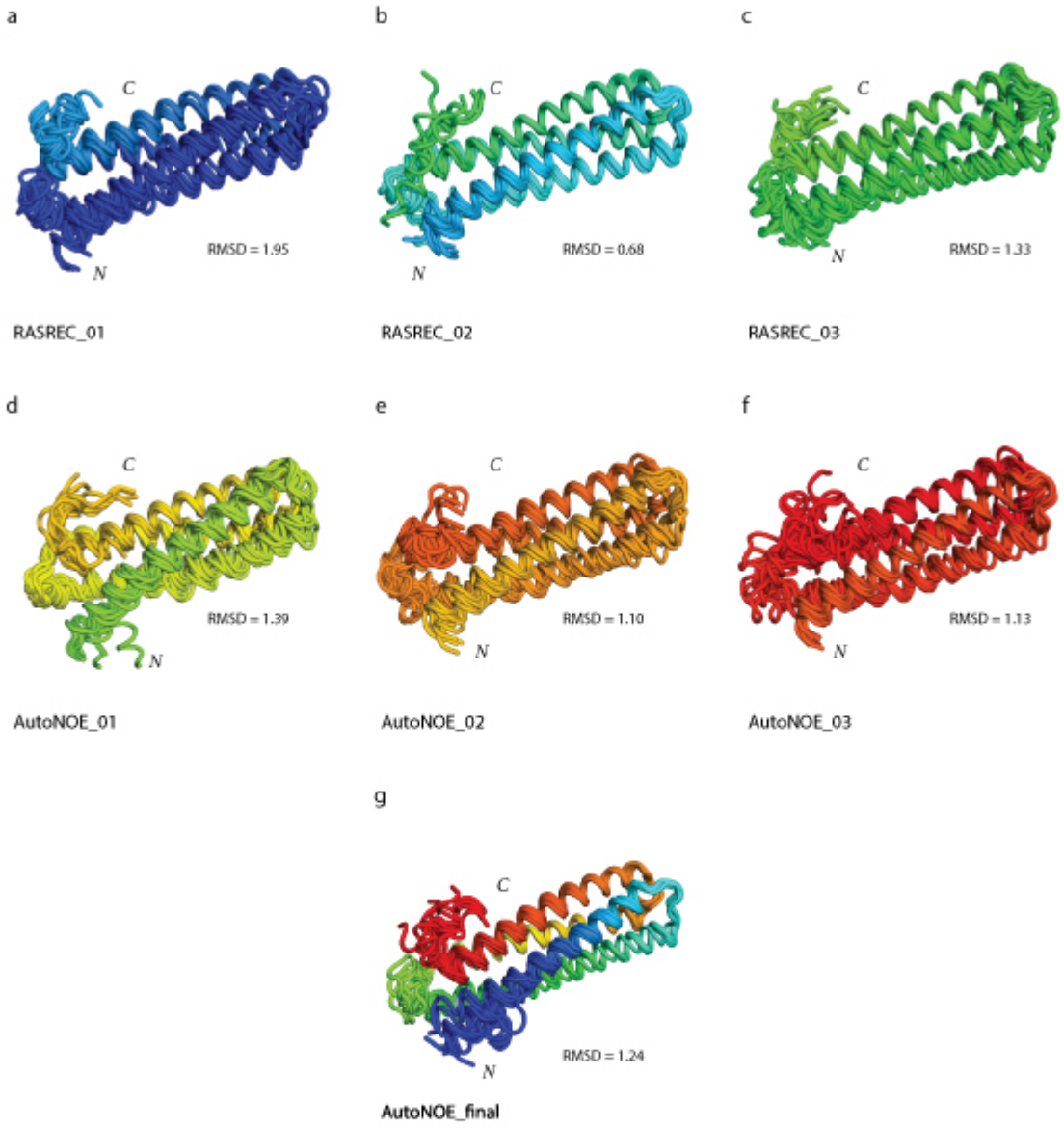


Extended Data Fig. 11 Biophysical and orthogonality characterization of hydrogen bond network permuted homodimers. a, SEC traces of all 6 homodimer designs. **b-g,** SAXS profiles of hydrogen bond network permuted homodimer designs. Black, experimental SAXS

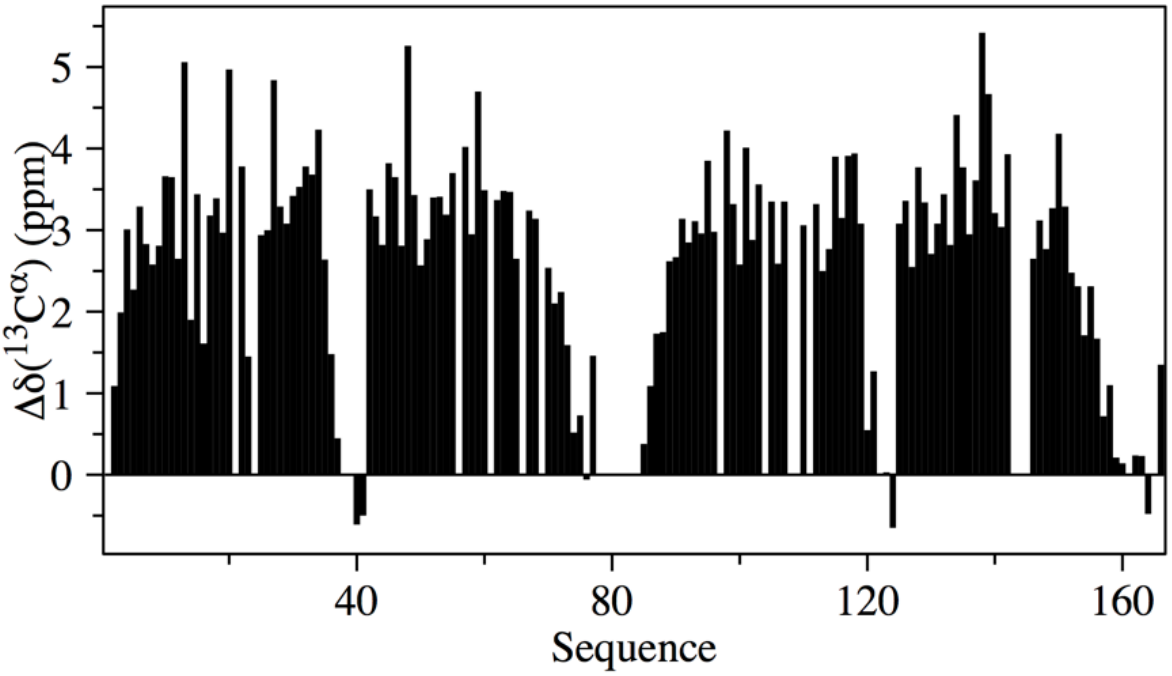
data; red, spectra computed from the designed backbones. **h,** Native MS spectra of the 6 homodimers, all but AAXX assembled to constitutive homodimers. **i,** Y2H all-against-all assay of 6 pairs of homodimers. **j,** Native MS mixing assay of all 6 homodimers, black square indicates an observation of a homodimer or heterodimer complex.



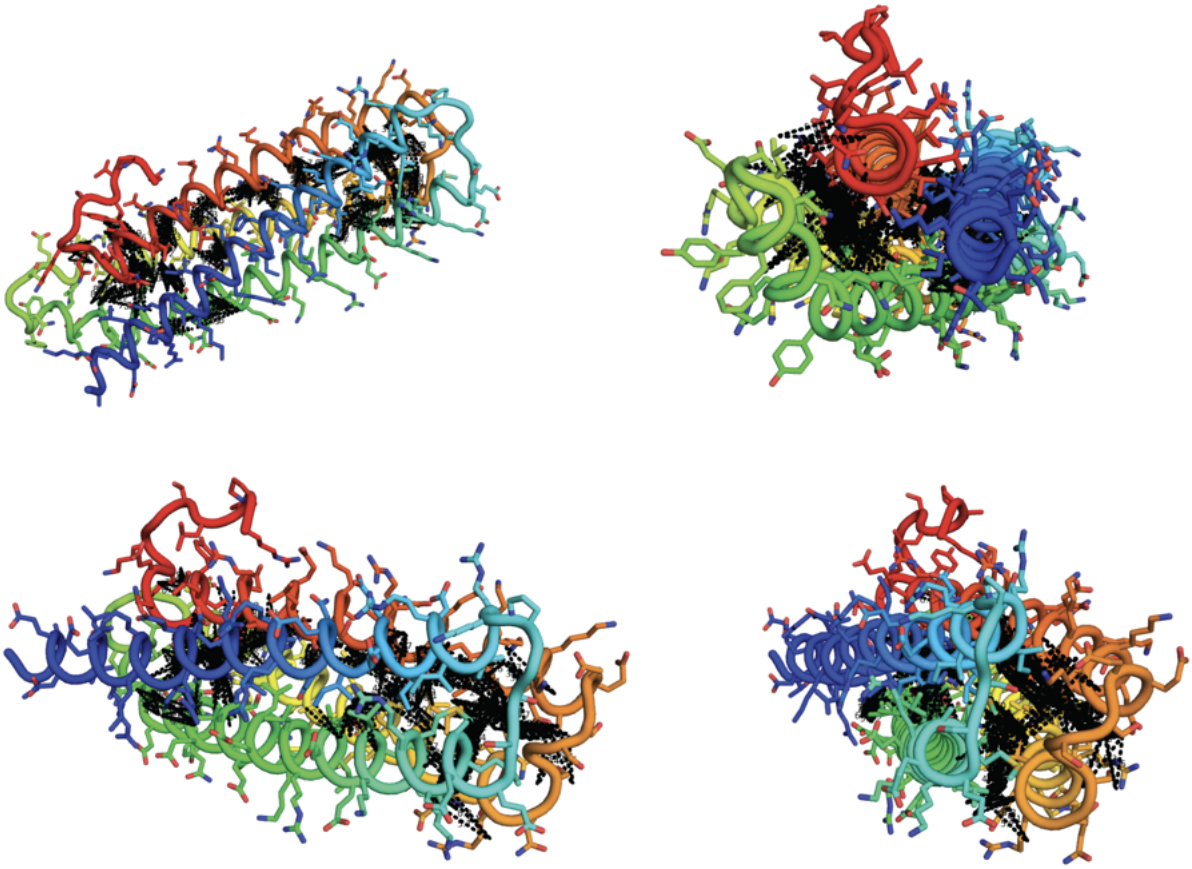
Extended Data Fig. 12 SAXS profiles of all tested DHDs. Black, experimental SAXS data; red, spectra computed from the designed backbones. **a,** SAXS profiles with χ values smaller than 6. **b,** SAXS profiles with χ values greater than 6. All tested designs showed close agreement to expected radius of gyration (R_g) and maximum distance (d_{max}).



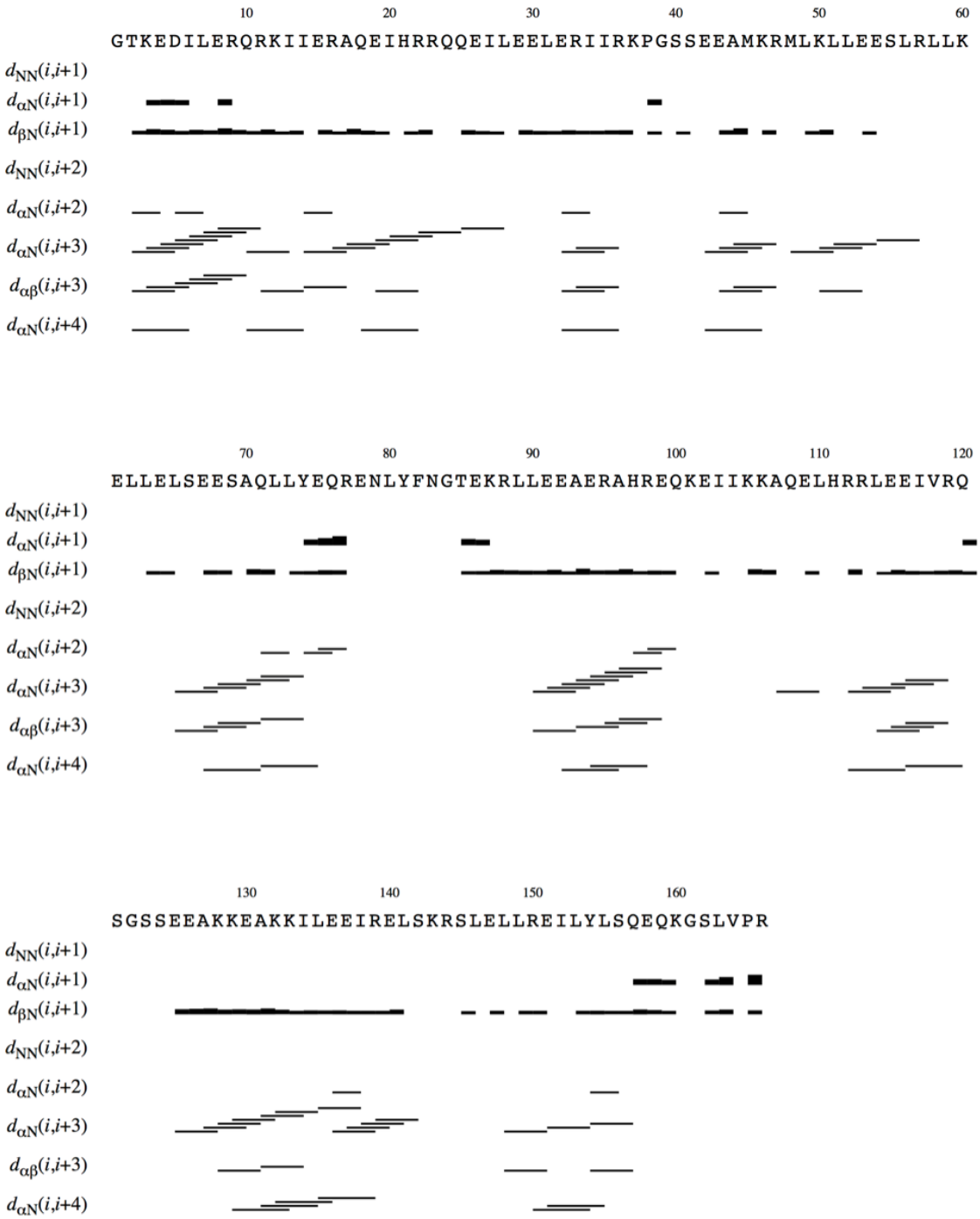
Extended Data Fig. 13 Models of the OPHD_13_XAAA protein used for NMR assignments. **a-g**, early NMR models obtained using different approaches.



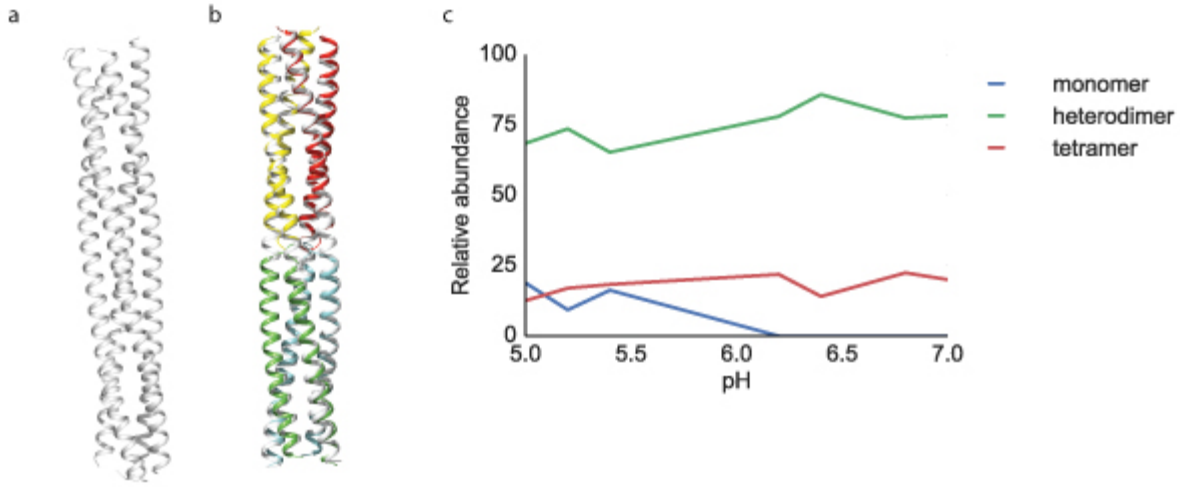
Extended Data Fig. 14 Carbon chemical shift index (dCexp-dCref). A value over 1 for 3 consecutive residues indicate a α-helical secondary structure. A value below -1 indicates a β-strand is more likely.



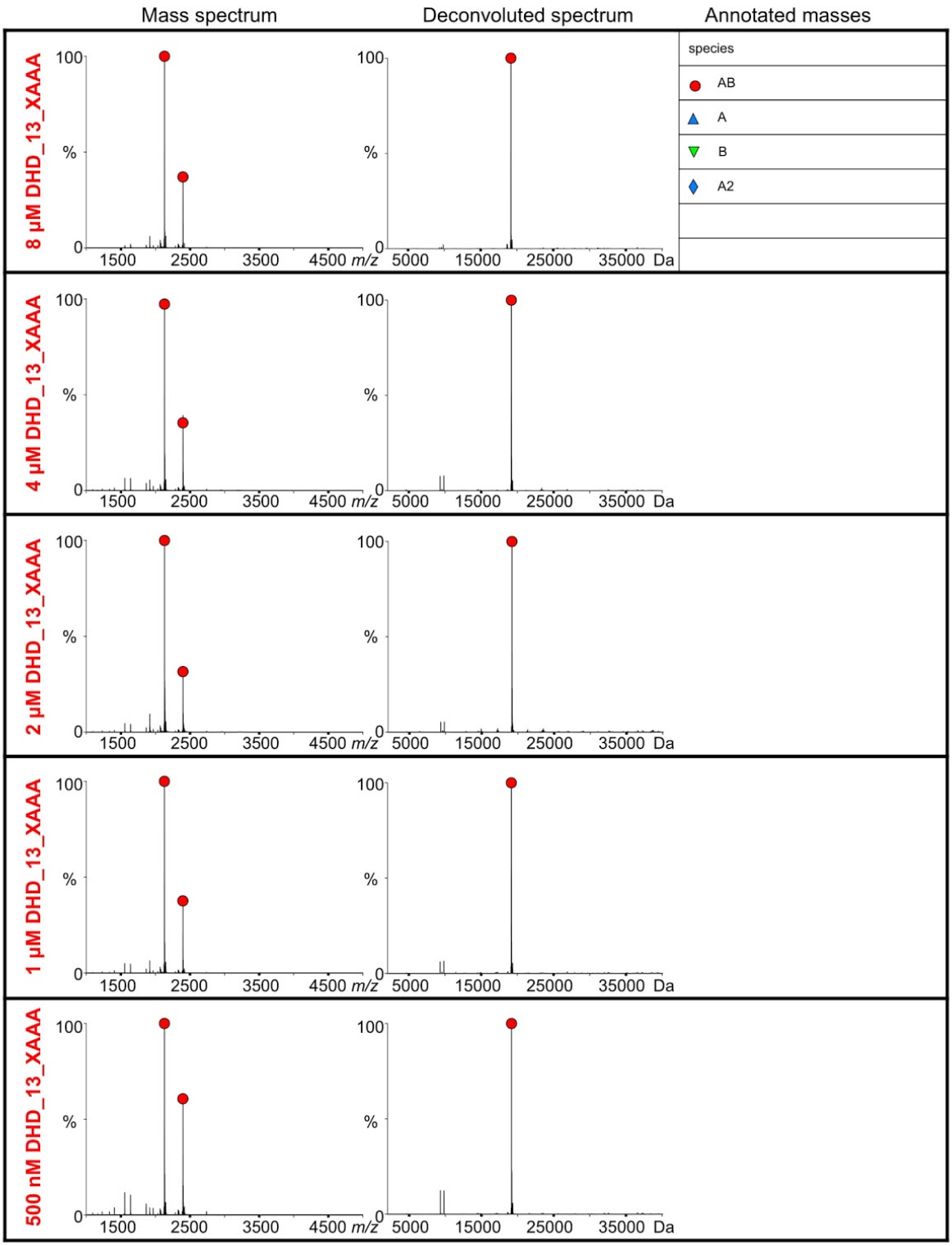
Extended Data Fig. 15 Initial NOE contacts loaded onto the model AutoNOE_final. Black dashed lines, NOE contacts.

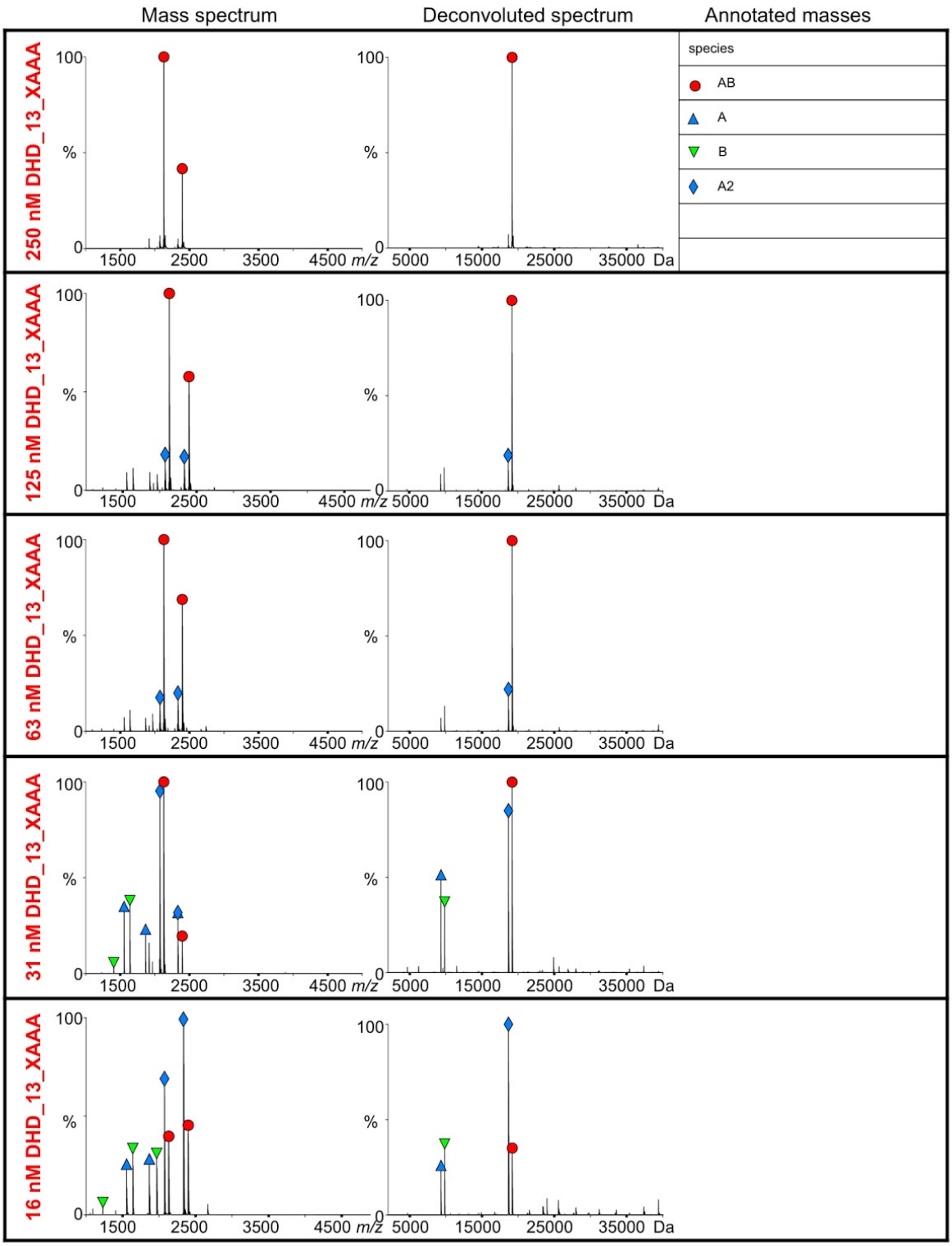


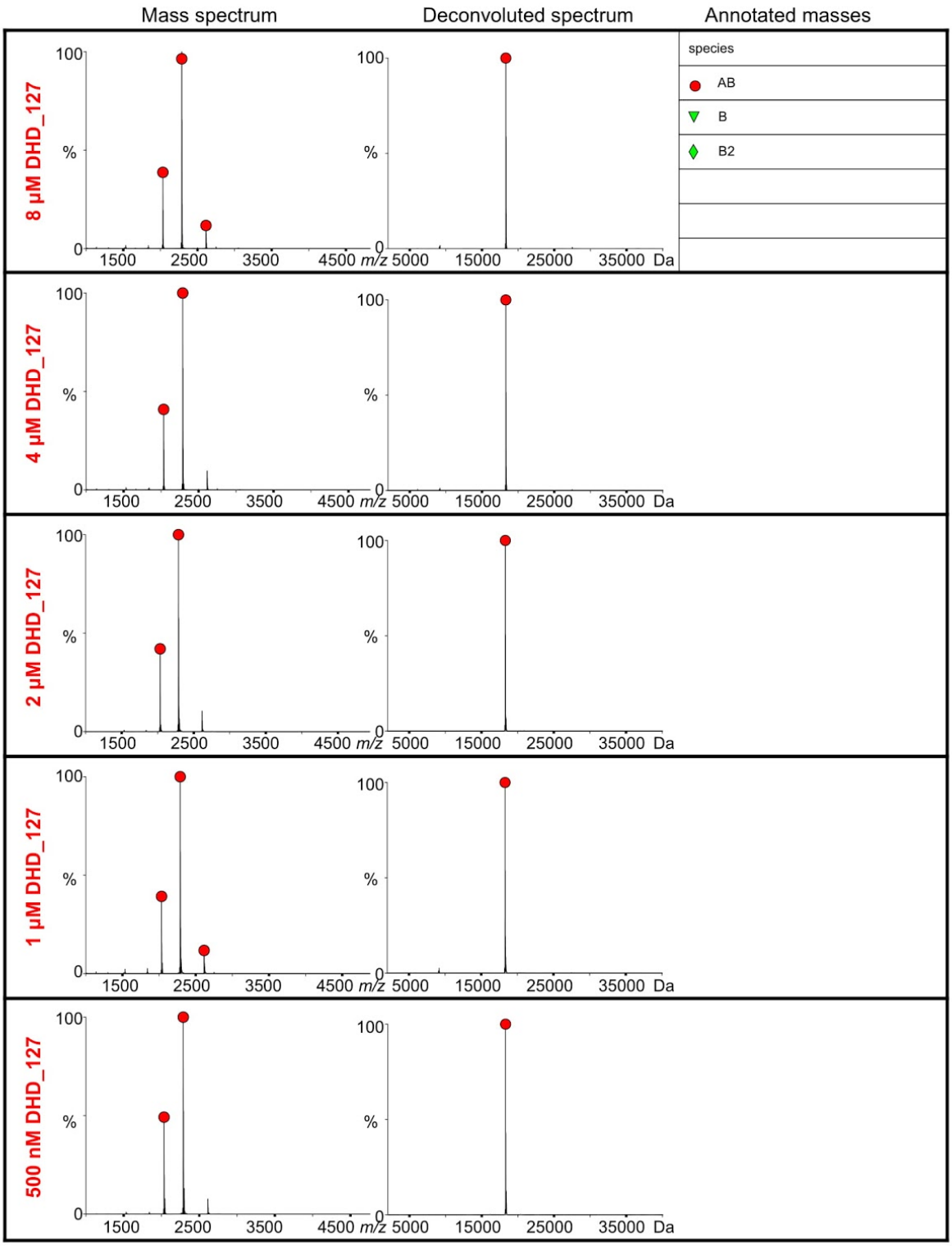
Extended Data Fig. 16 Short and medium range NOE contacts. The subscripts indicate the atom types involved in the contact. The data supports the alpha helix fold observed in independent structure calculations and the chemical shift analysis.

