HHS Public Access

Author manuscript

Science. Author manuscript; available in PMC 2019 June 21.

Published in final edited form as:

Science. 2018 February 23; 359(6378): 930–935. doi:10.1126/science.aam7229.

Structural principles that enable oligomeric small heat-shock protein paralogs to evolve distinct functions

Georg K.A. Hochberg[1,2,3], **Dale A. Shepherd**[1,2,4], **Erik G. Marklund**[1,2,5], **Indu Santhanagoplan**[6], **Matteo T. Degiacomi**[1,7], **Arthur Laganowksy**[1,8], **Timothy M. Allison**[1], **Eman Basha**[2,9], **Michael T. Marty**[1,10], **Martin R. Galpin**[1], **Weston B. Struwe**[1], **Andrew J. Baldwin**[1], **Elizabeth Vierling**[2], and **Justin L.P. Benesch**[1,*]

[1]Department of Chemistry, Physical & Theoretical Chemistry Laboratory, University of Oxford, Oxford, OX1 3QZ, U.K.

[3]Present addresses: Department of Ecology and Evolution, University of Chicago, Illinois 60637 U.S.

[4]Waters Corporation, Stamford Ave., Wilmslow, SK9 4AX, U.K.

[5]Department of Chemistry - BMC, Uppsala University, Box 576, 75123, Uppsala, Sweden.

[6]Department of Biochemistry & Molecular Biology, University of Massachusetts, Amherst, MA 01003, U.S.A.

[7]Department of Chemistry, Durham University, South Rd, Durham, DH1 3LE, U.K.

[8]Center for Infectious and Inflammatory Diseases, Institute of Biosciences and Technology, Texas A&M Health Science Center, Houston, Texas 77030, U.S.; Department of Chemistry, Texas A&M University, College Station, Texas 77842, U.S.; Department of Microbial Pathogenesis & Immunology, College of Medicine, Texas A&M Health Science Center, Bryan, Texas 77807, U.S.

[9]Botany Department, College of Science, Tanta University, Tanta, Egypt.

[10]Department of Chemistry and Biochemistry University of Arizona, 1306 E. University Blvd, Tucson, AZ 85721, U.S.

Abstract

Oligomeric proteins assemble with remarkable selectivity, even in the presence of closely related proteins, in order to perform their cellular roles. We show that most proteins related by gene duplication of an oligomeric ancestor have evolved to avoid hetero-oligomerization, and that this correlates with their acquisition of distinct functions. We report how co-assembly is avoided by two oligomeric small heat-shock protein paralogs. A hierarchy of assembly, involving intermediates that are populated only fleetingly at equilibrium, ensures selective oligomerisation. Conformational flexibility at non-interfacial regions in the monomers prevents co-assembly, allowing interfaces to remain largely conserved. Homomeric oligomers must overcome the

[*]Correspondence to justin.benesch@chem.ox.ac.uk, +44 1865 285420.

[2]These authors contributed equally.

Supplementary Materials

Materials and Methods; Supplementary Text; Figs. S1 to S15; Tables S1 to S2; References (19–77); Data S1 to S2.

entropic benefit of co-assembly and, accordingly, homomeric paralogs comprise fewer subunits than homomers that have no paralogs.

One sentence summary:

Small heat-shock proteins avoid dysfunctional co-assembly using mechanisms that cause minimal disruption to their conserved interfaces

---

Many proteins associate into selective homo- or heteromers in order tofunction (1). New assemblies are most often created by gene duplication of a pre-existing homomer (2). The resulting oligomeric paralogs initially co-assemble because both have the same sequence (and hence structure and interfaces) as their ancestor (Fig. 1A) (3). This co-assembly can easily become entrenched if evolution of the two resulting duplicates is functionally constrained to maintain the interaction (4, 5), implying that heteromerisation should be the most likely fate of oligomeric paralogs. However, when we interrogated the human, *Arabidopsis,* yeast, and *E. coli* interactomes (Supplementary Materials, Data S1), we found that the majority of oligomeric paralogs in fact do not form heteromers (i.e do not co-assemble) (Fig. 1B), despite overlapping localization and expression profiles (Fig S1A,B). Moreover, we found that when paralogs cannot co-assemble, they share lower sequence identity and fewer common functions than paralogs that can (Fig. 1C,D). This suggests that heteromerisation acts as a constraint on the functional divergence of oligomeric paralogs (6). Relieving this constraint is therefore a key step in the evolutionary trajectories of oligomeric proteins towards evolving novel functions.

To interrogate how this occurs, we examined the selective assembly of two paralogous small heat-shock proteins (sHSPs), molecular chaperones found across the tree of life that are key to the cell's ability to respond to stress (7, 8). A duplication event led to land plants having two classes of cytosolic sHSPs (class-1 and −2, Fig. 1E, S2) that both assemble as dodecamers but cannot form heteromers between classes (9). Both are required for thermo-tolerance *in vivo* (10), and have different mechanisms of action (11, 12). We chose one paralog of each class from *Pisum sativum,* HSP18.1 and HSP17.7 (hereafter WT-1 and WT-2, respectively). Both proteins comprise an N-terminal region, an α-crystallin domain and a C-terminal tail, and both form homo-12-mers (12) using three independent interfaces: the α-crystallin domain mediates the formation of an isologous α·α dimer, these dimers assemble into oligomers through heterologous contacts between the α - crystallin domain and the C-terminal tails from neighbouring dimers (α·C), and interactions between the N-terminal regions (N·N) (Fig. 1F) (13). Their complex, multi-interface architecture makes these proteins an ideal system to investigate how evolution acts to regulate the biophysical properties of oligomers to develop a set of selective interfaces that allows them to diverge functionally.

Small-angle X-ray scattering experiments indicated that both proteins form tetrahedral oligomers (Fig. S3), implying that there are no major differences in quaternary structure that prevent co- assembly. Nonetheless, when we obtained native mass spectra of a mixture of WT-1 and WT-2 after prolonged incubation (Fig. 1G, upper) or initiating re-assembly from their subunits (Fig. S4A), we failed to detect any hetero-12-mers in either case. However,

both homo-12-mers underwent continual dissociation and re-association, though WT-1 did so >10 times faster than WT-2 (Fig S4). These facile quaternary dynamics show that heteromers are in principle kinetically accessible and so, despite the similarity in quaternary architectures of WT-1 and WT-2, must be thermodynamically unfavourable.

