LETTER

https://doi.org/10.1038/s41586-018-0802-y

# Programmable design of orthogonal protein heterodimers

Zibo Chen[1,2,3], Scott E. Boyken[1,2], Mengxuan Jia[4], Florian Busch[4], David Flores-Solis[5], Matthew J. Bick[1,2], Peilong Lu[1,2], Zachary L. VanAernum[4], Aniruddha Sahasrabuddhe[4], Robert A. Langan[1,2,3], Sherry Bermeo[1,2,3], T. J. Brunette[1,2], Vikram Khipple Mulligan[1,2], Lauren P. Carter[1,2], Frank DiMaio[1,2], Nikolaos G. Sgourakis[5], Vicki H. Wysocki[4] & David Baker[1,2,6]*

**Specificity of interactions between two DNA strands, or between protein and DNA, is often achieved by varying bases or side chains coming off the DNA or protein backbone—for example, the bases participating in Watson–Crick pairing in the double helix, or the side chains contacting DNA in TALEN–DNA complexes. By contrast, specificity of protein–protein interactions usually involves backbone shape complementarity[1], 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[2]) limits the number of orthogonal pairs that can be created simply by varying side-chain interactions[3,4]. Here we show that protein–protein interaction specificity can be achieved using extensive and modular side-chain hydrogen-bond networks. We used the Crick generating equations[5] 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 to optimize the remainder of the sequence. Of 97 such designs expressed in *Escherichia coli*, 65 formed constitutive heterodimers, and the crystal structures of four designs were in close agreement with the computational models and confirmed the designed hydrogen-bond networks. In cells, six heterodimers were fully orthogonal, and in vitro—following mixing of 32 chains from 16 heterodimer designs, denaturation in 5 M guanidine hydrochloride and reannealing—almost all of the interactions observed by native mass spectrometry were between the designed cognate pairs. The ability to design 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.**

Orthogonal sets of protein–protein and protein–peptide interactions have important roles in biological systems[6]. It has proven difficult to use sequence redesign to create new specificities starting from naturally occurring interacting proteins, such as toxin–antidote pairs[7] (promiscuous binding has usually resulted[8]); the natural specificity results at

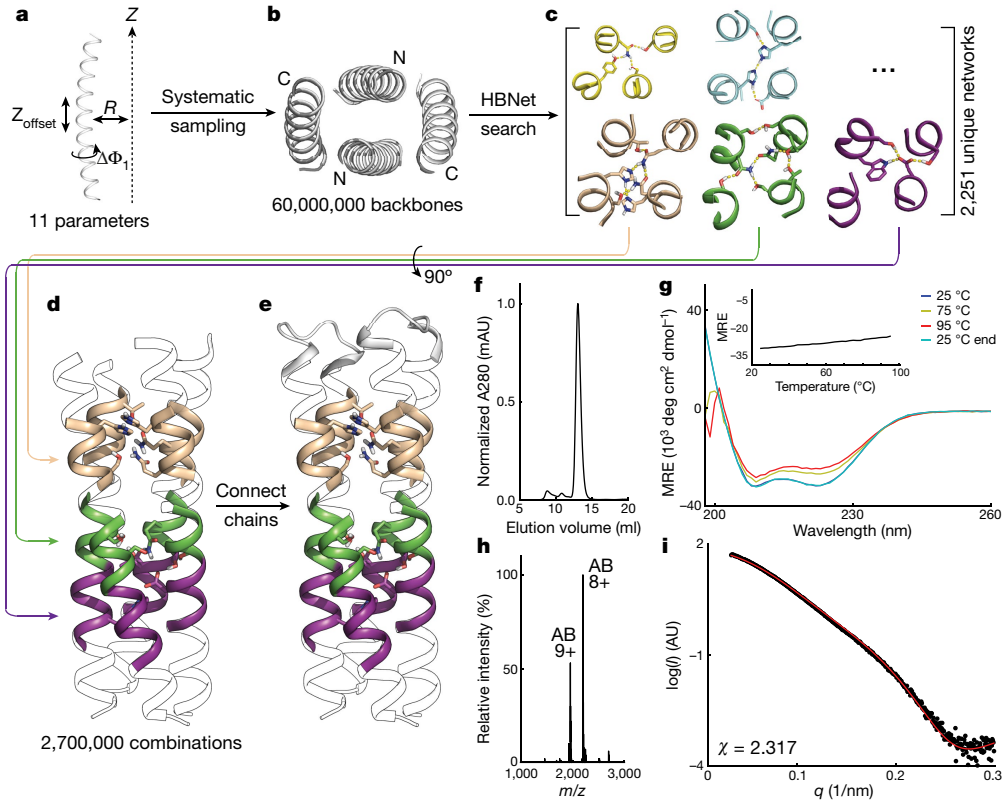



**Fig. 1 | Modular heterodimer design. a**, Individual helix generation: the helical phase ($\Delta\phi_1$), supercoil radius ($R$) and offset along the $Z$-axis ($Z_{offset}$) were exhaustively sampled; a total of 11 free parameters, because there is no $Z_{offset}$ for the first helix. **b**, Top-down view of a representative four-helix backbone. **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 four helices into two helix hairpins. **f**–**i**, SEC trace (**f**), circular dichroism spectra and (inset) temperature melt (**g**), native MS spectrum (**h**) and SAXS (**i**; black, experimental SAXS data; red, spectra computed from the designed backbones; $q$, scattering vector) profiles of the design DHD37_ABXB. Experiments were performed once. A280, absorbance at 280 nm; mAU, milli-absorbance units; MRE, molar residue ellipticity.

[1]Department of Biochemistry, University of Washington, Seattle, WA, USA. [2]Institute for Protein Design, University of Washington, Seattle, WA, USA. [3]Graduate Program in Biological Physics, Structure, and Design, University of Washington, Seattle, WA, USA. [4]Department of Chemistry and Biochemistry, The Ohio State University, Columbus, OH, USA. [5]Department of Chemistry and Biochemistry, University of California Santa Cruz, Santa Cruz, CA, USA. [6]Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA. *e-mail: dabaker@uw.edu

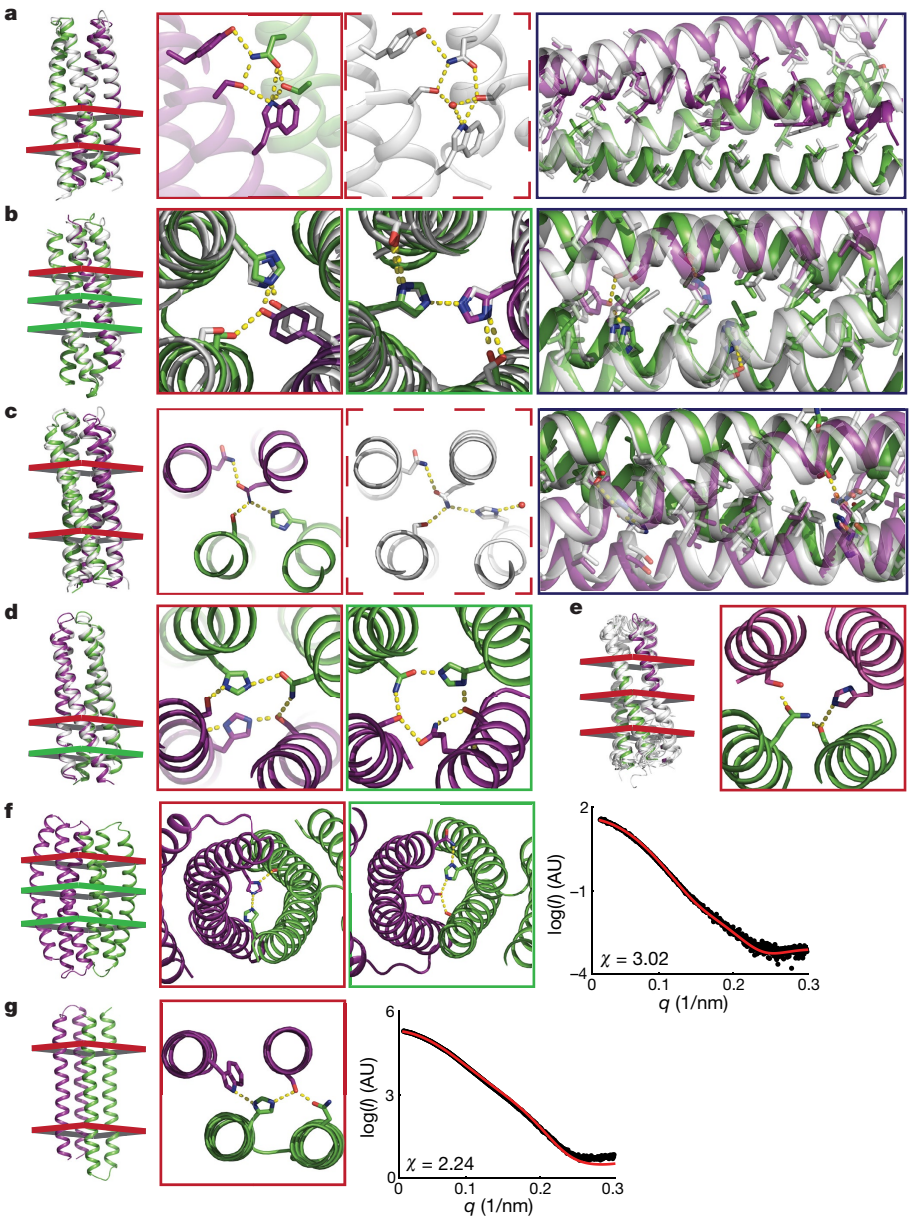



**Fig. 2 | Structural characterization of designed heterodimers.** **a**–**e**, Crystal and NMR structures (white) superimposed on design models with monomers coloured green and purple; coloured cross-sections of backbones (left) indicate locations of designed hydrogen-bond networks (middle panels). Solid and dashed red boxes compare networks in design model and crystal structure, respectively. Green and black boxes denote additional hydrogen-bond network and hydrophobic packing layers, respectively. **a**, DHD_131, 2.4 Å resolution with 1.0 Å Cα r.m.s.d. **b**, DHD37_1:234, 3.3 Å resolution with 1.4 Å r.m.s.d. **c**, DHD_127, 1.8 Å resolution with 1.7 Å r.m.s.d. **d**, DHD_15, 3.4 Å resolution with 0.9 Å r.m.s.d.; hydrogen-bond networks were not well-resolved. **e**, NMR ensemble (white) of DHD13_XAAA superimposed onto the design model; the assigned side-chain–side-chain NMR distance constraints were not sufficient to define hydrogen-bond networks. **f**, **g**, Backbones and designed hydrogen-bond networks of DHD_39 and DHD_120. Experimental SAXS data (black) are similar to spectra computed from the designed backbones (red).

least in part from complementary variation in backbone conformation[9]. Orthogonal sets of 2–4 interacting coiled-coil pairs have been created and experimentally validated[10,11], including the widely used SYNZIPs[12–18], but interaction promiscuity has again hampered the design of larger orthogonal sets.

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[5,19]. 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), restricting the supercoil phases of the helices to 0, 90, 180 and 270°, and the supercoil twist ($\omega_0$) and helical twist ($\omega_1$) to the ideal values for either a two-layer left-handed supercoil ($\omega_0 = -2.85$ and $\omega_1 = 102.85$), or a five-layer untwisted bundle ($\omega_0 = 0$ and $\omega_1 = 100$)[20] (Extended Data Fig. 1a–d). This 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. 1g).