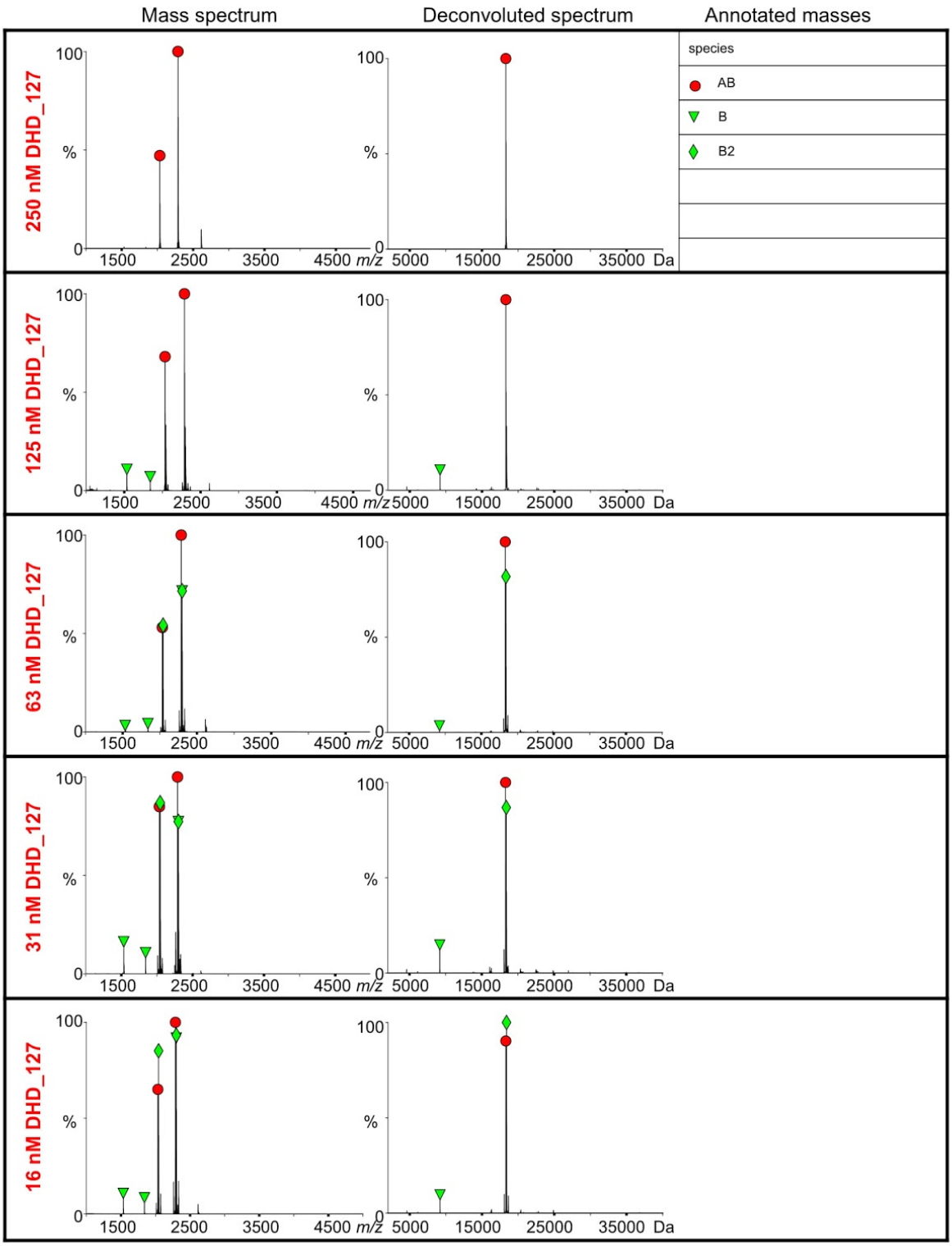


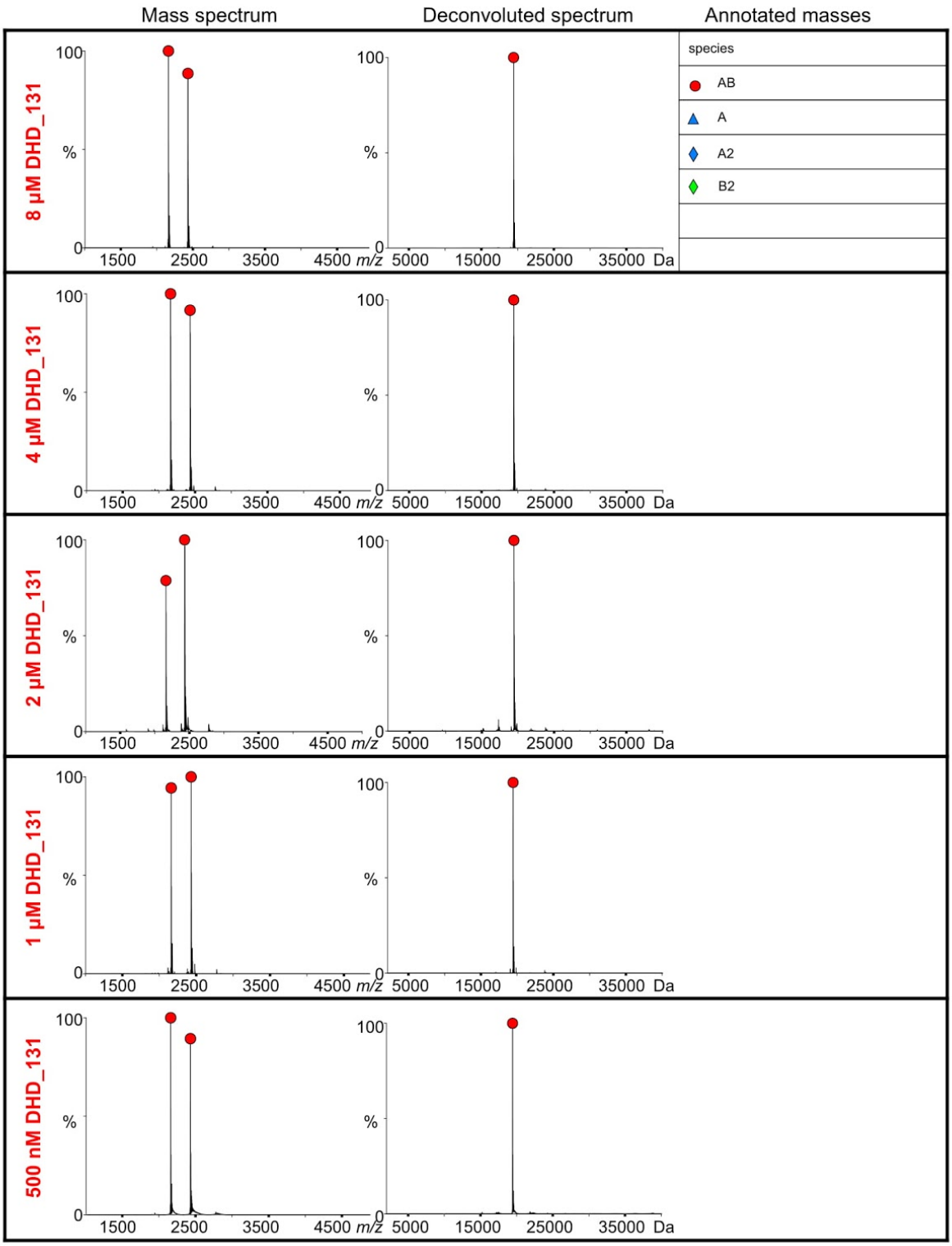
Extended Data Fig. 17 Crystal structure of the domain swapped DHD_15. a, Crystal structure of DHD_15 at pH 6.5, with 2.25 Å resolution. **b,** Superposition of design models in color onto both halves of the crystal structure in white, with backbone RMSD of 1.83 Å. **c,** Native MS study of DHD_15 at different pH indicates the heterodimers, rather than heterotetramers, are dominant in solution.

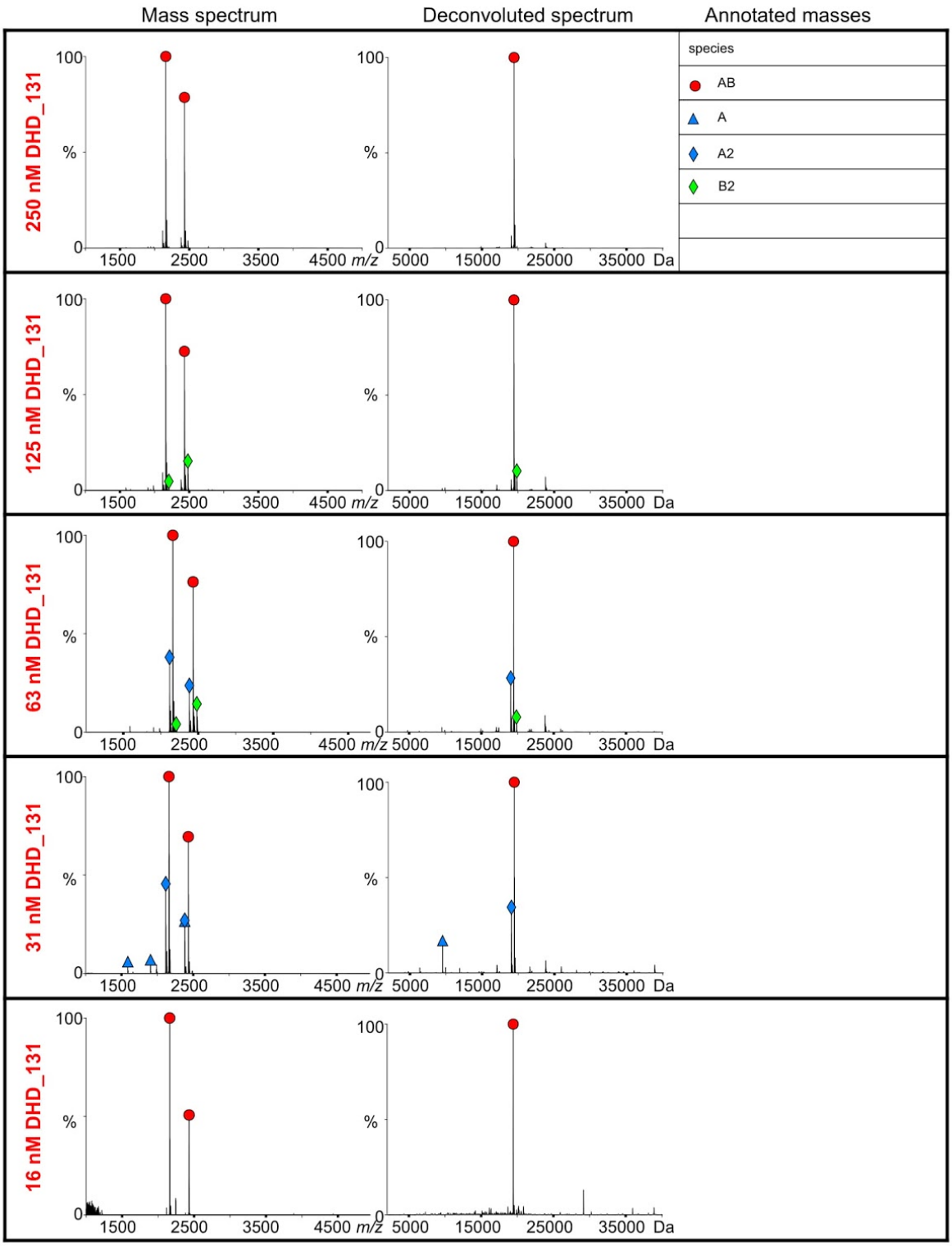

Mass spectrum
Deconvoluted spectrum
Annotated masses
8 µM DHD_13_XAAA
4 µM DHD_13_XAAA
2 µM DHD_13_XAAA
1 µM DHD_13_XAAA
500 nM DHD_13_XAAA
100
%
0
1500
2500
3500
4500
m/z
5000
15000
25000
35000
Da
species
AB
A
B
A2

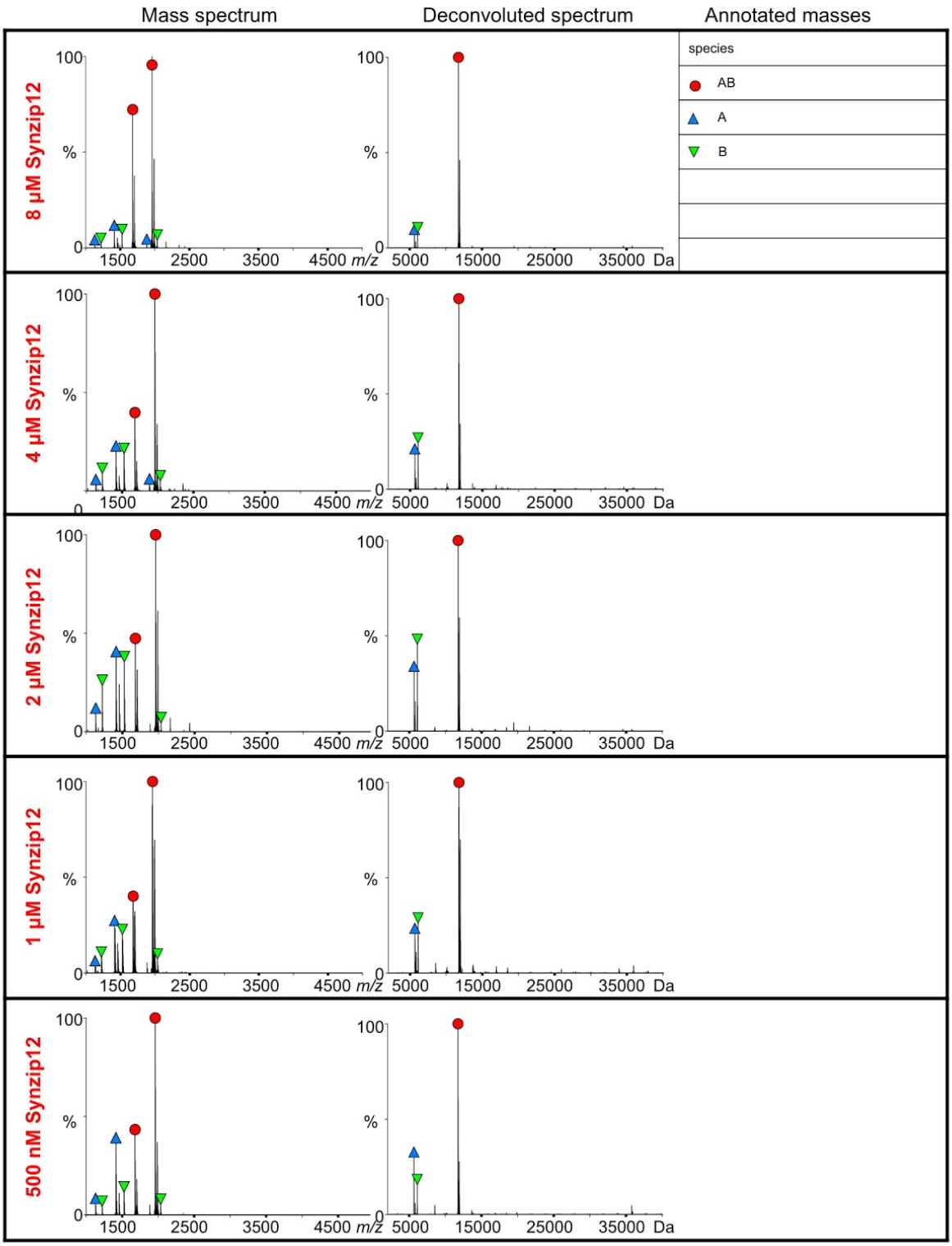

Mass spectrum
Deconvoluted spectrum
Annotated masses
250 nM DHD_13_XAAA
125 nM DHD_13_XAAA
63 nM DHD_13_XAAA
31 nM DHD_13_XAAA
16 nM DHD_13_XAAA
100
%
0
1500
2500
3500
4500 m/z
5000
15000
25000
35000 Da
species
AB
A
B
A2

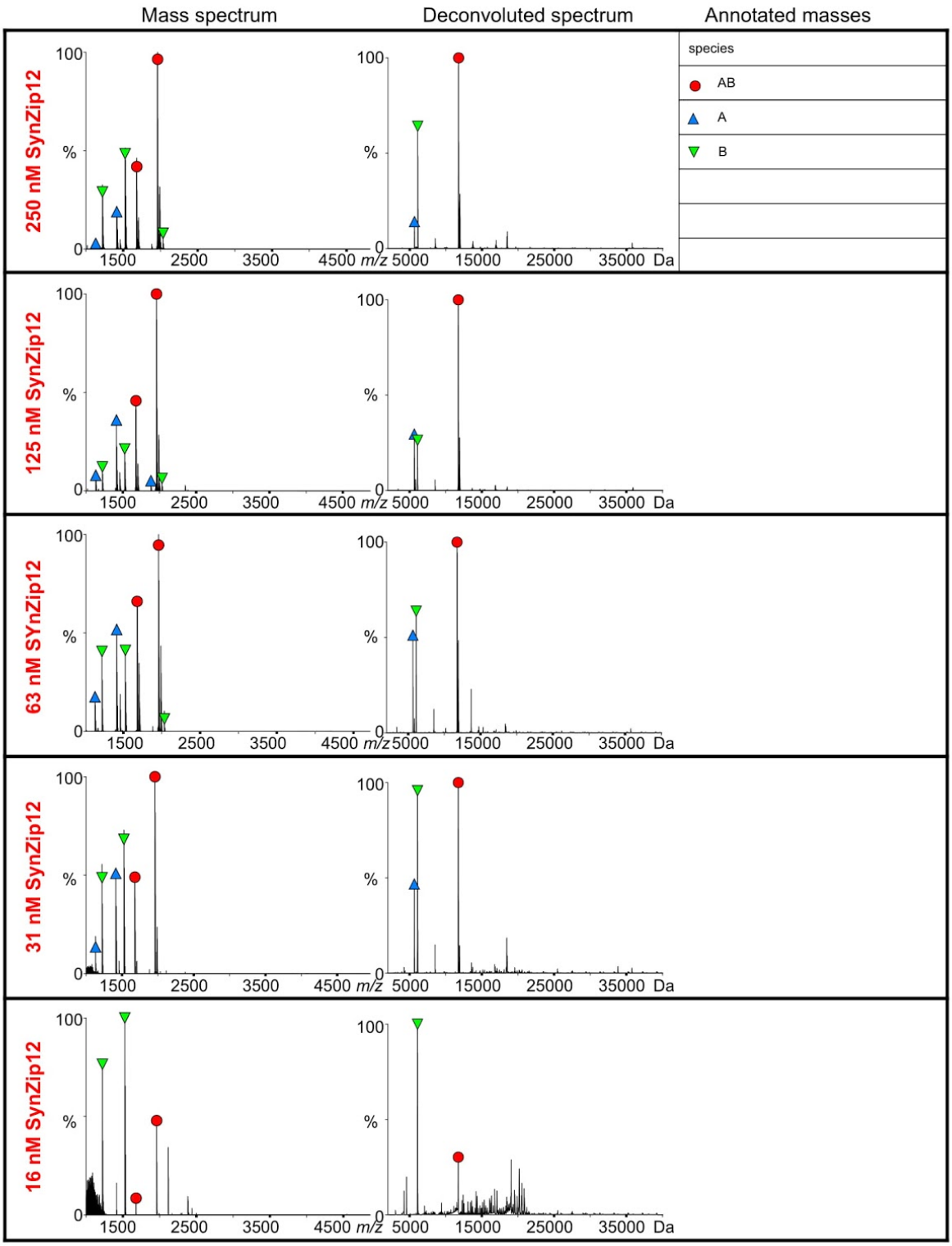

Mass spectrum
Deconvoluted spectrum
Annotated masses
species
AB
B
B2
8 µM DHD_127
4 µM DHD_127
2 µM DHD_127
1 µM DHD_127
500 nM DHD_127
100
%
0
1500
2500
3500
4500
m/z
5000
15000
25000
35000
Da


Mass spectrum
Deconvoluted spectrum
Annotated masses
species
AB
B
B2
250 nM DHD_127
125 nM DHD_127
63 nM DHD_127
31 nM DHD_127
16 nM DHD_127
100
%
0
1500
2500
3500
4500
m/z
5000
15000
25000
35000
Da


Mass spectrum
Deconvoluted spectrum
Annotated masses
8 µM DHD_131
4 µM DHD_131
2 µM DHD_131
1 µM DHD_131
500 nM DHD_131
100
%
0
1500
2500
3500
4500
m/z
5000
15000
25000
35000
Da
species
AB
A
A2
B2

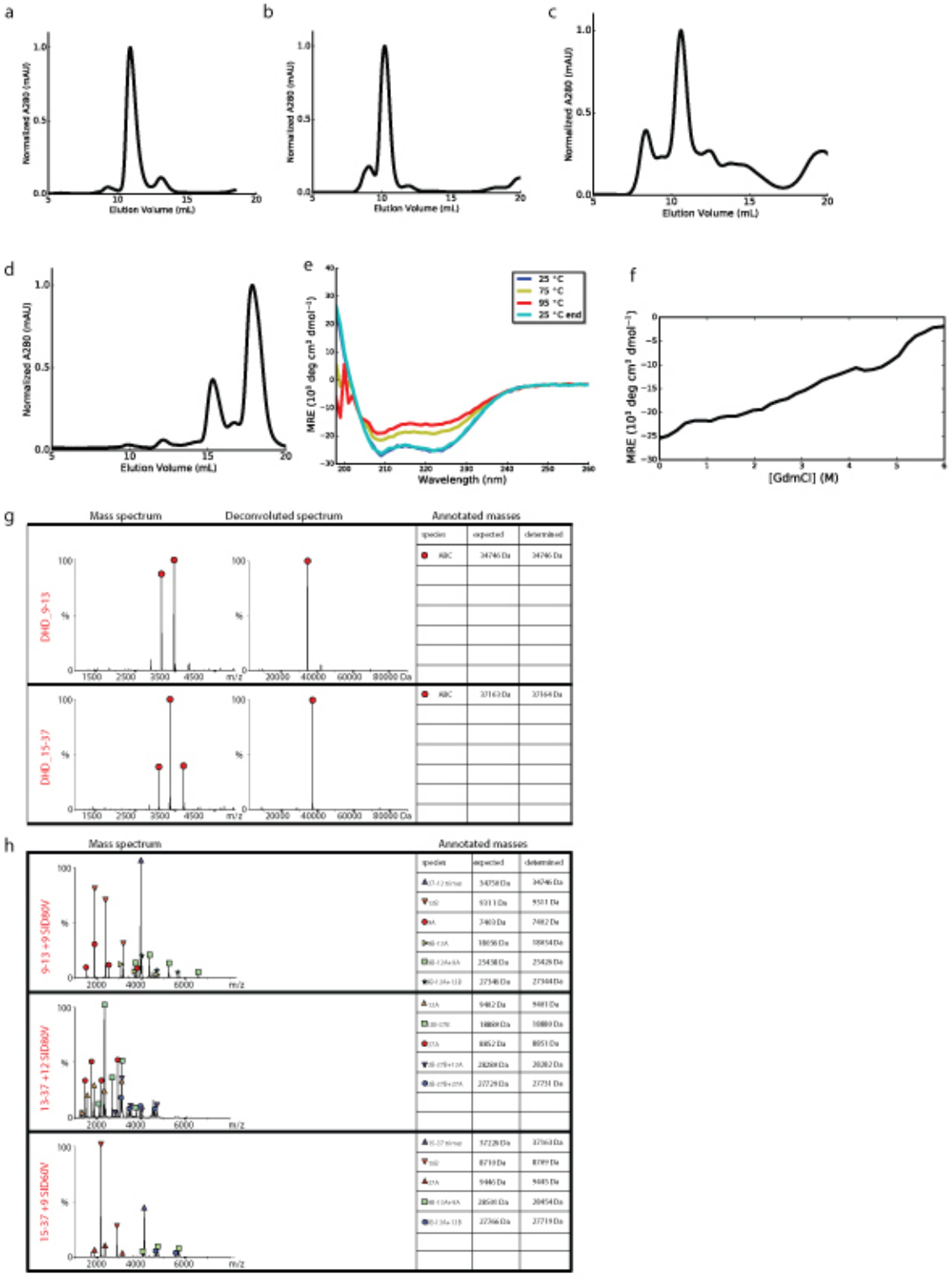

Mass spectrum
Deconvoluted spectrum
Annotated masses
250 nM DHD_131
125 nM DHD_131
63 nM DHD_131
31 nM DHD_131
16 nM DHD_131
100
%
0
1500
2500
3500
4500 m/z
5000
15000
25000
35000 Da
species
AB
A
A2
B2