To identify the sequence-determinants of selective assembly, we aligned class-1 and −2 sHSPs, and noted conserved differences in their C-terminal tails (Fig. S5). We then engineered a chimera with the class-1 N-terminal region and α–crystallin domain linked to the class-2 C-terminal tail ($^{N}1^{\alpha}1^{C}2$, see Table S1) and incubated it with WT-2. This small change in sequence produced a series of hetero-12-mers formed between WT-2 and $^{N}1^{\alpha}1^{C}2$ (Fig. 1G, lower). These represent a proxy for class-1 and −2 co-assembly, and allowed us to interrogate the functional consequences of heteromerisation. We incubated purified sHSPs with pea leaf lysate under heat-shock conditions toform reversible aggregates (14), mimicking their action *in vivo* (10, 11). WT-2 partitioned significantly faster into the insoluble fraction than WT-1 (Fig. 1H, S6). The rate measured for the heteromers of $^{N}1^{\alpha}1^{C}2$ and WT-2, however, was intermediate to WT-1 and WT-2 homomers. The functional differentiation of the two proteins therefore depends on their selective homomerisation, demonstrating the operational necessity of avoiding co-assembly.

The hetero-12-mers formed by swapping C-terminal tails comprised only even numbers of each type of subunit (Fig. 1G, lower), implying that either the α·α or the N·N interface must also be selective. To determine which, we engineered an N-terminal chimera, $^{N}2^{\alpha}1^{C}1$, and incubated it with WT-1. This produced a series of hetero-12-mers comprising odd and even numbers of each subunit (Fig. S7A). While N·N contacts therefore are not thermodynamically selective (and hence the α·α interface must be), we noticed that dissociation of $^{N}1^{\alpha}1^{C}2$ oligomers was as fast as WT-1 (Fig. S7B), whereas dissociation of $^{N}2^{\alpha}1^{C}1$, was slow (Fig. S7A). This means that the promiscuous N·N contacts, not the thermodynamically selective α·C and α·α interfaces, control the kinetic stability of the 12-mers.

Our subunit-exchange data indicate that, over the functional temperature range, hetero-12-mers formed via N·N contacts during assembly would decompose into homomers on the timescale of minutes to hours (Fig. S4E). Yet, we had observed no long-lived heteromers in our assembly experiment, even at low temperatures (Fig S4A). To resolve this apparent conflict, we generated constructs of WT-1 and WT-2 lacking the N-terminal region and measured their stoichiometries using native ion mobility mass spectrometry (IM-MS). Both were polydisperse, spanning dimers to 12-mers (Fig. 2A, S8A). Constructs instead lacking the C-terminus only formed monomers and dimers (Fig. 2B, S8B). α·C contacts therefore likely form early and ensure rapid self-selective oligomerisation, while N·N contacts subsequently stabilize the 12-meric fraction (Fig. S8C, see Supplemental Text). This hierarchy obviates the need for kinetically stable N·N contacts to be selective, and avoids long-lived heteromers that would compromise the rapid stress response of sHSPs in the cell.

To understand the thermodynamic basis of selectivity at the α·C interface, we examined chimeric versions of the N-terminal truncations. $^{\alpha}1^{C}2$ formed polydisperse oligomers, but $^{\alpha}2^{C}1$ did not assemble beyond a dimer (Fig. 2C, S8D–F). Selectivity in the α·C interface is

therefore directional, arising from an unfavourable association between the WT-1 C-terminal tail and WT-2. We quantified this effect directly by excising the core domains of both proteins ($^{\alpha}1$ and $^{\alpha}2$, Table S1) and measuring their affinity for each other's C-terminal tails. Whereas $^{\alpha}1$ bound peptides mimicking each tail equally well, $^{\alpha}2$ had a much lower affinity for a WT-1 than WT-2 peptide ($\Delta\Delta G$ >6 kJmol^{-1}, Fig. S9).

We next turned our attention to the α·α interface, which is selective (Fig. S10A) despite high sequence conservation (Fig. S5B). Crystal structures revealed $^{\alpha}1$ and $^{\alpha}2$ to be extremely alike (Fig. 3A,B, Table S2). The dimer interface is formed in both homodimers by salt bridges centred on the β8- β9 loop (L8/9) that are fully conserved between the two proteins; and by reciprocal strand-exchange between β6 and β2. The latter involves only one obvious class-specific contact: between the π-systems of a histidine on β6 and a tryptophan on β2 in WT-1 that is absent in WT-2 (Fig. 3C,D). In 2-μs molecular dynamics (MD) simulations, both homodimers and a modelled heterodimer were stable. The interfaces of the heterodimer featured equivalent overall numbers of interacting side-chains, hydrogen bonds, and level of structural flexibility compared to both homodimers (Fig. S10B–E, S11). Remarkably, the α-crystallin domain is therefore selective with only minimal differences in the number or type of contacts at its interface.

To investigate the origin of this selectivity, we performed calorimetric measurements and found that there are differences in the relative contributions from entropy and enthalpy to the favourable free energy of dimerization in $^{\alpha}1$ and $^{\alpha}2$ (Fig. S12A–C). This suggests subtle differences in their association mechanisms that may impart selectivity. To quantify which parts of the dimer are responsible for selectivity, we divided the core domain into three segments (Fig. 3E, Table S1): the β-sandwich (S), which includes the L8/9 interface and β2 from the β6·β2 interface; β6 (Bg; and the loop (L) connecting β6 to the β-sandwich. We shuffled these segments between $^{\alpha}1$ and $^{\alpha}2$ (Fig. 3E) and, for the 36 pairwise combinations of chimeric and wild-type constructs, determined the corresponding free energy of dimerization, $\Delta G_{\alpha\cdot\alpha}$ by performing quantitative IM-MS titration experiments (Fig. S12D–G). From the overall dataset, we identified statistically significant intermolecular interactions between β6 and the β-sandwich (B·S), and the loop and the β-sandwich (L·S). Summed (B+L·S, Fig 3F), these interactions contribute ≈11 kJmol^{-1} to the stability of the dimer, except when $^{S}2$ encounters $^{B}1^{L}1$, which unilaterally destabilizes the dimer by ≈7 kJmol^{-1} (Fig. 3F, left). The L·S and B·S components contribute nearly equally to dimer stability (Fig. 3F, middle and right), a surprising observation considering that the loop is not part of the interface.