To identify the modular hydrogen-bond network equivalents to DNA base pairs, we used Rosetta HBNet[21] to design buried hydrogen-bond networks in the central repeat units of each backbone, and obtained 2,251 hydrogen-bond networks involving at least four side-chain residues with all heavy-atom donors and acceptors participating in

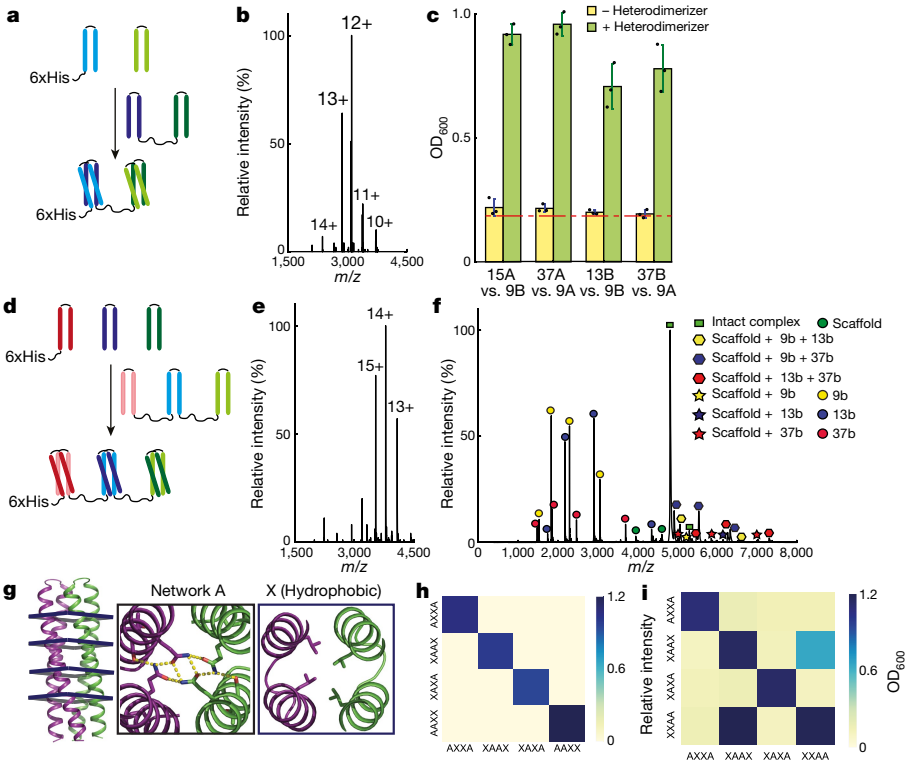



**Fig. 3 | New functionality from DHD combinations. a**, Induced dimerizer formed from *b* component of DHD13_XAAA (dark blue) fused to *b* component of DHD37_ABXB (dark green) with an intervening flexible linker. The *a* components of the two heterodimers (light blue and light green) are brought into close proximity by the heterodimerizer. **b**, Native MS of purified DHD13_XAAA:DHD37_ABXB heterotrimer complex; no heterodimers or monomers were observed. Molecular mass: designed 37,133 Da; observed 37,132 Da. **c**, Y2H data on four induced dimerization systems. Yellow, without heterodimerizer fusion; green, with heterodimerizer fusion. Red dashed line indicates background growth with unfused activation domain and DBD. Data are mean ± s.d. from three biological repeats. **d**, 9_a (pink), 13_XAAA_a (light blue) and 37_ABXB_a (light green) were covalently linked to form a scaffold, recruiting 9_b (red, hexahistidine tagged), 13_XAAA_b (dark blue) and 37_ABXB_b (dark green). **e**, Native MS of purified scaffold complex; no heterotrimers, heterodimers or monomers were observed. Molecular mass: designed 52,979 Da; observed 52,979 Da. **f**, SID of the 11+ peak in **e**; no cross binding between *b* monomers is detected. **g**, The backbone of 2L4HC2_23 can accommodate hydrogen-bond networks at four heptad positions. **h**, Native MS mixing data of four variants generated by hydrogen-bond-network shuffling; the interactions are orthogonal. **i**, Y2H data of four hydrogen-bond shuffling variants. Two biologically independent experiments were performed for **b**, **e**, **f**, **h**, **i**.

hydrogen bonds, and connecting all four helices (Fig. 1c, Extended Data Fig. 2, Supplementary Table 1). We then identified all of the geometrically compatible placements of these hydrogen-bond networks in each backbone (Fig. 1d), selected backbones that accommodated at least two networks, and connected pairs of helices with short loops (Fig. 1e). Low-energy sequences were identified using RosettaDesign[22] calculations in which the hydrogen-bond networks were held fixed. Designs with fully satisfied hydrogen-bond networks and tight hydrophobic packing were selected for experimental characterization, excluding those with networks with *C*2 symmetry to disfavour homodimerization of monomers (Extended Data Fig. 1e, f). Designed heterodimers (DHDs) are referred to by numbers with monomers labelled *a* or *b*; for example, DHD15_a refers to monomer *a* of design DHD15.

Of the 97 selected designs (Supplementary Table 2), 94 were well-expressed in *E. coli* with both monomers co-purifying by Ni-affinity chromatography (only one monomer contains a hexahistidine tag; Supplementary Table 3). For 85 of these 94, the dominant species observed in size-exclusion chromatography (SEC) had the expected size (Fig. 1f). Thirty-nine of these 85 were exclusive heterodimers at 15 μM by native mass spectrometry (MS)[23,24] (Fig. 1h), 13 were heterodimers with a minor population of heterotetramers, and 13 formed heterodimers with one monomer (but never both) also present as a homodimer arising from unbalanced expression in *E. coli* (Supplementary Tables 4–6). Native MS experiments with serially diluted samples suggested that the DHDs have affinities in the nanomolar range (Supplementary Table 7). Three designs characterized by circular dichroism spectroscopy were found to be all α-helical and stable at 95 °C (Fig. 1g, Extended Data Fig. 3).

We investigated 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 in which 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, whereas DHD103_1:423 indicates a heterodimer in which one monomer consists of the first helix of DHD103 and the other monomer consists of helices 2–4 (Extended Data Fig. 4). Thirteen of fourteen hydrogen-bond-network-permuted variants and nine of ten '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 (Fig. 2b, Supplementary Tables 8, 9). Using the hydrogen-bond network permutation approach, we were also able to generate a set of four orthogonal homodimers (Fig. 3g–i, Extended Data Fig. 5, Supplementary Table 10).

Small-angle X-ray scattering (SAXS) spectra collected for 44 designs that were constitutive heterodimers by native MS are consistent with the design models (Figs. 1i, 2f, g, Extended Data Fig. 6, Supplementary Table 11). The nuclear magnetic resonance (NMR) structure of DHD13_XAAA closely matched the design model: the root mean square deviation (r.m.s.d.) over all main-chain α-carbon (Cα) atoms was 2 Å between the designed structure and the lowest-energy NMR model (Fig. 2e). The X-ray crystal structures of

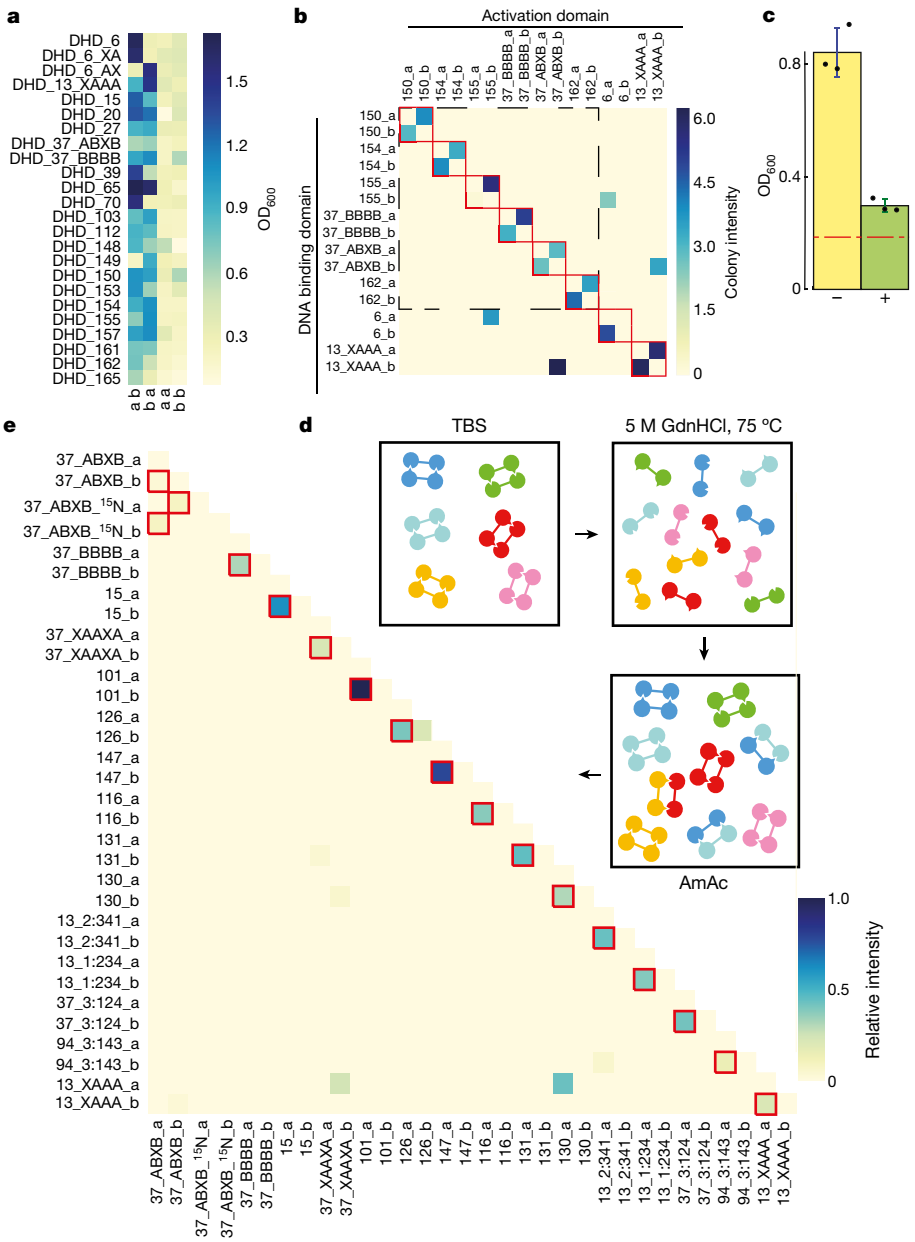



**Fig. 4 | All-against-all orthogonality assessment. a**, Y2H for 21 heterodimers shows heterodimer formation with little homodimer formation. First letter at bottom indicates monomer fused to activation domain; second letter indicates monomer fused to DBD. **b**, Y2H all-by-all testing of eight pairs of heterodimers; colours indicate growth. Red boxes indicate designed cognate heterodimer pairs, dashed black box indicates a set of six orthogonal heterodimers. **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. Data are mean ± s.d. Red dashed line indicates background growth with unfused activation domain and DBD. **d**, **e**, All-against-all orthogonality of 16 pairs of heterodimers assessed by native MS mixing assay. Red boxes indicate designed cognate pairs. Exchange of unlabelled and partially $^{15}$N-labelled DHD37_ABXB results in a distribution of overlapping species with low individual signal intensities. Two (**a**, **b**) or three (**c**) biologically independent or three (**e**) technically independent experiments were performed. AmAc, ammonium acetate.

DHD131, DHD37_1:234, DHD127 and DHD15 have backbone Cα atom r.m.s.d. values to the design models 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 bridging water molecule (Fig. 2a). The two designed hydrogen-bond networks in DHD37_1:234, which contain buried histidine and tyrosine aromatic side chains that sterically disfavour homodimers, are in close agreement with the crystal structure (Fig. 2b). In DHD127, the histidines in the two hydrogen-bond networks adopt a rotamer that differs 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), whereas a structure at pH 6.5 is of a domain-swapped, heterotetramer conformation; native MS at pH 6.5 suggests that the designed heterodimer—rather than the heterotetramer—is dominant in solution (Extended Data Fig. 7a–c).