Mass spectrum
Deconvoluted spectrum
Annotated masses
species
AB
A
B
8 µM Synzip12
4 µM Synzip12
2 µM Synzip12
1 µM Synzip12
500 nM Synzip12
100
%
0
1500
2500
3500
4500
m/z
5000
15000
25000
35000
Da



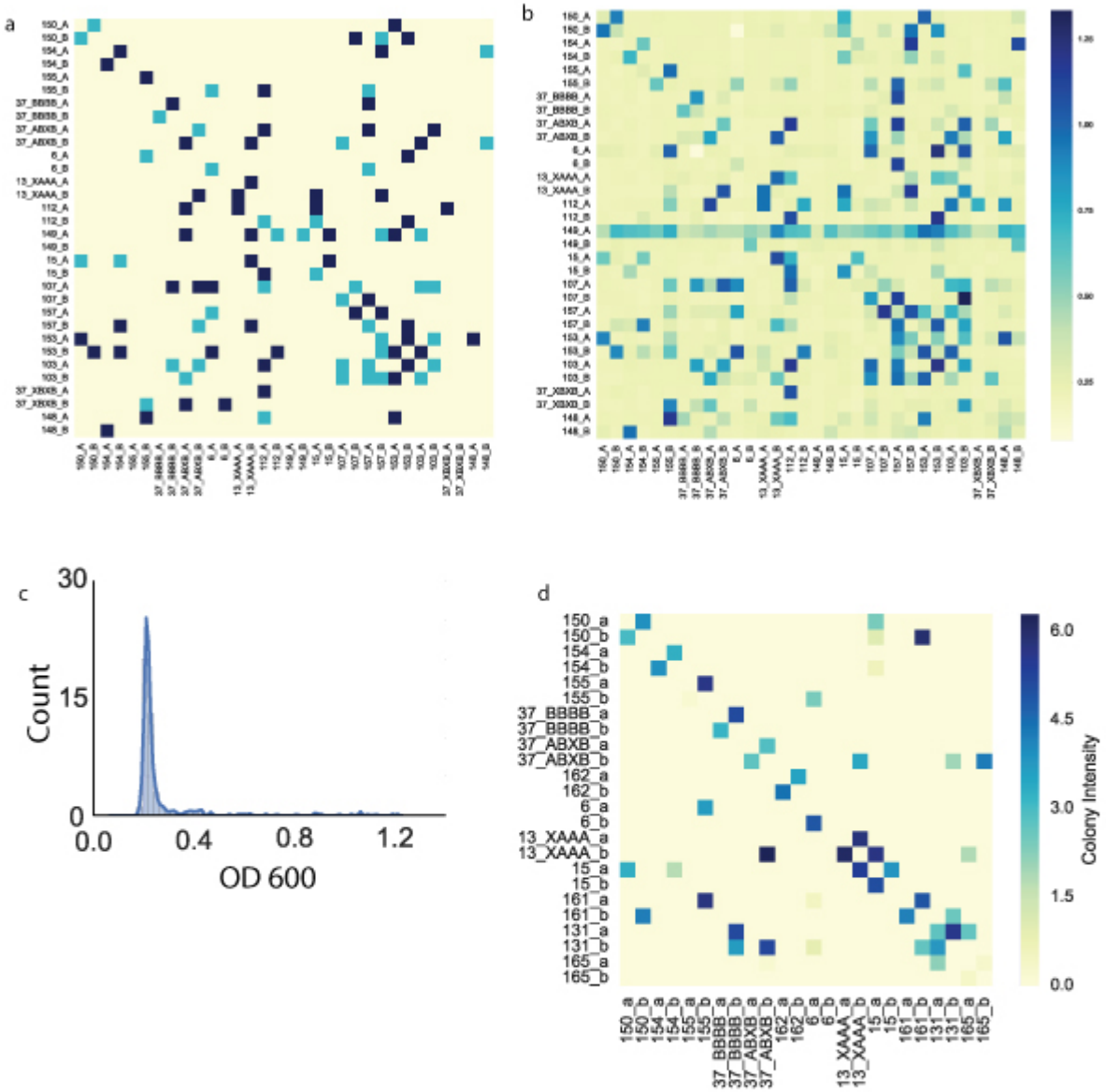
Extended Data Fig. 18 Native MS titration experiment on DHDs. Designs were incubated at indicated concentration for 24 hours before analyzed by native MS. For comparison, we also

carried out native MS experiments on the previously reported Synzip12 pair, which was found by fluorescence polarization to have an affinity below 10nM. The three DHDs examined (DHD_13_XAAA, DHD_127 and DHD_131) were observed to form heterodimers exclusively by native MS down to 125 nM, whereas Synzip12 monomers were observed at concentrations as high as 8 uM (the native MS conditions are clearly different from those in the solution fluorescence polarization experiment, so the discrepancy is not surprising).



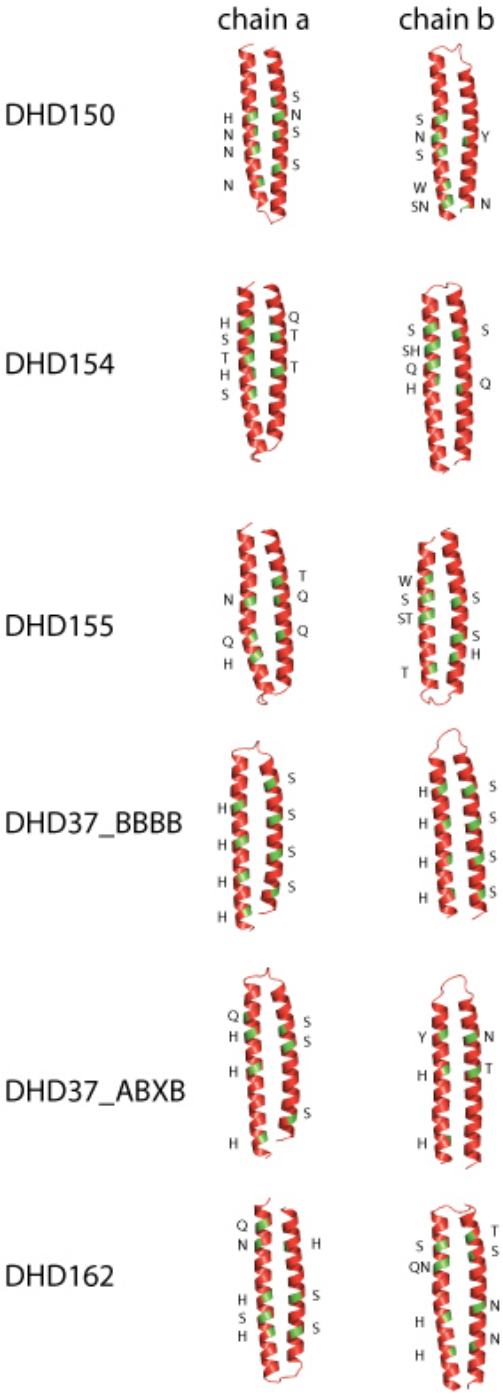
Extended Data Fig. 19 Biophysical characterization of induced dimerization and scaffolding proteins. a-d, SEC traces of the induced dimerization DHD_9-13 fusion, DHD_15-37 fusion, DHD_13-37 fusion, and the scaffolding complex in **Fig. 3d**. **e**, CD thermal

melt curves for the scaffolding complex in **Fig. 3d.** Wavelength scan was performed at 25°C, 75°C, 95°C, and final 25°C. Design was alpha helical and stable up to 95°C. **f,** CD chemical denaturation profile of the scaffolding complex in **Fig. 3d. g,** Native MS spectra for induced dimerization DHD_9-13 fusion and DHD_15-37 fusion. **h**, SID data of the induced dimerization complexes, showing no cross talk between each monomer complexes without the heterodimerizer fusion.

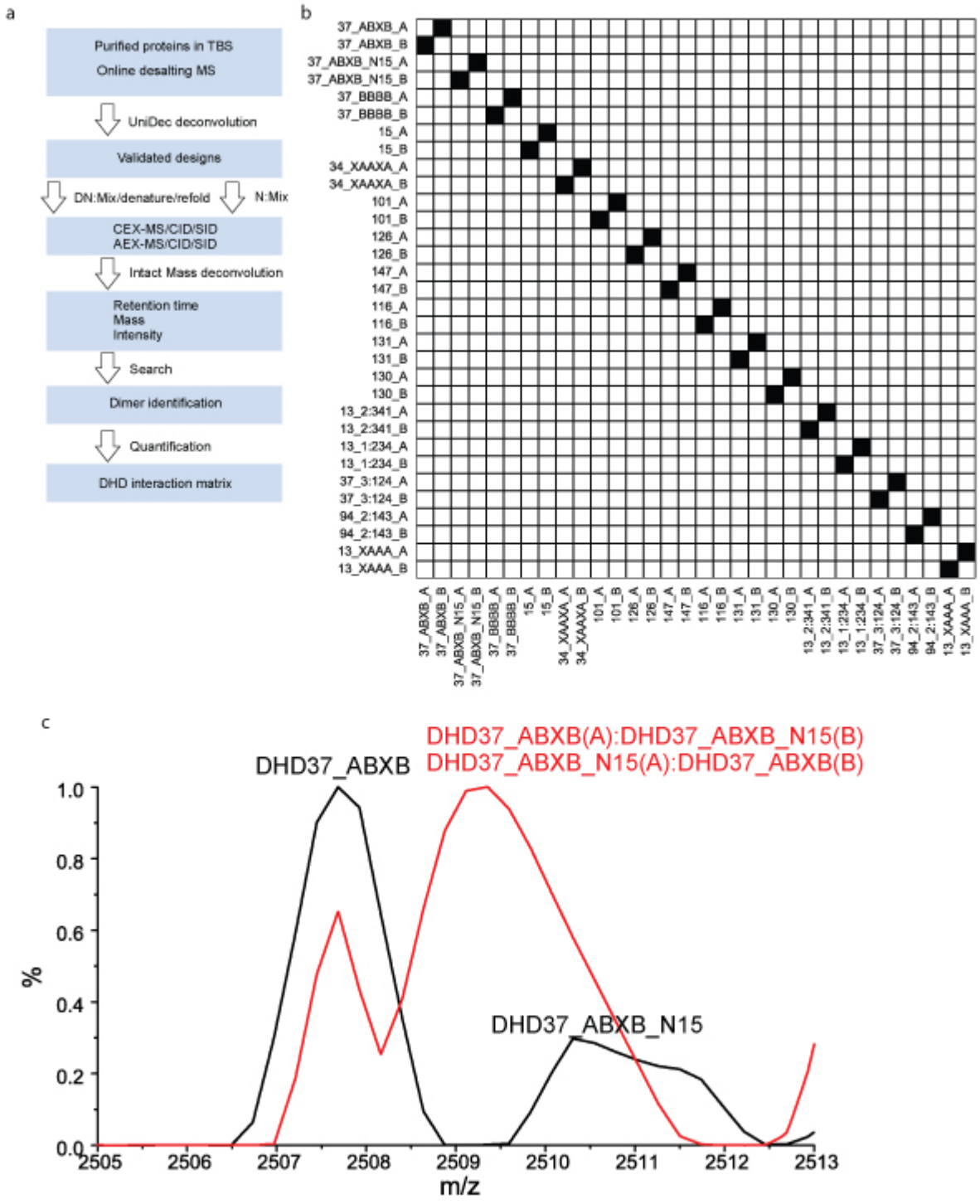


Extended Data Fig. 20 Y2H all-against-all assay of 16 DHDs. a, Y2H assay with cell growth on agar plates containing 100 mM 3-AT, lacking tryptophan, leucine and histidine. Plates were imaged at Day 5. Yellow, no growth on agar plates; light blue, weak growth forming non-circular

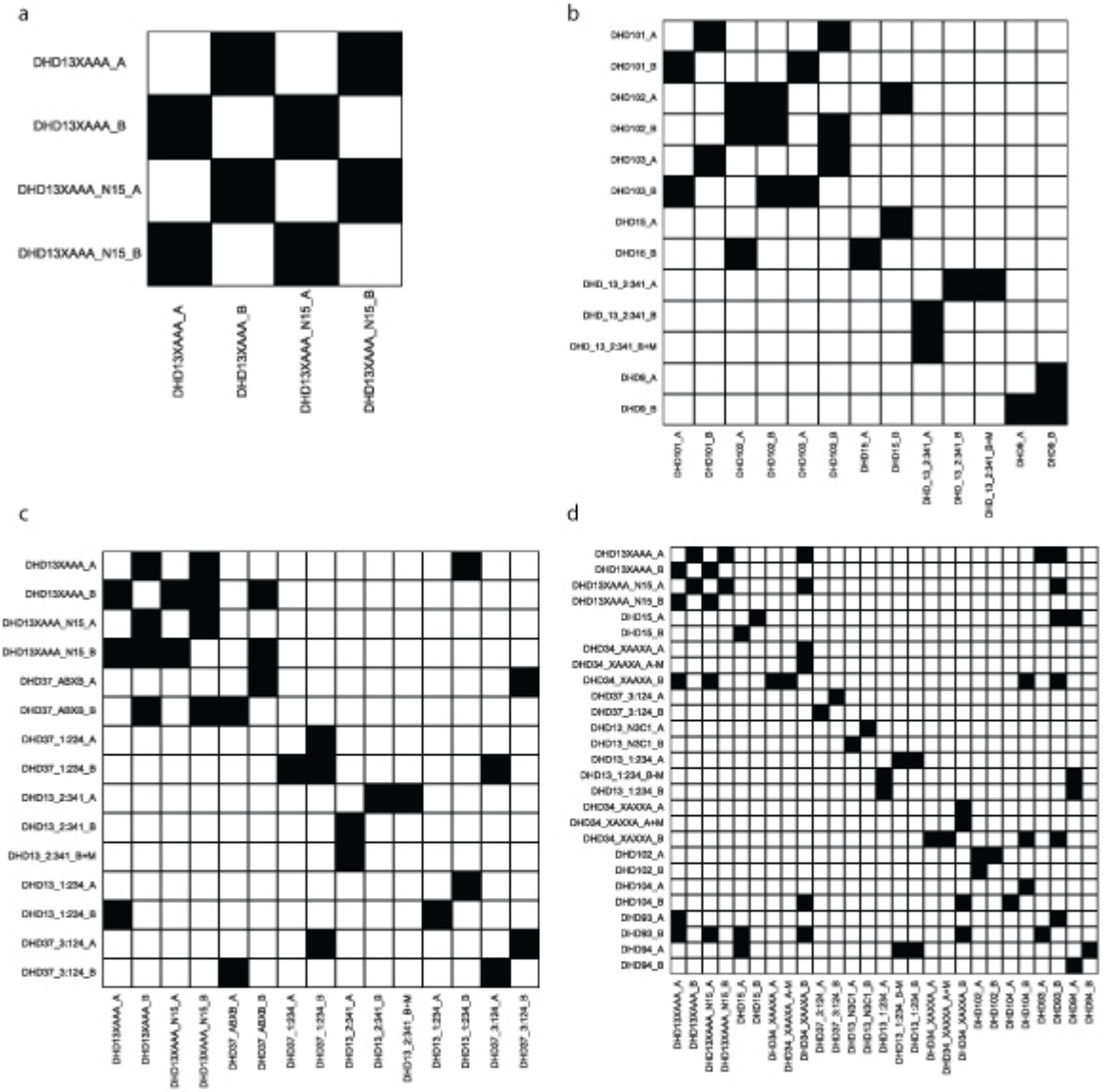
colonies; dark blue, strong growth. **b,** Y2H result by growing yeast culture in liquid media containing 100 mM 3-AT, lacking tryptophan, leucine and histidine. OD 600 values were measured at Day 2 to evaluate cell growth. **c**, Distribution of OD 600 values for non-cognate interactions in **b**, the majority of cells grew to OD 600 values less than 0.4, indicating weak interactions for non-cognate binding. **d,** An additional set of DHDs tested by Y2H showing improved orthogonality.

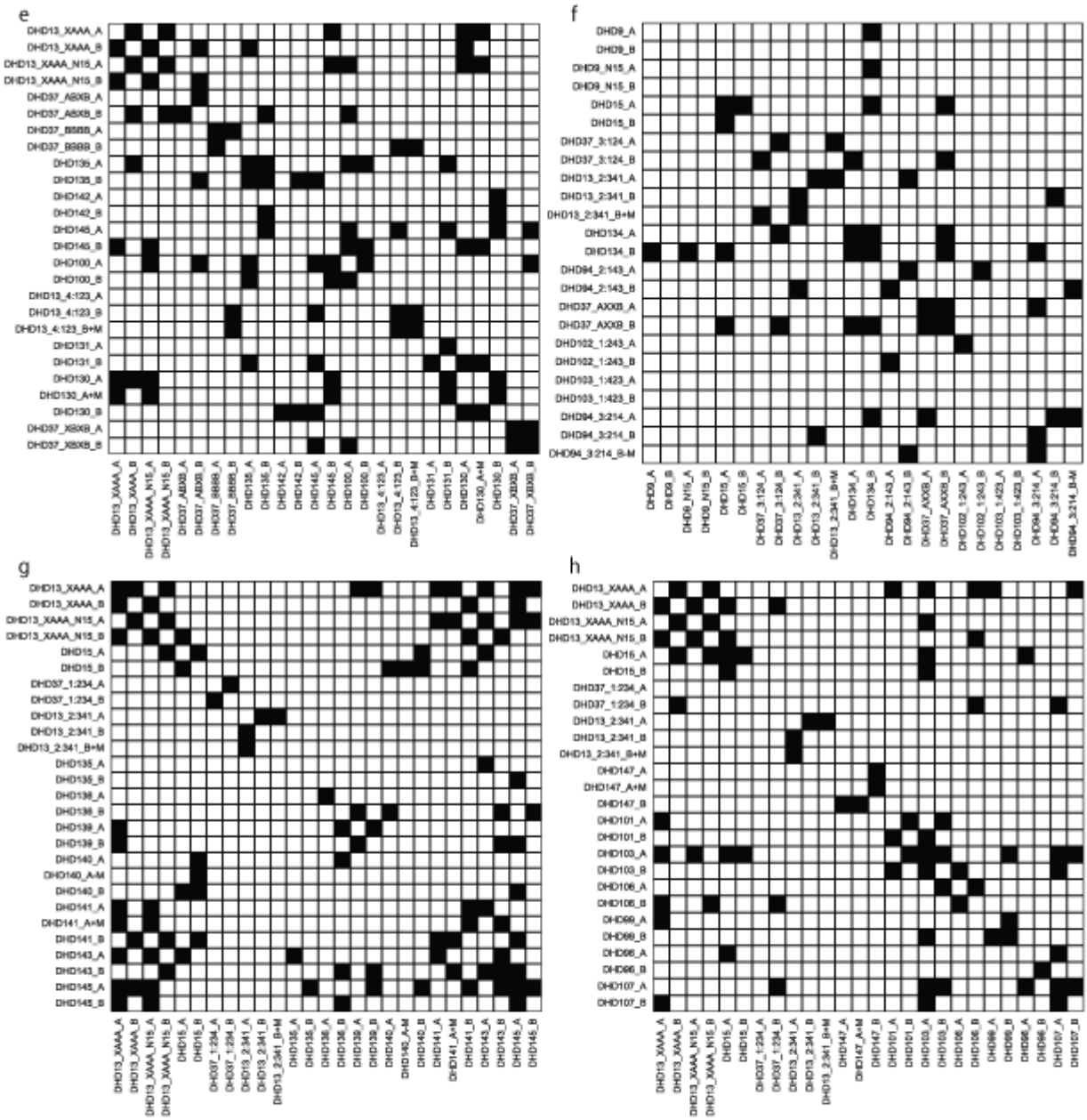


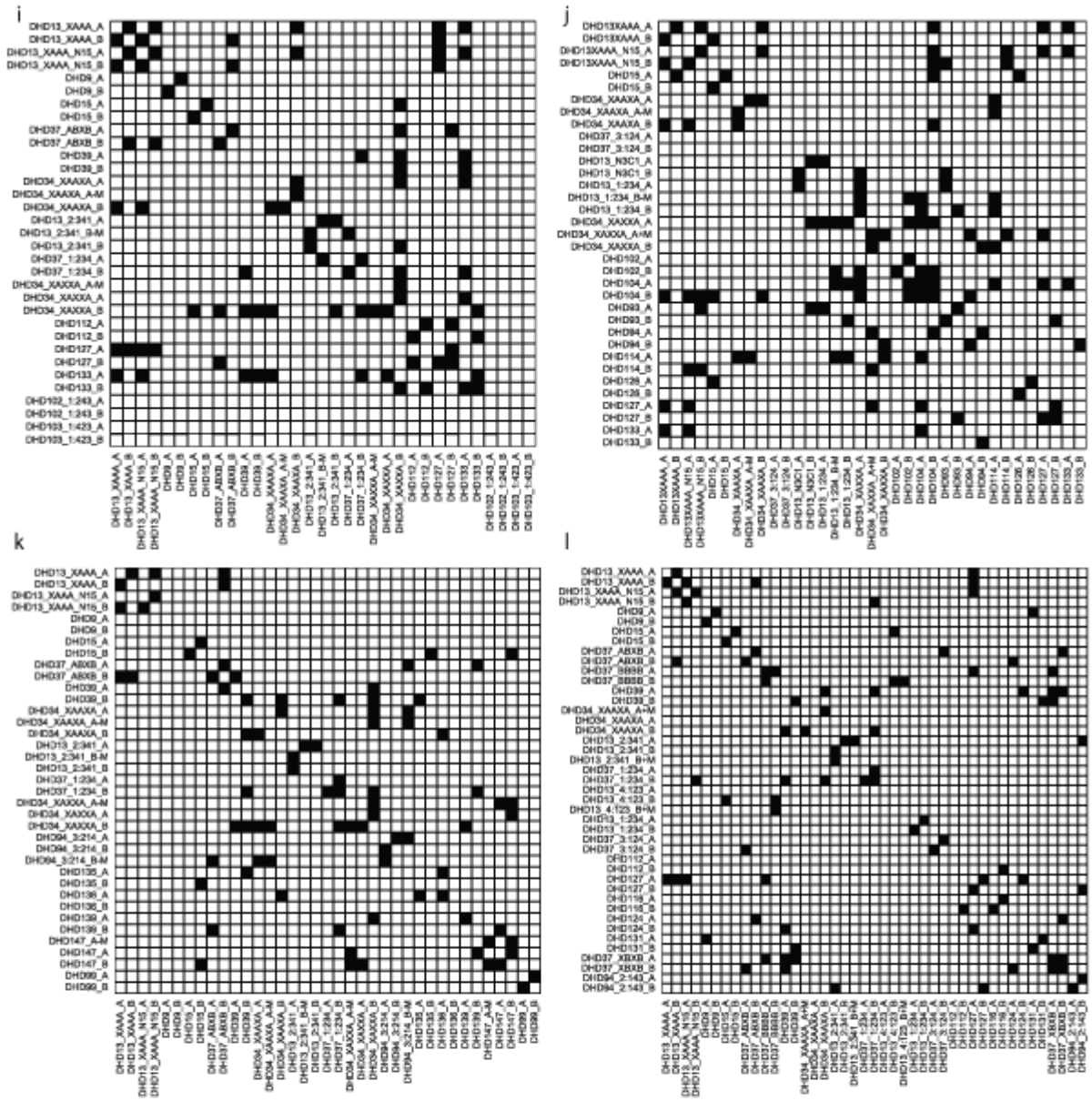
Extended Data Fig. 21 Hydrogen bond network sequence motifs of the set of 6 orthogonal pairs in Y2H experiments. Green patches mark the location of hydrogen bond network forming residues on the backbones.


a
Purified proteins in TBS
Online desalting MS
UniDec deconvolution
Validated designs
DN:Mix/denature/refold
N:Mix
CEX-MS/CID/SID
AEX-MS/CID/SID
Intact Mass deconvolution
Retention time
Mass
Intensity
Search
Dimer identification
Quantification
DHD interaction matrix
b
37_ABXB_A
37_ABXB_B
37_ABXB_N15_A
37_ABXB_N15_B
37_BBBB_A
37_BBBB_B
15_A
15_B
34_XAAXA_A
34_XAAXA_B
101_A
101_B
126_A
126_B
147_A
147_B
116_A
116_B
131_A
131_B
130_A
130_B
13_2:341_A
13_2:341_B
13_1:234_A
13_1:234_B
37_3:124_A
37_3:124_B
94_2:143_A
94_2:143_B
13_XAAA_A
13_XAAA_B
c
DHD37_ABXB
DHD37_ABXB(A):DHD37_ABXB_N15(B)
DHD37_ABXB_N15(A):DHD37_ABXB(B)
DHD37_ABXB_N15
%
m/z
1.0
0.8
0.6
0.4
0.2
0.0
2505
2506
2507
2508
2509
2510
2511
2512
2513

Extended Data Fig. 22 The workflow of Native MS mixing experiment. a, Proteins samples were characterized using online-desalting coupled to native MS and deconvoluted using UniDec software. Proteins shown expected masses were mixed in equimolar ratio, and the final mix was divided into two parts: in the experimental group (DN), proteins were denatured by 5M GdnHCl at 75°C and refoled into 150mM AmAc; in the control mixing experiment (N), denaturation and refolding steps were omitted. Sample mixtures in each group was further equally divided into 3 parts and was individually injected on LC-MS with WCX and WAX respectively. LC-MS analysis was performed for mixtures in full MS mode and MSMS mode with HCD and SID, respectively. Data were deconvoluted using Intact Mass™. The deconvoluted mass lists from Intact Mass™ were searched against a theoretical mass list of all possible monomer, dimer, trimer and tetramer combinations. Dimers were identified based on the full MS runs and MSMS runs with both subunits being detected at the same retention time. **b,** In the control mixing experiment (N), after mixing all 16 proteins in solution without the denaturation and renaturation steps, no exchange among proteins were observed. **c,** A mixing experiment of DHD_37_ABXB and ^{15}N labeled DHD_37_ABXB with (red) or without (black) the denaturation and refolding steps. MS peaks merged after subunit exchange due to the similarity in the masses of ^{15}N labeled and unlabeled subunits.