Because the $^{\alpha}1$ and $^{\alpha}2$ dimer structures did not reveal differences that account for our experimental thermodynamic data, we performed steered MD simulations in which we gradually detached β6 from β2, and estimated the resulting free-energy profile (Fig. 3G). As predicted by our thermodynamic data, we found that the heteromeric $^{B+L}1^{\cdot S}2$ interface was significantly easier to break than the other combinations. We also noticed that in unconstrained simulations of the $^{\alpha}1$ monomer (performed in triplicate) the β-sandwich remained rigid (Fig. 3H, S13A,C,D), while the loop distorted and formed intra-molecular contacts (Fig S13D). In the $^{\alpha}2$ monomer, the loop more closely retained its conformation

from the dimer (Fig. 3I, S13B–D), but β2 detached from the β-sandwich and became highly flexible (Fig S13C,D,E).

Our data imply that the loop in $^{\alpha}1$, and β2 in $^{\alpha}2$, have a propensity to sample conformations in the monomers that are limited upon formation of a dimer interface (Fig. S13D). In both homodimers only one side of each B+L·S interface is restrained in this way, while in the heterodimer both sides of the $^{B+L}1{\cdot}^{S}2$ interface are (Fig. 4), making it easier to break apart. Conversely, to dimerize, dynamic regions must undergo a structural transition from their monomeric conformations. In homodimers, only one side of each interface would have to do this, with the other being pre-ordered for dimerization. In a heterodimer, this conformational complementarity would be absent for the $^{B+L}1{\cdot}^{S}2$ interface, also leading to a slow association rate. These effects would therefore combine to discourage the formation of heterodimers and instead ensure self-selection.

If this mechanism is correct, with the loop making a large contribution to the instability of the heterodimer (Fig. 3E), it should be a major regulator of the monomeric structure. Indeed, the conformations of simulated chimeric monomers lie between the extremes occupied by $^{\alpha}1$ and $^{\alpha}2$, and the segment that shifts the structure the most is the loop, not the interfacial segments (Fig. S13F). Similarly, chimeric dimers incorporating segments that do not change conformations in our simulations ($^{S}1$, $^{B}2$, and $^{L}2$, Fig. 3E) should be more stable than both $^{\alpha}1$ and $^{\alpha}2$. This prediction is borne out in their experimental melting temperature being ≈5 °C higher (Fig. S13G).

We mined our MD trajectories for specific contacts that were more abundant in one class over the other, and identified 11 and 3 that involved residues that displayed class-specific evolutionary conservation in $^{\alpha}1$ and $^{\alpha}2$, respectively. Strikingly, we found that the majority of these are outside of the dimer interface: in $^{\alpha}1$ 7 out of 11 conserved sites either attach β2 to the sandwich or promote curling of the loop, while in $^{\alpha}2$ one maintains an extended loop conformation (Fig. S14), and another makes β2 prone to detach in the monomer. Thus non-interfacial regions, and their effects on the structure of dissociated monomers, determine selectivity in the α-crystallin domain of class-1 and −2 sHSPs across land plants. This is consistent with the observation that non-interfacial residues can affect interface stabilities (15).

To homomerize, paralogs must overcome a substantial entropic benefit of co-assembly arising from the number of ways distinguishable subunits can be arranged. This mixing entropy increases with the number of subunits in the oligomer such that the energetic cost of homomerization rises logarithmically (Fig. 4A, Supplementary Text). Combining this contribution with the strength of interactions we quantified experimentally, allowed us to generate a model predicting the stability of all possible combinations of the two sHSPs and their chimeras, dependent only on their stoichiometry and constituent α·C and α·α interfaces (Supplementary Text, Fig. S15). We used this model to calculate the difference in stability between every possible heteromer and the corresponding homomers along the assembly pathway (Fig. 4B). The selective interactions in the α·C and α·α interfaces narrowly overcome the entropic benefit of co-assembly for all stoichiometries (Fig. 4C),

resulting in a predicted population of hetero-12-mers at equilibrium that is just below detectable levels (Fig. 4C, right).

Homomers are therefore only marginally more stable than heteromers, even though the paralogs have diverged for >400 million years (16). The number and type of selective interactions we found is the minimum required for a tetrahedron (17), with half of the oligomeric interfaces (N·N and those involving $^{C}2$) remaining promiscuous. These observations imply that selectivity is difficult to evolve, perhaps because most substitutions that disfavour co-assembly, also disfavour self-assembly (18).

Our model predicts that this would be more problematic for oligomers with more subunits, for which the entropic barrier to self-assembly is higher (Fig. 4A). Using a dataset of oligomeric architectures based on curated crystal structures (17) and combining it with our list of paralogs (Fig. 1B, Data S2), we found that self-selective paralogs comprise fewer subunits than homomers that have no paralogs (Fig. 4D). The data are well explained by the probability that selectivity evolves after duplication being inversely proportional to the mixing entropy (Supplemental Text). Applying this relationship to scale the stoichiometry distribution of oligomers without paralogs renders it indistinguishable from the self-selective set (Fig. 4D). This indicates that this fundamental thermodynamic bias acts as a significant evolutionary constraint across oligomeric proteins. The mechanisms for selectivity we have uncovered for the sHSPs studied here are some, of possibly many, ways in which proteins have evolved to escape co-assembly.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Carol Robinson, Jason Schnell, Philipp Kukura, David Staunton (all University of Oxford) and Brian Metzger (University of Chicago) for helpful discussions. We acknowledge access to B21 and help from Mark Tully and James Doutch at the Diamond Synchrotron (JLPB for SM9384–2); and the ARCUS cluster at Advanced Research Computing, Oxford. We thank the following funding sources: Engineering & Physical Sciences Research Council (GKAH for a studentship, JLPB for EP/J01835X/1); Carl Trygger's Foundation (EM); Swiss National Science Foundation (MTD for P2ELP3_155339) Biotechnology & Biological Sciences Research Council (AJB for BB/J014346/1, JLPB for BB/K004247/1 and BB/J018082/1); National Institutes of Health (EV for R01 GM42761); Massachusetts Life Sciences Center (EV for a New Faculty Research Award); Royal Society (JLPB for a University Research Fellowship). All data to support the conclusions is available in the manuscript or the Supplementary Materials, and is deposited with DOI 10.5287/bodleian:54jBVeAzw.

References and notes

1. Marsh JA, Teichmann SA, Structure, dynamics, assembly, and evolution of protein complexes. Annu Rev Biochem, (2014).
2. Pereira-Leal JB, Levy ED, Kamp C, Teichmann SA, Evolution of protein complexes by duplication of homomeric interactions. Genome Biol 8, R51 (2007). [PubMed: 17411433]
3. Kaltenegger E, Ober D, Paralogue interference affects the dynamics after gene duplication. Trends Plant Sci 20, 814–821 (2015). [PubMed: 26638775]
4. Diss G et al., Gene duplication can impart fragility, not robustness, in the yeast protein interaction network. Science 355, 630–634 (2017). [PubMed: 28183979]

5. Finnigan GC, Hanson-Smith V, Stevens TH, Thornton JW, Evolution of increased complexity in a molecular machine. Nature 481, 360–364 (2012). [PubMed: 22230956]
6. Baker CR, Hanson-Smith V, Johnson AD, Following gene duplication, paralog interference constrains transcriptional circuit evolution. Science 342, 104–108 (2013). [PubMed: 24092741]
7. Balch WE, Morimoto RI, Dillin A, Kelly JW, Adapting proteostasis for disease intervention. Science 319, 916–919 (2008). [PubMed: 18276881]
8. Richter K, Haslbeck M, Buchner J, The heat shock response: life on the verge of death. Mol Cell 40, 253–266 (2010). [PubMed: 20965420]
9. Helm KW, Lee GJ, Vierling E, Expression and native structure of cytosolic class II small heat-shock proteins. Plant Physiol 114, 1477–1485 (1997). [PubMed: 9276957]
10. McLoughlin F et al., Class I and II small heat-shock proteins protect protein translation factors during heat stress. Plant Physiol, (2016).
11. Kirschner M, Winkelhaus S, Thierfelder JM, Nover L, Transient expression and heat-stress-induced co-aggregation of endogenous and heterologous small heat-stress proteins in tobacco protoplasts. Plant J 24, 397–411 (2000). [PubMed: 11069712]
12. Basha E, Jones C, Wysocki V, Vierling E, Mechanistic differences between two conserved classes of small heat shock proteins found in the plant cytosol. J Biol Chem 285, 11489–11497 (2010). [PubMed: 20145254]
13. Hilton GR, Lioe H, Stengel F, Baldwin AJ, Benesch JL, Small heat-shock proteins: paramedics of the cell. Top Curr Chem 328, 69–98 (2013). [PubMed: 22576357]
14. Wallace EW et al., Reversible, specific, active aggregates of endogenous proteins assemble upon heat stress. Cell 162, 1286–1298 (2015). [PubMed: 26359986]
15. Perica T et al., Evolution of oligomeric state through allosteric pathways that mimic ligand binding. Science 346, 1254346 (2014). [PubMed: 25525255]
16. Waters ER, Vierling E, The diversification of plant cytosolic small heat shock proteins preceded the divergence of mosses. Mol Biol Evol 16, 127–139 (1999). [PubMed: 10331257]
17. Ahnert SE, Marsh JA, Hernandez H, Robinson CV, Teichmann SA, Principles of assembly reveal a periodic table of protein complexes. Science 350, aaa2245 (2015). [PubMed: 26659058]
18. Aakre CD et al., Evolving new protein-protein interaction specificity through promiscuous intermediates. Cell 163, 594–606 (2015). [PubMed: 26478181]
19. Chatr-Aryamontri A et al., The BioGRID interaction database: 2017 update. Nucleic Acids Res 45, D369–D379 (2017). [PubMed: 27980099]
20. Rajagopala SV et al., The binary protein-protein interaction landscape of Escherichia coli. Nat Biotechnol 32, 285–290 (2014). [PubMed: 24561554]
21. The UniProt C., UniProt: the universal protein knowledgebase. Nucleic Acids Res 45, D158–D169 (2017). [PubMed: 27899622]
22. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ, Basic local alignment search tool. J Mol Biol 215, 403–410 (1990). [PubMed: 2231712]
23. C. The Gene Ontology, Expansion of the Gene Ontology knowledgebase and resources. Nucleic Acids Res 45, D331–D338 (2017). [PubMed: 27899567]
24. Uhlen M et al., Proteomics. Tissue-based map of the human proteome. Science 347, 1260419 (2015). [PubMed: 25613900]
25. Liu J et al., Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in Arabidopsis. Plant Cell 24, 4333–4345 (2012). [PubMed: 23136377]
26. Ghaemmaghami S et al., Global analysis of protein expression in yeast. Nature 425, 737–741 (2003). [PubMed: 14562106]
27. Orfanoudaki G, Economou A, Proteome-wide subcellular topologies of E. coli polypeptides database (STEPdb). Mol Cell Proteomics 13, 3674–3687 (2014). [PubMed: 25210196]
28. Korber BT et al., Genetic differences between blood- and brain-derived viral sequences from human immunodeficiency virus type 1-infected patients: evidence of conserved elements in the V3 region of the envelope protein of brain-derived sequences. J Virol 68, 7467–7481 (1994). [PubMed: 7933130]

29. Laganowsky A et al., Crystal structures of truncated alphaA and alphaB crystallins reveal structural mechanisms of polydispersity important for eye lens function. Protein Sci 19, 1031–1043 (2010). [PubMed: 20440841]
30. Kondrat FD, Struwe WB, Benesch JL, Native mass spectrometry: towards high-throughput structural proteomics. Methods Mol Biol 1261, 349–371 (2015). [PubMed: 25502208]
31. Sobott F, Hernandez H, McCammon MG, Tito MA, Robinson CV, A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. Anal Chem 74, 1402–1407 (2002). [PubMed: 11922310]
32. Bush MF et al., Collision cross sections of proteins and their complexes: a calibration framework and database for gas-phase structural biology. Anal Chem 82, 9557–9565 (2010). [PubMed: 20979392]
33. Hopper JT, Sokratous K, Oldham NJ, Charge state and adduct reduction in electrospray ionization-mass spectrometry using solvent vapor exposure. Anal Biochem 421, 788–790 (2012). [PubMed: 22086073]
34. Hilton GR et al., C-terminal interactions mediate the quaternary dynamics of alphaB-crystallin. Philos Trans R Soc Lond B Biol Sci 368, 20110405 (2013). [PubMed: 23530258]
35. Petoukhov MV et al., New developments in the program package for small-angle scattering data analysis. J Appl Crystallogr 45, 342–350 (2012). [PubMed: 25484842]
36. Baldwin AJ et al., The polydispersity of alphaB-crystallin is rationalized by an interconverting polyhedral architecture. Structure 19, 1855–1863 (2011). [PubMed: 22153508]
37. Degiacomi MT, Dal Peraro M, Macromolecular symmetric assembly prediction using swarm intelligence dynamic modeling. Structure 21, 1097–1106 (2013). [PubMed: 23810695]
38. Buchan DWA, Minneci F, Nugent TCO, Bryson K, Jones DT, Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Res 41, W349–W357 (2013). [PubMed: 23748958]
39. Svergun D, Barberato C, Koch MHJ, CRYSOL - A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. J Appl Crystallogr 28, 768–773 (1995).
40. Breiman L, Random forests. Mach Learn 45, 5–32 (2001).
41. Marty MT et al., Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. Anal Chem 87, 4370–4376 (2015). [PubMed: 25799115]
42. Baldwin AJ, Lioe H, Robinson CV, Kay LE, Benesch JL, alphaB-crystallin polydispersity is a consequence of unbiased quaternary dynamics. J Mol Biol 413, 297–309 (2011). [PubMed: 21839090]
43. Adams PD et al., PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66, 213–221 (2010). [PubMed: 20124702]
44. Abraham MJ et al., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1–2, 19 – 25 (2015).
45. Lindorff-Larsen K et al., Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins 78, 1950–1958 (2010). [PubMed: 20408171]
46. Horn HW et al., Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. J Chem Phys 120, 9665–9678 (2004). [PubMed: 15267980]
47. Bussi G, Donadio D, Parrinello M, Canonical sampling through velocity rescaling. J Chem Phys 126, (2007).
48. Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR, Molecular-Dynamics with Coupling to an External Bath. J Chem Phys 81, 3684–3690 (1984).
49. Parrinello M, Rahman A, Polymorphic transitions in single-crystals - a new molecular-dynamics method. J Appl Phys 52, 7182–7190 (1981).
50. Hess B, P-LINCS: A parallel linear constraint solver for molecular simulation. J Chem Theory Comput 4, 116–122 (2008). [PubMed: 26619985]
51. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM, LINCS: A linear constraint solver for molecular simulations. J Comput Chem 18, 1463–1472 (1997).

52. Ryckaert JP, Ciccotti G, Berendsen HJC, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 23, 327–341 (1977).
53. Feenstra KA, Hess B, Berendsen HJC, Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. J Comput Chem 20, 786–798 (1999).
54. Hockney RW, Eastwood JW, Comput simul using part (Hilger A, Bristol England; Philadelphia, ed. Special student, 1988), pp. xxi, 540 p.
55. Essmann U et al., A smooth particle mesh Ewald method J Chem Phys 103, 8577–8592 (1995).
56. Torrie GM, Valleau JP, Non-physical sampling distributions in Monte-Carlo free-energy estimaiton: Umbrella sampling. J Comput Phys 23, 187–199 (1977).
57. Hub JS, de Groot BL, van der Spoel D, g_wham—A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. J Chem Theory Comput 6, 3713–3720 (2010).
58. Pronk S et al., GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 29, 845–854 (2013). [PubMed: 23407358]
59. Daura X et al., Peptide Folding: When Simulation Meets Experiment. Angewandte Chemie Int Ed 38, 236–240 (1999).
60. Gelman A, Rubin DB, Inference from Iterative Simulation Using Multiple Sequences. Statist Sci 7, 457–472 (1992).
61. Gabelica V, Galic N, Rosu F, Houssier C, De Pauw E, Influence of response factors on determining equilibrium association constants of non-covalent complexes by electrospray ionization mass spectrometry. J Mass Spectrom 38, 491–501 (2003). [PubMed: 12794869]
62. Starr TN, Thornton JW, Epistasis in protein evolution. Protein Sci 25, 1204–1218 (2016). [PubMed: 26833806]
63. McWilliam H et al., Analysis tool web services from the EMBL-EBI. Nucleic Acids Res 41, W597–600 (2013). [PubMed: 23671338]
64. Abascal F, Zardoya R, Posada D, ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21, 2104–2105 (2005). [PubMed: 15647292]
65. Guindon S et al., New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59, 307–321 (2010). [PubMed: 20525638]
66. Jones DT, Taylor WR, Thornton JM, The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8, 275–282 (1992). [PubMed: 1633570]
67. Sobott F, Benesch JL, Vierling E, Robinson CV, Subunit exchange of multimeric protein complexes. Real-time monitoring of subunit exchange between small heat shock proteins by using electrospray mass spectrometry. J Biol Chem 277, 38921–38929 (2002). [PubMed: 12138169]
68. Ke H et al., Genetic investigation of tricarboxylic acid metabolism during the Plasmodium falciparum life cycle. Cell Rep 11, 164–174 (2015). [PubMed: 25843709]
69. Levy ED, Boeri Erba E, Robinson CV, Teichmann SA, Assembly reflects evolution of protein complexes. Nature 453, 1262–1265 (2008). [PubMed: 18563089]
70. Marsh JA et al., Protein complexes are under evolutionary selection to assemble via ordered pathways. Cell 153, 461–470 (2013). [PubMed: 23582331]
71. Akaike H, A new look at the statistical model identification. IEEE Trans Autom Control 19, 716–723 (1974).
72. Kimura M, Evolutionary Rate at the Molecular Level. Nature 217, 624–626 (1968). [PubMed: 5637732]
73. Tokuriki N, Stricher F, Schymkowitz J, Serrano L, Tawfik DS, The stability effects of protein mutations appear to be universally distributed. J Mol Biol 369, 1318–1332 (2007). [PubMed: 17482644]
74. Lynch M, Evolutionary diversification of the multimeric states of proteins. Proc Natl Acad Sci U S A 110, E2821–2828 (2013). [PubMed: 23836639]
75. van Montfort RL, Basha E, Friedrich KL, Slingsby C, Vierling E, Crystal structure and assembly of a eukaryotic small heat shock protein. Nat Struct Biol 8, 1025–1030 (2001). [PubMed: 11702068]

76. Kim KK, Kim R, Kim SH, Crystal structure of a small heat-shock protein. Nature 394, 595–599 (1998). [PubMed: 9707123]

77. Kabsch W, Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J Appl Crystallogr 26, 795–800 (1993).

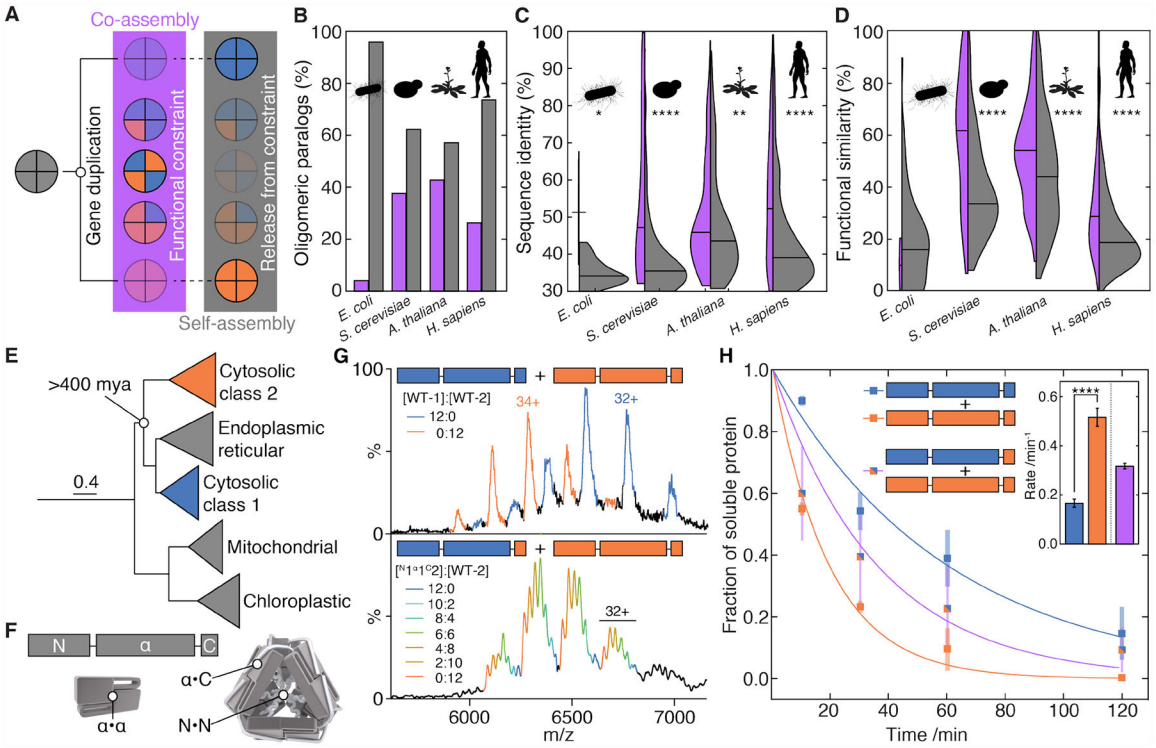

Figure 1. Self-selective assembly allows oligomeric paralogs to evolve distinct functions.

A) After gene duplication, oligomeric paralogs co-assemble into and predominantly populate heteromers, constraining their functions to be compatible with co-assembly. If they subsequently evolve the ability to assemble self-selectively into homomers, their functions are free to diverge.

B) Percentage of pairs of oligomeric paralogs that either co-assemble into heteromers (purple) or only self-assemble into homomers (grey) in *E. coli* (73 pairs in dataset), *Saccharomyces cerevisiae* (215 pairs), *Arabidopsis thaliana* (742 pairs), and *Homo sapiens* (1086 pairs).

C) Pairwise sequence identity is higher between co-assembling paralogs (purple) than between self-assembling paralogs (grey). Horizontal lines denote medians. * p<0.05, ** p<0.01, **** p<<0.0005, Mann-Whitney rank sums test.

D) Pairwise functional similarity of co-assembling (purple) and self-assembling (grey) pairs of paralogs as measured by the intersection over the union of their gene ontology annotations. Horizontal lines denote medians. **** p<<0.0005, Mann-Whitney rank sums test.

E) Maximum-likelihood phylogeny of select clades of plant sHSPs. Scale bar indicates average number of substitutions per site.

F) Schematic of the three different interfaces used by sHSP to assemble into oligomers.

G) Mass spectrum of WT-1 and WT-2 after prolonged incubation plotted in the mass-to-charge (m/z) dimension. WT-1 (blue) and WT-2 (orange) 12-mers are observed, with varying numbers of charges. No peaks corresponding to heteromers are detected (**upper**). Hetero-12-mers are formed via exchange of dimers if WT-2 is mixed with $^{N}1^{\alpha}1^{C}2$, resulting in additional peaks for each charge state (**lower**). One charge-state is labelled for each 12-mer.

H) When mixed prior to incubation with pea-leaf lysate at 42 °C, WT-1 and WT-2 partition into aggregates at different rates (**** p<<0.0005). When WT-2 is incubated with $^{N}1^{\alpha}1^{C}2$, subunits from both proteins partition at the same, intermediate rate (**inset**). Heteromers thus function differentially to segregated WT oligomers. Error bars in the raw data are standard deviations from three independent experiments; error bars in the inset are standard deviations calculated from 1000 bootstrap replicates of the fit.

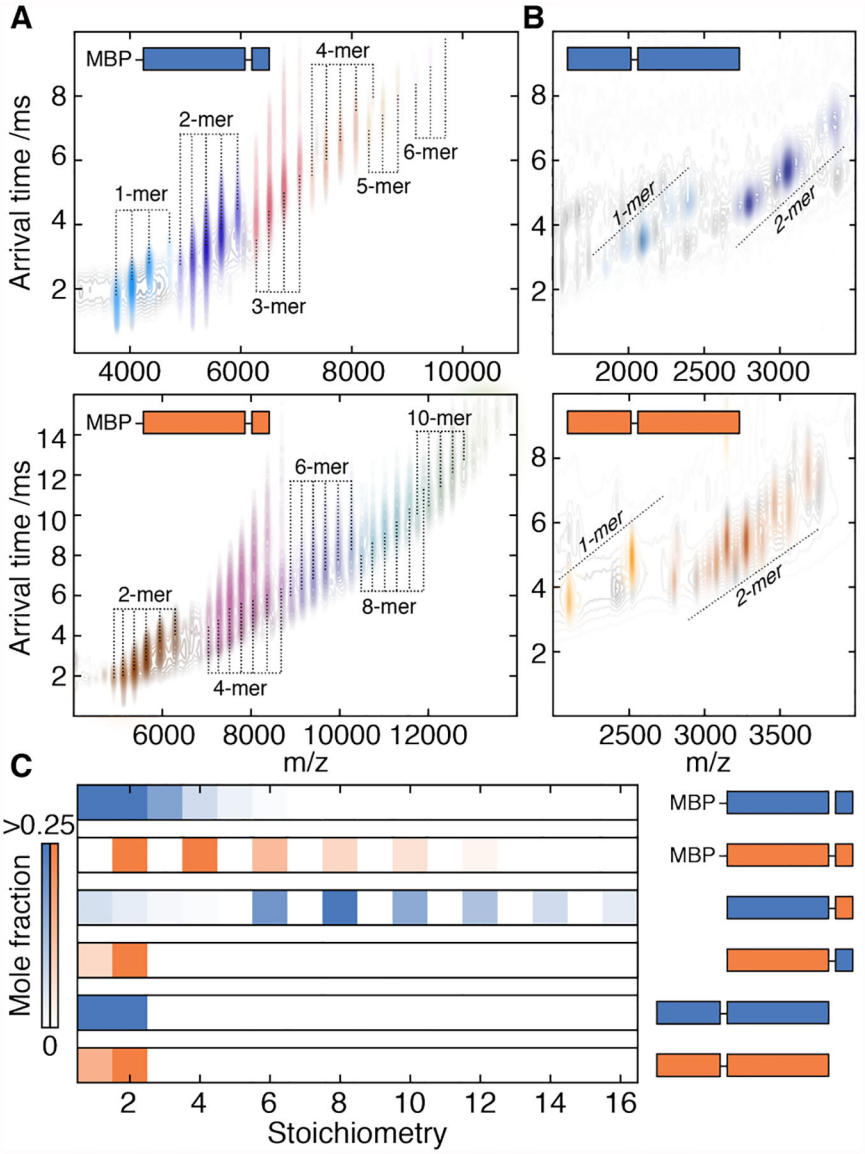

Figure 2. Oligomeric interfaces form in a hierarchical order.

A) IM-MS spectra of truncated constructs of WT-1 (**upper**) and WT-2 (**lower**) lacking the N-terminal region. The two dimensions of separation (m/z and arrival time, which depends on collision cross-section) separate charge-state series corresponding to a series of stoichiometries (coloured individually). Both truncated proteins assemble into polydisperse ensembles. MPB – maltose binding protein.

B) IM-MS spectra of truncated constructs of WT-1 (**upper**) and WT-2 (**lower**) lacking the C-terminal tail. Both proteins do not assemble beyond dimers. Truncations on the exposed N-terminus result in several charge-series for monomers and dimers that are separated in the arrival time dimension (see Fig. S8 for detailed assignments).

C) Distribution of stoichiometries populated by truncated constructs extracted from spectra in **A, B,** Fig S6. The C-terminal tail is required for assembly beyond dimers, whereas the N-

terminus is required for monodisperse 12-mers. The $^{\alpha}2^{C}1$ construct (Fig. S8E) does not oligomerize, indicating an unfavourable α·C interaction.

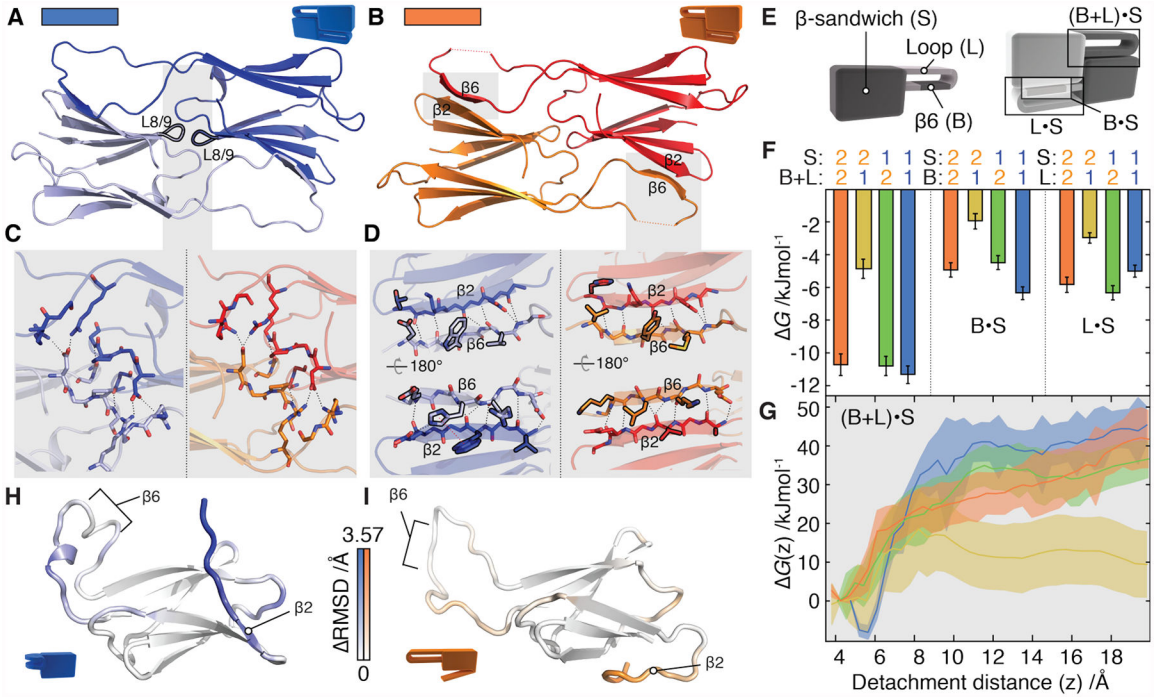

Figure 3. Selectivity in the structurally conserved α-crystallin domain.

A and **B)** $^{\alpha}1$ and $^{\alpha}2$ dimers have an identical fold (backbone RMSD = 1.2 Å) in which two highly similar interfaces (labelled L8/9 and β6· β2) connect monomers.

C) The L8/9 interface is centred on the loop between β8 and β9 (black outline) and is indistinguishable in the two proteins. Inter-chain hydrogen bonds are shown as dashed lines.

D) The two β6·β2 interfaces in the dimer are formed by exchange between the β6 and β2 strands. Side-chains that differ between $^{\alpha}1$ and $^{\alpha}2$ at homologous positions are outlined in black. The π-stacking interaction specific to $^{\alpha}1$ is shown as a dotted red line.

E) Constructs were designed by swapping the β-sandwich, loop, and β6 strand (**left**). These were used to assess the strength of the β6· β2 interface, and deconvolve the contribution from the loop and β6 strand (**right**).

F) Global thermodynamic model of dimerization based on experimentally determined $\Delta G_{\alpha.\alpha}$ values in Fig S12G. The combined loop and β6 from $^{\alpha}1$ interact less favourably with β2 from $^{\alpha}2$ than all other combinations (**left**). $^{\alpha}2$ and $^{\alpha}1$ partition contributions to $\Delta G_{\alpha.\alpha}$ differently (**shaded**). Error bars are standard deviations from 1000 bootstrap replicates of the model fit.

G) In a simulated heterodimer, the free energy barrier is significantly reduced for the $^{\alpha}2$·$^{\alpha}1$ pair (yellow), but indistinguishable from the homodimers in the case of $^{\alpha}1$·$^{\alpha}2$ (green) when the β6·β2 interface is disrupted along a reaction coordinate that separates them. Shaded area corresponds to the standard error of the mean.

H,I) Median monomeric conformations determined by principal component analysis coloured according to structural difference. This is calculated at each residue from the Cα RMSD between $^{\alpha}1$ and $^{\alpha}2$ monomers, minus the RMSD between repeats for each monomer. Positive ΔRMSD values indicate conformational differences between proteins that cannot be explained by the variations intrinsic to each protein, and only those with $p<0.05$ (after

Bonferroni correction, permutation test) are coloured. Differences are apparent in the loop surrounding β6 and in β2. In $^{\alpha}$1 the loop curls up, whereas in $^{\alpha}$2 the β2 strand detaches readily from the remainder of the β-sandwich.

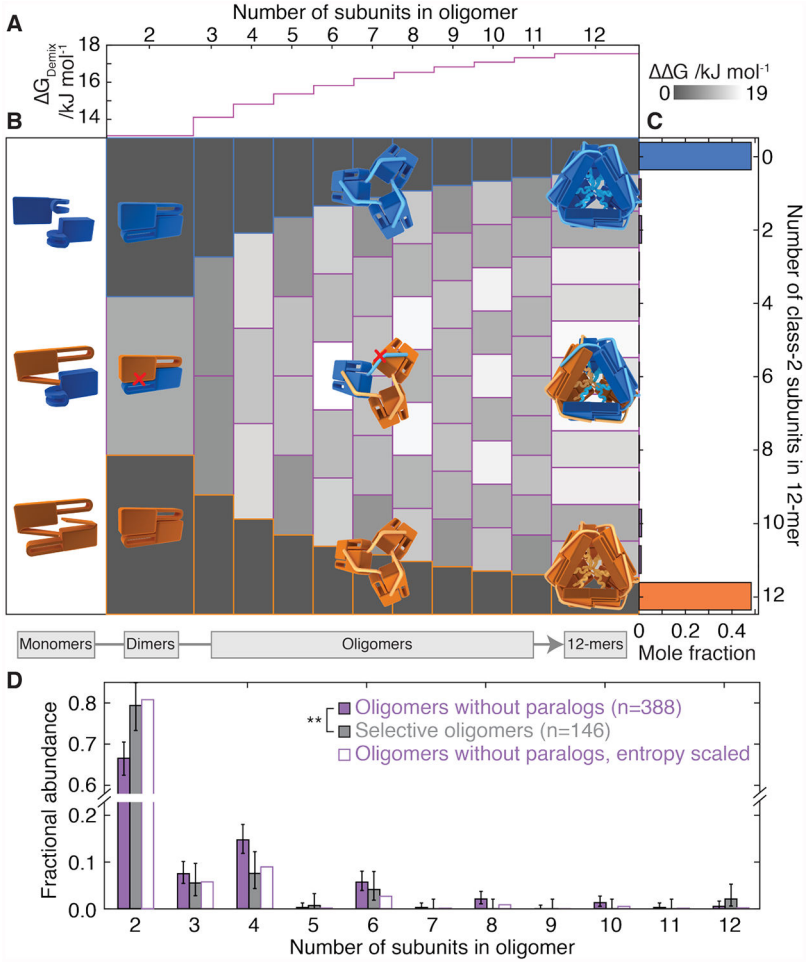

Figure 4. Selective interfaces overcome unfavourable entropy of homomerization.

A) Selective homomerization is entropically unfavourable and requires an energetic penalty upon forming heteromeric contacts to suppress heteromerization. Shown is the theoretical magnitude of this penalty per subunit ($\Delta G_{\mathrm{Demix}}$) required to populate heteromers at only 2% of all oligomers. It increases logarithmically with the size of the oligomer, making it more challenging for larger oligomers to be selective.

B) Empirically derived stabilities of all possible heteromers along the assembly pathway compared to homomers of the same size ($\Delta\Delta G = \Delta G_{\mathrm{heteromer}} - \Delta G_{\mathrm{homomer}}$). The upper and lower tiles of each column correspond to homomers of WT-1 and WT-2, respectively. Those in between represent heteromers, with increasing numbers of WT-2 subunits (downwards). The $\Delta\Delta G$ values are positive for all heteromers, meaning that energetic penalty to co-assembly we quantified in selective interactions is larger than the positive entropy of heteromerization.

C) The equilibrium population of homo- and hetero-12-mers calculated based on the values in **B** results in mole fractions of hetero-12-mers just below detectable levels. >96% of subunits partition into homomers, compared to only 0.05% based on the binomial distribution of hetero-oligomers that would arise in the absence of selective interfaces.

D) The oligomeric stoichiometries populated by selective oligomeric paralogs (grey fill) are smaller with a particular excess of dimers than for a control set of oligomers that have no paralogs (purple). ** $p<0.005$, Mann-Whitney rank sums test. Error bars represent 90% Clopper-Person confidence interval, n denotes sample size. Applying a scaling according to ΔG_{Demix} to the control set reproduces closely the observed selective distribution (purple outline, $p=0.0005$, Akaike information criterion).