We built three induced dimerization systems by fusing one monomer from each of two different heterodimers via a flexible linker, and testing whether the remaining two monomers from each pair could be brought together by the fusion (Fig. 3a). In each case, the three components copurified by Ni-nitrilotriacetic acid (Ni-NTA) chromatography (one monomer has a hexahistidine tag), and native MS showed they formed constitutive heterotrimers (Fig. 3b); no partial complexes (heterodimers) were observed. Surface-induced dissociation (SID) MS[25,26] resulted in binary complexes consisting of the heterodimerizer bound to either one of the monomers (Supplementary Tables 12, 13),

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

We constructed synthetic scaffolding systems[27] by covalently linking the *a* subunits of three DHDs via flexible linkers (Fig. 3d), and co-expressing this scaffold and the three separate *b* subunits, one with a hexahistidine tag, in *E. coli*. Native MS of the purified sample revealed a heterotetramer of all four proteins (Fig. 3e), with SID producing sub-complexes consisting of the scaffold with one or two of the three *b* subunits bound (Fig. 3f); no association between *b* subunits without the scaffold was detected. The scaffold plus monomer assembly is stable at 95 °C and has a guanidine denaturation midpoint of 4 M (Extended Data Fig. 7h, i).

By generating interfaces with many polar groups that are energetically costly to bury without geometrically matched hydrogen-bonding interactions, our design protocol implicitly disfavours non-cognate interactions (explicit negative design to disfavour non-cognate interactions is computationally intractable, given the very large number of possible off-target binding modes). We investigated the interaction specificity of the DHDs in cells using Y2H experiments. For 24 designs, strong interactions were observed by Y2H with the two partners fused to the DBD and activation domain, but not when either partner was fused to both domains; the designed heterodimers, but not the homodimers, form in cells (Fig. 4a). The 24 monomers in 12 of these designs were crossed in an all-by-all Y2H experiment; interactions were observed for all cognate pairs, and 27 of the 552 possible non-cognate interactions (Extended Data Fig. 8). Orthogonality was higher for an eight-DHD subset: of 240 possible non-cognate interactions, only four were observed (Fig. 4b; the interacting polar residues are depicted schematically in Extended Data Fig. 9). Coexpression of unfused monomers eliminated off-target interactions (Fig. 4c); the cognate interactions are evidently stronger than the non-cognate interactions.

To probe all-by-all interactions specificity of the designed monomers when allowed to associate freely in a single pot, purified DHDs were mixed, denatured in 5 M guanidine hydrochloride (GdnHCl) at 75 °C (Fig. 4d), and allowed to reanneal by dialysis. An $^{15}$N-labelled variant was added as a control for subunit exchange under denaturing conditions; the hybrid labelled–unlabelled complexes expected if full exchange was taking place were observed in all cases (Extended Data Fig. 10). The resulting mixture was analysed by online ion exchange chromatography coupled to high-resolution native MS (Supplementary Table 14). Sixteen designs (15 unique pairs and the $^{15}$N-labelled control) with the highest cognate specificity were pooled together. The native MS results on the 32-chain mixture are notable (Fig. 4e). All 16 designed pairs were recovered, and of the 512 non-cognate binary complexes possible, only 6 were observed. No non-cognate trimers or higher-order oligomers were observed. Several of the orthogonal pairs were generated using hydrogen-bond network shuffling and backbone permutation: the orthogonality of the former is 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 are likely to stem from the dependence of the former 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).

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 heterogeneity in the size of the residues at the designed interface (Extended Data Fig. 8d–i), analogous to the contribution of steric effects to the specificity of Watson–Crick base pairing. The power of native MS to determine interaction specificity in complex mixtures is highlighted by the 32-chain mixing experiment in Fig. 4e; of the large number of possible oligomeric complexes that can be formed from these chains (528 two-chain species, 5,984 three-chain species, and so on), only the 15 designed heterodimers and 6 off-target interactions were observed. The relatively simple encoding of specificity in DNA gave birth to a broad spectrum of new technology, from DNA origami[28] to artificial circuits[29]. Our 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 with faster response times and better integration with signalling inputs and outputs than current nucleic-acid-based circuitry.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-018-0802-y.

Received: 6 June 2018; Accepted: 8 November 2018;
Published online: 19 December 2018

1. Jones, S. & Thornton, J. M. Principles of protein–protein interactions. *Proc. Natl Acad. Sci. USA* **93**, 13–20 (1996).
2. Harbury, P. B., Zhang, T., Kim, P. S. & Alber, T. A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science* **262**, 1401–1407 (1993).
3. Diss, M. L. & Kennan, A. J. Orthogonal recognition in dimeric coiled coils via buried polar-group modulation. *J. Am. Chem. Soc.* **130**, 1321–1327 (2008).
4. Thomas, F., Boyle, A. L., Burton, A. J. & Woolfson, D. N. 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**, 5161–5166 (2013).
5. Crick, F. H. C. The Fourier transform of a coiled-coil. *Acta Cryst.* **6**, 685–689 (1953).
6. Zarrinpar, A., Park, S.-H. & Lim, W. A. Optimization of specificity in a cellular protein interaction network by negative selection. *Nature* **426**, 676–680 (2003).
7. Aakre, C. D. et al. Evolving new protein–protein interaction specificity through promiscuous intermediates. *Cell* **163**, 594–606 (2015).
8. Joachimiak, L. A., Kortemme, T., Stoddard, B. L. & Baker, D. 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**, 195–208 (2006).
9. Skerker, J. M. et al. Rewiring the specificity of two-component signal transduction systems. *Cell* **133**, 1043–1054 (2008).
10. Crooks, R. O., Baxter, D., Panek, A. S., Lubben, A. T. & Mason, J. M. Deriving heterospecific self-assembling protein–protein interactions using a computational interactome screen. *J. Mol. Biol.* **428**, 385–398 (2016).
11. Gradišar, H. & Jerala, R. De novo design of orthogonal peptide pairs forming parallel coiled-coil heterodimers. *J. Pept. Sci.* **17**, 100–106 (2011).
12. Thompson, K. E., Bashor, C. J., Lim, W. A. & Keating, A. E. SYNZIP protein interaction toolbox: *in vitro* and *in vivo* specifications of heterospecific coiled-coil interaction domains. *ACS Synth. Biol.* **1**, 118–129 (2012).
13. Reinke, A. W., Grant, R. A. & Keating, A. E. A synthetic coiled-coil interactome provides heterospecific modules for molecular engineering. *J. Am. Chem. Soc.* **132**, 6025–6031 (2010).
14. Acharya, A., Rishi, V. & Vinson, C. 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**, 11324–11332 (2006).
15. Grigoryan, G. & Keating, A. E. Structure-based prediction of bZIP partnering specificity. *J. Mol. Biol.* **355**, 1125–1142 (2006).
16. Gonzalez, L. Jr, Woolfson, D. N. & Alber, T. Buried polar residues and structural specificity in the GCN4 leucine zipper. *Nat. Struct. Biol.* **3**, 1011–1018 (1996).
17. Lumb, K. J. & Kim, P. S. A buried polar interaction imparts structural uniqueness in a designed heterodimeric coiled coil. *Biochemistry* **34**, 8642–8648 (1995).
18. Tatko, C. D., Nanda, V., Lear, J. D. & Degrado, W. F. Polar networks control oligomeric assembly in membranes. *J. Am. Chem. Soc.* **128**, 4170–4171 (2006).
19. Grigoryan, G. & Degrado, W. F. Probing designability via a generalized model of helical bundle geometry. *J. Mol. Biol.* **405**, 1079–1100 (2011).
20. Huang, P.-S. et al. High thermodynamic stability of parametrically designed helical bundles. *Science* **346**, 481–485 (2014).
21. Boyken, S. E. et al. De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity. *Science* **352**, 680–687 (2016).
22. Leaver-Fay, A. et al. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487**, 545–574 (2011).
23. Ruotolo, B. T. & Robinson, C. V. Aspects of native proteins are retained in vacuum. *Curr. Opin. Chem. Biol.* **10**, 402–408 (2006).

24. Sahasrabuddhe, A. et al. Confirmation of intersubunit connectivity and topology of designed protein complexes by native MS. *Proc. Natl Acad. Sci. USA* **115**, 1268–1273 (2018).
25. Zhou, M., Huang, C. & Wysocki, V. H. Surface-induced dissociation of ion mobility-separated noncovalent complexes in a quadrupole/time-of-flight mass spectrometer. *Anal. Chem.* **84**, 6016–6023 (2012).
26. Zhou, M. & Wysocki, V. H. Surface induced dissociation: dissecting noncovalent protein complexes in the gas phase. *Acc. Chem. Res.* **47**, 1010–1018 (2014).
27. Anderson, G. P., Shriver-Lake, L. C., Liu, J. L. & Goldman, E. R. Orthogonal synthetic zippers as protein scaffolds. *ACS Omega* **3**, 4810–4815 (2018).
28. Rothemund, P. W. K. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297–302 (2006).
29. Qian, L. & Winfree, E. Scaling up digital circuit computation with DNA strand displacement cascades. *Science* **332**, 1196–1201 (2011).

**Acknowledgements** We thank Rosetta@Home volunteers for contributing computing resources; A. Kang for protein crystallization support; B. Sankaran for assistance with diffraction data collection; K. Lau and B. Groves for assistance with Y2H assays; S. Rettie for MS support; S. Ovchinnikov for help with TMalign; M. Marty, M. Bern and A. Norris for assistance with native MS; S. Pennington for making media for Y2H assays; the SIBYLS mail-in SAXS program, supported by the DOE BER IDAT grant (DE-AC02-05CH11231) and ALS-ENABLE (P30 GM124169) for SAXS; and A. Keating, G. Rocklin and N. Woodall for feedback on the manuscript. Additional funding and computing resources are listed in the Supplementary Information.

**Reviewer information** *Nature* thanks G. Grigoryan, C. Robinson and the other anonymous reviewer(s) for their contribution to the peer review of this work.

**Author contributions** Z.C., S.E.B. and D.B. designed the research. Z.C. and D.B. wrote the manuscript. M.J., F.B., Z.L.V., A.S. and V.H.W. performed native MS experiments and analysed data. L.P.C. prepared proteins for NMR experiments. D.F.-S. and N.G.S. performed NMR experiments. Z.C. wrote the heptad stacking code. S.E.B. improved the HBNet method. V.K.M. wrote the parametric backbone generation code. T.J.B. wrote the loop closure code. Z.C. and S.E.B. carried out design calculations, and R.A.L. and S.B. helped. Z.C., M.J.B., P.L. and F.D. solved crystal structures. All authors discussed results and commented on the manuscript.

**Competing interests** Z.C., S.E.B., R.A.L., S.B. and D.B. are inventors on US provisional patent application no. 62755264 and patent application WO2017173356A1. D.B. and S.E.B. hold equity in Lyell Immunopharma.

**Additional information**
**Extended data** is available for this paper at https://doi.org/10.1038/s41586-018-0802-y.
**Supplementary information** is available for this paper at https://doi.org/10.1038/s41586-018-0802-y.
**Reprints and permissions information** is available at http://www.nature.com/reprints.
**Correspondence and requests for materials** should be addressed to D.B.
**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

**Computational design.** *Systematic sampling of parametric helical backbones.* We used a generalization of the Crick coiled-coil parameters[5] to independently sample all four helices of the heterodimers supercoiled around the same axis, as previously described[19–21]. The supercoil twist ($\omega_0$) and helical twist ($\omega_1$) were coupled and ideal values were used[20] with $\omega_0$ and $\omega_1$ held constant among the helices. A left-handed supercoil results from $\omega_0 = -2.85$ and $\omega_1 = 102.85$, and a straight bundle with no supercoiling from $\omega_0 = 0$ and $\omega_1 = 100$. 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 accommodate 5 heptad repeats. After removing redundant sample points from the overlapping regions of radius sampling, the supercoiled helical bundles contained more than 60 million unique backbones, and the straight helical bundles contained more than 27 million unique backbones.

*HBNet search.* For each parametrically generated backbone, HBNet[21] 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 the 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 were 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 five alanines in the middle heptad and no buried unsatisfied polar heavy atoms were selected for downstream design.