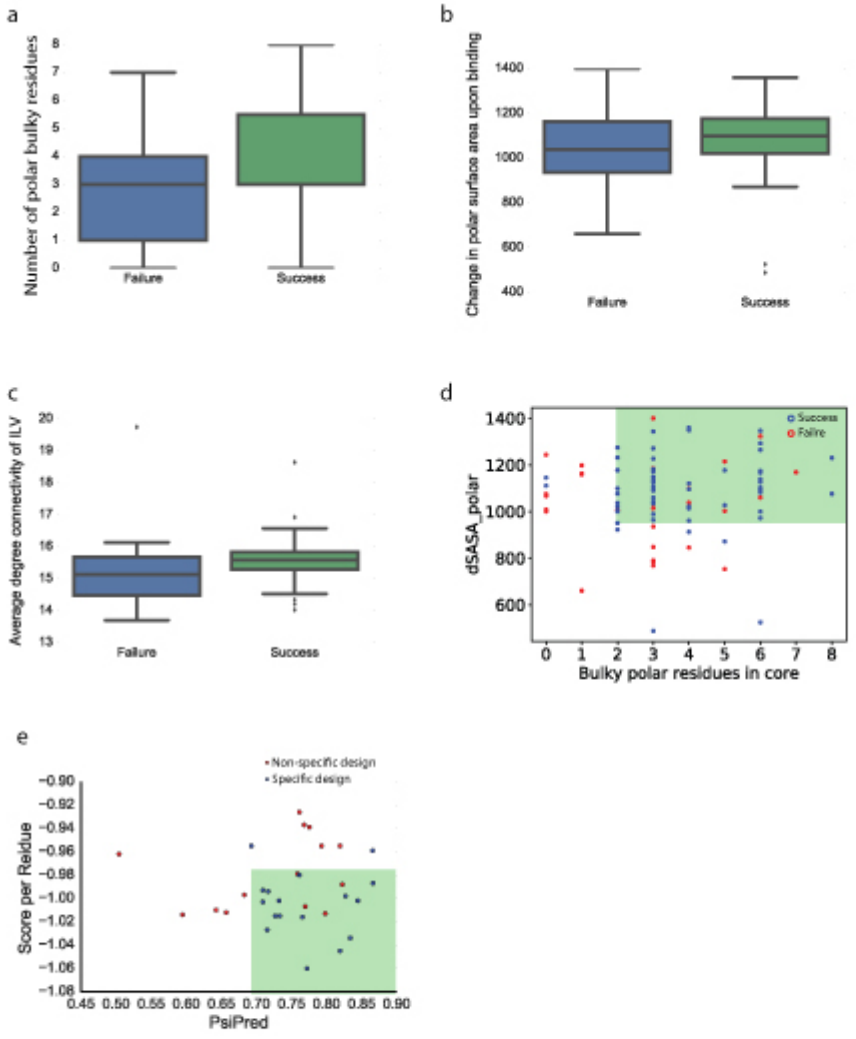

a
DHD13XAAA_A
DHD13XAAA_B
DHD13XAAA_N15_A
DHD13XAAA_N15_B
b
c
d


e
DHD13_XAAA_A
DHD13_XAAA_B
DHD13_XAAA_N15_A
DHD13_XAAA_N15_B
DHD37_ABXB_A
DHD37_ABXB_B
DHD37_BBBB_A
DHD37_BBBB_B
DHD135_A
DHD135_B
DHD142_A
DHD142_B
DHD145_A
DHD145_B
DHD100_A
DHD100_B
DHD13_4:123_A
DHD13_4:123_B
DHD13_4:123_B+M
DHD131_A
DHD131_B
DHD130_A
DHD130_A+M
DHD130_B
DHD37_XBXB_A
DHD37_XBXB_B
f
DHD9_A
DHD9_B
DHD9_N15_A
DHD9_N15_B
DHD15_A
DHD15_B
DHD37_3:124_A
DHD37_3:124_B
DHD13_2:341_A
DHD13_2:341_B
DHD13_2:341_B+M
DHD134_A
DHD134_B
DHD94_2:143_A
DHD94_2:143_B
DHD37_AXXB_A
DHD37_AXXB_B
DHD102_1:243_A
DHD102_1:243_B
DHD103_1:423_A
DHD103_1:423_B
DHD94_3:214_A
DHD94_3:214_B
DHD94_3:214_B+M
g
DHD13_XAAA_A
DHD13_XAAA_B
DHD13_XAAA_N15_A
DHD13_XAAA_N15_B
DHD15_A
DHD15_B
DHD37_1:234_A
DHD37_1:234_B
DHD13_2:341_A
DHD13_2:341_B
DHD13_2:341_B+M
DHD135_A
DHD135_B
DHD138_A
DHD138_B
DHD139_A
DHD139_B
DHD140_A
DHD140_A+M
DHD140_B
DHD141_A
DHD141_A+M
DHD141_B
DHD143_A
DHD143_B
DHD145_A
DHD145_B
h
DHD13_XAAA_A
DHD13_XAAA_B
DHD13_XAAA_N15_A
DHD13_XAAA_N15_B
DHD16_A
DHD15_B
DHD37_1:234_A
DHD37_1:234_B
DHD13_2:341_A
DHD13_2:341_B
DHD13_2:341_B+M
DHD147_A
DHD147_A+M
DHD147_B
DHD101_A
DHD101_B
DHD103_A
DHD103_B
DHD106_A
DHD106_B
DHD99_A
DHD99_B
DHD96_A
DHD96_B
DHD107_A
DHD107_B



Extended Data Fig. 23 Native MS Mixing experiments to identify a set of orthogonal DHDs. Black, interaction observed; white, no interaction observed. **a,** A control mixing experiment of DHD13_XAAA with or without ^{15}N labeling, all expected exchanges occurred after

denaturation and refolding. **b-l**, Scouting mixing experiments to identify orthogonal DHDs.


a
Number of polar bulky residues
Failure
Success
b
Change in polar surface area upon binding
Failure
Success
c
Average degree connectivity of ILV
Failure
Success
d
dSASA_polar
Bulky polar residues in core
Success
Failre
e
Non-specific design
Specific design
Score per Reidue
PsiPred

Extended Data Fig. 24 Sequence and structural properties that correlate with specificity. **a-c,** box plots of various properties for designs that assembled to off-target oligomeric states by native MS (failure) and that assembled into constitutive heterodimers (success). **a,** More buried bulky polar residues strongly correlates with design success. **b,** Successful designs tend to have bigger polar interface surface area. **c,** Designs with better hydrophobic packing (as reported by the Rosetta filter value Average Degree on Ile, Leu and Val residues) tend to have a higher chance of being constitutive heterodimers as assessed by native MS. **d,** Contribution of bulky residues and hydrogen bond networks to specific dimer formation. “dSASA_polar” measures interface hydrophilicity and positively correlates with the surface area of hydrogen bond networks at the interface. “Bulky polar residues in core” counts the total number of buried bulky residues that participate in hydrogen bond networks. Constitutive heterodimer formation (blue circles) or off-target oligomer formation (red circles) were determined with native MS. Filter cutoff values of dSASA_polar > 970 $Å^2$ and more than 1 polar bulky residues buried in the core includes most of the successful designs while excludes most of design failures. **e,** Based on the Y2H data in **Extended Fig. 20a**, all 32 monomers from the 16 pairs were categorized as being specific (blue, has fewer than or equal to 1 non-cognate binding), or non-specific (red, has more than 1 non-cognate binding). Applying secondary structure prediction scores (PsiPred (66)) and Rosetta centroid energy score per residue as filters, designs with higher PsiPred values and lower Rosetta centroid score per residue are more specific (green box).

Additional Tables

Table 1. Top 50 most frequent unique hydrogen bond networks identified during the systematic HBNet search.

HBNet composition	Frequency	Percentage (%)
(S/T)2Q1Y1	13954	3.870713616
(S/T)3Q1	9959	2.762536685
(S/T)2D1H1	8452	2.344508491
(S/T)1Q2Y1	7603	2.109003556
(S/T)1D1Q1Y1	7359	2.041320159
(S/T)3D1	6332	1.756439631
(S/T)3D1Q1	5525	1.532585117
(S/T)1D1Q3	5071	1.406649616
(S/T)1D1Q2	5062	1.404153098
(S/T)2N1Y1	5046	1.399714842
(S/T)1N1Q1Y1	4921	1.365040971
(S/T)2H2	4683	1.299021919
(S/T)2H1Q1	4572	1.268231522
(S/T)3H1	3955	1.097081292
(S/T)2D1Q2	3946	1.094584773
(S/T)1D1N1Q2	3862	1.071283932
(S/T)3N1	3783	1.049370045
(S/T)2D1Y1	3762	1.043544835
(S/T)2D1Q1	3669	1.017747474

(S/T)1D1H1Q1	3653	1.013309219
(S/T)1D1Q1W1	3409	0.945625822
(S/T)1D1Q2Y1	3342	0.927040627
(S/T)2Q3	3111	0.862963312
(S/T)2D1N1Q1	2999	0.831895523
(S/T)2Q2Y1	2850	0.790564269
(S/T)1D1W1Y1	2849	0.790286878
(S/T)2N1Q2	2741	0.760328653
(S/T)2D1Q1Y1	2723	0.755335615
(S/T)1D1N1Q1Y1	2684	0.744517367
(S/T)2Q1W1	2641	0.732589556
(S/T)2H1N1Q1	2591	0.718720007
(S/T)2Q2	2582	0.716223488
(S/T)2N1Q1	2554	0.708456541
(S/T)2D1W1	2467	0.684323527
(S/T)2H1N1	2377	0.659358339
(S/T)4Q1	2305	0.639386189
(S/T)1N1Q2Y1	2296	0.636889671
(S/T)1D2Q2	2285	0.63383837
(S/T)2D1H1Q1	2276	0.631341851
(S/T)2D1Q1W1	2267	0.628845332
(S/T)1H1Q1Y1	2222	0.616362739
(S/T)1D1N1Q1	2207	0.612201874
(S/T)2H1Y1	2150	0.596390589
(S/T)1D1N1Y1	2109	0.585017559

(S/T)1Q1Y2	1962	0.544241086
(S/T)1H1Q2	1957	0.542854131
(S/T)1Q1W1Y1	1954	0.542021958
(S/T)2N1Q1Y1	1935	0.53675153
(S/T)3H1Q1	1901	0.527320237
(S/T)1D1H1W1	1879	0.521217635

Table 2. Design Sequences

Design name	Oligomerization State	Chain	Design sequence
DHD9	Heterodimer	a	GSPKEEARELIRKQKELIKEQKKLIKEAKQKSDSRDAERIWKRSREINRESKKINKRIKELIKS
DHD9	Heterodimer	b	PKKEAEELAEESEELHDRSEKLHERAEQSSNSEEARKILEDIERISERIEEISDRIERLLRS
DHD13_XAAA	Heterodimer	a	GTKEDILERQRKIIERAQEIHRRQQEILEELERIIRKPGSSEEAMKRMLKLLEESLRLLKELLELSEESAQLLYEQR
DHD13_XAAA	Heterodimer	b	GTEKRLLEEAERAHREQKEIIKKAQELHRRLEEIVRQSGSSEEAKKEAKKILEEIRELSKRSLELLREILYLSQEQKGSLVPR
DHD13_XAXA	Heterodimer	a	TKEDILERQRKIIERAQEIIRRQQEILEELERIIRKPGSSEEAMKRMLKLLEESLRLLKELLELLEESAQLLYEQR
DHD13_XAXA	Heterodimer	b	GSTEKRLLEEAERAHREAKEIIKKAQELHRRLEEIVRQSGSSEEAKKEAKKILEEIRELSKRLLELLREILYLSQEQK
DHD13_XAAX	Heterodimer	a	TKEDILERARKIIERAQEIHRRQQEILEELERIIRKPGSSEEAMKRMLKLLEESLRLLKELLELSEELAQLLYEQR
DHD13_XA	Heterodimer	b	GSTEKRLLEEAERAIREQKEIIKKAQELHRRLEEIVRQSGSSEEAKKEAKKILEEIRELSKRSLELLREILYLLQEQK

AX

DHD13_2:341	Heterodimer	a	TKEDILERQRKIIERAQEIHRRQQEILEELEYIIR
DHD13_2:341	Heterodimer	b	MSEEAMKRMLKLLEESLRLLKELLELSEESAQLLYEQRKANNGSETEKRLLEEAERAHREQKEIIKKAQELHRRLEEIVRQSGSSEEAKKEAKKILEEIRELSKRSLELLREILYLSQEQK
DHD13_AAAA	Heterodimer	a	MTKEDILERQRKIIERAQEIHRRQQEILKEQEKIIRKPGSSEEAMKRSLKLIEESLRLLKELLELSEESAQLLYEQR
DHD13_AAAA	Heterodimer	b	GTEKRLLEEAERAHREQKEIIKKAQELHKELTKIHQQSGSSEEAKKRALKISQEIRELSKRSLELLREILYLSQEQK
DHD13_BAAA	Heterodimer	a	TKEDILERQRKIIERAQEIHRRQQEILKRSEEIIRKPGSSEEALETLRELQEESLRLLKELLELSEESAQLLYEQR
DHD13_BAAA	Heterodimer	b	GSTEKRLLEEAERAHREQKEIIKKAQELHRRTEEIIRQSGSSEEAKDELRRIQEEIRELSKRSLELLREILYLSQEQK
DHD13_4:123	Heterodimer	a	TTKRYLEEAERAHREQKEIIKKAQELHRRLEEIVRQ
DHD13_4:123	Heterodimer	b	GSSEEAKKEAKKILEEIRELSKRSLELLREILYLSQQVNDVDEKALERQRKIIERAQEIHRRQQEILEELERIIRKPGSSEEAMKRMLKLLEESLRLLKELLELSEESAQLLYEAR
DHD13_1:2	Heterodimer	a	EAMKRMLKLLEESLRLLKELLELSEESAQLLYEAR

34

DHD13_1:234	Heterodimer	b	TTKRYLEEAERAHREQKEIIKKAQELHRRLEEIVRQSGSSEEAKKEAKKILEEIRELSKRSLELLREILYLSQQVNDVDEKALERQRKIIERAQEIHRRQQEILEELERIIRKPGS
DHD15	Heterodimer	a	TREELLRENIELAKEHIEIMREILELLQKMEELLEKARGADEDVAKTIKELLRRLKEIIERNQRIAKEHEYIARERS
DHD15	Heterodimer	b	GTERKLLERSRRLQEESKRLLDEMAEIMRRIKKLLKKARGADEKVLDELRKIIERIRELLDRSRKIHERSEEIAYKEE
DHD20	Heterodimer	a	GDRQELIRRNIELLKEHIKILEEISQLIEELSELLDKSSSEEVVKRYKKILERYKQLLRKSQEIHKESSEIAKKES
DHD20	Heterodimer	b	GDEQKLIERSQRMQKESLELLKEIIKILDTIEKLLDKPDSEELLDTIKKLHDTLKKIHDRNKKLLKEHEEILRQRSGSLVPR
DHD21	Heterodimer	a	DKEEEYKRLLDEIKEILKESKEVLKDSKRVLEDIKRKVPDDDLVKLLEKHVRLLEEHVKLLEQLIREAEKSSK
DHD21	Heterodimer	b	QGSSAEELLKKIKESEKKIRDSLRKIKEIIKKSRKEGVDDKQLDLIRKVVESHRDLLRLHRDLLRLLREETS
DHD25	Heterodimer	a	DIDESIKEVEKLLEEVEQSLQKLDDSLKKLLEKVNQDPDVDDSVRKIVKRHVEILKRHEEVLKRLIEVVKEHTKTVK
DHD25	Heterodimer	b	GSDREEVHKEIVKLIREIIKIHKKILKIHEKIKNGEIDPSEILKLSEEIKKLTDTIIKIIEDLEQLTRDLRR
DHD27	Heterodimer	a	DRKEIVKRHQKVVELLKESSKLLRESSKLLQRLLDKTGDENLQKAVDDQDKAIKRQETAIRKSQEASKKLD
DHD27	Heterodim	b	DNSEEIKKVAKTSREVAEYSERVAKENDKVVKTLEEGKIDESELLRLLEESIKIFDTALKLHEEAYKLHQDLVRKVS

	er		
DHD30	Heterodimer	a	DESEAASVAIESVQILVESVKLLEESVRILLDAVKKNGVEDLLRVAQRWEKLVDEWLKVVKRWLDNVRDIQR
DHD30	Heterodimer	b	GSDKAEEVEKSVRKIEESIKKIRKSIKKAEDAVQLLKEGKIDAKDFLRIVREDLEVVKEDVEIVKEDVENVREFSS
DHD33	Heterodimer	a	SDKEVSDKLLKASKKLLKVSEELLEVVRRLLKALKDDELIKKIADLLRKIIDKDKKFIRTSEEIVKESR
DHD33	Heterodimer	b	GSDLKEVLKTVEEAVKEIIKSSEELLQISRKILEISRVGVDEHEYISAIREYLKALEKHIQILKKFIEILKELIRAVS
DHD34_XAAXA	Heterodimer	a	SKEEIDKIVKKHKKKIEEHKKKVDELKKLVEEHDKRVSQDKDDKVKKLSEEVKKIIKRLEEVSKRLEEVSKKLLKVISDKR
DHD34_XAAXA	Heterodimer	b	GSNDEELKKILETLDRILKKLDKILTRLIEVLKKSEDPNLDDKDYTELVKQFIELIKKYEEVVKEYEEVVRQLIRLFS
DHD34_XAXXA	Heterodimer	a	SKEEIDKIVKKHKKKIEELKKLVDELKKLVEEHDKRVSQDKDDKVKKLSEEVKKIIKRVEEVAKRLEEVSKKLLKVISDKR
DHD34_XAXXA	Heterodimer	b	GSNDEELKKILETLDRILKKLEKILTRLIEVLKKSEDPNLDDKDYTELVKQFIELIKKFEEVIKEYEEVVRQLIRLFS
DHD34_XA	Heterodimer	a	SKEEIDKIVKKHKKKIEEHKKKVDEHKKLVEEHDKRVSQDKDDKVKKLSEELKKISKRLEEVSKRLEEVSKKLLKVISDKR

AAA			
DHD34_XAAAA	Heterodimer	b	GSNDEELKKILETLDRILKKLDKILTRLDEVLKKSEDPNLDDKDYTELVKQYIELVKKYEEVVKEYEEVVRQLIRLFS
DHD36	Heterodimer	a	DHSRKLKEILDRLRKHVKRLKEHLDELRDLVRQVPEDKLLEHVVKLSDKILQISERAVREFTKSVDKDS
DHD36	Heterodimer	b	GSDKKDELERILDEIRRLIERLDEILSRLNKLLELLKHGVPNAKEVVKDYIRLLKEYLELVKEFLKLVKRHADLVS
DHD37_ABXB	Heterodimer	a	DSDEHLKKLKTFLENLRRHLDRLDKHIKQLRDILSENPEDERVKDVIDLSERSVRIVKTVIKIFEDSVRKKE
DHD37_ABXB	Heterodimer	b	GSDDKELDKLLDTLEKILQTATKIIDDANKLLEKLRRSERKDPKVVETYVELLKRHEKAVKELLEIAKTHAKKVE
DHD37_BBBB	Heterodimer	a	MDEEDHLKKLKTHLEKLERHLKLLEDHAKKLEDILKERPEDSAVKESIDELRRSIELVRESIEIFRQSVEEEE
DHD37_BBBB	Heterodimer	b	GDVKELTKILDTLTKILETATKVIKDATKLLEEHRKSDKPDPRLIETHKKLVEEHETLVRQHKELAEEHLKRTR
DHD37_XBXB	Heterodimer	a	DSDEHLKKLKTFLENLRRHLDRLDKLLKELRDILSENPEDERVKDVIDELERVIRIVKTVIKIFEDSVRKKE
DHD37_XB	Heterodimer	b	GSDDKELDKLLDTLEKILQTATKIIDDLNKVLEKLRRSERKDPKVIETVVELLKRHEKAVKELLEIAKTHAKKVE

XB			
DHD37_AXXB	Heterodimer	a	DSDEHLKKLKTFLENLRRLEDLLDKHIKQLRDILSENPEDERVKDVIDLSERVVRTVKTVIKIFEDSVRKKE
DHD37_AXXB	Heterodimer	b	GSDDKELDKLLDTLEKILQTATKVVDDANKLLEKLRRSERKDPKVVETYVELLKRLEKLIKELLEIAKTHAKKVE
DHD37_3:124	Heterodimer	a	DSDEHLKKLKTFLENLRRHLDRLDKHIKQLRDILSEN
DHD37_3:124	Heterodimer	b	EDERVKDVIDLSERSVRIVKTVIKIFEDSVRKLEKTKPDSKTAKELDKLLDTLEKILQTATKIIDDANKLLEKLRRSERKDPKVVETYVELLKRHEKAVKELLEIAKTHAKKVE
DHD37_1:234	Heterodimer	a	DSDEHLYKLKTFLENLRRHLDRLDKHIKQLRDILSENPEDERVKDAIDLSERSVRIVKTVIKIFEDSVRKKEKRPIDKRDDKELDKLLDTLEKILQTATKIIDDANKLLEYLRR
DHD37_1:234	Heterodimer	b	GDPKVVETYVELLKRHEKAVKELLEIAKTHAKKVE
DHD37_AXBB	Heterodimer	a	DSDEHLDRLDKHLKKLKTFLENLRRHIKQLRDILSENPEDERVKDVIDLSKTVIKIFEDSVRKKERSVRIVE
DHD37_AXBB	Heterodimer	b	GSDDKEATKIIDDLDKLLDTLEKILQTANKLLEKLRRSERKDPKVVETYVKAVKELLEIAKTHAELLKRHEKKVE
DHD37_XB	Heterodimer	a	DSDEHIKQLRDHLDRLDKHLKKLKTFLENLRRILSENPEDERVKTVIKIFEDSVRKKERSVRIVKDVIDLSE

BA			
DHD37_XBBA	Heterodimer	b	GSDDKEANKLLEKATKIIDDLDKLLDTLEKILQTLRRSERKDPKAVKELLEIAKTHAELLKRHEKVVETYVKKVE
DHD39	Heterodimer	a	DHSRKLEEILDRLRKHVKRLLEHLRELLSLVKENPEDKDLVEVLELSLAILRRSLEAVEAFLKSVTKKDPDDEDLRRKADEIRKEVEEIKKSLAEVEKEIYKLK
DHD39	Heterodimer	b	GSSADDVLEDILKIIRELIEILDQILSLLNQLLKLLRHGVPNAKKVVEKYKEILELYLQLVSLFLKIVKTHADAVSGKIDKKAEEEIKKEEEKIKEKLRQAKDILKKLQEEIDKTR
DHD40	Heterodimer	a	DRDAHLYKLLTFLEQLVRHLDRLVKHITQLRDIVKKDPEDERAVDVIRQSVRSLEIVITVLKIFVDSVSDAARSKEAEKIVRKIRKEIDEIRQKLREIDKEVKKTTS
DHD40	Heterodimer	b	GSNDKVLDKILDILDRILRLATRVIDLANKLLQVKKKSTHKDPRIVETYKELLKIHETAVRLLLELADLHRRLKSKDEEANKRVETELDRIRKKVKDIEDKVRKLEDKVRKTAS
DHD43	Heterodimer	a	NDLSKEVLKKLEKSVEELLRRVQKSVKEAQKRGLLSDELVDRHLKILNQLVKRHLELLQEVIKRSDKK
DHD43	Heterodimer	b	GSDEAVKRVVEKSLKILDEVIKKSLDILRELIELQIRHAKDDESVIRASKSALKDAIEALKKSLDEIKKALKRSADEG
DHD65	Heterodimer	a	SSEEVVKVHEKVVKLHKEILELLKKIIKIHETAARDPDDKDSIKKLSDEIKKIVKRIEDISDQAKRESSDAQRKQS
DHD65	Heterodimer	b	DKEEESKELLKKLKEILKRSEELLEESKELLKLAKNGEIDESELADADRKLNKKHEKLVQDIQDLLREHERQDR
DHD70	Heterodimer	a	DEKKKIDKIVKETEDLLQKSEKLLQQSKEAVKRIRSQVKENEIVDRLLRISEELLKISRRLVEISRRIASTLS
DHD70	Heterodim	b	GSSKEEVIRLLKENVRLIKENLELLTRNLKLITDLVRGSNGSEEKIKTLKELLKEYRELLKRYRKLVEDYKRLVDKHD

	er		
DHD88	Heterodimer	a	EIQELIKSSRRIIEESKELIKESEEVLRRIKEILDRIRNGVDNQEDLLREILKLLTKNLKIIQRNLKLLQDNAEILKRLVS
DHD88	Heterodimer	b	GSYIEDVIKKILDVSRELIKLSRTIIKISEEINKQLQQGRDTKDLVKKYDEIIKKYTRIVQHYTELIKELQKLLS
DHD89	Heterodimer	a	SPTEEAIQLSQRVIELSKRVIELSKEILKLLKRVLDLLPDLDKNEEKRLDDYDKELKEYDKELKKYEKRLKDLAS
DHD89	Heterodimer	b	GSEEEEILKIQKELLRIQSEILDKQKKILDTLRSNGAVTEEVRSILEKVERLSEEAKELSKEAKELTKEVSKLIS
DHD90	Heterodimer	a	SPLKELNNQLLRLLRELVKVSKKIVDLSKTIIEVLKHTDLDPRLLDSLEKSQQELDKSQKELDKVVKELTKVNKKLQ
DHD90	Heterodimer	b	GSPLEDLVRKYDELVKTYEKLVEEFKKAVDKYDKAVKKAPVSKEATDSLDLIRKVLELLDRNLKLIKENAKLIKELLK
DHD91	Heterodimer	a	SPTRENEKVIKENEKVISDNERVLEEVVKVVETATDRKEIQDAVDEVRKSVDKLRDSVRKLEESVRTLD
DHD91	Heterodimer	b	GSPIKDISKRLLEISKRLVEISDRIVELLQRIADSKDPNKDLQKEVKDVLEEYKRLVREYREVVKEYEKVVS
DHD92	Heterodimer	a	DEDEHVKQLIKNADLLRKHAELLKELVKLFQEIASQIPDDRVAKKVTDVVDRIDKILKQTEKLVRRTKQILDYSR
DHD92	Heterodimer	b	GSNLEELVKLLKEVLEMHERLLRIHEDLVEAHKSNASDKESERKLKKSDKDIKESLKKIKSIIDQVRYIQS
DHD93	Heterodimer	a	PVEDIIEESLRLLEESLKLLNRILKLLEDSLRKLPRSEEWRQRLDEFRKKLEDWKEELERWIEDVRYKKT

DHD93	Heterodimer	b	GSDEDYESREIIDEIRKLLDRSKKIVHRSQRLVERVKSTPLSEDQEDLIRRHEETINRHRELVKELEKVLEDHERHIR
DHD94	Heterodimer	a	PEEDSRRVLERFVRVSREVLKVLEEFLRVSEELLREADRDRDRRLEEYERQVDELREEIRRYKEEVDKFDKEVKYYKK
DHD94	Heterodimer	b	GSPEKDENRKLLDKVRKLVEKSRRLVEELRKLVDQSTKNGLIDEKALRKQQEVLRKVEEVLEKQERVLRELEEISYRVI
DHD94_3:214	Heterodimer	a	GSPERDENRKLLDKVRKLVEKSRRLVEELRKLVDQSTKN
DHD94_3:214	Heterodimer	b	GSDEKALRKQQEVLRKVEEVLEKQERVLRELEEISYRVITRGEDHKAEEDSRRVLERFVRVSREVLKVLEEFLRVSEELLREADRDRDRRLEEYERQVDELREEIRRYKEEVDKFDKEVKYYKK
DHD94_2:143	Heterodimer	a	GSDRRLEEYERQVDELREEIRRYKEEVDKFDKEVKYYKK
DHD94_2:143	Heterodimer	b	GSPERDENRKLLDKVRKLVEKSRRLVEELRKLVDQSTKNGLIDEKALRKQQEVLRKVEEVLEKQERVLRELEEISYRVITRGEDHKAEEDSRRVLERFVRVSREVLKVLEEFLRVSEELLREADR
DHD95	Heterodimer	a	DLSEESKKFVEKVKKLEKESRELEKQVKKIEEDSRSVENDVQKEFLELLKRLLDIQKKVVEVLREVVKVQQYVDS
DHD95	Heterodimer	b	GSDSEYESRQVLRELDTVLKDSHTVLEALRQVIRDSQDVVSKSDEESRRVIDDLEKVIQDSKKVLDDIKRLIDKSKSIKS
DHD96	Heterodimer	a	NEDELLKLLTENLKLLDENLKLLRENLSLLRQANNITDKNRIREIVKQSKEIVKQSREILKQSKEIVERIKYIVS
DH	Heter	b	GSSLYELTQRYEKLVQQYEELVKDYRRLVKKLEKLKRDNKPDKRLLKEIV

D96	odimer		DVIKKSVEIIDRSLKLLEESIKILEETD
DHD97	Heterodimer	a	SQERSLEILKRILDVLKESLEILKESLSILRQLASRIKNPNRKIEEILKESDKIIKESDKVLKEIEEVIRYSS
DHD97	Heterodimer	b	GSDIEYESKEILELIKELLKLSRELLKESRRALELVRKSRDDSIVEEVIQVHKKVLDIHKEVLKIVRKVVEVHRRVKS
DHD98	Heterodimer	a	SKKDESTKLERLAEKIDEITKRIEELVKDVKRKSSEGVDKDQQQKIDEVFQKLLDLQREILEILDRILKVQQYILD
DHD98	Heterodimer	b	GSDLEYLNRRLLQLIKTLIDLNRHLLKLIDKLKKLNSREGDEEKIKEESKQIQEQFKEIVERSKEIIKQIKEIIKRSQ
DHD99	Heterodimer	a	DFERSSRRLEKVVEDLRRSSDRLREVIDELRKSADEKDEDEDLRRARKEHRDLIEELKRALEKQEEIIKHLQELVYRQL
DHD99	Heterodimer	b	GSEESEEVRKVVERIKKISRELEEVVKELDRVSKEFDRHGETDEIVREHERIVEKLEEIVKKHTKIVEELAEIVYKQQ
DHD100	Heterodimer	a	SDDDSVRVLDEIVKILDESVKLLKESLKLLDDFLRTKPDDHLKEVVKESKKVVEQSKKVLDRIKKIIYESK
DHD100	Heterodimer	b	GSDLLYLSKELLKLVRELLKLSRELVELSRRLVNSTHKSPELVKKYDKLVKKYQDLLKKLADVADEYLRQRS
DHD101	Heterodimer	a	DEKDYHRRLIEHLEDLVRRHEELIKRQKKVVEELERRGLDERLRRVVDRFRRSSERWEEVIERFRQVVDKLRKSVE
DHD101	Heterodimer	b	GSDAYDLDRIVKEHRRLVEEQRELVEELEKLVRRQEDHRVDKKESHEILERLERIIRRSTRILTELEKLTDEFERRTR
DHD10	odim	a	DERYRAREHIRRVEEHTKRLRHILKRLREHEEKLRRELKPGDEITESVDRFKKIVDQFEESIKKFETVSEELRKSDS

2	er		
DHD102	Heterodimer	b	GSDRQRILDRLDKILEKLDDILKKLKDILETLSKDDVSDRRHKDLVEKFRELVDTHHKLVERYRELVYQNR
DHD102_1:243	Heterodimer	a	GSDEITESVDRFKKIVDQFEESIKKFETVSEELRKSIS
DHD102_1:243	Heterodimer	b	GSDPQRAADRLDKILEKLDDILKKLKDILETLSKDDVKDRRAKDLVEKFRELVDTHHKLVERYRELVYTATAGSDLARELIRRVEEHTKRLRHILKRLREHEEKLRR
DHD103	Heterodimer	a	NADDQLATSIKKLEDSIDQLIKIVRKFEESVKKLQKHGVDQHHVEILRKIVEIFRQHIEKLKKHLEKLRYTSS
DHD103	Heterodimer	b	GSDKEYLVTEHEKLVREHEKIVSEIEKLVKKHEAGVDESELEEILKKVEKLLRKLDEILEQLTQLLRKTE
DHD103_1:423	Heterodimer	a	GSDQHVVEILRKIVEIFRQHIEKLKKHLEKLRYTSS
DHD103_1:423	Heterodimer	b	GSDAEYLVTEHEKLVREHEKIVSEIEKLVKKHEKGVDESELEEILKKVEKLLRKLDEILEQLTQLLRKAEKHIDKHSKAADQLATSIKKLEDSIDQLIKIVRKFEESVKKLQKH
DHD104	Heterodimer	a	DEDDDIRRVLDESRRVLEHSRRVLKRSEEVLEKASRKKEKDTEEIEKHLKRLREHAKKLEKHRRELDDFLYKEI
DHD104	Heterodimer	b	GSRDKYLLERLNDILKKLDEIVDKLSDILKRLKDVRHDDRLQELVERYKEIVKEYKRIVEEYEKLVREFEEQQR
DHD105	Heterodimer	a	DRDYEDKEFKKIIKELEDVQEELKKLQEKIKRFSSELEEPNELLKEQLKVNEEQLEVNKKILKILRDQLKQNE

DHD105	Heterodimer	b	GSDAEYKVRESVKRSKESVKHSEDVVDKLNKSVKLSESGHSDAEKASRELVKLVREVVELSREVIKLSEKVLRVIS
DHD106	Heterodimer	a	DLQYKQEKLIRHFDRVVREWDKLVRKFSKVLEKQKHESKDKELEEASRRVDELIKRLREQLKRSKEILRRLKELSRKSS
DHD106	Heterodimer	b	GSDWEELLRRLEKVLQEYEEIVKELIDLIERLIKVSEDKSKDASEYKKLVTELEKLISKLEEISKKLEELVKEYEYKTE
DHD107	Heterodimer	a	DAKDELEKSLQEIEESLKELKKLLEELDKSLRELTSQGRNKKLEEHIKKVQKFIELVKKYIKAVQDYLKEVRYDNS
DHD107	Heterodimer	b	GSDKERAARATEEMVKLTKKLLKAVEDLVRDVRRLLKEGLISEKHARIAETILEVFKKHAKIIKKHVDIVKYDES
DHD108	Heterodimer	a	GSPLKERLLEIQRDLDRVLEEVVERLLRIQERLDSVVERKPPDVHEEYKYIVDEIREIVERVVREYEEIVKRIDEEVR
DHD108	Heterodimer	b	GSEEDERIRYDLDRIRKDVRRKLEEIRQRVRELEKKLRDAGHRRDEKELLRELIETSKDILRLVEELLKKIIDKSEDLLRKTE
DHD109	Heterodimer	a	GSDEEDYINENVEKDVRDIEDDVRRINERIRELLEKIRTEEVLQRVLEEHHELVERVLRKLVEILRKHEEENR
DHD109	Heterodimer	b	GSDEEEYYKEKLHKLLREIEELLKHYRELVRRLEELVKRGELDKDTAAHILERLSELLERIIRRVAHTLRRLSEERR
DHD110	Heterodimer	a	GSDEDEISYDSKRRVEEIVRQAREKSEKSRKDIEDVAEVLRKGDVSEKEVVDELVKVLEEQVKVLREAVERLREVLKKQVDDVR
DHD110	Heterodimer	b	GSDIVELVDHLLKRSLKLLEELAELVRRLLEKSTELLKRRTEEHKEEVVEESEYMVRELEERLRRVVDESEKLVRDADKHIR
DH	Heter	a	GSKEKDIVKTLVDLLRENLETLERLIEEVVRLLKENVDVRDEGRDDKDSER

D11 1	odimer		ILRDIKRRIDEAAKESREIIERIEKEVEYRSR
DHD11 1	Heterodimer	b	GSPEVDVLRRIVREILKASEELLRLLRKLIDEALKLSERKRDSQEYREVVDRVKKELERLLDEYRKLVEELKEKLRYDTR
DHD11 2	Heterodimer	a	GSDKRYESEKLKRRLDEAVEKVREVVERVERESDRVLEEVRRRRESKEVVDKVIEDNDKALEDVLRVVDEVAKVVRDVVRENTR
DHD11 2	Heterodimer	b	GSPREYHSKDILRKVDEILERIRRHADRVKKKSERLKRENVDVNEHSKDVKRVIRELLELVKELLRLAKKHSDDQQE
DHD11 3	Heterodimer	a	GSDEDEILYHSERLLQKLKKELDDLKEKSRELLEELKKEDPDDRLIERIIRLHDEVLKDLDEVLKNILEVHREVLERLR
DHD11 3	Heterodimer	b	DKLDRLLKIHEEALRRAEELIKRLLDIHRRALDLARRGELDDYLLKESERELREIIRRAREELKESRDRLEEISR
DHD11 4	Heterodimer	a	GSPKEELIRRVLEEVKRLNEKLLEIIRRAAELVKRANDELPETEKLREIDRELEKKLKEIEDELRRIDKELDDALYEIED
DHD11 4	Heterodimer	b	GSPKLDKLRELLERNLEKLREILEEVLKILRTNLERVREDIRDEDVLQEYERLIRKAEEDLRRVLKEYDDLLKKLVYELR
DHD11 5	Heterodimer	a	GSKEDESVKRAEEIVRTLLKLLEDSLREAERSLRDIKNGEDEHNLRRISEKLEELSKRITETIERLLRELQYTSR
DHD11 5	Heterodimer	b	GSPNQELLDRVRKILEDLLRLNEELVRLNKELLKRALEMRRKNRDSEEVLERLAEEYRKRLEEYRRELEKLLEELEETIYRYKR
DHD11 6	Heterodimer	a	GSDESEEAQHEVEKVLDDIRRLSEHLQKRLEEVLEEVYELRREGSDRTEVVELLKEVIREIVRVNREALERLLRVVEEAVKRNE
DHD11	Heterodimer	b	GSDEEELVETVKRIQKEILDRLTELAKLLVEIQREIKKLKDEGEDDKELKRLSDELEEKVRQVVEEIKRLSDELEETVEYVSR

6	er		
DHD117	Heterodimer	a	GSDEEEEVVRRAEELVKEHEELIERVIRTHEELVYKLEDQGADKKLVDVLKRVVEESERVAREIVKVSRELIRLLEEASR
DHD117	Heterodimer	b	GSSKEEILKELEDLQRRLIEELKKLQERVVELLEELIKRLRDRGRDDKHLKRLVKEVRRLSEEVLRSIKEVSDRVRYQLR
DHD118	Heterodimer	a	GSDKEEESEYLLRDLVRLLEKVKEKIEEVNREVEKLLKKVKDGRLDRREVLREILRLNRELAEIIKEVVDRIRHVVERSER
DHD118	Heterodimer	b	GSDLHEVVYETKELLKRIEEVVEELRKKSEDIIRKAERGEISEDELKRLQEEIAREAKKLLDEIKRVLERHLEQTL
DHD119	Heterodimer	a	GSPVEEIIKEVVKRVIEVQEKVLRIISHAVKRVVEVQKKYDPGSEESNRVVEEVKKTIEDAIRESDEVVDEVVKRIQYTVR
DHD119	Heterodimer	b	GSPEQEIADRILTEIRESQKELERLARKILKLLDESQEKAKRGRLSEEESDELLERIKKELDELLERSKELLKKIEYELR
DHD120	Heterodimer	a	GSDEDKEANRVLDEVLKTVRDLLETANEVLKEVLYRLKRTDDQEKVVRTLTEVLKEHLKLVEEIVRILDKVLKEHLETEK
DHD120	Heterodimer	b	GSPEDDVLRRLEEVSEKILRVAEDVARQLREVSEKITQGKVDRKEWEEDIKRLKRELEELLREWKEEIERLTYELR
DHD121	Heterodimer	a	GSRREEVVKRIRELLKRNKELIDRIRELLEENEYLDKDARDKDVLRRSVELLEELVRILEESVELAKEIIKLLREVVE
DHD121	Heterodimer	b	GSDEKEDNRRLQHKIERILEKNEDLQRKLEEILELLERGEADEEKIDRLRKAVEDYRRVVEEIKEDVKRHKYTVR
DHD122	Heterodimer	a	GSDEKEEAKKASEESVRTVERILEELLKASEESVELLRRGEDAKDVVERSKEALKRVKELLDEVVKRSDEILKYIHN

DHD122	Heterodimer	b	GSDEKKLINEVVETQKRLIKEAAKRLSEVVRHQTELIRELREKNVDDKDVEKLLKESLDLAEEIVRRIKELLDESKKLVEYVSN
DHD123	Heterodimer	a	GSPDMDEVKRVLDELIEIQEEILREIKRVLEKLIKIQEDNGSEYESREVVREIVEIARKLVERSRRVVKKITETLQ
DHD123	Heterodimer	b	GSDERYATREIVERIERIAREILKRTEEIVREVREVLSRDVDQEEVVRRLADLLRESVELVQHLVRRVEELLQESVERKK
DHD124	Heterodimer	a	GSPEREALREVLEDLKRVTDRLRELVERVLEELKKVTDHVDSERILRESRRVLKELKDIIEEILRESEKVLEKLKYTED
DHD124	Heterodimer	b	GSPAREILEEVVKKHLEVVEDAARILEEIIREHEKAVREDRDKKELEEISRDLLRKAREALKKVKDISDDLSREIEYVAS
DHD125	Heterodimer	a	GSPVEEAIKKVIDDLRDVQRKIRELVEELIRLLEEVQRDNDKRESEYVVERVEEILRRITETSREVVRKAVEDLS
DHD125	Heterodimer	b	GSDSDEKAEYLLKEMERVVRESDEVVKKILRDLEEVLERLRRGEISEDDVTEILKELAERHIRAIEELVRRLRELLERHKR
DHD126	Heterodimer	a	GSPVEEVLKELSEVNERVRDIAREIIERLSEVNEEVKETDDEDELKKISKKVVDEVEDLLRKILEVSEEVVRRVEYHDR
DHD126	Heterodimer	b	GSPKEDILREVLRRHKEIVREIVRLVREAVETHLELVKRNSDDRDAQDVIRKLEEDLERLVRHAQEVIEEIFYRLH
DHD127	Heterodimer	a	GSPRSYLLKELADLSQHLVRLLERLVRESERVVEVLERGEVDEEELKRLEDLHRELEKAVREVRETHREIRERSR
DHD127	Heterodimer	b	GSDREYIIKDILDSQEHLLRLIEELLETQKELLEILKRRPDSVERVRELVRRSKEIADEIRRQSDRNVRLLEEVSK
DH	Heter	a	GSDEKDEIRHVIESVERLIEDIKRLLKTLRELAHDDSDKKTVKEVLDRVKEM

D128	odimer		IERHRRELEEHRKELERAEYEVR
DHD128	Heterodimer	b	GSESEDRIKELLKRHIELVERHEELLHEIKKLIDLEEKDDKDREEAVKRIDDAIKESEEMLEESKEILEEIEYLNR
DHD129	Heterodimer	a	GSSLEDSVRLNDEVVKVVERVVRLNQEVVRLIKHATDVEDEETVKYVLERVREVLDESREVLKRVHELLEESERRLE
DHD129	Heterodimer	b	GSHEKDIVYKVEDLVRKSDRIAERAREIVKRSRDIMREIRKDKDNKKLSDDLLKVTRDLQRVVDELEELSRELLRVAEESRK
DHD130	Heterodimer	a	GSPELDEVKKLIDELKKSVERLEESIREVKESIKKLRKGDIDAEENIKLLKENIKIVRENIKIIKEIIDVVQYVLR
DHD130	Heterodimer	b	GSDEEEIEELLRELEKLLKKSEEALEESKKLIDESEELLRRDRLDKEKHVRASEEHVKLSEEHLRISREIVKILEKAVYSTR
DHD131	Heterodimer	a	GSDESDRIRKIVEESDEIVKESRKLAERARELIKESEDKRVSEERNERLLEELLRILDENAELLKRNLELLKEVLYRTR
DHD131	Heterodimer	b	GSDEDDELERLLREYHRVLREYEKLLEELRRLYEEYKRGEVSEEESDRILREIKEILDKSERLWDLSEEVWRTLLYQAE
DHD132	Heterodimer	a	GSDKKDASRRAIRVLHEFVRVSEEVLEVLRKSVESLKRLDVDEKIKRTHDRIEEELRRWKRELEELIERLREWEYHQD
DHD132	Heterodimer	b	GSDDEEEDKRLLEEVKRSLDTDERILEKLRHSLERQLEDVDKDEDSRRVLRELDEITKRSREVVKRLRKLAYESK
DHD133	Heterodimer	a	GSDKEYKLDRILRRLDELIKQLSRILEEIERLVDELEREPLDDKEVQDVIERIVELIDEHLELLKEYIKLLEEYIKTTK
DHD13	Heterodim	b	GSPSKEYQEKSAERQKELLHEYEKLVRHLRELVEKLQRRELDKEEVLRRLVEILERLKDLHKKIEDAHRKNEEAHKENK

3	er		
DHD134	Heterodimer	a	GSRDRKISEELIKALEDHIRMLEELIRAIEEHIKLAERGVDEKELRESLEELKKIVDELEKSLEELRKLAERYKYETR
DHD134	Heterodimer	b	GSPKEESVEELKRVIDKHEEILRELKRVLEEHERVSHDEDENELRRSLERLKHILDRLHESLKELHELLKKNEYTER
DHD135	Heterodimer	a	GSDHEYWVKIVERILRVMEKHAEIVKKHLEIVERVVREGPSEDLRRKLKESLREIEESLRELKELLDELDELSEKTR
DHD135	Heterodimer	b	GSDEEYVTRSQRRLKRLLEEYIKVVEEHARLVERNERDDKELKRSIDELDKLTKELLELVKRYKELVDKTET
DHD136	Heterodimer	a	GSDKEEIVKLQDEVIKTLERHLDILRKHIDLLEKLKDHLSEELKERVDRSIKKLEESIKRLERIIEELQELAEYSL
DHD136	Heterodimer	b	GSREEELKESAEELERSVRELKKEADKYKEEVDRLHYRGKVDKDWVRVVEKLIKLVEEHLELIREHLELLKEERR
DHD137	Heterodimer	a	GSDMEYELKKSAEELRKSLEELKRILDELHKSLRELRRHGDDEEYVQTVEELRKELEEHAKKLEEHLKELERVAT
DHD137	Heterodimer	b	PEYELKKSVDDLKRDVDRLVEEVEEVFELSKERLREDRKHLELVEEMVRLIEKHLELIKEHLKLADDHVR
DHD138	Heterodimer	a	GSREKDESKELNDEYKKLLEEYERLLRRSEELVKRAKGPRDEKELKRILEENEDILRRTKEILERTKEISEEQKYRRR
DHD138	Heterodimer	b	GSDKDERQERLNEESDKSNEESERSNRESEELNRRARGPNDEKELQEILDRHLELLERNQRLLDENKEILRESQYLND
DHD139	Heterodimer	a	GSENKYILKEILKLLRENLKLLHDILRLLDENLEELEKHGAKDLDDYRRKIEEIRKKVEDYREKIEEIEKKVERDR

DHD139	Heterodimer	b	GSESEYTQEEILELLKESIKLLREILRLLEESEELWRRENTKSERSEEIKERAKEAIKRSEEILERVKRLSDHSR
DHD140	Heterodimer	a	GSDEEEANYVSDKAVKIAEDVQELLKELLELSEVVRRGEVDEDEYDRVLRKLQEVMKEYEEVLKEYEEVSRKHE
DHD140	Heterodimer	b	GSPEKYLIKTQEELLRRHAEILEDLIRKVERQVDLRRKVDERDEDLKRELERSLRELERLVRESSRLVEEIRELSKEIKR
DHD141	Heterodimer	a	GSDEEYELERISRESKELLERYKRLLREYQELLKELRHVKDLDRAVKIIHELMRVSKELVEISHRLLELHERLVRRRK
DHD141	Heterodimer	b	GSEKEYIEKLSRKIEEDIRRSEERAKDSERLVRRLEELAKRKRLDLDDVLRVAEENLEILEDNLRILEEILKEQDKSNR
DHD142	Heterodimer	a	GSPHEEVVELHERVMEISERAVELIQRIIDIIRRIREDDKDIEKLVKTIRDLVREYEELHRELEEIDEEIYKKSE
DHD142	Heterodimer	b	GSDHEDVVRLHEDLVRKQEDARRVLEEIVRLAEEIVEVIKKDEKDKDRVTRLVEEIEKLVEEYKKKVDEMRKISDEIKYRSR
DHD143	Heterodimer	a	GSRAREVVKRAKRIIEEWQKILEEWRRILEEWRRLLEDERVDDRDNERIIRENERVIRENEKIIRDVIRLLEELLYERR
DHD143	Heterodimer	b	GSREDEELEEEIDRIRQMVEEYEELVKEYEELTEKYKQGKVDKEESKKIIEKSERLLDLSQDAVRKVKEIIRRILYTNR
DHD144	Heterodimer	a	GSPKEEIVKLHDESAELHRRSVEVADEILKMHERSKDVDDERESRELSKEIERLIREVEEVSKRIKRLSEEVEYLVR
DHD144	Heterodimer	b	GSPLEEILKIQRRINKIQDDINKILHEILRMQEKLNRSSDKDEVEESLRRIRELIKRIKDLSKEIEDLSREVKYRTT
DH	Heter	a	GSPEDEHVYVVREIYEVLREHAEVLEENREVIERLLEAKKRGDKSEELVKE

D145	odimer		LKKSIDKLKEISRKLEEIVKELEKVSEKLK
DHD145	Heterodimer	b	GSDEDETSYRILELLREIVRASRELIRLSEELLEVARRDDKDETVLETLIREYKELLDRYRRLIEELTRLVEEYEERSR
DHD146	Heterodimer	a	GSTQEEINRIQHEVLRIQEEIDEILRDIVEKLKAISRGELDHEVVKDVEDKVREALEKSEELLDKSRKVEYKSE
DHD146	Heterodimer	b	GSDEEELNRELLEKSKRLVDINRDIIRTAQELIEMLKDSKDGRVDEDTKRELRDKLRKLEEKLERVREELRKYEELLRYVQR
DHD147	Heterodimer	a	GSDEKDRVYEILKEVQRLVKEYRDISKEIEDLVKHYEHITDDEAQEVSKELIDKSLRASEIVRELIRLIKELLDELE
DHD147	Heterodimer	b	GSDEEDVLYHLRELLEELKRVSDDYERLVREIKETSERKDRDTKENKDMLDELVKAHREQEKLLERLVRLLEELFERKR

Table 3. Summary of biophysical data for tested designs

Design name	Expression	SEC sigle peak	Native MS	M=Multiple peaks in SEC
DHD_100	Y	Y	S	S=Strep-His double purification resulted in clean heterodimer
DHD_101	Y	Y	Y	N=SEC showing multiple peaks or Native MS failed
DHD_102	Y	Y	Y	H=homodimer of one monomer present
DHD_102_1:243	Y	M	Y	4=heterotetramer present
DHD_103	Y	Y	Y	
DHD_103_1:423	Y	M	4	
DHD_104	Y	Y	Y	
DHD_105	Y	M	N	
DHD_106	Y	Y	Y	
DHD_107	Y	M	Y	
DHD_108	Y	Y	S	
DHD_109	Y	M	Y	
DHD_110	Y	M	N	
DHD_111	Y	Y	N	
DHD_112	Y	Y	Y	
DHD_113	Y	N		
DHD_114	Y	Y	Y	
DHD_115	Y	Y	4H	
DHD_116	Y	Y	S	

DHD_117	Y	M	N
DHD_118	Y	Y	N
DHD_119	Y	Y	N
DHD_120	Y	Y	S
DHD_121	Y	Y	S
DHD_122	Y	Y	N
DHD_123	Y	Y	N
DHD_124	Y	Y	SH
DHD_125	Y	Y	N
DHD_126	Y	Y	4
DHD_127	Y	Y	4
DHD_128	Y	Y	N
DHD_129	Y	M	N
DHD_13_XAAA	Y	Y	Y
DHD_13_XAXA	Y	Y	N
DHD_13_XAAX	Y	Y	4
DHD_13_AAAA	Y	Y	Y
DHD_13_2:341	Y	Y	Y
DHD_13_BAAA	Y	Y	H
DHD_13_4:123	Y	Y	4
DHD_13_1:	Y	Y	Y

234

DHD_130	Y	M	4
DHD_131	Y	Y	4
DHD_132	Y	Y	N
DHD_133	Y	Y	Y
DHD_134	Y	Y	H
DHD_135	Y	Y	S
DHD_136	Y	Y	4
DHD_137	Y	N	
DHD_138	Y	Y	N
DHD_139	Y	Y	Y
DHD_140	Y	Y	Y
DHD_141	Y	Y	Y
DHD_142	Y	Y	SH
DHD_143	Y	Y	4
DHD_144	Y	Y	Y
DHD_145	Y	M	S
DHD_146	Y	Y	N
DHD_147	Y	Y	Y
DHD_15	Y	Y	Y
DHD_20	Y	Y	N
DHD_21	N		
DHD_25	Y	Y	Y
DHD_27	Y	N	
DHD_30	Y	Y	N

DHD_33	Y	Y	N
DHD_34_XAAXA	Y	Y	Y
DHD_34_XAXXA	Y	Y	Y
DHD_34_XAAAA	Y	Y	Y
DHD_36	N		
DHD_37_ABXB	Y	Y	Y
DHD_37_XBXB	Y	Y	Y
DHD_37_AXXB	Y	Y	4
DHD_37_3:124	Y	Y	Y
DHD_37_1:234	Y	Y	Y
DHD_37_AXBB	Y	Y	4H
DHD_37_XBBA	Y	Y	Y
DHD_37_BBBB	Y	Y	Y
DHD_39	Y	Y	Y
DHD_40	N		
DHD_43	Y	N	
DHD_65	Y	N	
DHD_70	Y	N	

DHD_88	Y	M	H
DHD_89	Y	M	N
DHD_9	Y	Y	Y
DHD_90	Y	Y	Y
DHD_91	Y	Y	4
DHD_92	Y	Y	N
DHD_93	Y	Y	H
DHD_94	Y	Y	Y
DHD_94_3:214	Y	Y	Y
DHD_94_2:143	Y	Y	Y
DHD_95	Y	N	
DHD_96	Y	M	Y
DHD_97	Y	N	
DHD_98	Y	N	
DHD_99	Y	M	Y

Table 4. SAXS data collection and analysis. R_g is the radius of gyration, R_c is the cross-sectional radius of gyration determined from Guinier fitting, and P_x is the Porod exponent. The R_c value for most designs cluster around 12 Å, in a close agreement with design models.