*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 two 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 that the heptad stacking method would remain the same if we simply searched for HBNets in the middle heptad on all full backbones, extracted the middle heptads, and aligned 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[30]. All alignments with r.m.s.d. less than 0.3 were identified and full backbones that could accommodate at least two middle heptads were selected for final design.

*Connecting parametric helical backbones.* Helical backbones are connected with short 2–5-residue loops such that the r.m.s.d. of each loop is less than 0.4 to a 9-residue stretch in a native protein. The distance and directionality between helices limit what loops can connect, so our closure extends and shrinks helices by up to three residues. We then superimpose all short loops from the PDB onto the first and last two helical residues. The loops with the lowest stub-r.m.s.d. are minimized using the Rosetta score function onto the helical endpoints to ensure near-perfect closure. Loop quality is assessed by measuring the distance in r.m.s.d. to the closest nine stretch in the PDB. The loop with the lowest r.m.s.d. is returned as the solution. We repeat this procedure to connect all helices and report the solution with the lowest r.m.s.d.

*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, three methionines and three phenylalanines 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 described in 'HBNet search' above.

*Selection criteria and metrics used to evaluate designs.* Designs were selected using 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 chain 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 optical density at 280 nm ($OD_{280}$). Surface charge residues for a few of the designs were redesigned to shift the theoretical isoelectric point away from buffer pH.

**Calculations of r.m.s.d.** Crystal structures and the corresponding design models were superimposed with TMalign using all heavy atoms. From this alignment, r.m.s.d. was calculated across all α-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 inputs to a logistic regression method[31] to find correlations between computational metrics and experimental observations.

**Visualization and Figures.** All structural images for figures were generated using PyMOL[32].

**Buffer and medium recipes.** 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 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 three and four 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 three proteins) or first two (in the case of four proteins) genes were cloned between NcoI and HindIII sites, with a ribosome binding site separating the two genes 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 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 r.p.m. for 18 h 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 r.p.m. for 24 h at 37 °C.

For expression of $^{13}C^{15}N$- or $^{15}N$-labelled proteins, 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 r.p.m. for 18 h at 37 °C. Starter culture (10 ml) was then transferred to a 2-l baffled flask containing 500 ml Terrific Broth (Difco), with 25 mM $Na_2HPO_4$, 25 mM $KH_2PO_4$, 50 mM $NH_4Cl$, 5 mM $Na_2SO_4$, and 100 μg/ml carbenicillin. The culture was grown at 37 °C to an $OD_{600}$ of approximately 1.0, then centrifuged at 5,000 relative centrifugal force (r.c.f.) for 15 min 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 labelled medium (25 mM $Na_2HPO_4$, 25 mM $KH_2PO_4$, 50 mM $^{15}NH_4Cl$, 5 mM $Na_2SO_4$, 0.2% (w/v) $^{13}C$ glucose), and 100 μg/ml carbenicillin. The cells were allowed to grow at 37 °C for 2 h, before isopropyl β-D-1-thiogalactopyranoside (IPTG; Carbosynth) was added to 1 mM and the temperature was reduced to 18 °C. The labelled glucose and $NH_4Cl$ were obtained from Cambridge Isotopes.

**Affinity purification.** Cells were collected by centrifugation for 15 min at 5,000 r.c.f. 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 min. For immobilized metal affinity chromatography, lysates were clarified by centrifugation at 4 °C and 18,000 r.p.m. for at least 30 min and applied to Ni-NTA (Qiagen) columns pre-equilibrated with lysis buffer. The column was washed twice with five column volumes (CV) of wash buffer, followed by 5 CV of elution buffer. For Strep tag purification, elution fractions from immobilized metal affinity chromatography 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. The mass and purity of eluted proteins were confirmed using electrospray ionization mass spectrometry on a Thermo Scientific TSQ Quantum Access mass spectrometer.

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

**Circular dichroism measurements.** Circular dichroism (CD) wavelength scans (260–195 nm) and temperature melts (25–95 °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. 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 1-min mixing time for each step. Titrant solution consisted of the same concentration of protein in PBS + GdmCl.

**Nuclear magnetic resonance.** SEC-purified $^{13}C^{15}N$-labelled protein was concentrated to >1 mM and buffer exchanged into 50 mM NaCl, 20 mM sodium phosphate, 10% $D_2O$, 0.01% $NaN_3$ at pH 6.3. Sample was loaded into a 5.0-mm Shigemi tube and four NMR experiments were recorded and analysed: 2D transverse relaxation optimized spectroscopy-heteronuclear single quantum coherence, 4D HNCH nuclear Overhauser effect spectroscopy (NOESY), 4D HNCH total correlated spectroscopy and 4D HCCH NOESY. The data were acquired with a non-uniform sampling (NUS) scheme 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 were recorded with 0.3% and 3.0% of sparsity for the HNCH and HCCH experiments, respectively. The final spectra were loaded and analysed on Sparky 3.115.

The spin systems were identified using supervised NMR data analysis and 148 residues were successfully assigned (93.67%). The completeness in terms of protons assigned was 87.4% (Supplementary Table 4). For the structural determination, 3,423 peaks were extracted from the NOESY data. Twenty-four long-range contacts ($|i-j|>4$) were manually assigned with the 4D HNCH NOESY experiment and 29 in 4D HCCH. Owing 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, centre and end of both sequences. The assignments, chemical shifts and proton–proton constraints were used for RASREC or AutoNoe structural calculations in ROSETTA 3. Summaries of refinement statistics are provided in Supplementary Table 16.

**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, using 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 1–14 days by the hanging drop vapour diffusion method with the drops consisting of a 1:1, 2:1 or 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 datasets were collected at the Advanced Light Source at Lawrence Berkeley National Laboratory with beamlines 8.2.1 and 8.2.2. Datasets were indexed and scaled using either XDS[33] or HKL2000[34]. Initial models were generated by the molecular-replacement method with the program PHASER[35] within the Phenix software suite[36], using the design models as the initial search models. Efforts were made to reduce model bias through refinement with simulated annealing using Phenix.refine[37], or, if the resolution was sufficient, by using Phenix.autobuild[38] with rebuild-in-place set to false, simulated annealing and prime-and-switch phasing. Iterative rounds of manual building in COOT[39] and refinement in Phenix were used to produce the final models. Owing 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 reported by Phenix.Xtriage, which complicated structure refinement and may explain the higher-than-expected $R$ values reported. The r.m.s.d. values of bond lengths, angles and dihedrals from ideal geometries were calculated with Phenix[36]. The overall quality of all final models was assessed using the program MOLPROBITY[40]. Summaries of diffraction data and refinement statistics are provided in Supplementary Table 17.

**Small-angle X-ray scattering.** 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 ($\lambda$) 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 $2\theta$ 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-s exposures for 10 s resulting in 32 frames per sample. For each sample, data were collected for two different concentrations to test for concentration-dependent effects; 'low' concentration samples ranged from 2 to 3 mg/ml and 'high' concentration samples ranged from 5 to 7 mg/ml. Data were processed using the SAXS FrameSlice online serve and analysed using the ScÅtter software package[41,42]. FoXS[43,44] was used to compare design models to experimental scattering profiles and calculate quality of fit ($\chi$) 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 DBDs or activation domains, using the LiAc/SS carrier DNA/PEG method[45]. 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) medium lacking tryptophan and leucine for 48 h with shaking at 1,000 r.p.m. at 30 °C. The resulting culture was diluted 1:100 and grown for 16 h in fresh SDO medium lacking tryptophan and leucine, before being transferred to a 96-well plate and diluted 1:100 into SDO medium containing 100 mM 3-amino-1,2,4-triazole (3-AT), lacking tryptophan, leucine and histidine (5 mM 3-AT in the case of induced dimerization). The culture was incubated with shaking at 1,000 r.p.m. at 30 °C. As it is necessary to bring the DBD and the transcription activation domain into proximity for the growth of yeast cells in medium lacking histidine, binding of two proteins was indicated by the growth of yeast cells[46,47]. The optical density of yeast cells was recorded after 48 h. For Y2H assay on agar plates, the 1:100 diluted overnight culture was transferred onto a 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 analysed by the ColonyArea[48] 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 24 h at 4 °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[25,26,49].

**Native MS of individual heterodimers.** Sample purity and integrity were first analysed using a self-packed buffer exchange column[50] (P6 polyacrylamide gel, BioRad), 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 or dimer masses were selected for mixing experiments.

**Native MS mixing assay and data analysis.** In the mixing experiment, heterodimers were mixed in equimolar ratio of 10 μM. GdnHCl was added to a final concentration of 5 M and the mixture was incubated at 75 °C for 30 min to ensure complete denaturation. To allow relative quantification of exchanged species, a control mixing experiment was performed in which the denaturation and refolding steps were omitted. The mixtures were then dialysed against 150 mM AmAc solution for refolding and subsequent formation of protein–protein interactions. Eight microlitres 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 1,000 mM AmAc over a period of 55 min. The eluting proteins were detected online using 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[51]. Details of instrument settings are in Supplementary Table 15. Data were deconvoluted using Xcalibur (Thermo Scientific), UniDec[52] and Intact Mass (Protein Metrics[53]). The detailed deconvolution parameters are listed in Supplementary Table 15. 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 heterotrimer 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 using 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

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

in which $I_r(N_aN_b)$ is the relative intensity of a dimer $N_aN_b$ identified in the mixing experiment. $I_{DN}(N_aN_b)$ is the intensity of the $N_aN_b$ species in the run involving denaturation and refolding. $I_N(N_aN_b)$ and $I_N(M_aM_b)$ are the intensities 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. 10.

**Native MS of higher-order hetero-oligomers.** Samples were buffer exchanged into 200 mM ammonium acetate using Micro Bio-Spin 6 columns (Bio-Rad). Twenty per cent (v/v) 200 mM triethylammonium acetate (Sigma) was added for charge reduction. SID was performed on an in-house modified SYNAPT G2 HDMS (Waters) with a SID device incorporated between a truncated trap travelling wave ion guide and the ion mobility cell[25]. 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, 45 V. The SID settings are listed in Supplementary Table 15.

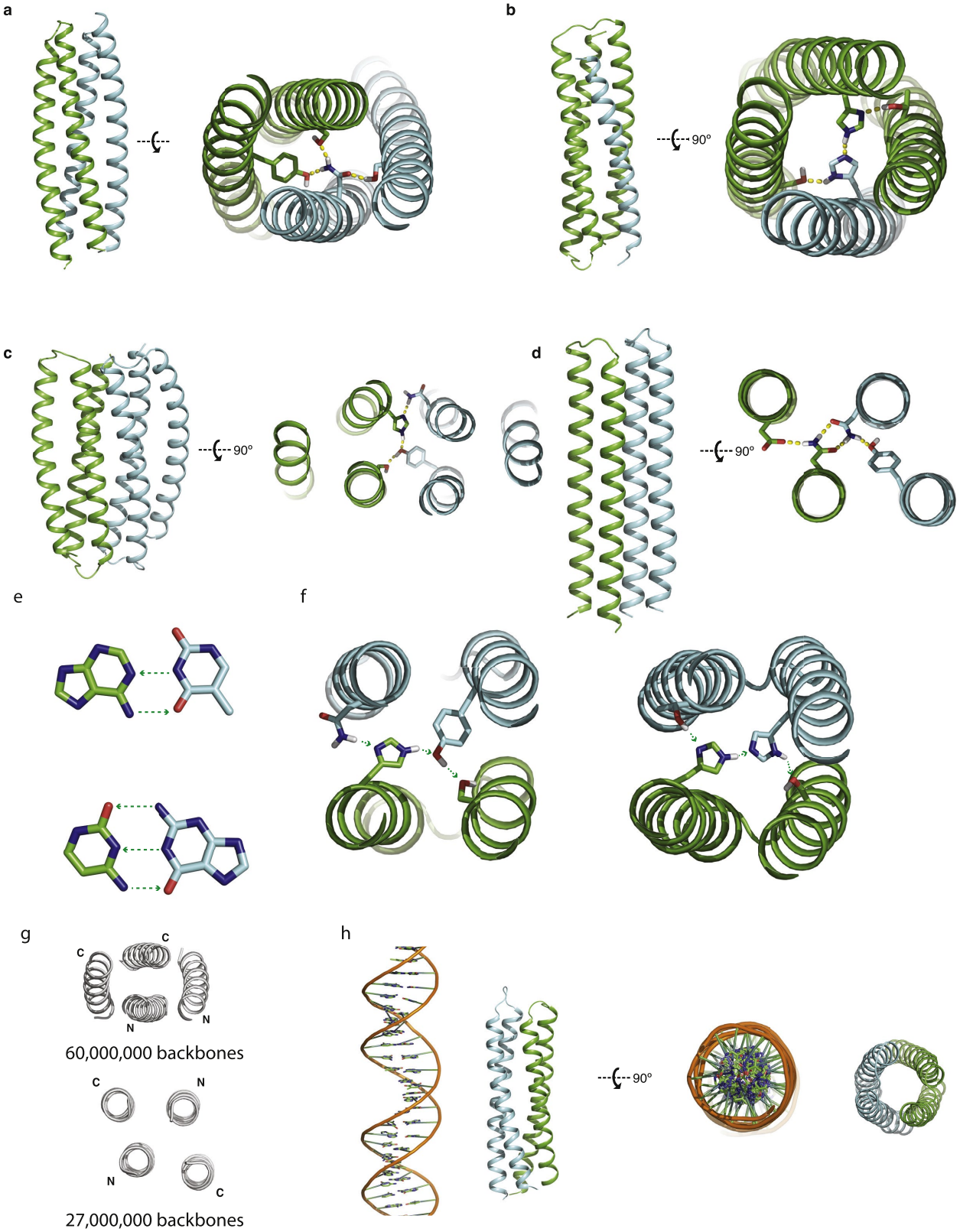
**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