SAXS data collection and analysis. Rg is the radius of gyration, Rc is the cross-sectional radius of gyration determined from Guinier fitting, and Px is the Porod exponent. The Rc value for most designs cluster around 12 Å, in a close agreement with design models

Design name	I(0) (cm^{-1}) [Reciprocol space]	I(0) (cm-1) [Real space]	R_g [Reciprocol space]	R_g [Real space]	Porod volume estimate ($Å^3$)	Dmax (Å)	R_c (Å)	P_x
AAXX	154.30	141.00	20.51	20.3	42007	65	13.28	3.147
AXAX	166.00	162.80	19.19	19.73	37997	63	12.28	2.964
AXXA	148.40	145.00	19.62	19.76	39020	61	12.5	2.916
XAAX	273.80	270.70	17.73	18.6	33006	58	12.14	3.094
XAXA	210.90	210.50	18.32	18.95	35786	59	12.3	2.935
XXAA	178.60	167.60	18.88	18.98	35384	58	12.11	2.969
DHD13_2:341	203.00	186.90	21.83	20.85	40700	70	12.15	2.91
DHD13_XAXA	36680.00	34610.00	21.35	21.05	52316	66	13.42	3.172
DHD13_XAAX	32710.00	30950.00	20.52	20.32	36204	66	12.27	3.052
DHD13_1:234	206.70	194.30	20.08	19.78	36830	65	12.64	2.936
DHD13_AAAA	13.41	12.69	20.59	20.27	37810	68	12.35	3.021

DHD15	24.22	22.40	23.6	21.74	46929	67	12.47	3.012
DHD17	103.50	94.14	22.37	20.62	41958	65	11.91	2.972
DHD18	14.11	12.65	25.74	23.07	52001	68	13.06	3.113
DHD37_1:234	169.70	158.20	20.9	20.11	37935	65	12.67	2.923
DHD37_AXBB	19000.00	16000.00	26.5	22.12	49500	66	13.23	3.2
DHD37_XBXB	142.50	124.70	21.67	20.57	41001	65	12.56	2.931
DHD37_3:124	129.80	118.50	21.26	19.99	35463	64	11.98	2.985
DHD37_XBBA	8872.00	8187.00	21.5	20.79	38741	65	12.53	3.162
DHD37_ABXB	6.50	6.35	19.82	19.88	35105	65	12.28	3.125
DHD39	5.18	4.87	23.06	22.24	54736	69	13.91	3.722
DHD65	15.94	14.33	20.61	20.24	40142	65	13.05	3.1
DHD89	18.35	17.19	22.11	21.03	39345	65	12.76	3.322
DHD90	116.00	107.90	21.89	20.85	39783	64	13.01	3.114
DHD91	111.90	102.50	20.29	19.73	34402	65	11.9	3.045
DHD93	304.10	287.20	21.28	20.81	43624	65	13.75	3.095
DHD94_3:214	222.40	198.30	25.99	23.03	53133	66	14.23	3.034

DHD94_2:143	103.30	94.84	22.24	20.81	41546	64	12.63	2.983
DHD101	67.09	64.94	20.02	20.36	38014	65	12.43	3.094
DHD102	138.10	128.40	19.7	19.59	33522	65	12	3.106
DHD103_1:423	78.91	74.47	19.43	19.5	32617	65	11.87	3.115
DHD103	137.60	129.70	18.92	19.16	31650	65	11.88	3.163
DHD104	136.80	128.00	20.75	20.26	36970	65	12.65	3.08
DHD108	107.30	92.63	22.11	21.22	42377	65	12.89	3.002
DHD112	36.15	31.79	27.36	22.63	54509	65	13.08	3.006
DHD114	134.50	116.90	21.61	20.84	40893	65	12.77	2.975
DHD119	233.20	218.90	21.42	20.99	39869	65	13.13	3.121
DHD120	202.20	182.90	20.75	20.04	38997	65	12.8	3.086
DHD121	109.10	96.30	20.46	19.54	35694	61	12.14	2.983
DHD126	203.30	188.00	20.39	20.36	37377	64	12.79	3.134
DHD127	382.70	343.30	25.77	22.07	50166	66	13.18	3.156
DHD128	206.70	198.10	18.37	18.99	34571	60	12.93	3.224
DHD133	226.10	208.20	20.64	20.45	39221	65	13.2	3.142

DHD134	170.00	160.80	20.32	20.17	37044	65	12.77	3.225
DHD136	193.90	184.20	19.64	19.7	36281	64	12.8	3.185
DHD138	294.90	249.70	27.01	25.35	53449	75	13.26	2.941
DHD139	171.20	155.00	21.01	20.21	38730	65	13	3.155
DHD141	141.20	123.80	21.16	20.39	38516	65	12.95	3.072
DHD143	239.60	223.00	24.21	22.17	47863	65	13.95	3.126
DHD147	789.30	768.80	16.55	17.55	29771	55	12.56	3.146

Table 5. NMR and refinement statistics.

	DHD13_XAAA
NMR distance and dihedral constraints	
Distance constraints	**217**
Total NOE	**2076**
Intra-residue	**1402**
Inter-residue	**674**
Sequential (|*i* – *j*| = 1)	**377**
Medium-range (|*i* – *j*| < 4)	**193**
Long-range (|*i* – *j*| >= 5)	**104**
Intermolecular	**49**
Hydrogen bonds	**-**
Total dihedral angle restraints	**-**
f	**-**
y	**-**
Structure statistics	
Violations (mean and s.d.)	**0.1 ± 0.30**
Distance constraints (Å)	**5.0**
Dihedral angle constraints (º)	**-**
Max. dihedral angle violation (º)	**-**
Max. distance constraint violation (Å)	**2.6**
Deviations from idealized geometry	

Bond lengths (Å)	**-**
Bond angles (º)	**-**
Impropers (º)	**-**
Average pairwise r.m.s. deviation (Å)**	
Heavy	**1.016**
Backbone	**0.763**

Table 6. Manually assigned NMR constraints.

AmbiguousNMRDistance QD1 6 QQD 72 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HCCH.peaks
AmbiguousNMRDistance HG2 6 HE21 76 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HNCH.peaks
AmbiguousNMRDistance QQD 7 HE21 157 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HNCH.peaks
AmbiguousNMRDistance HG2 13 QQD 65 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HCCH.peaks
AmbiguousNMRDistance QB 17 QQD 149 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HCCH.peaks
AmbiguousNMRDistance QG2 27 H 51 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HNCH.peaks
AmbiguousNMRDistance QG2 27 QQD 51 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HCCH.peaks
AmbiguousNMRDistance QD1 27 H 55 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HNCH.peaks
AmbiguousNMRDistance QG2 27 H 55 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HNCH.peaks
AmbiguousNMRDistance QQD 28 H 139 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HNCH.peaks
AmbiguousNMRDistance QQD 28 HA 139 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HCCH.peaks
AmbiguousNMRDistance H 31 QE 48 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HNCH.peaks
AmbiguousNMRDistance H 32 QQD 135 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HNCH.peaks
AmbiguousNMRDistance HA 32 QQD 135 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HCCH.peaks
AmbiguousNMRDistance HG2 34 QB 44 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HCCH.peaks
AmbiguousNMRDistance QG2 35 H 128 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HNCH.peaks
AmbiguousNMRDistance QD1 35 QB 131 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HCCH.peaks
AmbiguousNMRDistance QD1 35 H 132 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HNCH.peaks
AmbiguousNMRDistance QD1 35 QQD 135 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HCCH.peaks
AmbiguousNMRDistance QE 45 QQG 118 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HCCH.peaks

AmbiguousNMRDistance H 49 QQG 118 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HNCH.peaks
AmbiguousNMRDistance QQD 49 QQG 118 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HCCH.peaks
AmbiguousNMRDistance QQD 52 H 115 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HNCH.peaks
AmbiguousNMRDistance QQD 52 QQG 118 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HCCH.peaks
AmbiguousNMRDistance H 70 QG2 152 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HNCH.peaks
AmbiguousNMRDistance QQD 73 HG 90 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HCCH.peaks
AmbiguousNMRDistance QQD 73 QQD 90 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HCCH.peaks
AmbiguousNMRDistance QQD 73 QB 93 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HCCH.peaks
AmbiguousNMRDistance QQD 73 H 156 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HNCH.peaks
AmbiguousNMRDistance HA 73 HE21 159 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HNCH.peaks
AmbiguousNMRDistance QQD 73 HE21 159 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HNCH.peaks
AmbiguousNMRDistance QQD 89 H 159 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HNCH.peaks
AmbiguousNMRDistance QQD 90 HE21 159 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HNCH.peaks
AmbiguousNMRDistance QB 93 QG2 152 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HCCH.peaks
AmbiguousNMRDistance H 93 QQD 155 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HNCH.peaks
AmbiguousNMRDistance H 96 QQD 148 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HNCH.peaks
AmbiguousNMRDistance QB 96 QQD 148 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HCCH.peaks
AmbiguousNMRDistance QB 96 H 152 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HNCH.peaks
AmbiguousNMRDistance QB 96 HD1 152 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HCCH.peaks
AmbiguousNMRDistance H 97 HD1 152 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HNCH.peaks
AmbiguousNMRDistance H 100 QQD 148 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HNCH.peaks

AmbiguousNMRDistance QD1 103 H 145 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HNCH.peaks
AmbiguousNMRDistance QB 106 QQD 141 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HCCH.peaks
AmbiguousNMRDistance H 107 QQD 141 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HNCH.peaks
AmbiguousNMRDistance QQD 110 H 138 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HNCH.peaks
AmbiguousNMRDistance HG2 117 QB 127 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HCCH.peaks
AmbiguousNMRDistance QD1 117 QB 127 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HCCH.peaks
AmbiguousNMRDistance HG2 117 H 131 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HNCH.peaks
AmbiguousNMRDistance QD1 117 H 131 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HNCH.peaks
AmbiguousNMRDistance QD1 117 QB 131 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HCCH.peaks
AmbiguousNMRDistance HE21 120 QB 127 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HNCH.peaks
AmbiguousNMRDistance H 121 QB 127 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HNCH.peaks
AmbiguousNMRDistance QA 122 HA 128 BOUNDED 1.5 5.00 0.30 automatic NOE Peak HCCH.peaks

Table 7. Results of the Rosetta calculations (10 lowest structure models for calculation based on "atom_pair_constraint" score formula). The violations are calculated over the ensemble and the total number of long- range NOE constraints.

Structure Calculation	**NOE type**	**Total no. of NOEs [1]**	**Viol (< 1 Å)**	**Viol (1-2 Å)**	**Viol (2-5 Å)**	**Viol (> 5 Å)**	**Convergence[2]**			
							FC	**RMSD AVS**	**MP RMSD**	**AVS* MP**
Rasrec without initial restraints	All assigned	71 (25)	10.30 ± 1.95	5.90 ± 1.45	8.80 ± 5.79	1.20 ± 1.17	0.67	1.95	2.74	4.12
Rasrec 46 initial csts (pseudo-atoms and upl **5.0**)	All assigned	79 (30)	16.50 ± 3.56	4.80 ± 2.18	3.00 ± 2.00	0.30 ± 0.46	1.00	0.68	1.02	1.02
Rasrec 46 initial csts (pseudo-atoms and upl **7.0**)	All assigned	79 (28)	15.9 ± 2.43	4.50 ± 2.06	4.70 ± 2.41	1.00 ± 1.34	0.87	1.33	1.92	2.19
autoNOE without restraints	All assigned	79 (24)	11.20 ± 4.02	7.40 ± 4.05	3.30 ± 1.95	0.00	0.90	1.39	2.06	2.28
autoNOE 46 initial csts (pseudo-atoms and upl **5.0**)	All assigned	81 (28)	16.70 ± 1.90	8.50 ± 3.50	4.40 ± 1.20	0.10 ± 0.30	0.98	1.10	1.64	1.68

autoNOE 46 initial csts (pseudo-atoms and upl **7.0**)	All assigned	72 (25)	16.00 ± 3.52	5.70 ± 3.38	5.40 ± 2.42	0.00	0.96	1.13	1.68	1.75
autoNOE 56 initial csts (pseudo-atoms and upl **5.0**)	All assigned	104 (31)	3.80 ± 1.89	1.89 ± 1.02	0.50 ± 0.92	0.10 ± 0.30	0.92	1.24	1.87	2.04

[1] Number of total long-range constraints at the end of the automatic NOE assignments based on the 10 lowest structures (HNCH restraints in brackets)

[2] Using a RIG file defining the structured region between the residues 3-36, 42-75, 87-120 and 125-157. The values reported are: Fraction of residues converged (FC); RMSD to average structure (RMSD AVS), mean pairwise RMSD (MP RMSD) & mean pairwise RMSD * superposed_fraction-of-atoms^{-1} (AVS*MP)

Table 8 Violations on Original Model. Violations were computed using experimental restrains listed on the second column, derived from different structure calculations (first column), and fit on the original model for DHD13_XAAA.

Structure Calculation	**Total no. of NOEs[1]**	**Viol (< 1 Å)**	**Viol (1-2 Å)**	**Viol (2-5 Å)**	**Viol (> 5 Å)**
Rasrec without initial restraints	71 (25)	10	9	4	1
Rasrec 46 initial csts (pseudo-atoms and upl **5.0**)	79 (30)	10	8	13	0
Rasrec 46 initial csts (pseudo-atoms and upl **7.0**)	79 (28)	10	13	8	0
autoNOE without restraints	79 (24)	9	7	10	1
autoNOE 46 initial csts (pseudo-atoms and upl **5.0**)	81 (28)	10	9	10	2
autoNOE 46 initial csts (pseudo-atoms and upl **7.0**)	72 (25)	7	9	7	2
autoNOE 56 initial csts (pseudo-atoms and upl **5.0**)	104 (31)	8	4	3	1

[1] Long-range NOE restrains from both peak lists (HNCH restraints) at the end of each calculation.

Table 9. Data collection and refinement statistics (molecular replacement)

	DHD127	DHD131	DHD1:234	DHD15_extended	DHD15_closed
Data collection					
Space group	P 21 21 21	P 1 21 1	P 64	P 21 21 21	P 32
Cell dimensions					
***a*, *b*, *c* (Å)**	44.918, 48.755, 64.152	39.199,74.709,105.257	96.885, 96.885, 27.846	27.854, 92.739, 115.333	57.615, 57.615, 130.861
a, b, g (°)	90, 90, 90	90, 92.747, 90	90, 90, 120	90, 90, 90	90, 90, 120
Resolution (Å)	36.8 - 1.753 (1.816 - 1.753)	42.99 - 2.38 (2.465 - 2.38)	48.44 - 3.261 (3.378 - 3.261)	43.02 - 2.25 (2.331 - 2.25)	39.68 - 3.363 (3.483 - 3.363)
R_{sym} or R_{merge}	0.123 (0.734)	0.1795 (1.514)	0.1247 (1.765)	0.05055 (0.5611)	0.1496 (1.389)
I* / s*I	18.06 (2.16)	7.26 (1.01)	11.09 (1.51)	22.79 (3.95)	7.79 (1.05)
Completeness (%)	98.94 (97.56)	87.19 (73.42)	99.72 (100.00)	94.59 (78.26)	98.12 (94.81)

Redundancy	**7.7 (7.7)**	**4.1 (4.2)**	**7.6 (8.0)**	**6.9 (7.0)**	**6.5 (6.4)**
Refinement					
Resolution (Å)	**36.8 - 1.753 (1.816 - 1.753)**	**42.99 - 2.38 (2.465 - 2.38)**	**48.44 - 3.261 (3.378 - 3.261)**	**43.02 - 2.25 (2.331 - 2.25)**	**39.68 - 3.363 (3.483 - 3.363)**
No. reflections	**14631 (1397)**	**24450 (1798)**	**2481 (246)**	**14920 (1159)**	**6782 (639)**
R_{work} / R_{free}	**0.1968 (0.2715)/0.2085 (0.3368)**	**0.2408 (0.3434)/0.2832 (0.4433)**	**0.2583 (0.3908)/0.2855 (0.4405)**	**0.2609 (0.3359)/0.3097 (0.3961)**	**0.2923 (0.4283)/0.3332 (0.4203)**
No. atoms					
Protein	**1283**	**4622**	**932**	**2446**	**3360**
Ligand/ion	**0**	**0**	**0**	**5**	**0**
Water	**149**	**63**	**0**	**30**	**0**
***B*-factors**					
Protein	**26.05**	**44.55**	**103.82**	**56.01**	**110.95**
Ligand/ion	**N/A**	**N/A**	**N/A**	**102.87**	**N/A**
Water	**31.89**	**35.32**	**N/A**	**48.77**	**N/A**

R.m.s. deviations					
	0.011	**0.002**	**0.003**	**0.002**	**0.002**
Bond lengths (Å)					
	1.16	**0.44**	**0.64**	**0.41**	**0.37**
Bond angles (°)					

Table 10 OD 600 values for yeast culture in Y2H assays. Yeast cells were grown for 2 days at 30°C before OD 600 values were taken. a_AD-b_DBD, monomer a was fused to activation domain (AD), monomer b was fused to DNA binding domain (DBD); a_DBD-b_AD, monomer a was fused to DBD, monomer b was fused to AD; a_AD-a_DBD, monomer a was fused to AD or DBD; b_AD-b_DBD, monomer b was fused to AD or DBD.

0 mM 3-AT

Design	A_AD-B-DBD	A_DBD-B-AD	A_AD-A-DBD	B_AD-B-DBD
DHD1	0.407	0.36	0.081	0.337
DHD2	1.499	0.256	1.505	0.345
DHD3	0.445	0.452	0.373	0.451
DHD4	1.701	1.225	0.326	1.719
DHD5	0.388	0.34	0.294	0.08
DHD6	1.683	0.308	0.265	0.394
DHD7	1.629	1.6	1.574	1.507
DHD8	0.181	0.242	0.231	1.298
DHD13	0.908	1.491	0.391	0.303
DHD15	1.286	0.858	0.259	0.395
DHD16	1.078	0.413	0.618	0.297
DHD18	1.408	1.578	1.093	1.731
DHD19	0.506	0.373	0.953	0.316
DHD20	1.318	1.221	0.077	0.427
DHD21	0.753	1.579	1.13	1.298
DHD22	1.614	1.683	1.571	0.092
DHD23	0.289	0.217	0.267	1.109
DHD24	1.067	0.231	1.511	0.163
DHD25	1.409	1.263	1.531	0.401
DHD26	0.257	0.376	1.275	1.443

DHD27	0.9	0.955	0.332	0.321
DHD28	0.077	1.435	0.077	1.502
DHD29	1.658	1.25	0.103	1.021
DHD30	1.786	1.782	1.843	1.314
DHD31	0.276	0.192	0.793	0.291
DHD32	1.137	0.695	1.566	0.365
DHD33	1.641	1.604	1.638	1.624
DHD34	1.629	1.741	1.633	0.353
DHD36	1.722	1.84	1.896	1.73
DHD37	0.538	0.654	0.219	0.254
DHD38	0.32	1.577	1.513	0.174
DHD39	1.413	0.546	0.27	0.292
DHD40	1.374	1	1.334	0.284
DHD60	1.618	0.422	0.454	1.417
DHD63	0.396	0.354	0.414	0.412
DHD65	1.731	1.633	0.291	0.185
DHD66	1.363	1.442	1.482	0.255
DHD67	0.163	0.101	1.28	0.081
DHD69	1.15	0.078	1.383	0.079
DHD70	1.565	0.283	0.275	0.314
DHD71	0.497	0.471	0.661	0.395
DHD72	1.439	0.759	1.432	1.25
DHD73	0.084	0.126	0.081	0.08
100 mM 3-AT				
DHD101	0.29	1.18	1.245	0.819

DHD104	1.273	1.099	1.216	0.217
DHD106	1.11	1.168	1.294	0.204
DHD107	1.137	1.019	0.584	0.218
DHD112	0.899	0.84	0.225	0.206
DHD126	0.399	1.094	0.211	1.01
DHD139	0.302	0.836	0.96	0.583
DHD99	1.084	0.976	0.944	0.291
DHD102	0.353	0.46	0.198	0.838
DHD103	0.823	1.009	0.215	0.209
DHD127	1.022	1	0.215	0.824
DHD131	0.879	1.092	1.072	0.211
DHD148	0.601	0.407	0.442	0.277
DHD149	0.218	0.946	0.288	0.24
DHD150	1.074	1.017	0.327	0.608
DHD151	0.972	1.06	0.207	1.113
DHD152	1.08	0.233	0.718	1.08
DHD153	1.082	0.652	0.223	0.387
DHD154	0.934	1.094	0.205	0.201
DHD155	0.695	1.104	0.22	0.236
DHD156	0.211	0.224	0.201	0.23
DHD157	1.011	1.109	0.381	0.211
DHD37_XBXB	0.333	0.276	0.204	0.23
DHD37_BBBB	0.978	1.113	0.216	0.598
DHD_131	0.584	0.678	0.669	0.196
DHD_158	0.241	0.287	0.192	0.139
DHD_159	0.341	0.887	0.128	0.708

DHD_160	0.114	0.391	0.29	0.177
DHD_161	0.764	0.742	0.219	0.195
DHD_162	0.774	0.516	0.203	0.165
DHD_163	0.191	0.203	0.19	0.105
DHD_164	0.178	0.114	0.267	0.179
DHD_165	0.616	0.323	0.164	0.113
DHD_166	0.587	0.197	0.52	0.093

Table 11 Instrument settings for EMR Orbitrap Mass Spectrometer

	MS	HCD 100V	SID 85V
Source temperature/°C	250		
Spray voltage/kV	3.7		
Sheath gas	50		
Sweep gas	0		
AGC target	500,000		
AGC mode	prescan		
HCD/V	0	100	0
Source Fragmentation/V	0	0	0
Max injection time/ms	200	200	400
Resolution	8,750	140,000	140,000
Trap gas	4	4	4
Source DC offset/V	15	15	7
Injection Flatapole DC/V	14	14	7
Inter Flatapole Lens/V	12	12	6
Bent Flatapole DC/V	10	10	5
C-Trap Entrance Lens- Inject/V	-10	-10	-69
C-Trap Exit Lens- Inject/V	20	20	-40
C-Trap offset/V	19	19	-79

Table 12 Deconvolution settings for UniDec and Intact Mass

UniDec		Intact mass	
m/z range/Th	1000 to 5000	Mass range/Da	3000 to 32000
Background subtraction	0	m/z range/Th	1000 to 5000
Bin every	1	Minimum mass difference/Da	5
Charge range	2 to 20	Maximum peaks	20
Mass range	1000 to 40000	Vectors spacing	0.5
Sample mass every/Da	1	Baseline radius	15
Peak FWHM	1	Smoothing sigma	0.02
Peak shape function	Gaussian	Spacing	0.04
		Mass smoothing	2
		Mass spacing	0.5
		Interation max	10
		Charge	2 to 20
		Advanced command[1]	[Intact] UseEvenExcess=1

[1] This command is used to distinguish monomer and homodimer which overlap at the same m/z.

Table 13 SID parameters on SYNAPT G2

Voltage/V	Transfer Mode	SID 80V	SID 100V
Trap bias	45	140	160
Entrance 1	-64	49	69
Entrance 2	-49	-46.5	-46.5
Front top	-53	-80	-80
Front bottom	-57	19	19
Middle bottom	-56	-46.3	-46.3
Surface	-54	-30	-30
Back top	-57	-170.6	-170.6
Back bottom	-58	-55	-55
Exit 1	-58	-77.2	-77.2
Exit 2	-54	-54.3	-54.3

Chapter 2: Self-assembling 2D arrays with de novo protein building blocks

Abstract

Modular self-assembly of biomolecules in two dimensions (2D) is straightforward with DNA but has been difficult to realize with proteins, due to the lack of modular specificity provided by Watson-Crick base pairing. Here we describe a general approach to design 2D arrays using *de novo* designed pseudosymmetric protein building blocks. A homodimeric helical bundle was reconnected into a single chain building block, and the surface was redesigned in Rosetta to enable self-assembly into a 2D assembly in the C 1 2 layer symmetry group. The designed arrays assembled to μm scale under negative stain electron microscopy (EM), and displayed the designed lattice geometry. The design of 2D arrays with pseudosymmetric building blocks is an important step toward the design of programmable protein self-assembly via pseudosymmetric patterning of orthogonal binding interfaces.

Introduction

Programmatic molecular self-assembly -- the ability of complex assemblies to be built from simple building blocks -- is ubiquitous in Nature. In particular, naturally occurring protein assemblies combine symmetry with pseudosymmetry (67), use conformational flexibility to break symmetry (68), or bind scaffolding molecules such as nucleic acids (69) to construct large but bounded assemblies from simple protein building blocks. Mimicking native assemblies, we aim to design a modular scheme for programmatically controlling protein assemblies. Due to the simplicity of design, we focus on the use of pseudosymmetry -- through sequence variations on a fixed backbone -- as a preferred route of engineering programmable protein self-assemblies.

Several prior efforts have attempted to design such assemblies. Drawing off the specificity of Watson-Crick base pairing, a series of methods have been developed to accurately pattern DNA in two dimensions (40, 70, 71). For many applications, protein is a more attractive building material due to its inherent bio-reactivity. Previous efforts in making patterned 2D materials with native proteins have resulted in some success (72–75). These previous efforts all made use of native proteins as building blocks, which is practically limiting: the geometry and overall symmetry of the final assembly is limited by the native oligomeric state and symmetry of the native building block. *De novo* protein building blocks can be custom designed to have the desired symmetry (25, 76) and high stability (27), properties that should enable more robust building blocks for designing 2D self-assembling materials.

Here we describe a general approach for generating pseudosymmetric 2D assemblies based on a C 1 2 symmetric layer group. Starting from a *de novo* designed homodimer, we first design a new loop to monomerize the backbone of our building block, then identify configurations of this backbone capable of forming 2D arrays with pseudo[1]-C 1 2 symmetry, and finally redesign the interface so that the building block will be programmatically assembled into 2D arrays with the prescribed unit cell dimensions and subunit configuration. This monomerization of the multimeric protein building block allows unique sequences to be designed on each of the 4 binding interfaces, ultimately enabling the modular assembly of higher order interactions through the design of mutually orthogonal interfaces with the same subunit placement and unit cell dimensions (Fig. 1 A, B). This study experimentally characterizes a key step towards this overall design goal, which will enable various applications including patterned enzymatic reactions.

[1] The resulting layer group symmetry is pseudo-C 1 2 because each building block has pseudo-C2 symmetry due to the presence of an additional loop.

Results

Building off Gonen *et al.* (72), we developed a general strategy for the design of pseudosymmetric 2D protein assemblies using *de novo* designed proteins as building blocks, fully described in *Methods*. Figure 1 presents a high-level overview of the approach. Briefly, a previously designed helical bundle homodimer 2L4HC2_23 (PDB ID 5J0K (25), Fig. 1C) was connected into a single chain via a designed loop, resulting in a pseudo-C2 symmetric building block (SC_2L4HC2_23, Fig. 1D). We solved the X-ray crystal structure of the building block, revealing a backbone nearly identical to the design model and the original dimer structure (PDB ID 5J0K), with a Cα root mean square deviation (RMSD) of 1.08 Å between the design and crystal structure (Fig. 1H).

Using this monomerized building block as a starting point for pseudosymmetric assembly, we subsequently enumerated all possible pseudo-C 1 2 symmetric layer assemblies compatible with this design, exhaustively sampling three degrees of freedom: two parameters describing the lattice dimensions, and one parameter controlling rotation of the building block around its central axis (Fig. 1E). We sampled 576,000 settings of these three parameters, and removed those which were not capable of forming a connected, non-clashing 2D assembly. The remaining designs had their surfaces redesigned to self-assemble into the corresponding lattice arrangement. Two different design methodologies were explored: standard Rosetta fixed backbone design (29), as well as the Rosetta HBNet algorithm (25) to design buried hydrogen bonds at the interface; the latter potentially enables greater specificity at the binding interfaces. Using computationally predicted interface energies as well as visual inspection, ten designs were selected for experimental characterization.

Examination by negative-stain EM revealed regular arrays on the μm scale for two of the ten designs: one with exclusively hydrophobic packing across the designed interface (2D-HP, Fig. 2 A, B); and one with a designed hydrogen bond network at the new interface (2D-HBN, Fig. 2 D, E). The 2D assemblies of 2D-HBN (Fig. 2F) appeared to be more extended and regular compared to that of 2D-HP (Fig. 2C), likely due to better binding specificity conferred by hydrogen bond networks. As a result of the small interfaces, in both cases, the resulting 2D assemblies exhibited some flexibility under negative stain EM. To rule out the possibility of domain swapping from the single chain building block contributing to the final assembly, we expressed the building block protein of 2D-HBN as individual homodimers of helix hairpins, which similarly assembled into 2D arrays of the same morphology under the same condition (Supp Fig. S1). Finally, to verify that the array was forming a regular 2D grid, we collected a larger negative stain (NanoW) dataset of the best-behaved arrays (Supp Fig. S2). Subsequent 2D classification and averaging of 1,893 boxed ~20 nm regions yielded an image showing an ordered two-dimensional assembly (Fig. 2G) with a power spectrum indicating first-order spots (Fig. 2G, inset).

Surprisingly, while the resulting images was consistent with a C 1 2-symmetric complex, the unit-cell dimensions were somewhat different than designed: while the design had a 66 by 45 Å unit cell, the experimental images indicated approximately a 68 by 22 Å unit cell. The unit cell dimension of 68 Å agrees well with the designed dimension of 66 Å, however the observed 22 Å is half the designed 45 Å. Docking our protein building block into a C 1 2 lattice with the observed 68 and 22 Å dimensions is unable to identify a non-clashing orientation (Supp. Fig. S3); given that the crystal structure of the building block has a diameter of 23 Å, the observed 22 Å lattice repeating unit is unlikely to be a result of the assembly.

Given the inability of the building block to dock in the observeed unit cell, coupled with the observation that both monomeric and dimeric building blocks give rise to the same assembly, we believe the observed 22 Å value most likely arises from the averaging of multiple orientations into the final image, either by: a) ambiguity in individual images giving rise to systematic misalignments during averaging, or b) regular stacking of assembly layers on the EM grids. As evidence, we performed a computer simulation of two misaligned arrays of our designed assembly. Figure 2H-I shows how this might lead to the observed assembly: the sum of the two patterns indicated in Figure 2H, which consist of an image misaligned (or stacking) onto an offset of itself, results in an average image quite similar to the observed one (Fig. 2I). A Fourier transform of this composite density similarly shows very close agreement to the observed data (Fig 2G, 2I, insets), with spots at most one pixel away. Therefore we believe the arrays are forming as designed, and the observed 22 Å repeating distance is likely due to the stacking of multiple configurations in experimental characterization.

Discussion

We showed that by systematically sampling lattice dimensions followed by computational interface design, the same *de novo* designed helical bundle building block can be modularly self-assembled into two arrays with unique cell dimensions. Limited by available *de novo* designed protein building blocks, the designed interfaces are much smaller than those in previously designed 2D protein assemblies, resulting in a high degree of structural flexibility. Future work could address this in two ways: 1) using bigger *de novo* building blocks, more rigid arrays should be possible to realize with bigger binding interfaces; 2) designing in non-polar layer groups, which have a rotation about the layer plane (e.g., P 3 2 1 and P 4 2_1 2), effectively canceling out any “curvature” errors in binding along the z axis, further flattening out the 2D assembly. As more *de novo* building blocks are designed, particularly with higher-order symmetry, a variety of 2D assemblies with unique layer group symmetries should be achievable with the same design protocol.

The monomerization of the homodimer building block coupled with designed hydrogen bond networks should allow orthogonal interfaces to be designed at each intermolecular binding site, paving the way for the programmatic self-assembly of proteins into finite shapes (Fig. 1 B), which requires the design of multiple such interfaces on a single configuration of a building

block. Such interfaces can be applied modularly, by plugging designed sequence on to the corresponding helical bundle. Our work represents a key step toward this goal, and we show that *de novo* designed proteins can serve as building blocks for 2D assemblies.

Main Text Figures

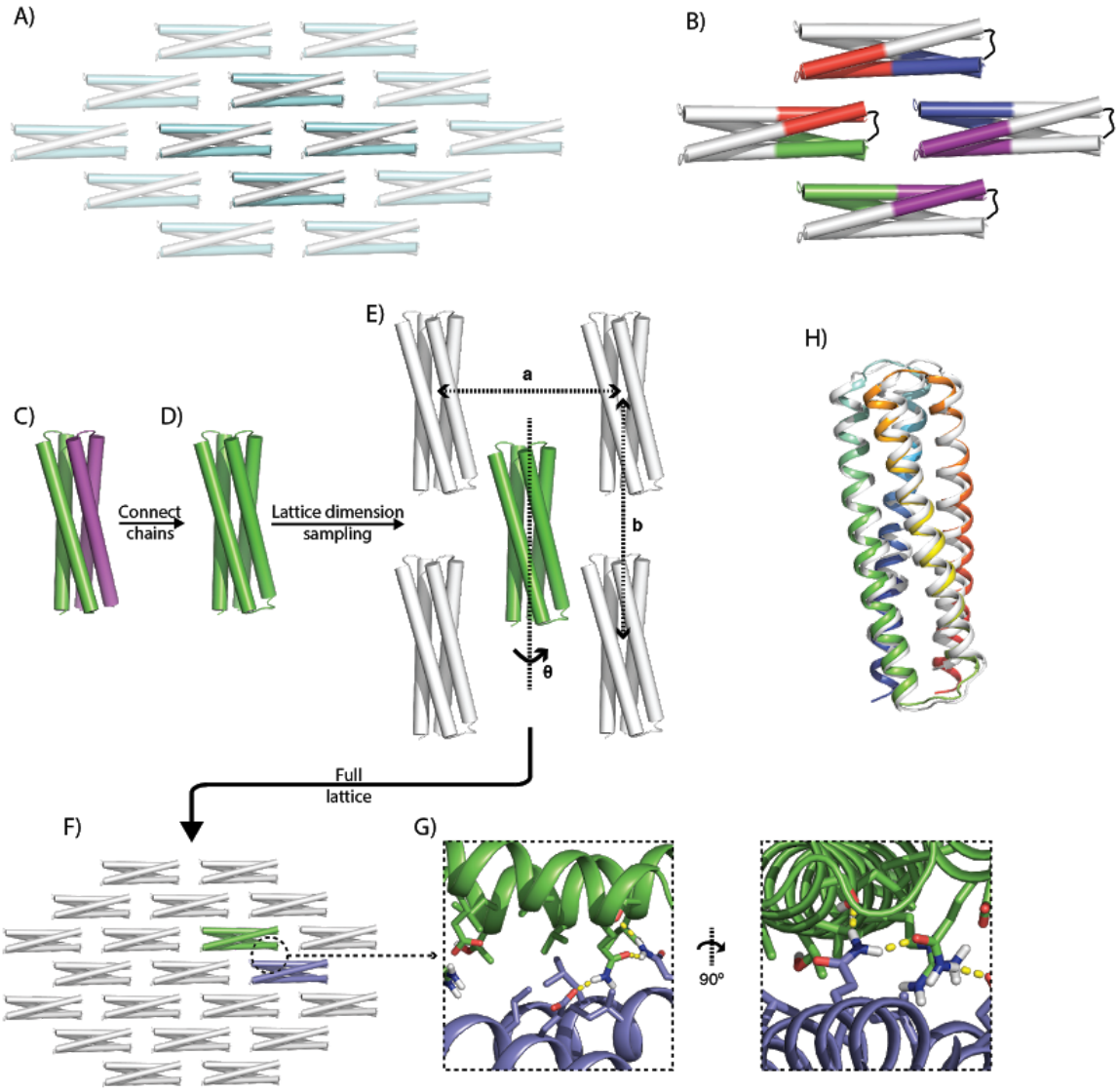


Figure 1. Overview of the design concept. (A) 2D self-assembly using homodimers as building blocks. Designed inter-building block binding interfaces are highlighted in cyan. In this scenario the assembly process will result in an infinite 2D lattice. (B) By using designed loops to monomerize the building block, and modularly mixing orthogonal interfaces (different colors), we enable programmatic assembly design (in this case, a heterotetramer). (C) - (G), Overview of the design process. A *de novo* designed homodimer (C) is connected into a single chain (D), and docked in a C 1 2 layer group symmetry with three parameters a, b, and θ (E), resulting in a 2D lattice (F). Binding interfaces are designed with hydrogen bond networks to confer specificity

(G). (H) A 1.74 Å resolution crystal structure of the design SC_2L4HC2_23 (PDB ID 6EGC, white) superimposed onto the design model (rainbow); the design model deviates from the crystal by 1.08 Å RMSD.

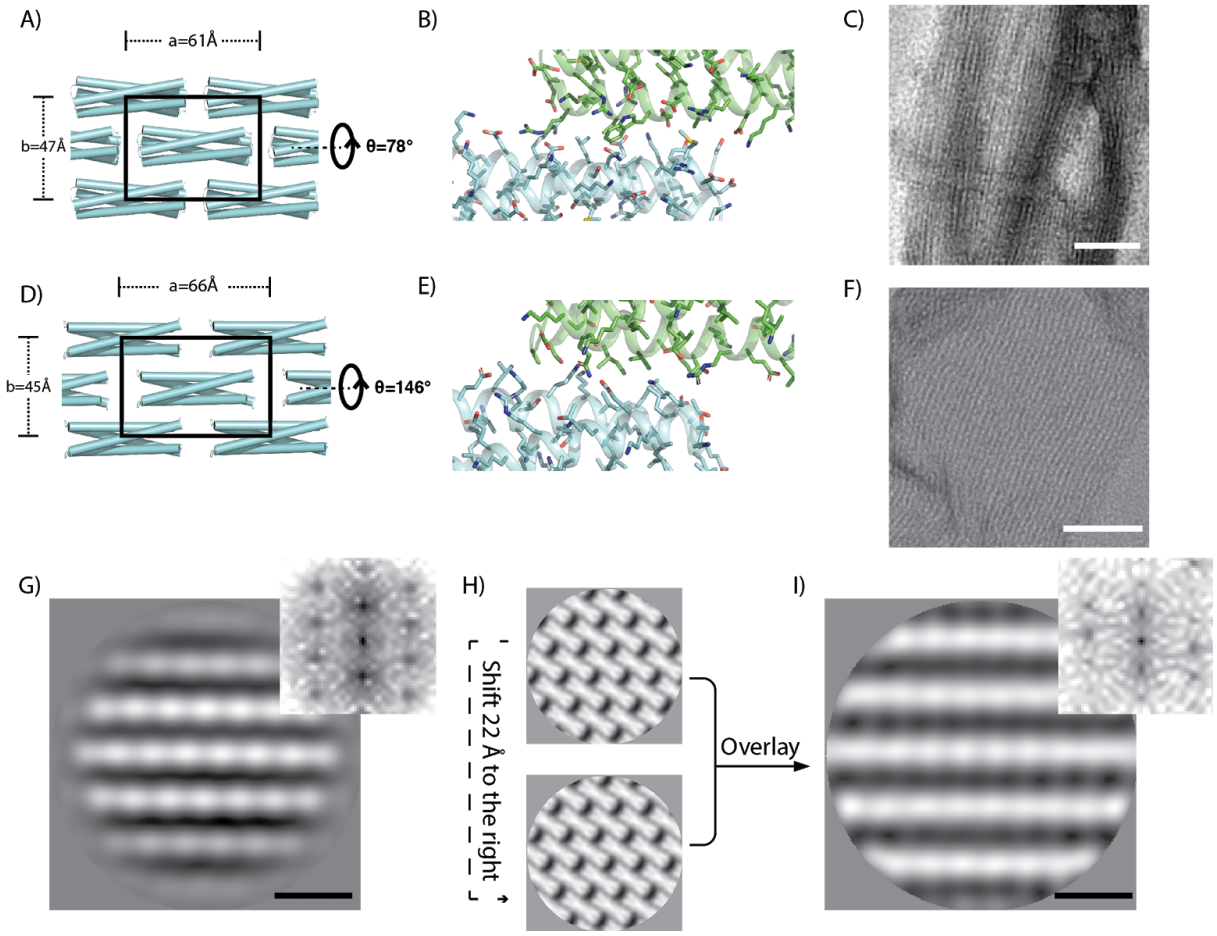


Figure 2. Negative stain electron microscopy analysis of designed 2D assemblies. (A) Lattice design of 2D-HP, with the black box showing unit cell. (B) Designed interface of 2D-HP with exclusive hydrophobic packing across the interface. (C) Negatively stained array of 2D-HP. (D) Lattice design of 2D-HBN, and its designed interface with a hydrogen bond network (E). (F) Negatively stained array of 2D-HBN showing an extensive and flexible 2D assembly. (G) 2D class average of 2D-HBN assembly with homodimer building blocks, inset shows its Fourier transform. (H) Design model placed in the assumed orientation of the class average. When shifted 22 Å to the right and superimposed onto the original image, it gives a simulated 2D class average in (I), and its Fourier transform in the inset, both in close agreement with the observed data in (G). All scale bars: black, 5 nm; white, 50 nm.

Program Code

All of the computational tools described in the manuscript are available with Rosetta, and can be downloaded at rosettacommons.org. Rosetta design xmls are attached as supplemental files. Command line options for the design calculations can be found below.

Step 1. Rapid generation of connecting and non-clashing 2D lattices from protein building blocks

~/Rosetta/main/source/bin/flatland.static.linuxgccrelease
-in:file:s [input pdb model]
-database [path to Rosetta database]
-ignore_unrecognized_res
-mh:path:scores_BB_BB
/gscratch/baker/zibochen/utilities/aa_count_ACDEFHIKLMNQRSTVWY_resl1_ang15_msc0.2_smooth1.3_ROSETTA/aa_count_ACDEFHIKLMNQRSTVWY_resl1_ang15_msc0.2_smooth1.3_ROSETTA -mh:score:use_ss1 true -mh:score:use_ss2 true -mh:score:use_aa1 false -mh:score:use_aa2 false #motif score specific options
-symmetry_definition dummy
-output_virtual
-tag [user defined name tag for the job]
-rot_step [search step size for the self-rotation of the building block, takes a real number]
-Cn [internal cyclic symmetry of the building block, 2]
-wallpaper [layer symmetry of the final 2D lattice, C211]
-dump_silent [dump a silent file containing all the lattices, boolean]
-C211_B [lattice parameter B for the C 1 2 layer group, takes a real number]
-cell_upper [upper limit for the cell dimensions, takes a real number]
-single_chain_version [if the input model is monomerized, the code accomondates for this psudeo-symmetry. Boolean]

-cell_step [search step size for the lattice cell dimensions, takes a real number]

Step 2. HBNet search at the interfaces of extracted adjacent building blocks

~/Rosetta/main/source/bin/rosetta_scripts.static.linuxgccrelease
-in:file:s [input pdb model]
-out::file::pdb_comments
-run:preserve_header
-use_input_sc
-out:prefix HBNet_
-beta
-missing_density_to_jump true
-parser:protocol 2D_HBNet.xml
-database [path to Rosetta database]
-chemical:exclude_patches LowerDNA UpperDNA Cterm_amidation VirtualBB ShoveBB VirtualDNAPhosphate VirtualNTerm CTermConnect sc_orbitals pro_hydroxylated_case1 pro_hydroxylated_case2 ser_phosphorylated thr_phosphorylated tyr_phos phorylated tyr_sulfated lys_dimethylated lys_monomethylated lys_trimethylated lys_acetylated glu_carboxylated cys_acetylated tyr_diiodinated N_acetylated C_methylamidated MethylatedProteinCterm
-in:file:fullatom
-multi_cool_annealer 10
-no_optH false
-optH_MCA true
-flip_HNQ

Step 3. Regenerate the complete 2D lattice and map newly designed interfaces to all symmetric copies

~/Rosetta/main/source/bin/symm_seq_gen_2D.default.linuxgccrelease
-database [path to Rosetta database]
-s [input pdb model]
-cn [symmetry of the building block, 2]

Step 4. Symmetric design of the 2D lattice in the context of its symmetry

~/Rosetta/main/source/bin/symm_seq_gen_2D.default.linuxgccrelease
-database [path to Rosetta database] \ -in:file:silent [input Rosetta silent file containing the 2D lattice]
-parser:script_vars resfile=[input resfile to enfore newly designed interfaces stay intact]
-out::file::pdb_comments
-run:preserve_header
-multi_cool_annealer 10
-use_input_sc
-symmetry_definition dummy
-out:prefix packed_
-beta -missing_density_to_jump true
-symmetry:detect_bonds false
-parser:protocol 2D_final_design.xml

Data Availability

Coordinates and structure files have been deposited to the Protein Data Bank with accession code: 6EGC (SC_2L4HC2_23).

Supplemental Materials

Computational design methods

1. Connecting the homodimer into monomer

The two monomers from the homodimer 2L4HC2_23 are connected into a single chain monomer with a 5-residue loop using the method described previously (25). Briefly, a database of backbone samples composed of fragments spanning two helical regions via a loop of five or less residues was generated from high resolution crystallographic structures. Loops in this database were then structurally aligned to terminal residues of the design backbone, and those that aligned within 0.35 Å RMSD were carried forward with full Rosetta design restricted to the loop and its neighborhood residues within 6 Å. The lowest-scoring candidate selected as the final loop design.

2. Systematic sampling of lattice parameters

A custom Rosetta protocol was developed to dock the building block into pseudo-C 1 2 layer group and systematically sample the three parameters that control lattice geometry: two parameters describing the lattice dimensions, and one parameter controlling rotation of the building block around its central axis (Fig. 1E). Taking into account the dimension of the building block, lattice parameter "*a"* was sampled from 60 Å to 100 Å, with a step size of 0.5 Å; lattice parameter "*b"* was sampled from 30 Å to 50 Å, with a step size of 0.5 Å; rotation of the building block around its central axis, θ, was sampled from 0° to 180° with a step size of 1°, resulting in 576,000 possible docked conformations. A rapid evaluation protocol in Rosetta was applied to remove lattices that have either clashes of building blocks or inter building block distance greater than 10 Å, resulting in 4,139 candidate lattices for further design. Two adjacent building blocks were extracted from the lattice for interface design calculations.

3. Design calculations

RosettaDesign (29) calculations were carried out on the interfaces between adjacent building blocks, while keeping the rest of the sequences fixed. To enhance the binding specificity among subunits, we optionally used the Rosetta HBNet algorithm (25) to design buried hydrogen bond networks at the interface between subunits, selecting for networks that involve at least 3 side chain residues with all heavy-atom donors and acceptors participating in hydrogen bonds. Low energy sequences were identified using RosettaDesign calculations in which the hydrogen bond networks were held fixed. A final step of minimization and side chain repacking without atom pair constraints was applied to identify the movement of HBNets, filtering out designs with significantly moved HBNets. The complete 2D lattice was then regenerated using the adjacent building blocks (now with designed interfaces), with the newly designed sequences applied to all building blocks. A final round of Rosetta design was carried out in the context of the C 1 2 layer group symmetry with the newly designed sequences fixed, to resolve potential side chain clashes in the final lattice.

4. Selection criteria and metrics used to evaluate designs

Fully designed models were selected based on the shape-complementarity of the designed interface (SC > 0.6), size of the designed interfaces (dSASA > 500 Å), average binding energy (ddG/dSASA < -0.02 Rosetta Energy Unit/Å^2) and no buried unsatisfied hydrogen bonds introduced at the new interfaces. Selected designs were then visually inspected for good packing of hydrophobic side chains at the interfaces.

Visualization and figures

All structural images for figures were generated using PyMOL (46).

Experimental methods

Buffer and media recipe

TBM-5052
1.2% [wt/vol] tryptone, 2.4% [wt/vol] yeast extract, 0.5% [wt/vol] glycerol, 0.05% [wt/vol] D-glucose, 0.2% [wt/vol] D-lactose, 25 mM Na2HPO4, 25 mM KH2PO4, 50 mM NH4Cl, 5 mM Na2SO4, 2 mM MgSO4, 10 µM FeCl3, 4 µM CaCl2, 2 µM MnCl2, 2 µM ZnSO4, 400 nM CoCl2, 400 nM NiCl2, 400 nM CuCl2, 400 nM Na2MoO4, 400 nM Na2SeO3, 400 nM H3BO3

TBS buffer
20 mM Tris pH 8.0, 100 mM NaCl

Construction of synthetic genes

Synthetic genes were ordered from Genscript Inc. (Piscataway, N.J., USA) and delivered in pET28b(+) *E. coli* expression vector, inserted between the NdeI and XhoI sites.

Protein expression

Plasmids were transformed into chemically competent *E. coli* expression strains BL21(DE3)Star (Invitrogen) for protein expression. Single colonies were picked from agar plates following transformation and growth overnight, and 5 ml starter cultures were grown at 37°C in Luria-Bertani (LB) medium containing 100 µg/mL kanamycin with shaking at 225 rpm for 18 hours at 37°C. Starter cultures were diluted into 500 ml TBM-5052 containing 100 µg/mL kanamycin, and incubated with shaking at 225 rpm for 24 hours at 37°C.

Protein purification

Cells were harvested by centrifugation for 15 minutes at 5000 rcf at 4°C and resuspended in 20 ml lysis buffer. Lysozyme, DNAse, and EDTA-free cocktail protease inhibitor (Roche) were added to the resuspended cell pellet before sonication at 70% power for 5 minutes. All 10 designs expressed and precipitated into cell pellet after clearing the cell lysate at 12,000g for 1 hour. Pellets were twice resuspended in 10 ml TBS followed by centrifugation at 12,000g for 20

min. The resulting pellet was resuspended in 1 M GdmHCl followed by centrifugation at 12,000g for 20 min. The supernatant was dialyzed overnight into TBS buffer.

X-ray crystallography and structure determination

Crystals of SC_2L4HC2_23 were grown by mixing 0.1ul of protein at 20 mg/ml plus 0.1ul of crystallization condition Morpheus H9 (Molecular Dimensions, 0.1M Amino acids, 0.1M Buffer System 3 pH 8.5, 50% (v/v) Precipitant Mix 1). As this solution is already a suitable cryoprotectant, crystals were flash-frozen directly in liquid nitrogen prior to data collection. Diffraction data was collected at the Advanced Light Source, Lawrence Berkeley National Laboratory, beamline 8.2.1. Diffraction data was indexed and scaled using HKL2000 (48). Initial models were generated by the molecular-replacement method using the program PHASER (49) within the Phenix software suite (50), with the computational design serving as the search model. Efforts were made to reduce model bias by using simulated annealing and prime-and-switch phasing within Phenix.autobuild (52). Iterative rounds of manual building in COOT (53) and refinement in Phenix were used to produce the final model. Due to the high degree of self-similarity inherit in coiled-coil-like proteins, datasets for the reported structures suffered from a high degree of pseudo translational non-crystallographic symmetry, as report by Phenix.Xtriage, which complicated structure refinement and may explain the higher than expected R values reported. RMSDs of bond lengths, angles and dihedrals from ideal geometries were calculated with Phenix (50). The overall quality of all final models was assessed using the program MOLPROBITY (54). Summaries of diffraction data and refinement statistics are provided in Supplementary Table 2.

Negative stain EM

Samples were applied to glow-discharged EM grids and stained with either uranyl formate (UF) or NanoW (Nanoprobes, Inc, Yaphank, NY, USA) for screening or analysis, respectively. Data was collected using a Tecnai T12 equipped with a Gatan Orius CCD. CTF estimation was performed using GCTF (77), and all other image processing steps were completed via Relion 2.1 (78). For the analysis in Figure 2G, 421 2D array segments were picked manually from 42 micrographs, and the resulting 2D class average used as a template for Relion autopicking, which yielded 5,823 2D array segments. After subsequent 2D classification and alignment, the dominant 2D class consisted of 1,893 array segments (each approximately 20 nm diameter).

Supp. Figure 1. Compare single chain and homodimer lattices

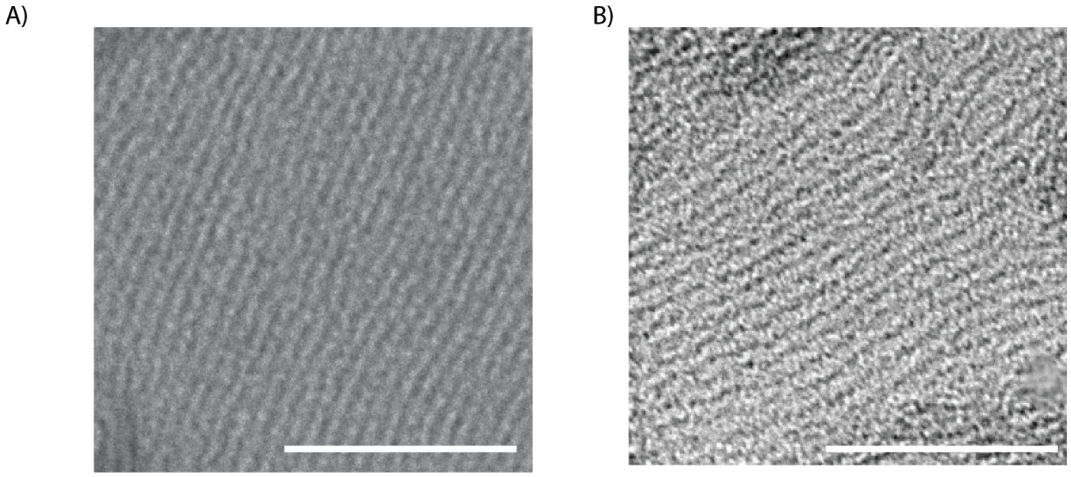


Figure S1. Comparison of 2D-HBN lattices using the monomer SC_2L4HC2_23 (A) or the homodimer 2L4HC2_23 (B) as building blocks. Negative stain EM shows similar patterns of 2D lattice formation in both cases. Scale bar: 50 nm.

Supp. Figure 2. Image to do 2D class average on

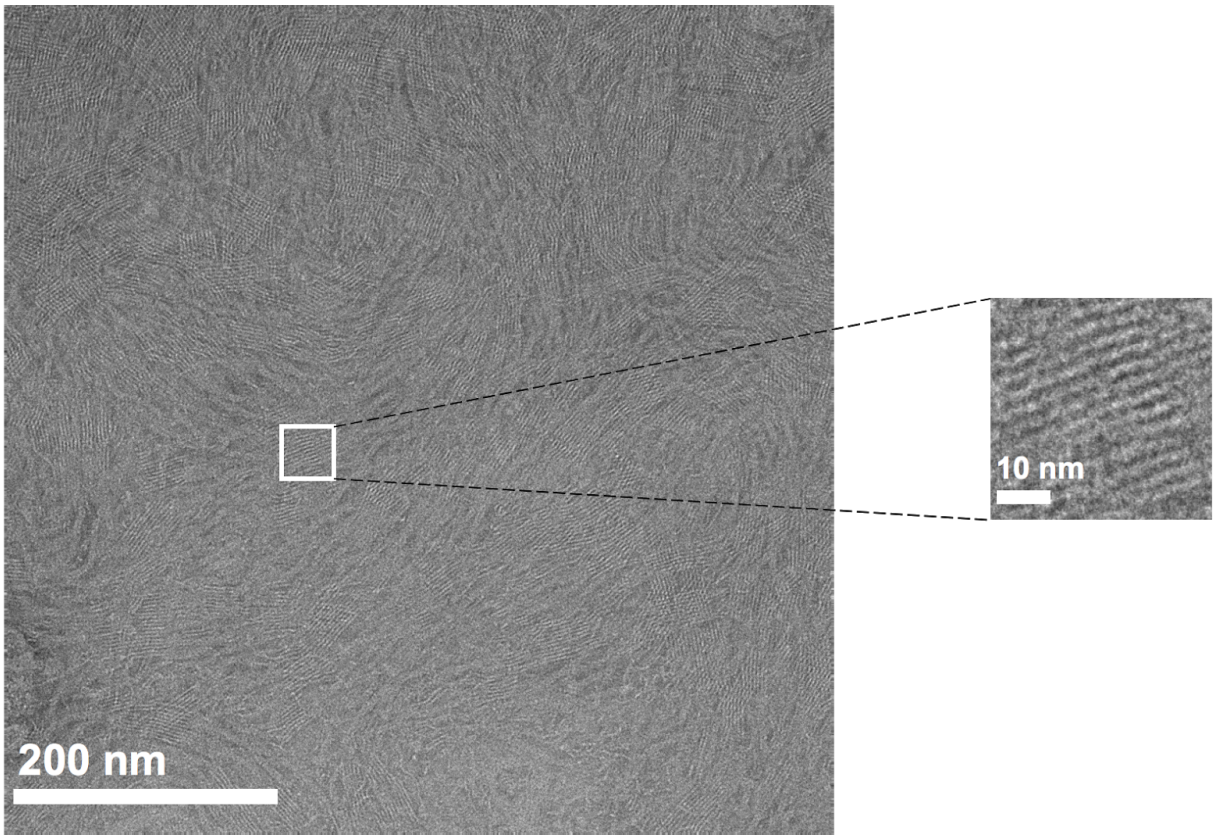


Figure S2. Representative image of 2D-HBN used for 2D class averaging. Inset on the right represents one of the 1,893 boxed sections picked for 2D averaging.

Supp. Figure 3. Clashes with observed parameters

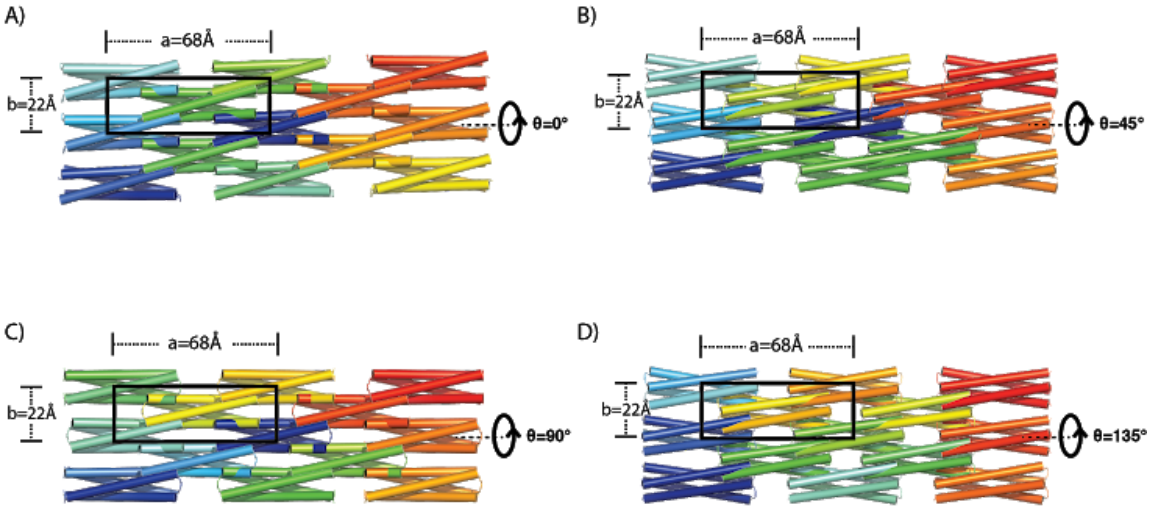


Figure S3. Docking the building block into observed lattice dimensions of 68 and 22 Å results in clashes. Black box marks the unit cell, with dimensions shown outside the box. A) - D): 0°, 45°, 90°, and 135° rotation of the building block around its central axis.

Table 1. Design Sequences

>2D-HP
MGSRTMYIRALEQSLREQEELAKRLKELLRELERLQREGSSDRDVKVLLWEIEALVEEIEKLARL
QKELVEKLKRQGSGNMYIRALEQSLREQEELAKRLKELLRELERLQREGSSDRDVKVLLWEIEA
LVEEIEKLARLQKELVEKLKRQD

>2D-HBN
GELTDIILKLIKSLQTQKLLAERLKTLLKVLEISQDSGADDKQVKKLLDEIRKLVEKIEKLARKQTKL
VEKLLKKGPGNDIILKLIKSLQTQKLLAERLKTLLKVLEISQDSGADDKQVKKLLDEIRKLVEKIEKL
ARKQTKLVEKLLKKD

>2D-HBN_Homo (using the homodimer as building block)
GELTDIILKLIKSLQTQKLLAERLKTLLKVLEISQDSGADDKQVKKLLDEIRKLVEKIEKLARKQTKL
VEKLLKKD

>2D-3
MGTRTMIIRHLEESLREQEELAKRLKELLRELERLQREGSSDREVKSLLSEIEALVEEIEKLARNQ
KKLVEKLKRQGTGNMIIRHLEESLREQEELAKRLKELLRELERLQREGSSDREVKSLLSEIEALV
EEIEKLARNQKKLVEKLKYQD

>2D-4
MGTRTEIIEELQDSLAEQQQLAKLLKALLSWLRWLQEMGSSDEQVRELLRKIKELVEEIEKLARE
QEELVEELTQQGPGNEIIEELQDSLAEQQQLAKLLKALLSWLRWLQEMGSSDEQVRELLRKIKE
LVEEIEKLAREQEELVEELTQQD

>2D-5
MGTRTEIIEELNRSMAEQAELAKLLKALLQALESLQRMGTSDEDVRELLREIKELVEEIEKLMDE
QRYLVMELQQQGPGNEIIEELNRSMAEQAELAKLLKALLQALESLQRMGTSDEDVRELLREIKE
LVEEIEKLMDEQRYLVMELQQQD

>2D-6
MGTRTRAIRQLERSLREQEELAKRLKELLRELQRLQEEGSSDADVNALLILIQALVNEIEKLARE
QKATVEALKRQGPGNRAIRQLERSLREQEELAKRLKELLRELQRLQEEGSSDADVNALLILIQAL
VNEIEKLAREQKATVEALKYQD

>2D-7
ELTDIILKLIKSLQTQKLLAERLKTLLKVLEISQDSGADDKQVKKLLDEIRKLVEKIEKLARKQTKLV
EKLLKKGPGNDIILKLIKSLQTQKLLAERLKTLLKVLEISQDSGADDKQVKKLLDEIRKLVEKIEKLA
RKQTKLVEKLLKKD

>2D-8

PLQTIISLLIASLLLQEALAKLLKLHLEILEELQRKGENDEAVKVQLDLIRKLVEKIEKLARKQKKLV
NELLKQGPGNTIISLLIASLLLQEALAKLLKLHLEILEELQRKGENDEAVKVQLDLIRKLVEKIEKLA
RKQKKLVNELLKQD

>2D-9
GDQTRLIQQIFELVARLIVLIKKLKHLLALLKILVKVGADDETIKKLIKEIEKLVKEIEKAIKKLRELVN
KLSKQGPGNRLIQQIFELVARLIVLIKKLKHLLALLKILVKVGADDETIKKLIKEIEKLVKEIEKAIKKL
RELVNKLSKRA

>2D-10
SREDIILKLEKSLAKQQLLAKRLKLLLKALEIWQRMGADDETVREALREIKELVEEIEKLAREQQY
LSDELLKQGPGNDIILKLEKSLAKQQLLAKRLKLLLKALEIWQRMGADDETVREALREIKELVEEI
EKLAREQQYLSDELLKQA

Table 2. Data collection and refinement statistics

	SC_2L4HC2_23
Wavelength	0.9998
Resolution range	21 - 1.74 (1.802 - 1.74)
Space group	P 1 21 1
Unit cell	41.253 49.36 41.239 90 104.303 90
Total reflections	59303 (5381)
Unique reflections	16336 (1241)
Multiplicity	3.6 (3.4)
Completeness (%)	92.01 (76.04)
Mean I/sigma(I)	9.11 (1.26)
Wilson B-factor	36.92
R-merge	0.05331 (0.8799)
R-meas	0.06269 (1.034)
R-pim	0.03271 (0.5389)
CC1/2	0.998 (0.669)
CC*	1 (0.895)
Reflections used in refinement	15268 (1241)
Reflections used for R-free	1461 (121)
R-work	0.2266 (0.3887)
R-free	0.2657 (0.4216)
CC(work)	0.939 (0.792)

CC(free)	0.913 (0.658)
Number of non-hydrogen atoms	1185
macromolecules	1134
solvent	51
Protein residues	147
RMS(bonds)	0.019
RMS(angles)	1.45
Ramachandran favored (%)	98.6
Ramachandran allowed (%)	1.4
Ramachandran outliers (%)	0
Rotamer outliers (%)	0
Clashscore	3.62
Average B-factor	54.64
macromolecules	54.26
solvent	63.1
Number of TLS groups	4

References

1. Jones S, Thornton JM (1996) Principles of protein-protein interactions. *Proc Natl Acad Sci U S A* 93(1):13–20.

2. Schreiber G, Keating AE (2011) Protein binding specificity versus promiscuity. *Curr Opin Struct Biol* 21(1):50–61.

3. Harbury PB, Zhang T, Kim PS, Alber T (1993) A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science* 262(5138):1401–1407.

4. Diss ML, Kennan AJ (2008) Orthogonal recognition in dimeric coiled coils via buried polar-group modulation. *J Am Chem Soc* 130(4):1321–1327.

5. Bromley EHC, Sessions RB, Thomson AR, Woolfson DN (2009) Designed alpha-helical tectons for constructing multicomponent synthetic biological systems. *J Am Chem Soc* 131(3):928–930.

6. Thomas F, Boyle AL, Burton AJ, Woolfson DN (2013) A set of de novo designed parallel heterodimeric coiled coils with quantified dissociation constants in the micromolar to sub-nanomolar regime. *J Am Chem Soc* 135(13):5161–5166.

7. Fletcher JM, et al. (2012) A Basis Set of de NovoCoiled-Coil Peptide Oligomers for Rational Protein Design and Synthetic Biology. *ACS Synth Biol* 1(6):240–250.

8. Crick FHC (1953) The Fourier transform of a coiled-coil. *Acta Cryst (1953) Q6, 685-689 [doi:101107/S0365110X53001952]* 6(8):1–5.

9. Harris BZ, Lim WA (2001) Mechanism and role of PDZ domains in signaling complex assembly. *J Cell Sci* 114(Pt 18):3219–3231.

10. Zarrinpar A, Park S-H, Lim WA (2003) Optimization of specificity in a cellular protein interaction network by negative selection. *Nature* 426(6967):676–680.

11. Aakre CD, et al. (2015) Evolving new protein-protein interaction specificity through promiscuous intermediates. *Cell* 163(3):594–606.

12. Joachimiak LA, Kortemme T, Stoddard BL, Baker D (2006) Computational design of a new hydrogen bond network and at least a 300-fold specificity switch at a protein-protein interface. *J Mol Biol* 361(1):195–208.

13. Meenan NAG, et al. (2010) The structural and energetic basis for high selectivity in a high-affinity protein-protein interaction. *Proc Natl Acad Sci U S A* 107(22):10080–10085.

14. Skerker JM, et al. (2008) Rewiring the specificity of two-component signal transduction systems. *Cell* 133(6):1043–1054.

15. Gradišar H, Jerala R (2011) De novo design of orthogonal peptide pairs forming parallel coiled-coil heterodimers. *J Pept Sci* 17(2):100–106.

16. Crooks RO, Baxter D, Panek AS, Lubben AT, Mason JM (2016) Deriving Heterospecific Self-Assembling Protein-Protein Interactions Using a Computational Interactome Screen. *J Mol Biol* 428(2 Pt A):385–398.

17. Thompson KE, Bashor CJ, Lim WA, Keating AE (2012) SYNZIP protein interaction toolbox: in vitro and in vivo specifications of heterospecific coiled-coil interaction domains. *ACS Synth Biol* 1(4):118–129.

18. Reinke AW, Grant RA, Keating AE (2010) A synthetic coiled-coil interactome provides heterospecific modules for molecular engineering. *J Am Chem Soc* 132(17):6025–6031.

19. Harbury PB, Kim PS, Alber T (1994) Crystal structure of an isoleucine-zipper trimer. *Nature* 371(6492):80–83.

20. Acharya A, Rishi V, Vinson C (2006) Stability of 100 homo and heterotypic coiled-coil a-a' pairs for ten amino acids (A, L, I, V, N, K, S, T, E, and R). *Biochemistry* 45(38):11324–11332.

21. Grigoryan G, Keating AE (2006) Structure-based prediction of bZIP partnering specificity. *J Mol Biol* 355(5):1125–1142.

22. Gonzalez L Jr, Woolfson DN, Alber T (1996) Buried polar residues and structural specificity in the GCN4 leucine zipper. *Nat Struct Biol* 3(12):1011–1018.

23. Lumb KJ, Kim PS (1995) A buried polar interaction imparts structural uniqueness in a designed heterodimeric coiled coil. *Biochemistry* 34(27):8642–8648.

24. Tatko CD, Nanda V, Lear JD, DeGrado WF (2006) Polar Networks Control Oligomeric Assembly in Membranes. *J Am Chem Soc* 128(13):4170–4171.

25. Boyken SE, et al. (2016) De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity. *Science* 352(6286):680–687.

26. Grigoryan G, Degrado WF (2011) Probing designability via a generalized model of helical bundle geometry. *J Mol Biol* 405(4):1079–1100.

27. Huang P-S, et al. (2014) High thermodynamic stability of parametrically designed helical bundles. *Science* 346(6208):481–485.

28. Zhang Y, Skolnick J (2005) TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* 33(7):2302–2309.

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

30. Ruotolo BT, Robinson CV (2006) Aspects of native proteins are retained in vacuum. *Curr*

Opin Chem Biol 10(5):402–408.

31. Heck AJR (2008) Native mass spectrometry: a bridge between interactomics and structural biology. *Nat Methods* 5(11):927–933.

32. Sahasrabuddhe A, et al. (2018) Confirmation of intersubunit connectivity and topology of designed protein complexes by native MS. *Proc Natl Acad Sci U S A* 115(6):1268–1273.

33. Dyer KN, et al. (2014) High-throughput SAXS for the characterization of biomolecules in solution: a practical approach. *Methods Mol Biol* 1091:245–258.

34. Classen S, et al. (2013) Implementation and performance of SIBYLS: a dual endstation small-angle X-ray scattering and macromolecular crystallography beamline at the Advanced Light Source. *J Appl Crystallogr* 46(Pt 1):1–13.

35. Zhou M, Huang C, Wysocki VH (2012) Surface-induced dissociation of ion mobility-separated noncovalent complexes in a quadrupole/time-of-flight mass spectrometer. *Anal Chem* 84(14):6016–6023.

36. Zhou M, Wysocki VH (2014) Surface induced dissociation: dissecting noncovalent protein complexes in the gas phase. *Acc Chem Res* 47(4):1010–1018.

37. Anderson GP, Shriver-Lake LC, Liu JL, Goldman ER (2018) Orthogonal Synthetic Zippers as Protein Scaffolds. *ACS Omega* 3(5):4810–4815.

38. Reinke AW, Grant RA, Keating AE (2010) A Synthetic Coiled-Coil Interactome Provides Heterospecific Modules for Molecular Engineering. *J Am Chem Soc* 132(17):6025–6031.

39. Gradišar H, Jerala R (2011) De novo design of orthogonal peptide pairs forming parallel coiled-coil heterodimers. *J Pept Sci* 17(2):100–106.

40. Rothemund PWK (2006) Folding DNA to create nanoscale shapes and patterns. *Nature* 440(7082):297–302.

41. Qian L, Winfree E (2011) Scaling up digital circuit computation with DNA strand displacement cascades. *Science* 332(6034):1196–1201.

42. Testa OD, Moutevelis E, Woolfson DN (2009) CC+: a relational database of coiled-coil structures. *Nucleic Acids Res* 37(Database issue):D315–22.

43. Moutevelis E, Woolfson DN (2009) A periodic table of coiled-coil protein structures. *J Mol Biol* 385(3):726–732.

44. Crick FHC (1953) The Fourier transform of a coiled-coil. *Acta Crystallogr* 6(8-9):685–689.

45. Rocklin GJ, et al. (2017) Global analysis of protein folding using massively parallel design, synthesis, and testing. *Science* 357(6347):168–175.

46. Schrödinger, LLC (2015) The PyMOL Molecular Graphics System, Version 1.8.

47. Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):125–132.

48. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276:307–326.

49. McCoy AJ, et al. (2007) Phaser crystallographic software. *J Appl Crystallogr* 40(Pt 4):658–674.

50. Adams PD, et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213–221.

51. Afonine PV, et al. (2010) Joint X-ray and neutron refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* 66(Pt 11):1153–1163.

52. Terwilliger TC, et al. (2008) Iterative model building, structure refinement and density modification with the \it PHENIX AutoBuild wizard. *Acta Crystallogr D Biol Crystallogr* 64(1):61–69.

53. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126–2132.

54. Davis IW, et al. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 35(Web Server issue):W375–83.

55. Rambo RP, Tainer JA (2011) Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers* 95(8):559–571.

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

57. Schneidman-Duhovny D, Hammel M, Tainer JA, Sali A (2013) Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophys J* 105(4):962–974.

58. Schiestl RH, Gietz RD (1989) High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. *Curr Genet* 16(5-6):339–346.

59. Chien CT, Bartel PL, Sternglanz R, Fields S (1991) The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. *Proc Natl Acad Sci U S A* 88(21):9578–9582.

60. Bartel PL, Roecklein JA, SenGupta D, Fields S (1996) A protein linkage map of Escherichia coli bacteriophage T7. *Nat Genet* 12(1):72–77.

61. Guzmán C, Bagga M, Kaur A, Westermarck J, Abankwa D (2014) ColonyArea: an ImageJ plugin to automatically quantify colony formation in clonogenic assays. *PLoS One*

9(3):e92444.

62. Dyachenko A, et al. (2015) Tandem Native Mass-Spectrometry on Antibody-Drug Conjugates and Submillion Da Antibody-Antigen Protein Assemblies on an Orbitrap EMR Equipped with a High-Mass Quadrupole Mass Selector. *Anal Chem* 87(12):6095–6102.

63. Waitt GM, Xu R, Wisely GB, Williams JD (2008) Automated in-line gel filtration for native state mass spectrometry. *J Am Soc Mass Spectrom* 19(2):239–245.

64. Marty MT, et al. (2015) Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. *Anal Chem* 87(8):4370–4376.

65. Bern M, et al. (2018) Parsimonious Charge Deconvolution for Native Mass Spectrometry. *J Proteome Res* 17(3):1216–1226.

66. Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 292(2):195–202.

67. Takenoya M, Nikolakakis K, Sagermann M (2010) Crystallographic insights into the pore structures and mechanisms of the EutL and EutM shell proteins of the ethanolamine-utilizing microcompartment of Escherichia coli. *J Bacteriol* 192(22):6056–6063.

68. Guo F, et al. (2014) Capsid expansion mechanism of bacteriophage T7 revealed by multistate atomic models derived from cryo-EM reconstructions. *Proc Natl Acad Sci U S A* 111(43):E4606–14.

69. Klug A (1999) The tobacco mosaic virus particle: structure and assembly. *Philos Trans R Soc Lond B Biol Sci* 354(1383):531–535.

70. Wei B, Dai M, Yin P (2012) Complex shapes self-assembled from single-stranded DNA tiles. *Nature* 485(7400):623–626.

71. Tikhomirov G, Petersen P, Qian L (2017) Programmable disorder in random DNA tilings. *Nat Nanotechnol* 12(3):251–259.

72. Gonen S, DiMaio F, Gonen T, Baker D (2015) Design of ordered two-dimensional arrays mediated by noncovalent protein-protein interfaces. *Science* 348(6241):1365–1368.

73. Suzuki Y, et al. (2016) Self-assembly of coherently dynamic, auxetic, two-dimensional protein crystals. *Nature* 533(7603):369–373.

74. Liu Y, Gonen S, Gonen T, Yeates TO (2018) Near-atomic cryo-EM imaging of a small protein displayed on a designed scaffolding system. *Proc Natl Acad Sci U S A*. doi:10.1073/pnas.1718825115.

75. Jiang T, Xu C, Zuo X, Conticello VP (2014) Structurally homogeneous nanosheets from self-assembly of a collagen-mimetic peptide. *Angew Chem Int Ed Engl* 53(32):8367–8371.

76. Fallas JA, et al. (2017) Computational design of self-assembling cyclic protein homo-oligomers. *Nat Chem* 9(4):353–360.

77. Zhang K (2016) Gctf: Real-time CTF determination and correction. *J Struct Biol* 193(1):1–12.

78. Scheres SHW (2012) RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* 180(3):519–530.