**Code availability.** All program code is in Rosetta or can be downloaded from the Github repository at https://github.com/uagaug/DeNovoHeterodimers.

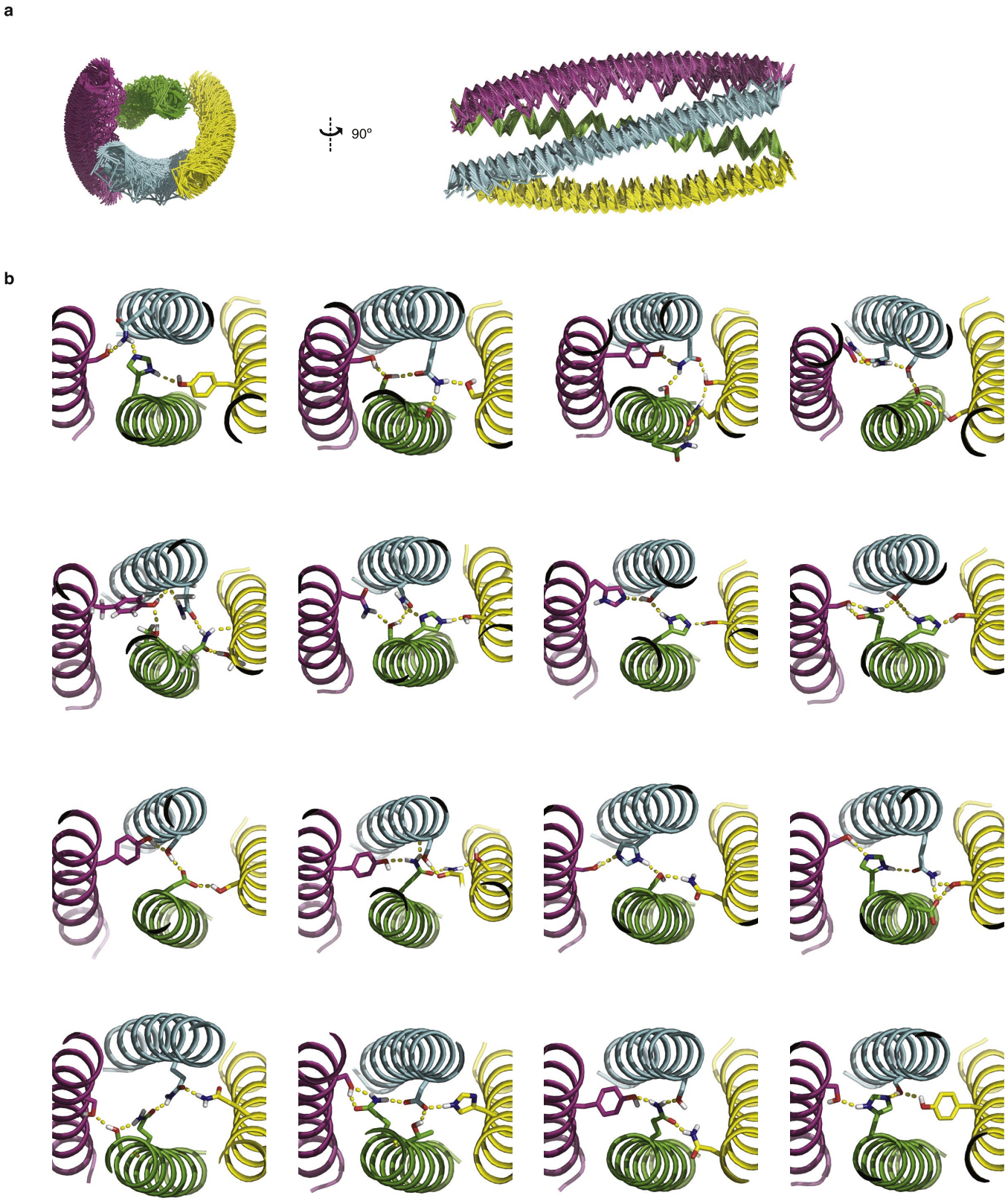
## Data availability

Coordinates and structure files have been deposited in the Protein Data Bank with accession codes: 6DMP (DHD13_XAAA), 6DKM (DHD131), 6DLC (DHD37_1:234), 6DLM (DHD127), 6DMA (DHD15 heterodimer) and 6DM9 (DHD15 heterotetramer). The native MS spectra generated and analysed during the current study are available at http://files.ipd.uw.edu/pub/de_novo_heterodimers_2018/180813_native_ms_raw.zip. Raw X-ray diffraction images have been deposited at https://proteindiffraction.org/. All source data are available upon request.

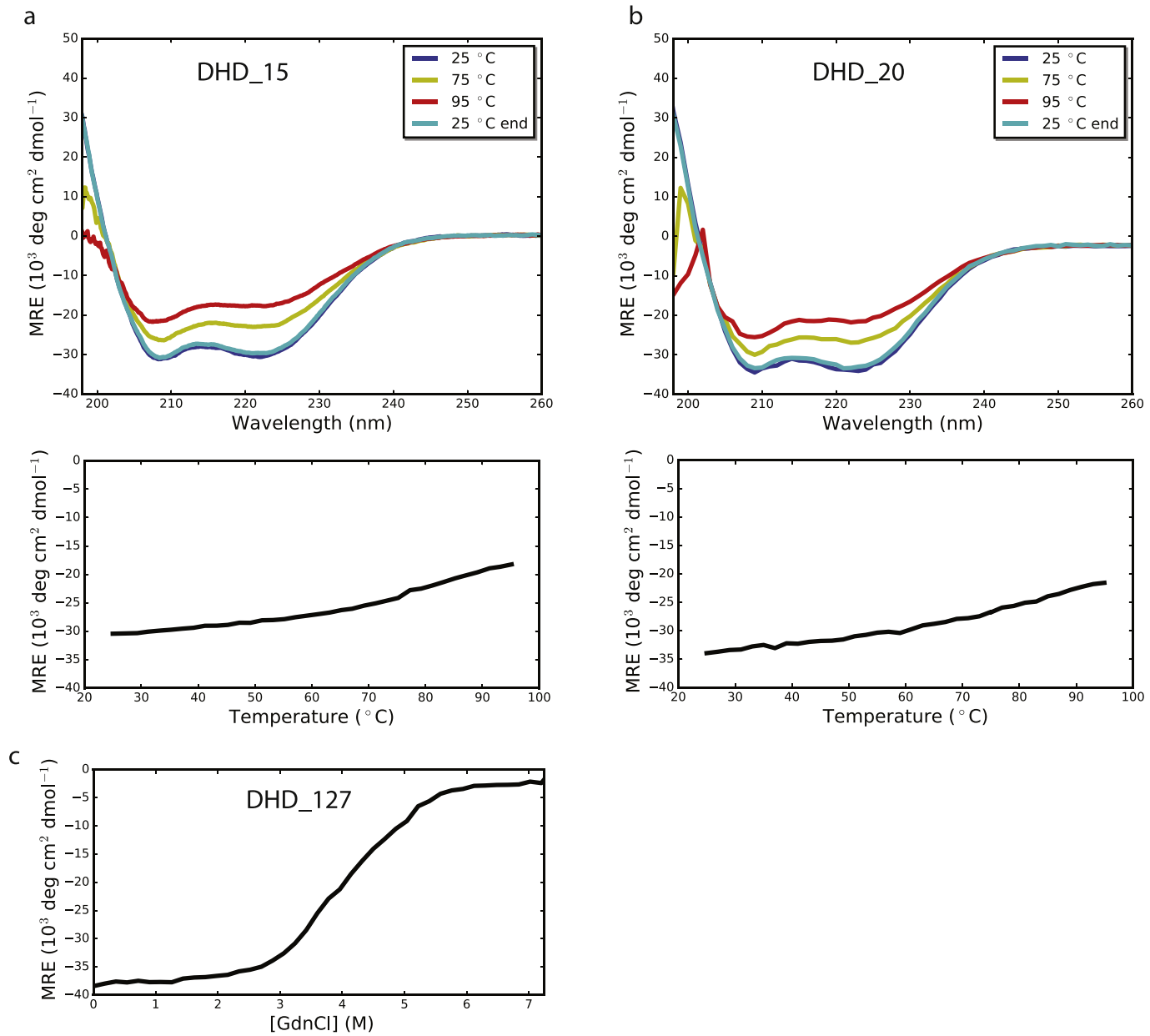

30. Zhang, Y. & Skolnick, J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33**, 2302–2309 (2005).
31. Rocklin, G. J. et al. Global analysis of protein folding using massively parallel design, synthesis, and testing. *Science* **357**, 168–175 (2017).
32. Schrödinger. The PyMOL Molecular Graphics System, Version 1.8. (2015).
33. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
34. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
35. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
36. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
37. Afonine, P. V. et al. Joint X-ray and neutron refinement with phenix.refine. *Acta Crystallogr. D* **66**, 1153–1163 (2010).
38. Terwilliger, T. C. et al. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. D* **64**, 61–69 (2008).
39. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
40. Davis, I. W. et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383 (2007).
41. Dyer, K. N. et al. High-throughput SAXS for the characterization of biomolecules in solution: a practical approach. *Methods Mol. Biol.* **1091**, 245–258 (2014).
42. Rambo, R. P. & Tainer, J. A. Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers* **95**, 559–571 (2011).
43. Schneidman-Duhovny, D., Hammel, M. & Sali, A. FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res.* **38**, W540–W544 (2010).
44. Schneidman-Duhovny, D., Hammel, M., Tainer, J. A. & Sali, A. Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophys. J.* **105**, 962–974 (2013).
45. Schiestl, R. H. & Gietz, R. D. High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. *Curr. Genet.* **16**, 339–346 (1989).
46. Chien, C. T., Bartel, P. L., Sternglanz, R. & Fields, S. The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. *Proc. Natl Acad. Sci. USA* **88**, 9578–9582 (1991).
47. Bartel, P. L., Roecklein, J. A., SenGupta, D. & Fields, S. A protein linkage map of *Escherichia coli* bacteriophage T7. *Nat. Genet.* **12**, 72–77 (1996).
48. Guzmán, C., Bagga, M., Kaur, A., Westermarck, J. & Abankwa, D. ColonyArea: an ImageJ plugin to automatically quantify colony formation in clonogenic assays. *PLoS ONE* **9**, e92444 (2014).
49. Dyachenko, A. et al. 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**, 6095–6102 (2015).
50. Waitt, G. M., Xu, R., Wisely, G. B. & Williams, J. D. Automated in-line gel filtration for native state mass spectrometry. *J. Am. Soc. Mass Spectrom.* **19**, 239–245 (2008).
51. VanAernum, Z. et al. Surface-induced dissociation of noncovalent protein complexes in an extended mass range Orbitrap mass spectrometer. Preprint available at https://doi.org/10.26434/chemrxiv.7415603.v1 (2018)
52. Marty, M. T. et al. Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. *Anal. Chem.* **87**, 4370–4376 (2015).
53. Bern, M. et al. Parsimonious charge deconvolution for native mass spectrometry. *J. Proteome Res.* **17**, 1216–1226 (2018).
54. Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195–202 (1999).



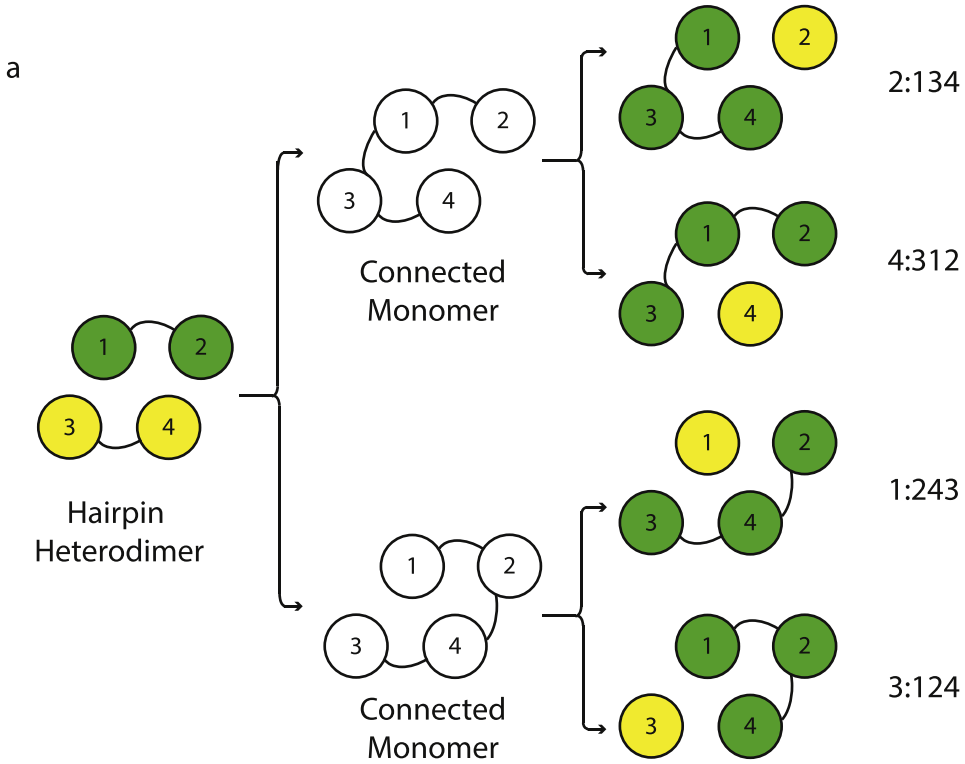

**Extended Data Fig. 1 | Overview of different topologies designed.** **a**–**d**, Overall topologies on the left and example HBNets on the right. **a**, A left-handed supercoiled backbone, with each monomer being helix hairpins. **b**, A backbone-permuted '3 + 1' design; one monomer is a single helix and the other is a three-helix bundle. **c**, A left-handed supercoiled backbone, with each monomer being a three-helix bundle. **d**, A straight, untwisted backbone, with each monomer being a helix hairpin. **e**, Hydrogen-bond pairing in DNA bases. Top, A–T base pairing. Bottom, C–G base pairing. Green arrows point from hydrogen-bond donors to acceptors. **f**, Two examples of hydrogen-bond pairing in designed protein hydrogen-bond networks. **g**, Top-down view of antiparallel twisted (top) and parallel untwisted (bottom) backbones sampled in this study. **h**, Comparison of a designed protein heterodimer (right) with B-form DNA (left) on the same scale.

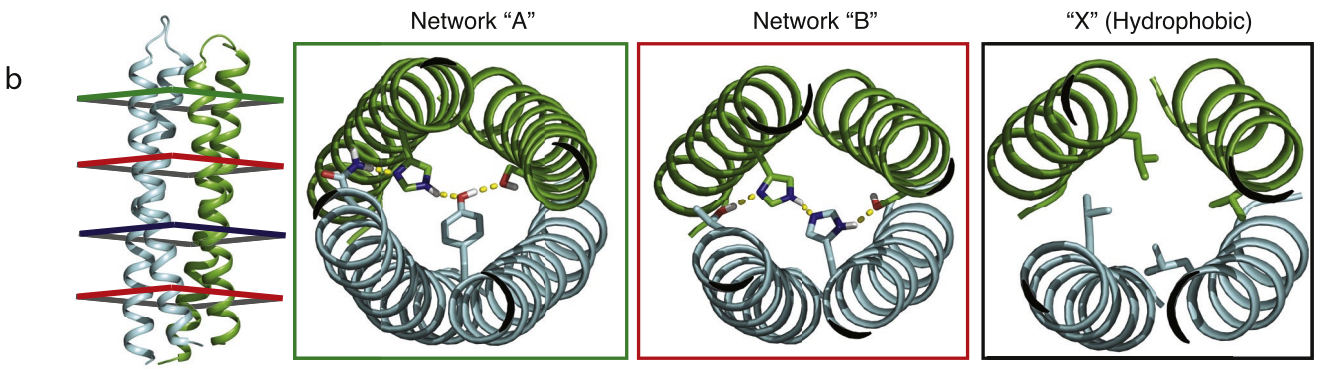



**Extended Data Fig. 2 | Example HBNets resulting from the systematic search. a**, Overlay of 50 backbones with different Crick parameters for each helix. **b**, Example hydrogen-bond networks from the systematic search, each involving at least four residues and contacting all four helices.

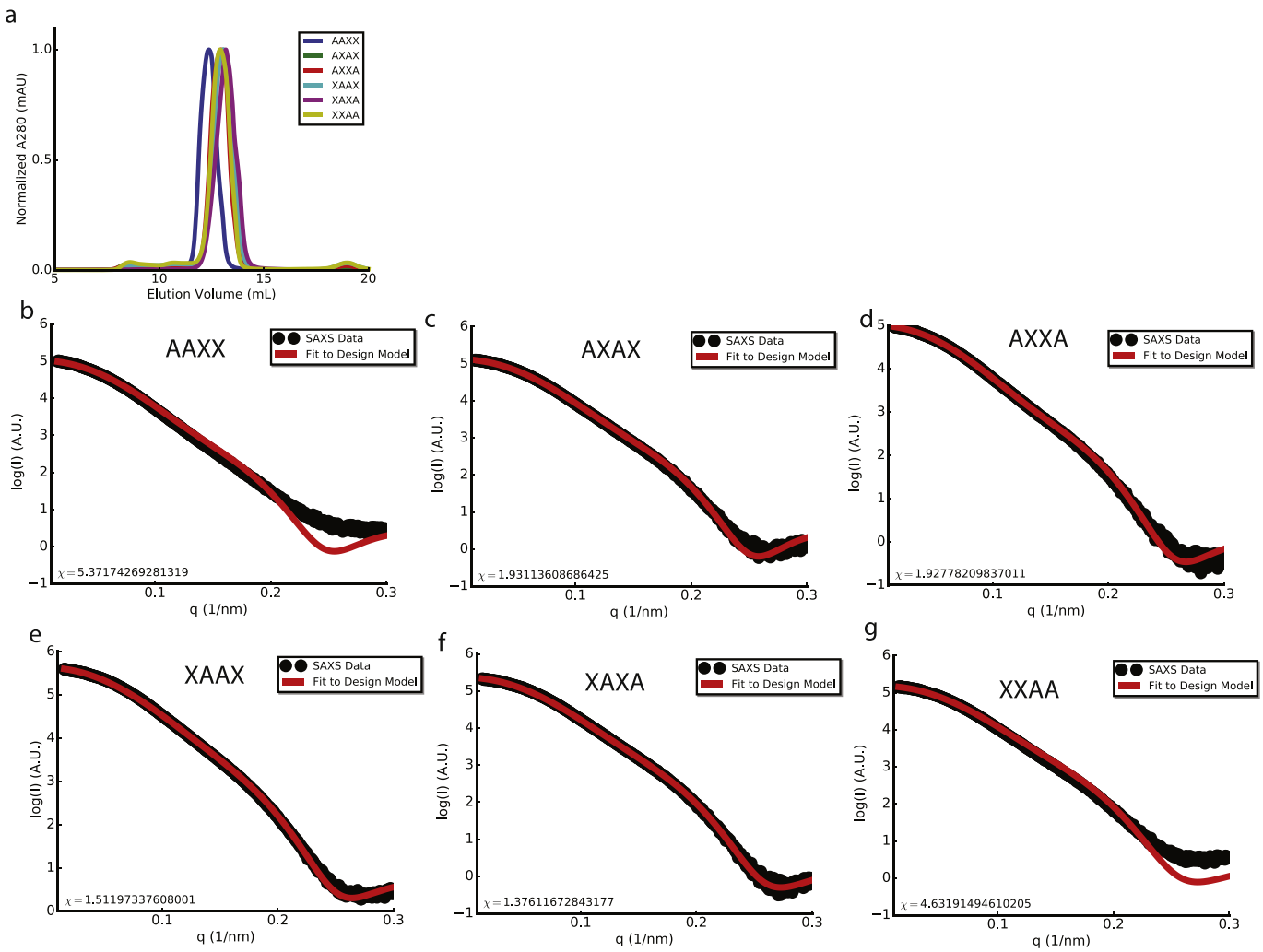



**Extended Data Fig. 3 | Thermal and chemical denaturation of DHDs.** **a**, **b**, CD spectra for thermal denaturation of DHD_15 and DHD_20, respectively. Top, wavelength scan at 25 °C, 75 °C, 95 °C and final 25 °C. Designs were α-helical and stable up to 95 °C. Bottom, CD temperature melts, monitoring absorption at 222 nm as temperature was increased from 25 °C to 95 °C. **c**, GdnHCl denaturation of DHD_127 measured by CD monitoring absorption at 222 nm. All CD experiments were performed once.

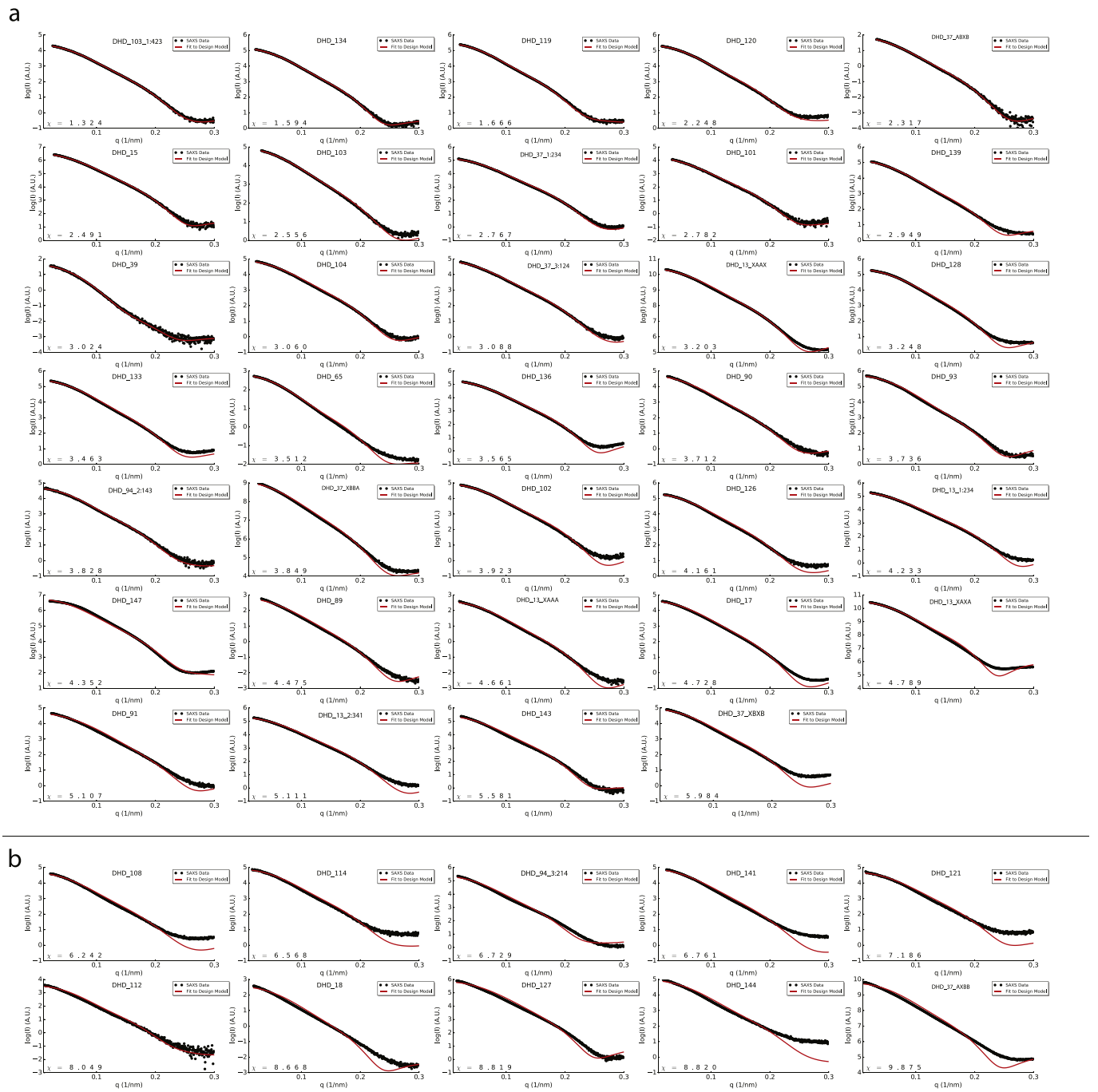


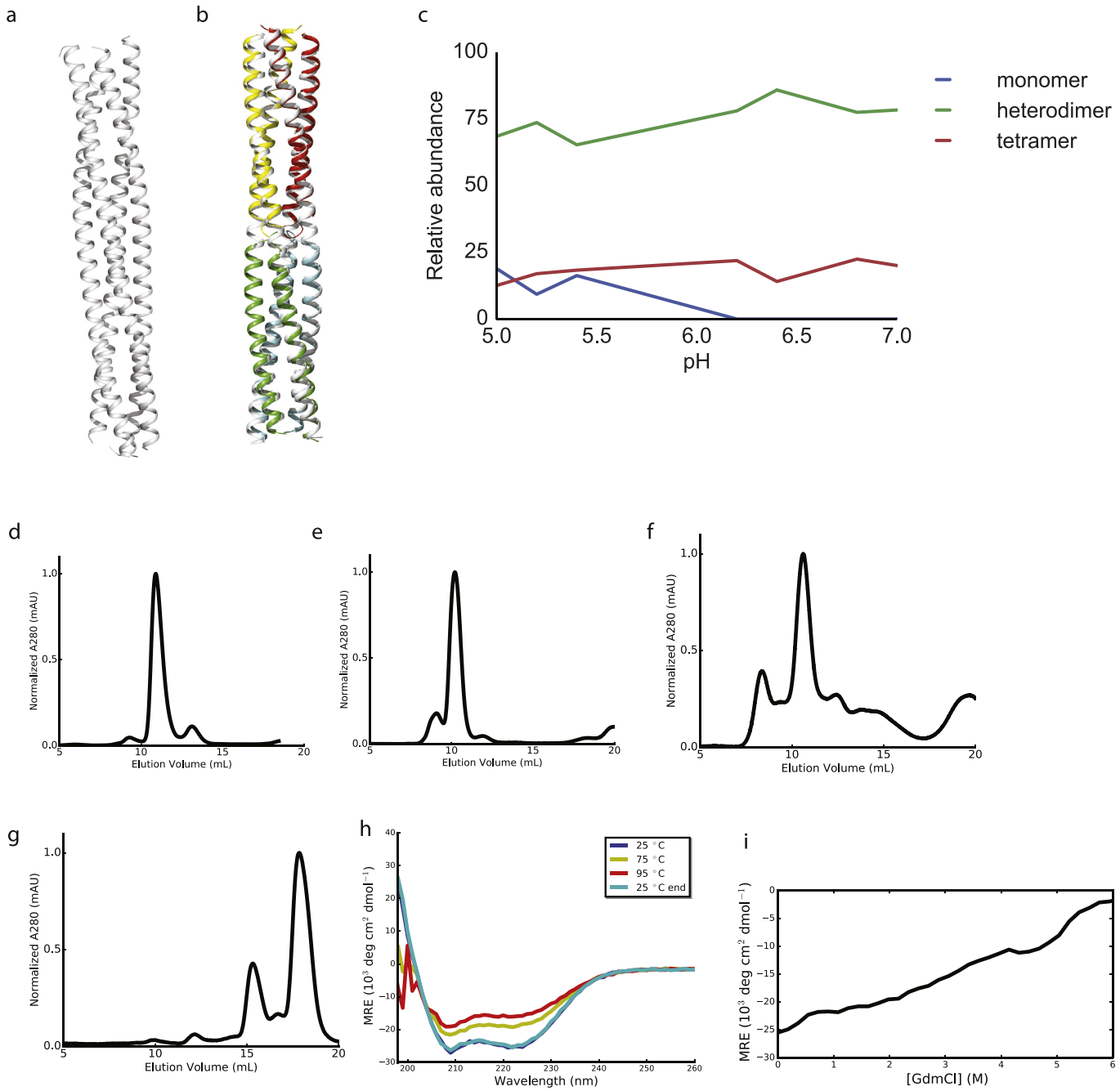

**Extended Data Fig. 4 | Backbone and hydrogen-bond network permutations. a**, On a 2 + 2 backbone (left), two loops were designed to connect the four helices into a single monomer in two different ways (middle), after which four different cut points were introduced to generate four possible backbone-permuted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



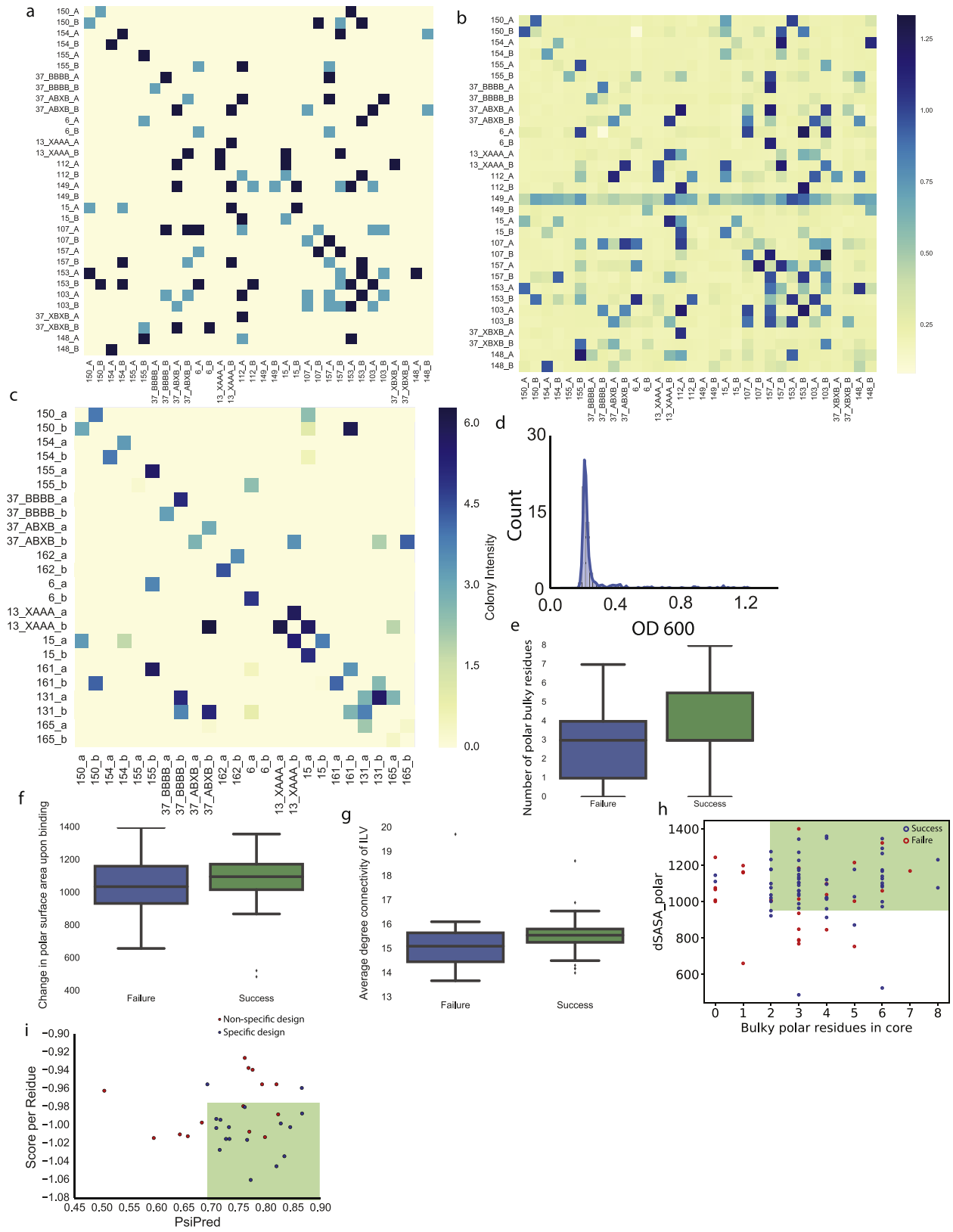

**Extended Data Fig. 5 | Biophysical characterization of hydrogen-bond-network-permuted homodimers. a**, SEC traces of all six homodimer designs. **b**–**g**, SAXS profiles of hydrogen-bond network-permuted homodimer designs. Black, experimental SAXS data; red, spectra computed from the designed backbones. Two (**a**) or one (**b**–**g**) biologically independent repeats were performed.




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

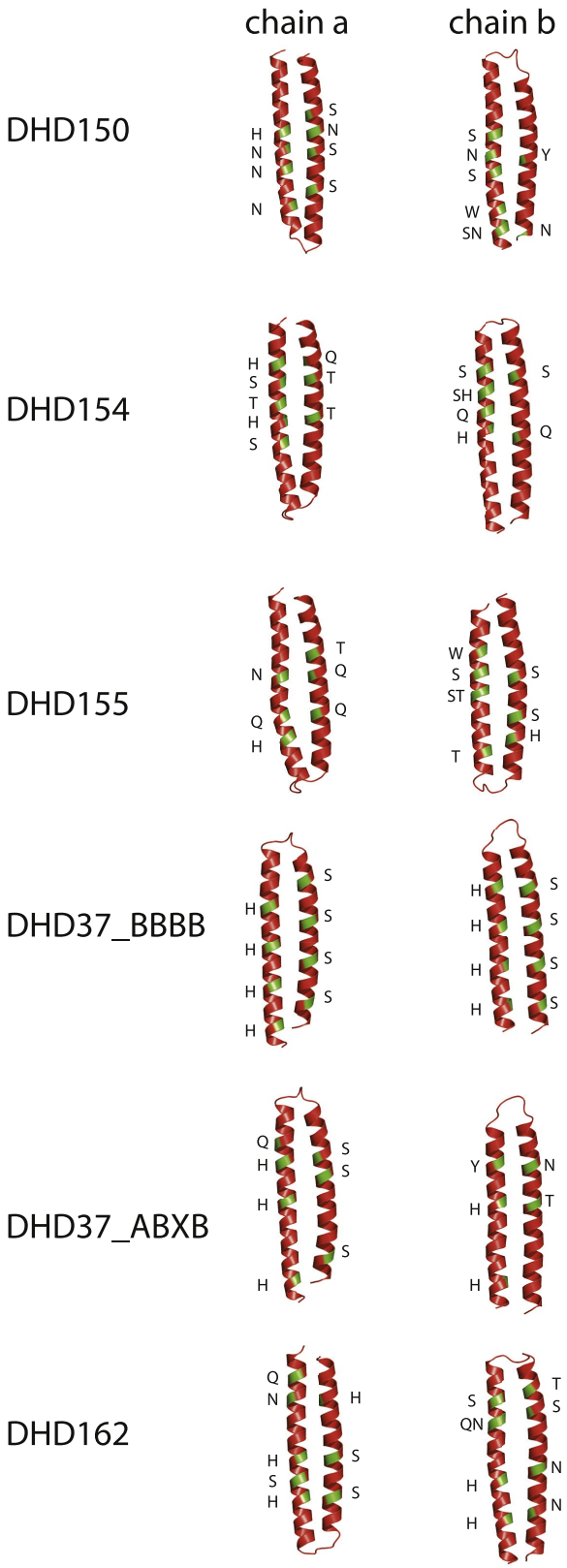



**Extended Data Fig. 7 | Crystal structure of the domain-swapped DHD_15 and biophysical characterization of higher-order oligomers.** **a**, Crystal structure of DHD_15 at pH 6.5, with 2.25 Å resolution. **b**, Superposition of design models (in colour) onto both halves of the crystal structure (in white), with backbone r.m.s.d. of 1.83 Å. **c**, Native MS study of DHD_15 at different pH values indicates that heterodimers, rather than heterotetramers, are dominant in solution. **d**–**g**, SEC traces of the induced dimerization DHD_9-13 fusion (**d**), DHD_15-37 fusion (**e**), DHD_13-37 fusion (**f**), and the scaffolding complex in Fig. 3d (**g**; the peak at around 15 ml corresponds to the fully assembled complex, followed by a peak representing an excess of individual components). **h**, 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 α-helical and stable up to 95 °C. **i**, CD chemical denaturation profile of the scaffolding complex in Fig. 3d. Two (**c**–**g**) or one (**h**, **i**) biologically independent repeats were performed.

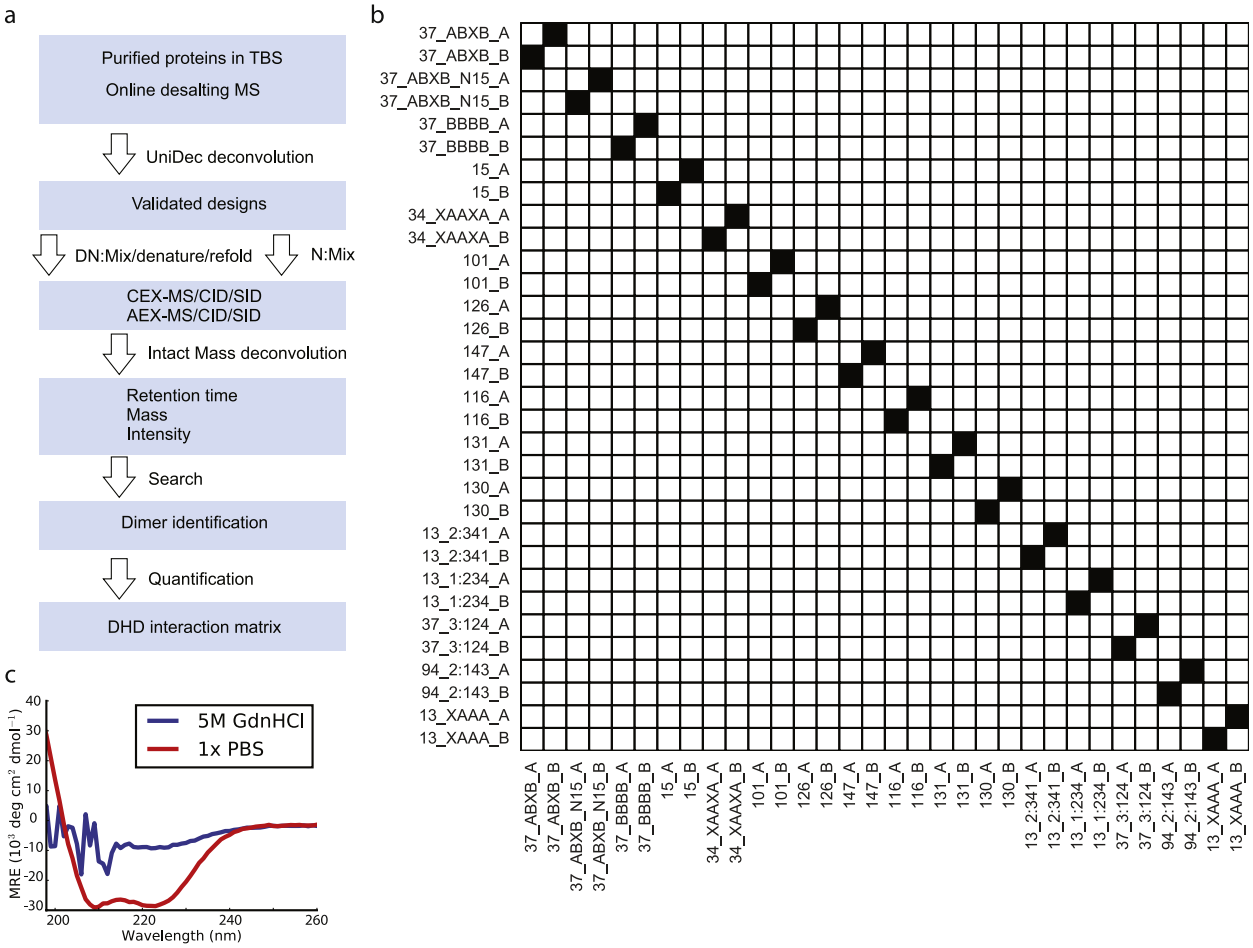



**Extended Data Fig. 8** | See next page for caption.

**Extended Data Fig. 8 | 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 on 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 medium containing 100 mM 3-AT, lacking tryptophan, leucine and histidine. $OD_{600}$ values were measured on day 2 to evaluate cell growth. **c**, An additional set of DHDs tested by Y2H showing improved orthogonality. **d**, Distribution of $OD_{600}$ values for non-cognate interactions in **b**. The majority of cells grew to $OD_{600} < 0.4$, indicating weak interactions for non-cognate binding. **e**–**g**, 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). $n = 88$; 25th, 50th and 75th percentiles are shown in the box with the centre being median, extended to 1.5 × interquartile range (IQR) beyond the box. **e**, The number of buried bulky polar residues correlates strongly with design success. **f**, Successful designs tend to have a bigger polar interface surface area. **g**, 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. **h**, Contribution of bulky residues and hydrogen-bond networks to specific dimer formation. dSASA_polar measures interface hydrophilicity and correlates positively 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 one polar bulky residue buried in the core includes most of the successful designs and excludes most of the design failures. **i**, On the basis of the Y2H data in **b**, all 32 monomers from the 16 pairs were categorized as being specific (blue, has $\leq$1 non-cognate binding), or non-specific (red, has $>$1 non-cognate binding). With application of secondary structure prediction scores (PsiPred[54]) 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). Two independent experiments were performed (**a**–**c**).




**Extended Data Fig. 9 | Hydrogen-bond network sequence motifs of the set of six orthogonal pairs in Y2H experiments.** Green patches mark the locations of hydrogen-bond network-forming residues on the backbones. Letters along the backbones indicate residue identities.




**Extended Data Fig. 10** | See next page for caption.

**Extended Data Fig. 10 | The workflow of native MS mixing experiments. a**, Protein samples were characterized using online desalting coupled to native MS and deconvoluted using UniDec software. Proteins showing 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 5 M GdnHCl at 75 °C and refoled into 150 mM AmAc; in the control mixing experiment (N), denaturation and refolding steps were omitted. Sample mixtures in each group were further equally divided into three parts that were individually injected on LC–MS with cation exchange and anion exchange, respectively, coupled with CID or SID. 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 using 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**, CD data for a mixture of purified DHDs in PBS (red) or 5 M GdnHCl and 75 °C (blue). Protein mixture was fully denatured under the latter conditions. **d**, A mixing experiment of DHD_37_ABXB and $^{15}$N-labelled DHD_37_ABXB with (red) or without (black) the denaturation and refolding steps. MS peaks merged after subunit exchange owing to the similarity in the masses of $^{15}$N-labelled and unlabelled subunits. Two biologically independent experiments were performed (**b**–**d**).

nature research

Corresponding author(s): David Baker

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

## Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☒ | ☐ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☒ | ☐ | A description of all covariates tested |
| ☒ | ☐ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☒ | ☐ | A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☒ | ☐ | For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ | Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |
| ☐ | ☒ | Clearly defined error bars<br>*State explicitly what error bars represent (e.g. SD, SE, CI)* |

*Our web collection on statistics for biologists may be useful.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | Rosetta software suite was used to perform protein design calculations, Rosetta is freely available for academic users on GitHub; TMalign (Version 5/21/2016) was used to perform structural alignment. |
| Data analysis | Custom Python code was written to perform data analysis and can be downloaded from https://github.com/uagaug/DeNovoHeterodimers, along with other command line scripts; Crystallographic data were analyzed with Phenix (release 1.101.1-2155) and Coot (v0.8.7 EL); Protein models were visualized using PyMOL 2.0; ImageJ 1.48v was sued to analyze Y2H colonies; Sparky 3.115 was used to analyze NMR spectra. |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Coordinates and structure files have been deposited to the Protein Data Bank with accession codes: 6DMP (DHD13_XAAA), 6DKM (DHD131), 6DLC (DHD37_1:234), 6DLM (DHD127), 6DMA (DHD15 heterodimer), 6DM9 (DHD15 heterotetramer).
The native MS spectra generated during and/or analysed during the current study are available at http://files.ipd.uw.edu/pub/de_novo_heterodimers_2018/180605_native_ms_raw_spectra.tar.gz
Raw X-ray diffraction images have been deposited at proteindiffraction.org

# Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Sample size | Heterodimers were designed by the Rosetta software suite and designs passing computational selection criteria were experimentally tested. Sample size was determined by the estimated work load. In total, 158 designed proteins were tested. Sample size was considered sufficient since no significant improvement of the success rate was observed for subsequent iterations of design/characterization. |
| Data exclusions | No data were excluded from the analysis. |
| Replication | At least two biological replicates were done for yeast-2-hybrid (Y2H) assay of individual proteins. Three technical replicates were performed for the all-against-all Y2H assays. Three technical replicates were performed for the all-against-all native MS mixing assays. The replicated experiments produced similar results. CD experiments were performed once due to the expected overall stability of these proteins. |
| Randomization | No randomized samples allocation in this work. Because of the complexity of the designs, it is unlikely that the folding amino acids sequences reported in this paper are due to chance rather than design calculation. Negative and positive controls were well defined and validated. |
| Blinding | Not relevant to the study since the experiments were well-defined by the computational design. |

# Reporting for specific materials, systems and methods

### Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ Unique biological materials |
| ☒ | ☐ Antibodies |
| ☒ | ☐ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology |
| ☒ | ☐ Animals and other organisms |
| ☒ | ☐ Human research participants |

### Methods

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ ChIP-seq |
| ☒ | ☐ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |