Imperfect symmetry facilitated the evolution of specificity and high-order stoichiometry in vertebrate hemoglobin

Carlos R. Cortez-Romero (1), Jixing Lyu (4), Arvind S. Pillai (2,5), Arthur Langanowsky (4), Joseph W. Thornton (2,4)

Department of (1) Cell and Molecular Biology, (2) Ecology and Evolution, and (3) Human Genetics, University of Chicago, Chicago, IL, 60637. (4) Department of Chemistry, Texas A&M University, College Station, TX, 77843. (5) Institute of Protein Design, University of Washington, Seattle, WA, 98195

C**orrespondence**: joet1@uchicago.edu

Competing Interest Statement: The authors declare no competing interests.

Classification: Major classification - Biological Sciences **Minor classifications –** Biochemistry/Evolution

Keywords: protein evolution, molecular complexes, multimeric proteins, heterodimers, evolution of specificity, gene duplication, isology

This PDF file includes:

Main Text
Figures 1 to 6
Material and methods
Citations

ABSTRACT

Many proteins form paralogous multimers – molecular complexes in which evolutionarily related proteins are arranged into specific quaternary structures. Little is known about the mechanisms by which they acquired their stoichiometry (the number of total subunits in the complex) and heterospecificity (the preference of subunits for their paralogs rather than other copies of the same protein). Here we use ancestral protein reconstruction and biochemical experiments to study historical increases in stoichiometry and specificity during the evolution of vertebrate hemoglobin (Hb), a $\alpha_2\beta_2$ heterotetramer that evolved from a homodimeric ancestor after a gene duplication. We show that the mechanisms for this evolutionary transition was simple. One hydrophobic substitution in subunit β after the gene duplication was sufficient to cause the ancestral dimer to homotetramerize with high affinity across a new interface. During this same interval, a single-residue deletion in subunit α at the older interface conferred specificity for the heterotetrameric form and the *trans*-orientation of subunits within it. These sudden transitions in stoichiometry and specificity were possible because the interfaces in Hb are isologous – involving the same surface patch on interacting subunits, rotated 180° relative to each other – but the symmetry is slightly imperfect. This architecture amplifies the impacts of individual mutations on stoichiometry and specificity, especially in higher-order complexes, and allows single substitutions to differentially affect heteromeric vs homomeric interactions. Many multimers are isologous, and symmetry in proteins is always imperfect; our findings therefore suggest that elaborate and specific molecular complexes may often evolve via simple genetic and physical mechanisms.

Significance statement

Many molecular complexes are made up of proteins related by gene duplication, but how these assemblies evolve is poorly understood. Using ancestral protein reconstruction and biochemical experiments, we dissected how vertebrate hemoglobin, which comprises two copies each of two related proteins, acquired this architecture from a homodimeric ancestor. Each aspect of this transition – from dimer to tetramer and homomer to heteromer – had a simple genetic basis: a single-site mutation in each protein drove the changes in size and specificity. These rapid transitions were possible because hemoglobin's architecture is symmetric, which amplified the effect of small biochemical changes on the assembly of the complex. Many protein complexes are symmetrical, suggesting that they too may have evolved via simple genetic mechanisms.

INTRODUCTION

Protein multimers – associations of multiple protein subunits arranged in specific quaternary architectures – carry out most biochemical functions in living cells (1, 2). The mechanisms by which these complexes evolved their stoichiometry and specificity present some puzzling questions (2-10). Multimers assemble via interfaces that typically contain dozens of sterically and electrostatically complementary residues, and higher-than-dimeric stoichiometries (tetramers, octamers, etc.) use several such interfaces on each subunit (11). This seems to imply that many sequence substitutions would be required for a new multimeric assembly to originate during evolution.

A second complication is that many multimers are composed of paralogs -- proteins related to each other by gene duplication. Paralogs are genetically and structurally indistinguishable when generated by duplication, so initially they assemble indiscriminately into homomers and heteromers. Most complexes, however, have evolved specificity for either the homomeric or heteromeric form, with the latter being the most common outcome (12). How specificity evolves is unclear, because mutations that affect multimerization are expected to cause correlated effects on the affinities of homomerization and heteromerization (6,12,13). The structural similarity of paralogs seems to imply that substitutions in both paralogs are required to confer any specificity at all. This complication is magnified for higher-order paralogous multimers, in which one might expect that every interface must evolve specificity to mediate assembly into the complex's particular architecture.

A critical factor in the evolution of specificity and high-order stoichiometry may be whether a multimer assembles through symmetrical interfaces. In many complexes, identical or paralogous subunits bind each other using an isologous interface – a form of symmetry in which a surface patch on one subunit binds to the same patch on its partner but rotated 180 degrees relative to each other (1). Isologous complexes might, in principle, have the potential to evolve changes in stoichiometry and specificity through simpler mechanisms than nonisologous head-to-tail interfaces. A single substitution appears twice across the interface(s) of an isologous homodimer or heterotetramer, four times in a homotetramer, etc. (Fig. 1A). Mutations that weakly affect affinity on their own can therefore confer large effects on the assembly of isologous multimers (1,5,9,14-16). Isology also changes the way that mutations can affect specificity. In a nonisologous interface, specificity requires mutations on both surfaces so that the tails are recognizably different from each other and each head prefers one tail over the other. In an isologous interface, however, a substitution on the surface of just one subunit has the potentially to differentially affect the affinity of each kind of complex, because it will appear twice in the interface of a homomer, once in the heteromer, and not at all in the other homomer (Fig. 1A).

Little is known about the historical evolution of heterospecific complexes or the role of symmetry in this process, especially in high-order complexes. Biochemical and protein engineering studies have addressed the determinants of binding affinity in both homomeric and heteromeric interfaces of extant proteins (19-25). But the genetic and structural mechanisms by which those

interactions were acquired long ago are often different from their derived forms in the present (26). Ancestral sequence reconstruction (ASR) can address this limitation by experimentally characterizing the effects of historical sequence changes when introduced into ancestral proteins. ASR has been used to understand the evolution of specificity after duplication in head-to-tail paralogous heteromers (17,18) and in multimers composed of unrelated proteins, which are by definition asymmetrical (25). But we know of no studies that have addressed how isologous heteromers historically evolved their specificity or how specificity in high-order complexes was acquired. A recent in silico analysis predicted that it should be possible for specificity in heterodimers to evolve rapidly after gene duplication through small perturbations in binding energy (27), but the underlying mechanisms and historical relevance of this phenomenon are unknown.

Here we use ASR to study the evolution of higher-order stoichiometry and specificity in vertebrate hemoglobin (Hb), the major carrier of oxygen in the blood of jawed vertebrates. Hb is a paralogous $\alpha_2\beta_2$ heterotetramer ((16), Fig. 1B), assembly of which is mediated by two distinct and isologous interface patches (IF1 and IF2). Each subunit of the tetramer uses its IF1 to bind IF1 of a paralogous subunit; two of these heterodimers bind to each other using the IF2 on each subunit ((28), Fig. 1B). Hbα and Hbβ descend from a gene duplication deep in the vertebrate lineage (Fig. 1C), and their sequences retain sufficient phylogenetic signal to allow high-confidence reconstruction of ancestral Hb protein sequences. Using ASR, we recently showed experimentally that extant Hb evolved its heterotetrameric architecture in two phases from a monomeric precursor via a homodimeric intermediate (16). In the first phase, prior to the gene duplication that yielded paralogous α and β lineages, a monomeric ancestor evolved the capacity to homodimerize with moderate affinity across IF1. In the second phase – after the gene duplication but before the last common ancestor of all vertebrates – binding across IF2 was acquired, yielding the tetrameric stoichiometry, and specificity for the heteromeric form $\alpha_2\beta_2$ also evolved (Fig. 1C).

Here we characterize the genetic and physical mechanisms that mediated the evolutionary transition from homodimer to heterotetramer in this second phase. By experimentally characterizing reconstructed ancestral hemoglobin subunits and the effects of historical sequence changes on them, we address the following questions: 1) How many substitutions were required to confer tetrameterization across IF2, and what thermodynamic and structural mechanisms mediated their effects? 2) Did the evolution of specificity for the heterotetrameric form require sequence changes at one or both interfaces, in one or both subunits, and what physical mechanisms drove the acquisition of this specificity? 3) How did the symmetry of Hb's two interfaces affect this evolutionary transition to a high-order, heterospecific architecture? Does a mutational propensity favor increased molecular complexity during the evolution of isologous complexes?

RESULTS

Evolution of tetrameric stoichiometry. We first sought to identify the historical substitutions that conferred tetramerization after duplication of the ancestral homodimer Ancαβ. We

previously identified two potentially important substitutions, both of which occurred on the branch leading from the duplication of Ancαβ to Ancβ (the Hbβ subunit in the last common ancestor of jawed vertebrates), which heterotetramerizes with Ancα (the Hbα subunit in the jawed vertebrate ancestor); like extant Hbβs, Ancβ also homotetramerizes with itself. Introducing these substitutions together into Ancαβ was sufficient to confer high-affinity assembly into homotetramers (16). One of these (q40W) is buried in the IF2 interface, whereas the other (t37V) makes contacts across both IF1 and IF2 (Fig. 1D. 4, using lower and upper case to denote ancestral and derived amino acids, respectively). Here we isolated the individual contributions of each amino acid by introducing them singly into Ancαβ and characterizing their effect on assembly into tetramers using size-exclusion chromatography (SEC) and native mass spectrometry (nMS) (29,30).

We found that substitution q40W alone is sufficient to recapitulate the evolution of Hb's tetrameric stoichiometry. Ancαβ forms only dimers in SEC at 100 μM of total protein subunits; by contrast, the mutant Ancαβ$_{q40W}$ is tetrameric, with occupancy of the tetramer similar to that observed in the derived Ancα + Ancβ complex and human Hb (Fig. 1F). We then used nMS across a titration series to measure the affinity with which dimers associate into tetramers and found that the tetramerization affinity of Ancαβ$_{q40W}$ (Kd 10 μM) is stronger than that of Ancα + Ancβ (61 μM) and human Hb (41 μM) (Fig. 1E). Substitution q40W is therefore sufficient to confer biologically relevant tetramerization on the ancestral Hb complex. This conclusion is robust to statistical uncertainty about the ancestral reconstruction, because the same experiments using alternative ancestral proteins yield almost identical results (Fig. S1).

The other historical substitution, t37V, is not sufficient to confer tetramerization. Mutant Ancαβ$_{t37V}$ confers no detectable tetramer occupancy by SEC, even at 1 mM (Fig. 1G), and it displays no measurable affinity to form tetramers using nMS (Fig. 1H). When combined with substitution q40W, however, t37V does increase affinity of the dimer-tetramer transition by a factor of 6 compared to the effect of q40W alone (Fig. 1D; Fig. S2).

Substitutions could also facilitate tetramerization by increasing affinity of the monomer-to-dimer transition, thus increasing the effective concentration of dimers, which would produce more tetramers even if affinity of the dimer-tetramer transition were unchanged. Using nMS, we found that t37V improves the monomer-dimer affinity of Ancαβ by >100-fold (Fig. 1H; Fig. S2). Substitution q40W, in contrast, has no effect on monomer-dimer affinity. These findings are consistent with the structural location of t37V at both IF1 and IF2 and that of q40W at IF2 only, and they explain why t37V further increases the impact of q40W but on its own does not confer tetramerization.

A likely physical mechanism for the effect of q40W is that tryptophan's bulky hydrophobic side chain nestles into a hydrophobic divot on the IF2 surface of the facing subunit, and further strengthened by a hydrogen bond to 102D (31). To test this hypothesis, we identified alternative amino acid replacements with similar biochemical properties and measured whether they also could have caused Ancαβ to evolve into a tetramer. Like tryptophan, the bulky hydrophobic residues phenylalanine or tyrosine at this position confer tetramerization, albeit at affinity slightly

worse than q40W but similar to that of Ancα+Ancβ and human Hb. Leucine, in contrast, which has a smaller volume and no hydrogen bonding capacity, confers no measurable tetramerization (Fig. 1I). High-affinity homotetramerization could therefore have evolved via any of three different aromatic replacements at site 40. The greater affinity of tryptophan may be due to its longer side chain, which buries more hydrophobic surface area across the interface. Our results suggest that the hydrogen bond with 102D is not necessary, because phenylalanine confers tetramerization but provides no hydrogen bond donor.

Taken together, these data indicate that a substitution at a single amino acid position was sufficient to confer tetramerization, and numerous alternative mutations at this site that could have caused this increase in stoichiometry during Hb evolution.

Isology facilitated IF2 evolution. How could a single amino acid replacement cause such a dramatic change in stoichiometry? The Hb tetramer can be viewed as two heterodimers, each of which is mediated by isologous assembly across IF1 (the larger interface); these heterodimers then bind to each other isologously across IF2. We hypothesized that this doubly symmetrical architecture allowed substitution q40W to confer the dimer-tetramer evolutionary transition, because isology causes the derived amino acid to appear four times in the homotetramer and twice in the heterotetramer.

If this hypothesis is correct, then assembly across IF2 by the derived Hb protein should require assembly across IF1 to multiply the intrinsic affinity of IF2 (Fig. 1A). We tested this prediction by introducing q40W into Ancαβ but doing so under conditions that prevent assembly across IF1. We first compromised dimerization across IF1 genetically by reverting the IF1 surface to the ancestral states of the monomeric ancestor AncMH; these mutations abolish dimer occupancy, leaving a monomers-only population at 20mM (Fig. 2A). We then introduced q40W into these IF1-ablated mutants and assessed stoichiometry using nMS. As predicted, these proteins do not form any observable dimers or tetramers (detection limit ~1μM) (Fig. S3). The dependence of IF2 formation on an effective IF1 is also apparent when using t37V/q40W to confer IF2-mediated assembly, and when IF1 is compromised by introducing mutation P127R, a non-historical mutation that introduces an unsatisfied charged residue into IF1 (Fig. 2B). The IF2 mutations do not compromise heme binding or solubility, because the mutant proteins are purifiable and heme-bound in nMS.

We also tested whether assembly across IF2 could have been acquired before dimerization across IF1 evolved. We introduced t37V/40W into the ancestral monomer AncMH – which existed before the evolution of dimerization -- and tested whether dimer assembly across IF2 can be conferred in this background. As predicted, only monomers were observed, with no dimers or higher stoichiometries detected (Fig. 2C). These data establish that acquisition of multimerization across IF2 by q40W and by the pair t37V/q40W depends on the prior evolution of dimerization via IF1.

Our observations can be explained by a simple model that arises from the symmetrical structure of the hemoglobin tetramer. A single iteration of IF2 is too weak to confer significant binding of

two monomers into a dimer; however, if the stronger IF1 mediates dimer assembly, each such dimer presents two iterations of the IF2 surface patch, and these are sufficient to mediate assembly of dimers into tetramers. This simple model implies that the energy of dimer-tetramer binding using IF2 should be twice that of monomer-dimer binding using the same interface, and the Kd of tetramerization should be the square of the Kd of dimerization (Fig. 2D). Our results are consistent with this prediction. The Kd of the dimer-tetramer transition by Ancαβ $_{t37V/q40W}$ across IF2 is 1 mM, which predicts that the affinity of IF2-mediated monomer-dimer transition when IF1 is compromised should be ~ 1mM. Consistent with this prediction, we detected no dimer occupancy by Ancαβ $_{t37V/q40W;\ IF1reverted}$ using an assay that can quantify Kd up to 400 μM (see Methods). We cannot rule out the possibility that IF1 binding may also allosterically modify IF2 and increase its affinity beyond the additive effect conferred by isologous repetition alone; however, any such effect must be relatively small, because the simple additive model explains most – and possibly all -- of the difference in affinity conferred when IF2 is doubled in the symmetrical tetramer.

Taken together, these data indicate that the isologous architecture of IF1 and IF2 facilitated the evolution of the Hb tetramer via substitution q40W. Without this doubly symmetrical architecture, IF2 would have been too weak to mediate multimerization. The dependence of q40W's effect on the presence of IF1 also creates contingency and order-dependence in the evolution of the Hb complex. We previously showed that IF1 evolved before the duplication of the dimeric ancestor Ancαβ (16). Our present results show that if that IF1-mediated dimer had never evolved, substitution q40W at IF2 would not have been sufficient to drive the acquisition of the tetrameric stoichiometry, and the ancestral Hb protein would have remained a monomer. If events had occurred in the opposite order – with the affinity-enhancing substitution at IF2 occurring first – this intermediate ancestor would have been a monomer; when the substitutions that confer binding across IF1 did occur, they would have triggered an immediate evolutionary transition from monomer to tetramer.

Heteromeric specificity evolved at a single interface. We next focused on understanding the evolution of Hb's specificity for the heterotetrameric form, which was acquired during the same phylogenetic interval after the duplication of Ancαβ. Our first question was whether specificity for heteromeric interactions was conferred by sequence changes at IF1, IF2, or both. Our previously published experiments suggest that evolutionary changes at IF2 confer no specificity: when all historical substitutions that occurred at the IF2 surface during the post-duplication interval are introduced into Ancαβ and this protein is coexpressed with Ancα, an indiscriminate mixture of homotetramers, $\alpha_1\beta_3$ heterotetramers, and $\alpha_2\beta_2$ heterotetramers is produced (16). We therefore hypothesized that heterospecificity of the Hb tetramer is encoded entirely by IF1, such that Ancα and Ancβ specifically heterodimerize across IF1, and these heterodimers then bind to each other via a nonspecific IF2, yielding $\alpha_2\beta_2$ heterotetramers.

This hypothesis makes two predictions: 1) IF1 mediates specific assembly of α and β subunits into heterodimers, and 2) this specificity is sufficient to account for the heterospecificity of $\alpha_2\beta_2$ heterotetramer. To test the first hypothesis, we characterized the specificity of hetero- vs

homodimer assembly by IF1 under two different conditions in which no binding across IF2 occurs. First, we diluted a coexpressed mixture of Ancα and Ancβ to concentrations at which dimers rather than tetramers assemble: at 50 μM, only heterodimers and heterotetramers form; at 5 μM, only heterodimers are observed (Fig. 3A). IF2 does not mediate assembly of monomers into dimers in the absence of IF1 (Fig. 2A, 2B), so these heterodimers must be IF1-mediated, indicating that IF1 is heterospecific (Fig. 3A). Second, we expressed Ancα and Ancβ separately and mixed them at equal and moderate concentration (rather than coexpressing them); under these conditions, only IF1 dimers form, and these are predominantly heterodimers (Fig. 3B, Fig. S4). Finally, we engineered protein Ancβ' – a variant of Ancβ in which all IF2 residues that were substituted between Ancαβ and Ancβ are reverted to the ancestral state, thus abolishing binding across IF2– and found that it also forms predominantly heterodimers when mixed with Ancα (Fig. 3C, Fig. S5). Together, these data indicate that the derived IF1 is specific, preferentially mediating assembly into heterodimers.

To test the second prediction – that the degree of heterospecificity mediated by IF1 is sufficient to drive specific assembly of $\alpha_2\beta_2$ heterotetramers even if IF2 is nonspecific – we measured the affinities of homomerization and heteromerization across IF1 and used these measurements to predict their effects on tetramer specificity in the absence of any specificity at IF2. Using nMS and Ancβ' ,we found that IF1's heterodimerization affinity (Kd=0.5 μM) is slightly worse than its homodimerization affinity (0.2 μM), but both are far better than the Ancα homodimer (21 μM) (Fig. 3D, S5, S6, S7). We then used these IF1 affinities to predict the occupancy of hetero- and homodimers and tetramers as the concentration of globin subunits changes, assuming that IF2 mediates tetramerization at Kd=30 μM, as measured in Ancα + Ancβ, with no preference for homomeric or heteromeric binding (Fig. 1D). At low concentrations, the system produces only IF1-mediated dimers, because the affinity of IF2 is weak, and these are almost all heterodimers. The predominance of heterodimers is attributable of Ancα's weak propensity to homodimerize; the excess of unbound Ancα subunits causes Ancβ subunits to preferentially heterodimerize rather than homodimerize at equilibrium, even though Ancβ's homodimerization affinity is slightly stronger than its heterodimerization affinity (Fig. 3D). As protein concentration increases, these dimers begin to assemble with each other across IF2 into tetramers, and the strong excess of heterodimers over homodimers means that the vast majority of these are heterotetramers, even though IF2 itself does not distinguish between these forms. At physiologically relevant concentrations of 3mM total Hb subunits (32), the population is dominated by $\alpha_2\beta_2$ heterotetramers, with a small fraction of heterodimers and virtually no homotetramers (Fig. 3d; right panel).

Taken together, these data establish that the measured specificity of IF1 alone mediates highly specific assembly of Ancα+ Ancβ into heterotetramers, even when IF2 is entirely nonspecific -- which our previous experiments suggest is the case – because IF1 is a much stronger interface than IF2. The historical acquisition of heterospecificity across IF1 after the Ancαβ gene duplication is therefore sufficient to account for the evolution of Hb's heterotetrameric architecture.

Heteromeric specificity evolved primarily by reducing homodimerization affinity of Ancα. Given our finding that heterospecificity evolved at the IF1 interface, we next sought to characterize whether the acquisition of specificity was driven by evolutionary changes in the α subunit, the β subunit, or both.

The heterospecificity of a pair of dimerizing proteins can be quantified in energetic terms as the difference in the ΔG of binding between the heterodimer and the mean of the two homodimers ($\Delta\Delta G_{spec}$). If $\Delta\Delta G_{spec} = 0$, then the fractional occupancy of the heterodimer at saturating and equal concentrations of subunits will be 50%, as will the sum of the homodimers; if the homodimer ΔGs are very different from each other and the heterodimer ΔG is halfway between them, then the two homodimers may have different occupancies but will still add to 50%. By contrast, if $\Delta\Delta G_{spec}<0$, then heterodimers will account for the majority of dimers, and if $\Delta\Delta G_{spec}>0$, homodimers together will predominate (Fig. 4A-C). Hetero- or homospecificity thus arises when two paralogs contribute nonadditively to dimerization.

We used this approach to quantify the heterospecificity of Ancα and Ancβ at IF1. We used nMS to measure the homodimer and heterodimer affinities of Ancα and Ancβ‘; the latter protein contains all substitutions that occurred along the Ancβ branch except those that mediate tetramerization across IF2, which allows us to isolate specificity effects at IF1 by preventing tetramerization. We calculated the ΔG of binding and expected fractional occupancy of each dimer at high and equal concentration of subunits. We found that $\Delta\Delta G_{spec}= -1.54$ (in units of kT) and heterodimer occupancy of 82% (Fig. 4D). This represents the total specificity acquired by the two diverging paralogs after the duplication of of Ancαβ, which by definition had no specificity. This specificity was acquired because of evolutionary changes in all three relevant affinities. Relative to the ancestral dimerization affinity of Ancαβ, Ancα‘s energy homodimerization became worse ($\Delta\Delta G = 0.85$) while homodimerization by Ancβ improved substantially ($\Delta\Delta G = -3.72$). The heterodimer affinity improved less than the Ancβ homodimer did ($\Delta\Delta G = -2.97$) but by more than the average of the two homodimers ($\Delta\Delta G = -1.44$), yielding the observed strong preference for the heterodimer.

We next sought to isolate the contribution to this acquired specificity of the evolutionary changes that occurred along each of the two branches. To measure the specificity acquired along the branch leading to Ancα, we measured affinities and calculated $\Delta\Delta G_{spec}$ when Ancα is mixed with the ancestor Ancαβ. This pair of proteins is heterospecific, with $\Delta\Delta G_{spec}= -1.19$ (expected heterodimer occupancy 76%). Changes in the α subunit alone therefore account for ~77% of the total specificity that was acquired by the entire Ancα+Ancβ system. This specificity was acquired via a 2.2-fold reduction in homodimerization affinity by Ancα relative to the Ancαb ancestor and a 1.8-fold improvement in heterodimer affinity (Fig. 4E; Fig. S4C & D).

To isolate the contribution to IF1 specificity of evolutionary changes that occurred along the branch to Ancβ, we measured affinities when Ancβ’ is mixed with Ancαβ. This pair of proteins is weakly heterospecific, with $\Delta\Delta G_{spec}= -0.34$ and expected heterodimer occupancy of just 58%. This small change arises because both the homodimer and heterodimer improved in affinity,

and the deviation of the heterodimer from the average of the homodimers is small. (Fig. 4F; Fig. S4A &B).

Finally, we assessed whether the evolutionary changes in the α subunit and those in the β subunit interacted with each other nonindependently. If the changes affect specificity entirely independently, $\Delta\Delta G_{spec}$ should equal the sum of the $\Delta\Delta G_{spec}$ acquired on each of the two branches, or –1.19 + –0.34 = –1.53. The observed $\Delta\Delta G_{spec}$ = –1.54, indistinguishable from this expectation (Fig. 4G).

Taken together, these data indicate that the specificity acquired by the derived complex Ancα + Ancβ is primarily attributable to substitutions in the α subunit, with substitutions in the β subunit making a much smaller contribution and nonadditive interactions between the two sets of changes having no role. The most important factor was that Ancα became much worse at binding itself than at binding Ancβ. Ancβ, by contrast, became slightly worse at binding Ancα than binding itself (Fig. 4G).

A one-residue deletion was the primary evolutionary cause of heterospecificity. We next sought to identify the particular historical substitutions in Ancα that conferred this heteromeric specificity on IF1. Only three sequence changes occurred on the branch from Ancαβ to Ancα: a single-residue deletion of a histidine at site 2 (ΔH2), a five-residue deletion in helix D (ΔD), and an amino acid replacement (v140A). ΔH2 is on the protein's N-terminal loop near IF1, and ΔD directly contributes to the interface. Substitution v140A is biochemically conservative and far away from the interface. The deletions are strictly conserved in Hbα subunits throughout the jawed vertebrates, whereas the amino acid at site 140 varies. We therefore focused first on the effects of the deletions.

To isolate the contribution of each deletion to the evolution of specificity, we introduced each one singly into Ancαβ and measured its effect on affinity and specificity when the mutant protein is mixed with Ancαβ. We found that introducing ΔH2 alone confers substantial specificity, recapitulating >80% of Ancα's acquired heterospecificity for Ancαβ ($\Delta\Delta G_{spec}$ = –0.99 out of a total $\Delta\Delta G_{spec}$ = –1.19 acquired along this branch) and about two-thirds of the total specificity acquired by the entire Ancα+Ancβ complex (Figs. 5A, C). ΔH2 enhances specificity by improving heterodimer affinity and reducing homodimer affinity, with both Kds very similar to those of Ancα (Fig. 5A; Fig. S8A & B).

The other deletion, ΔD, removes several residues that directly interact with the other subunit across IF1, but introducing this change into Ancαβ had a much weaker effect on specificity ($\Delta\Delta G_{spec}$ = –0.39, Fig. 5B; Fig. S8C & D). When the contributions of ΔH2 and ΔD to specificity are added together, they slightly exceed the specificity of Ancα, suggesting either a weak negative epistatic interaction between them or a small countervailing effect of the third change v104A. Taken together, these results indicate that ΔH2 was a large-effect historical sequence change that accounted for most of the specificity historically acquired by the derived Hb complex.

Structural mechanisms for the gain in specificity. We next considered the structural mechanisms by which ΔH2 conferred specificity by increasing heterodimer affinity and reducing homodimer affinity. For a mutation to have these opposite effects, it must yield favorable interactions when introduced into one side of the interface (in the heterodimer) but have deleterious effects when introduced twice (in the homodimer). Two kinds of mechanisms could cause these opposite effects. Either 1) the mutated residue interacts directly with the same residue on the other subunit favorably when one is in the derived state but unfavorably when both are, or 2) the symmetry of the interface is imperfect, such that introducing the mutation on one side of the interface is favorable but introducing it again onto the other side is net-unfavorable. The first scenario does not pertain in this case. Residue H2 is part of the N-terminal loop, which does not participate directly in IF1 but instead packs against helix H, which does contribute to IF1. But neither helix H nor the N-terminal loop contact the same elements in the other subunit across the interface (Fig. 5D). Asymmetry in the interface is therefore the likely of cause ΔH2's differential effects on heterodimer vs. homodimer specificity.

To gain insight into the possible nature of this asymmetry and the mechanism by which ΔH2 affects specificity, we modeled the structures of the Ancαβ homodimer, the Ancαβ$_{\Delta H2}$ homodimer, and the heterodimer of these two proteins. The Ancαβ homodimer itself begins with a subtle asymmetry: on one end of IF1, residue 130H on helix H sits close to 33R on the opposite subunit, which allows a cross-interface hydrogen bond to form; on the other end of the interface, the two residues are slightly further away from each other, leaving their hydrogen-bonding potential unsatisfied when bound (Fig. 5E). In the heterodimer, deleting His2 from one subunit repairs this unfavorable interaction. Specifically, the deletion shortens the N-terminal loop and changes its packing interaction against helix H, which causes helix H to slide along the interface by ~1 Å compared to its position in the unmutated Ancαβ homodimer (Figs. 5D, 5G). 130H moves closer to 37T on the other subunit, allowing it to form a new hydrogen bond across the interface, and several other interactions across the interface are also enhanced. On the other end of the isologous interactions, the favorable interactions found in the homodimer remain intact. This provides a potential structural explanation for how ΔH2 improves heterodimer affinity (Figs. 5D, G).

The modeled Ancαβ$_{\Delta H2}$ homodimer structure is notably asymmetric and suggests why introducing ΔH2 into both subunits reduces affinity (Fig. 5H). One side displays the favorable new cross-interface interactions caused by ΔH2 in the heterodimer, including the 130H-37T hydrogen bond. On the other side, however, the effect of the deletion is very different: ΔH2 again causes helix H to slide along the interface, but on this side the movement of 130H breaks the ancestral 130H-33R hydrogen bond, and 37T is also too far away to interact favorably. This leaves the side chains of both 130H and 33R unsatisfied, reducing homodimer affinity. In total, the homodimer of Ancαβ$_{\Delta H2}$ contains three unsatisfied hydrogen-bond donors/acceptors at these sites, whereas only one and two are unsatisfied in the heterodimer and the ancestral homodimer, respectively.

This mechanism appears to have persisted over time. The same pattern of interactions are found in the modeled structures of the hetero- and homodimers of Ancα + Ancβ (Fig. S9). It is also partially present in the crystal structure of the human Hb heterotetramer, where 33R also hydrogen bonds across IF1 to residue 130, but this interaction is again lacking in the homodimer of human Hbα, leaving 33R unsatisfied and explaining the weak homomeric affinity of Hbα (Fig. S9). At least some of the mechanisms of heterodimer specificity suggested by the structural models of the ancestral proteins are therefore present in the known structures of its present-day descendants.

Multiple historical sets of substitutions could have conferred heterospecificity. If specificity in an isologous interface can evolve simply by causing nonadditive impacts on the binding energies of heterodimer and homodimers, then there should be many mutations that have the potential to make the interface specific in one direction or another. Indeed, if the interface's symmetry is imperfect, then most mutations that affect affinity should impart specificity to some degree.

To test this hypothesis, we measured the effect on specificity of subsets of changes that occurred along the Ancβ lineage, which the results above show had strong effects on affinity when introduced all together. First, we tested the five substitutions that that occurred at the IF1 surface (Fig. 5E & 5F). We introduced these changes into Ancαβ (creating protein $Anc\alpha\beta_{IF1}$) and measured affinity and specificity when this protein is mixed with Ancαβ. These substitutions yield a highly heterospecific complex ($\Delta\Delta G_{spec}$ = –2.18, heterodimer occupancy 90%, Fig. 6A; Fig. S10A-C). Unlike the Ancα substitutions, the $Anc\beta_{IF1}$ substitutions confer heterospecificity by improving both homodimer and heterodimer affinity, but they improve the latter by more than the former.

Because $Anc\alpha\beta_{IF1}$ is specific in complex with Ancαβ, we wondered whether it would also be specific with Ancα. We found that this complex is weakly heterospecificity ($\Delta\Delta G_{spec}$ = –0.07, Fig. 6B), implying that other substitutions on the branch leading to Ancβ but not on the interface must have contributed to the evolution of specificity between $Anc\alpha\beta_{IF1}$ and Ancα. We therefore introduced an additional set of five historical substitutions that occurred in Ancβ but one structural layer away from IF1 (see ref. 16). This protein ($Anc\alpha\beta_{IF1+Adjacent}$) has strong heterospecificity when mixed with Ancα ($\Delta\Delta G_{spec}$ = –1.93, heterodimer occupancy >85%, Fig. 6D; Fig. S11D-F), because these mutations together in both heterodimer and the homodimer affinity, but with a larger improvement in the heterodimer. It is also moderately heterospecific when mixed with Ancαβ ($\Delta\Delta G_{spec}$ = –0.89).

Finally, we tested the effect of the adjacent substitutions on their own and found that they confer moderate specificity when mixed with Ancαβ ($\Delta\Delta G_{spec}$ = –0.85). These mutations impart specificity by causing almost identical changes in homo- and heterodimer affinity. They also confer some heterospecificity when $Anc\alpha\beta_{Adjacent}$ is mixed with Ancα ($\Delta\Delta G_{spec}$ = –0.57, Fig. S11A-D).

There are therefore several distinct sets of substitutions that occurred during history, and which can be sufficient to confer heterospecificity on their own (and in various combinations), and they do so via distinct patterns of effects on affinity. This degeneracy of mechanisms for evolving specificity arises because there are many ways in which the energy of binding can change nonadditively between heterodimer and homodimer. In every case, heteromeric specificity rather than preference for the homomer was the result.

DISCUSSION

This work provides a mechanistic history of the evolutionary transition from the ancestral Ancαβ homodimer to the derived Hb heterotetramer, illuminating the mechanisms for the evolution of tetramerization and the acquisition of heterospecificity from the ancestral non-specific dimer. Each transition was driven by a very simple genetic mechanism: a single substitution at IF2 conferred high affinity tetramerization, and a single amino acid deletion at IF1 conferred heteromeric specificity. Both evolutionary transitions were facilitated by the isologous architecture of Hb's two interfaces, which creates a mutational propensity to increase stoichiometry and acquire heterospecificity.

Symmetry facilitated evolution of the tetrameric stoichiometry. We found that tetramerization across IF2 was driven primarily by a single replacement to a bulky hydrophobic amino acid (q40W). In biochemical studies of extant protein interfaces, much of the free energy change in protein-protein binding is attributable to interactions of bulky hydrophobic residues with hydrophobic surface indentations (33), and mutations to bulky hydrophobic amino acids can drive assembly into high-order multimers (9,34-37). Similar substitutions during history may have been driving mechanisms during the evolution not only of Hb but of other molecular complexes, as well.

The majority of complexes assemble through isologous interfaces (38), and it has been suggested that this must means that isology confers some selective benefit by improving protein function (1). Our results suggest an alternative explanation. If mutations are much more likely to produce isologous complexes than nonisologous ones, then isologous complexes will predominate in nature, even if there is no systematic fitness difference between the two types of multimer. We found that although IF2 is intrinsically weak and mutation q40W cannot confer dimerization on its own, it can drive tetramerization if its effects are multiplied in an isologous higher-order complex. By contrast, If the interfaces were non-isologous -- with q40W interacting with a hydrophobic divot on some other surface of the facing subunit – then this favorable interaction would appear only once, and it would be insufficient to substantially improve binding energy and confer meaningful tetramer occupancy. Mutational propensity favors acquisition of isologous interfaces compared to head-to-tail associations. Isologous complexes are simply easier to produce by mutation, not more likely to be fixed by selection once they are generated, as has been suggested (1).

It has been observed that in high-order multimers, the interface with higher affinity usually evolves before the lower-affinity interface(s) (39-42), and a leading proposal is that this pattern reflects selection on complexes to assemble via pathways that prevent misassembly into anomalous high-order architectures (39,40). Hb evolution displays this pattern -- with the stronger interface IF1 evolving before IF2 (16) – but our work here suggests a different explanation: a low-affinity interface can mediate assembly into a higher-order stoichiometry only if a high-affinity interface is already present to multiply its effects. Mutational propensity therefore favors the evolution of complexes in which one interface is stronger than others; trajectories in which the stronger interface evolves first are far more likely than interfaces being acquired in the opposite order.

One interface confers specificity on a higher-order multimer. Our experiments show that evolutionary change at just one of Hb's interfaces was sufficient to confer specific assembly into heterotetramers. Specificity at IF1 alone was sufficient to mediate the heterospecificity of the tetramer because this interface is so much stronger than IF2: IF1 mediates the specific assembly of heterodimers, which assemble into heterotetramers across IF2, even though IF2 itself confers little or no specificity.

The specificity of IF1 and the isology of the complex also explains the *trans* conformation of Hb's quaternary structure, in which each Hbα subunit binds one Hbβ subunit across IF1 and a different Hbβ across IF2. The alternative *cis* conformation -- in which Hbα is paired with an Hbα (and Hbβ with Hbβ) across one of the interfaces – is never observed. Although IF2 imposes little or no specificity, its isologous orientation necessarily means that the two IF1-mediated heterodimers must be rotated 180° relative to each other, placing each Hbα across IF2 from the Hbβ of the other heterodimer. In the *cis* conformation, the heterodimers would not be rotated 180° relative to each other, and all the favorable interactions that IF2 comprises would not form; residue 40W, for example, would not face the hydrophobic divot on IF2 across the interface. Given the heterospecificity of IF1, isology constrains the Hb tetramer to its *trans* $\alpha_2\beta_2$ architecture.

These observations suggest a simple and potentially general mechanism for the evolution of specificity in the quaternary structures of high-order multimers. Specificity need not evolve at every interface in the complex, especially if the interfaces are isologous. Rather, mutations need only make the stronger interface specific to confer assembly into particular high-order architectures.

Imperfect symmetry allowed specificity to evolve in one subunit. We found that a single genetic change in one paralog – a one residue deletion in Ancα -- was sufficient to confer IF1's heterospecificity. This result contrasts with prior studies of nonisologous complexes, in which heterospecificity evolved because of genetic changes in both interacting subunits (7,12,17,18,20,23-25).

This difference in historical genetic mechanism reflects the opportunities presented by the two different types of multimeric architecture. In asymmetric complexes, a mutation in the "head" of one duplicate gene will not be sufficient to distinguish between its own tail and that of its paralog (unless it somehow changes the conformation of both distinct surfaces). In an isologous complex like Hb, however, a change in one subunit can confer specificity, because it makes the interface different between the heterodimer, the mutated homodimer, and the unmutated homodimer.

Acquiring specificity in an isologous interface does require the mutation to nonadditively change the affinity of the heterodimer relative to the homodimers. If the symmetry of such interfaces were perfect, a mutation in one subunit would affect interactions across the interface identically on each side of the interface, resulting in additive effects on affinity. Nonadditivity would arise only if mutations affect sites that interact with each other across the rotated interface. This would require either a mutation at the precise axis of rotational symmetry or multiple mutations at several sites.

If the symmetry is imperfect, however, a single mutation (like ΔH2 in Ancα) can affect interactions differently when it appears twice in the homodimer versus when it occurs once in the heterodimer. This observation is likely to have general relevance to the evolution of specificity. Virtually all isologous interfaces contain subtle asymmetries (43). This imperfection arises for two reasons: perfect symmetry is entropically unfavorable, and amino acids near the axis seldom face each other with perfect symmetry, because each amino acid itself is asymmetrical, and this asymmetry propagates elsewhere in the interface (43,44). Extant human hemoglobin is one of many examples of isologous interfaces in which asymmetry is imperfect (45). Isologous interfaces therefore provide a nearly universal starting point for homo- or heterospecificity to be acquired by substitutions in a single subunit.

Specificity evolved through a single mutation. We found that a single mutation – deletion of residue His2 in the alpha subunit – conferred most of the heterospecificity of Ancα + Ancβ. This simple mechanism was possible because only a small change in relative binding energy is required to yield substantial changes in specificity. The IF1 of Ancα+Ancβ occupies 90% heterodimer at equal and saturating concentrations, but its $\Delta\Delta G_{spec} = -1.54$; ΔH2 alone caused most of this shift, conferring heterodimer occupancy of almost 80% via a $\Delta\Delta G_{spec} = -0.99$. These differences in binding energy are less than that of a typical hydrogen bond or burial of a large hydrophobic residue. Our structural models reveal differences in hydrogen bonding and other interactions across the homodimer vs. heterodimer interfaces that could easily yield energetic differences of this magnitude. These results are consistent with recent in silico findings that small differences in ΔG can cause large differences in occupancy between homodimers and heterodimers (27).

Why do such subtle changes in energy cause large effects on specificity? Mutations that cause a modest deviation in binding energies can cause large changes in occupancy because of the nonlinear Boltzmann relationship between these quantities (Fig. 4C). Moreover, specificity is

548 determined by the deviation from additivity between homodimers and heterodimer, so small
549 differences in the free energy of binding propagate into even larger changes in specificity. We
550 therefore predict that the evolution of specificity in paralogous complexes with symmetrical
551 interfaces will often be attributable to one or a few genetic changes with relatively subtle
552 structural and energetic effects. That specificity can evolve so easily also implies that paralog
553 interference after gene duplication (46,47) may often be easily resolved through one or a few
554 mutations.

555 If specificity can be acquired by small deviations from energetic additivity in either direction, one
556 might expect that homomeric and heteromeric specificity would be equally likely to evolve. But
557 empirical observations suggest that heteromers evolve much more frequently after gene
558 duplication (12,48). Our findings suggest a plausible explanation for this pattern. We observed
559 that the critical mutation for conferring specificity on Hb does so because imperfect asymmetry
560 in the interface creates a kind of antagonistic pleiotropy: a favorable interaction occurs when the
561 mutation is introduced once in the heteromer, but it fails to produce the same favorable contact
562 and even disrupts a different favorable contact when introduced again on the other side of the
563 interface in the homomer. Heterospecificity will result whenever asymmetry causes this kind of
564 antagonistic pleiotropy, such that a favorable interaction can be optimized when it is iterated
565 once but not twice. In contrast, homomeric specificity requires a mutation to be even more
566 favorable the second time it is introduced on the other side of an interface. For this to occur,
567 imperfect symmetry must synergistically enhance the interactions caused by the two iterations
568 of the mutation in the homodimer. This scenario seems far less likely than an antagonistic
569 effect, because favorable interactions are constrained in many ways, requiring fairly precise
570 compatibility of polarity, size, angle, etc. The imperfect symmetry of isologous interfaces may
571 therefore create a mutational propensity that favors the evolution of heteromeric over
572 homomeric specificity.

573 Taken together, our observations contribute to a growing body of evidence that complex
574 multimeric complexes can evolve through simple genetic mechanisms (5, 14, 34, 36,49-53). In
575 Hb evolution, a single substitution in one of the duplicated genes was sufficient to cause a
576 doubling in stoichiometry from dimer to tetramer, and a single-residue deletion at one interface
577 in the other subunit was sufficient to confer strong preference for the $\alpha_2\beta_2$ heterotetrameric form.
578 Although other substitutions enhanced these effects, and others may have permitted or
579 entrenched them (5,54), our data indicate that discrete evolutionary increases in complexity can
580 occur by very short mutational paths from simpler ancestral forms. The single-mutation
581 evolutionary jumps in the stoichiometry and specificity of the Hb complex were possible
582 because they took place in the context of an isologous complex in which symmetry is slightly
583 imperfect. Because many multimers share similar structural properties, we predict that, when
584 other multimeric complexes are studied in detail, simple mechanisms will be found to have
585 driven their historical elaboration.

METHODS

Sequence data, alignment, phylogeny, and ancestral sequence reconstruction. The reconstructed ancestral sequences used here are the same as those reported previously (16). Briefly, 177 amino acid sequences of hemoglobin and related paralogs were collected and aligned. The maximum likelihood (ML) phylogenetic tree was inferred using the AIC best-fit model, LG+G+F (55,56). The phylogeny was rooted using as outgroups neuroglobin and globin X, which are found in both deuterostomes and protostomes and diverged prior to the gene duplications that produced vertebrate myoglobin and the hemoglobin subunits. Ancestral sequence reconstruction was performed using the empirical Bayes method (57), given the alignment, ML phylogeny, ML branch lengths, and ML model parameters. Reconstructed ancestors that were used in this study have been deposited previously in GenBank (IDs MT079112, MT079113, MT079114, MT079115).

The historical mutations that we introduced into those ancestral proteins are the following. For the set *IF1-reverted*, all sites in IF1 that were substituted on the branch leading to Ancb are reverted to the ancestral state found in Ancab; the mutations introduced are V36t, Y38h, V115a, V119e, H130r, D134e. For the set *IF2-reverted*, all sites that were substituted in IF2 on the branch leading to Ancb are reverted to the ancestral state found in Ancab; the mutations introduced are T37v, W40q, R43t, H100r, E104h. For the set *IF1*, all sites at IF1 that were substituted between Ancab and Ancb are changed to the derived state found in Ancb; the mutations introduced are t37V, k58M, r107K, h130Q, d134Q4. For the set *Adjacent*, five sites adjacent to IF1 that were substituted between Ancab and Ancb are changed to the derived state found in Ancb; the mutations introduced are h47S, s60N, q62K, a96S, h97E. The set *IF1+Adjacent* is the union of the sets *IF1* and *Adjacent*. Deletion DD deletes residues a54, e55, a56, i57, and k58 from Ancab.

Recombinant protein expression. Coding sequences for reconstructed ancestral proteins were optimized for expression in *Escherichia coli* using IDT Codon Optimization and synthesized *de novo* as gBlocks (IDT). Coding sequences were cloned by Gibson assembly into vector pLIC (58) under control of a T7 polymerase promoter. For co-expression of Ancα+Ancβ, a polycistronic operon was constructed under control of a T7 promoter and separated by a spacer containing a stop codon and ribosome binding site, as described in (59).

BL21 (DE3) *Esherichia coli* cells (New England Biolabs) were heat-shock transformed and plated onto Luria broth (LB) containing 50 ug/mL carbenicillin. For the starter culture, a single colony was inoculated into 50 mL of LB with 1:1000 dilution of working-stock carbenicillin and grown overnight. 5 mL of the starter culture were inoculated into a larger 500-mL terrific broth (TB) mixture containing the appropriate antibiotic concentration. Cells were grown at 37° C and shaken at 225 rpm in an incubator until they reached an optical density at 600 nm of 0.6-0.8.

For expression of single globin proteins, 100 uM of isopropyl-β-D-1-thiogalactopyranoside (IPTG) and 25 mg/500 mL of hemin were added to each culture. Expression of single proteins in culture were done overnight at 22° C. Cells were collected by centrifugation at 4,000*g* and stored at -80° C until protein purification. Coexpressed proteins were induced using 500 mM

IPTG expression with 25 mg/500 mL hemin for 4 hours at 37°C. Cells were collected by centrifugation at 4,000*g*, immediately followed by purification.

Human hemoglobin was bought commercially (Sigma-Aldridge) and resuspended in PBS.

We attempted to co-express and purify Ancα$\beta_{\Delta H2}$ in complex with Ancαβ_{40W}, but we were not able to identify conditions at which the two species could be expressed and purified to near-equal concentrations.

Protein purification by ion exchange. All singly expressed proteins (all ancestral globins except Ancα+Ancβ) were purified using ion exchange chromatography. All buffers were vacuum filtered through a 0.2 µM PFTE membrane (Omnipore). After expression, cells were resuspended in 30 mL of 50 mM Tris-Base (pH 6.88). The resuspended cells were placed in a 10 mL falcon tube and lysed using a FB505 sonicator (1s on/off for three cycles, each 1 minute). The lysate was saturated with CO, transferred to a 30 mL round bottom tube, and centrifuged at 20,000*g* for 60 minutes to separate supernatant from non-soluble cell debris. The supernatant was collected and syringe-filtered using HPX Millex Durapore filters (Millipore) to further remove debris. A HiTrap SP cation exchange (GE) column was attached to an FPLC system (Biorad) and equilibrated in 50 mM Tris-Base (pH 6.88). The lysate was passed over the column. 50 mL of 50 mM Trise-Base (pH 6.88) was run through the SP column to remove weakly bound non-target soluble products. Elution of bound ancestral Hbs was performed with 100-mL gradient of 50mM Tris-Base 1 M NaCl (pH 6.88) buffer which was run through the column from 0% to 100%. 1.5 mL fractions were captured during the gradient process, all fractions containing red eluant were put into an Amicon ultra-15 tube and concentrated by centrifugation at 4,000g to a final volume of 1 mL. For additional purification, concentrated sample was injected into a HiPrep 16/60 Sephacryl S-100 HR size exclusion chromatography (SEC) column. The column was equilibrated in phosphate buffered saline (PBS) at pH 7.4. Purified ancestral globins elute at different volumes depending on the protein's complex stoichiometry: 48-52 for tetramers, 56-60 for dimers, and 65-67 for monomers. The purified proteins were concentrated as mentioned above and then flash frozen with liquid nitrogen.

Protein purification by zinc affinity chromatography. Coexpressed proteins Ancα + Ancβ were purified using zinc-affinity chromatography, which was performed using a HisTrap metal affinity column (GE) on a Biorad NGC Quest. Nickle ions were stripped from the column (buffer 100 mM EDTA, 100 mM NaCl, 20 mM TRIS, pH 8.0), followed by five column volumes of water. To attach zinc to the column, 0.1 M $ZnSO_4$ was passed over until conductance was stable, approximately 5 column volumes, followed by five column volumes of water. After expression, cells were resuspended in a 50 mL lysis buffer (20 mM Tris, 150 mM Nacl, 10% glycerol (v/v), 1mM BME, 0.05% Tween-20, and 1 Roche Protease EDTA-free inhibitor tablet, pH 7.40), sonicated as described above, and the lysate passed through the prepared column. To remove non-specifically bound protein, the column was washed with 50 mL of lysis buffer. Bound protein was then eluted across a gradient of imidazole concentrations (0 to 500 mM) in a total of 100 mL elution buffer (20 mM Tris, 150 mM NaCl, 500 mM imidazole, 10% glycerol, and 1 mM BME, pH 7.4). 1 mL fractions were collected. The fraction corresponding to the second peak of UV

absorbance at 280 nm has a visible red color and was collected and concentrated as described above. The concentrated solution was injected into a Biorad ENrich 650 10 x 300 columns for additional purification and eluted in PBS buffer.

Size exclusion chromatography assay. For protein concentrations from 0 to 500 μM, size exclusion chromatography was performed using a Superdex 75 increase 10/300 GL column (GE) equilibrated in PBS, then injected with 250 μL of sample using a 2 mL injection loop on an Biorad NGC Quest FPLC and monitored by absorbance at 280 nm. For proteins at concentration 1 mM, a HiPrep 16/60 Sephacryl S-100 HR was equilibrated in PBS using an AKTAprime FPLC, then injected with 1mL sample and monitored by absorbance at 280 nm.

Native Mass Spectrometry. Protein samples were buffer exchanged into 200mM ammonium acetate using either a centrifugal buffer exchange device (Micro Bio-Spin P-6 Gel, Bio-Rad) or a dialysis device (Slide-A-Lyzer MINI Dialysis Unit, 10000 MWCO, Thermo) prior to native MS experiments. Samples were loaded into gold-coated glass capillaries made in-house and introduced to Synapt G1 HDMS instrument (Waters corporation) equipped with a 32k RF generator (29). The instrument was set to a source pressure of 5.47 mbar, capillary voltage of 1.75 kV, sampling cone voltage of 20 V, extractor cone voltage of 5.0 V, trap collision voltage of 10 V, collision gas (Argon) flow rate of 2 mL/min (2.65 x 10 -2 mbar), and T-wave settings (velocity/height) for trap, IMS and transfer of 100 ms -1 /0.2 V, 300 ms -1 /16.0 V, and 100 ms -1 /10.0 V, respectively. The source temperature (70 °C) and trap bias (30 V) were optimized. Part of the native MS experiments were conducted by Thermo Scientific Exactive Plus Orbitrap with Extended Mass Range (EMR) with tuning as follow: source DC offset of 15 V, injection flatapole DC to 13 V, inter flatapole lens to 5, bent flatapole DC to 4, transfer multipole DC to 3 and C trap entrance lens to 0, trapping gas pressure to 5.0 with the CE to 10, spray voltage to 1.50 kV, capillary temperature to 100 °C, maximum inject time to 100 ms. Mass spectra were acquired with a setting of 8750 resolution, microscans set to 1 and averaging set to 100. Mass spectra were deconvoluted using Unidec (60).

Calculating multimerization affinity of homomers. To estimate Kd of the monomer-to-homodimer transition of singly expressed proteins, we performed nMS at variable protein concentrations (P_{tot}). The occupancy of each oligomeric state at each concentration was calculated as the proportion of all globin subunits in that state, based on the summed areas under the corresponding peaks in the native MS spectrum. The fraction of subunits assembled into dimers (*Fd*) includes dimers and tetramers and is defined as

$$Fd = \frac{2x_d+4x_t}{(x_m + 2x_d+4x_t)},$$

where x_m, x_d, and x_t are the total signal intensities of all peaks corresponding to the monomeric, dimeric and tetrameric stoichiometries, respectively. Nonlinear regression was used to find the best-fit value of Kd of dimerization using the equation:

$$F_d = \frac{1}{P_{tot}} * \frac{(4P_{tot} + K_d) - \sqrt{(4P_{tot} + K_d)^2 - 16{P_{tot}}^2}}{4}$$

To estimate the Kd of the dimer–tetramer transition, the fraction of subunits assembled into tetramers is defined as

$$F_t = \frac{4x_t}{(2x_d + 4x_t)}.$$

The concentration of all dimers is defined as

$$P_d = F_d \times P_{tot}.$$

Nonlinear regression was then used to find the Kd of tetramerization using the equation:

$$F_t = \frac{1}{P_d} * \frac{(4P_d + K_d) - \sqrt{(4P_d + K_d)^2 - 16{P_d}^2}}{4}$$

Calculating multimerization affinity of heteromers. To determine the Kd of heterodimerization, we used nMS to measure stoichiometries across a titration series in which one protein's concentration was held constant at 50 mM and the other was added at variable concentration (1 to 50 mM). From the nMS spectrum, we estimated the proportion of the heterodimer and the two homodimers as

$$F_{\alpha\alpha} = \frac{2x_{\alpha\alpha}}{(2x_{\alpha\alpha} + 2x_{\alpha\beta} + 2x_{\beta\beta} + x_\alpha + x_\beta)}$$

$$F_{\alpha\beta} = \frac{2x_{\alpha\beta}}{(2x_{\alpha\alpha} + 2x_{\alpha\beta} + 2x_{\beta\beta} + x_\alpha + x_\beta)}$$

$$F_{\beta\beta} = \frac{2x_{\beta\beta}}{(2x_{\alpha\alpha} + 2x_{\alpha\beta} + 2x_{\beta\beta} + x_\alpha + x_\beta)}$$

where each *x* represents the signal intensity of all peaks corresponding to the species denoted in the subscript. The dissociation constant for each dimer is defined as $Kd_1 = \frac{{x_\alpha}^2}{x_{\alpha\alpha}}$, $Kd_2 = \frac{{x_\beta}^2}{x_{\beta\beta}}$, and $Kd_3 = \frac{x_\alpha x_\beta}{x_{\alpha\beta}}$. By substitution, $F_{\alpha\beta}$ can be expressed as

$$F_{\alpha\beta} = \frac{\sqrt{Kd_1 * Kd_2 * F_{\alpha\alpha} * F_{\beta\beta}}}{Kd_3}$$

Kd_3 was estimated using this equation by nonlinear regression, where $F_{\alpha\alpha}, F_{\alpha\beta}$ and $F_{\beta\beta}$ were measured using the titration series, and the affinities Kd_1 and Kd_2 were assigned the values estimated in the homodimerization experiments described above.

Prediction of homodimer and heterodimer occupancy at high concentrations. The occupancy of each dimer at physiologically relevant concentrations (1 mM total globin subunits) was predicted as follows, because nMS is limited to concentrations <100mM. In a mixture of two types of globins *A* and *B*, the total concentration of each subunit can be expressed in terms of the concentration of monomers [A] and [B] in the mixture:

$$[A]_{tot} = [A] + [AB] + 2[AA] = [A] + \frac{[A][B]}{Kd_3} + \frac{2[A]^2}{Kd_1}$$

$$[B]_{tot} = [B] + [AB] + 2[BB] = [B] + \frac{[A][B]}{Kd_3} + \frac{2[B]^2}{Kd_2}$$

We used these equations to predict [A] and [B] at any value of C_A and C_B given the experimentally estimated Kds. The concentration of each dimer was then estimated using the equations $[AA] = \frac{[A]^2}{Kd_1}$, $[BB] = \frac{2[A][B]}{Kd_3}$, and $[BB] = \frac{[B]^2}{Kd_2}$.

Establishing the upper limit of IF2 Kd. We estimated the minimum Kd of assembly across IF2 by Ancαβ $_{37V+40W;\ IF1\ removed}$, because no homotetramer was observed using nMS at a protein concentration of 20 mM. The minimum detection limit for dimers in the nMS assay is 1 mM. Kd is defined as $Kd = \frac{[M]^2}{[D]}$, where $[M]$ and $[D]$ are the concentrations of monomer and dimer, respectively. Therefore

$$Kd_{min} = \frac{(20 * 10^{-6})^2 M}{1 * 10^{-6} M} = 400\ uM$$

Determining ΔΔG of specificity. Specificity for heterodimer assembly between two paralogs can be defined as the difference between the additive affinity of the heterodimer and the measured affinity of the heterodimer, using ΔGs derived measured dimerization affinity for two homodimers and their respective heterodimer. The additive affinity of the heterodimer is defined as the averaged ΔG of both homodimers:

$$\Delta G^{additive}_{heterodimer} = \frac{\Delta G_{homodimer\ 1} + \Delta G_{homodimer\ 2}}{2}$$

Specificity is then the difference between the additive and measured heterodimer ΔG.

$$\Delta\Delta G_{spec} = \Delta G^{measured}_{heterodimer} - \Delta G^{additive}_{heterodimer}$$

This metric is analogous to the coupling energy, which expresses the deviation of the measured DG for a double mutant from that expected given the DGs of two single mutants assuming additivity (62-64).

Quantifying non-additive effect on specificity between Ancα and Ancβ. The non-additive effect on specificity can be defined as the difference between the predict and measured ΔΔG of the derived complex Ancα + Ancβ.

$$\Delta\Delta\Delta G = \Delta\Delta G_{\alpha+\beta} - \left(\Delta\Delta G_{\alpha} + \Delta\Delta G_{\beta}\right).$$

Prediction of monomer, dimer, and tetramer occupancies with no IF2 specificity. The occupancy of monomers, dimers, and tetramers between 1 mM and 4 mM predicted was calculated as follows. The concentration of subunit in each stoichiometric species can be expressed in terms of the concentration of monomers [A] and [B]:

$$[\mathrm{A}]_{\mathrm{tot}} = [\mathrm{A}] + [\mathrm{AB}] + 2[\mathrm{AA}] + [\mathrm{ABBB}] + 2[\mathrm{AABB}]$$

$$= [\mathrm{A}] + \frac{[\mathrm{A}][\mathrm{B}]}{\mathrm{Kd}_3} + \frac{2[\mathrm{A}]^2}{\mathrm{Kd}_1} + \frac{\frac{[\mathrm{A}][\mathrm{B}]^3}{\mathrm{Kd}_2 * \mathrm{Kd}_3}}{\mathrm{Kd}_4} + \frac{\frac{2[\mathrm{A}]^2[\mathrm{B}]^2}{\mathrm{Kd}_2^2}}{\mathrm{Kd}_4}$$

$$[\mathrm{B}]_{\mathrm{tot}} = [\mathrm{B}] + [\mathrm{AB}] + 2[\mathrm{BB}] + 2[\mathrm{AABB}] + 3[\mathrm{ABBB}] + 4[\mathrm{BBBB}]$$

$$= [\mathrm{B}] + \frac{[\mathrm{A}][\mathrm{B}]}{\mathrm{Kd}_3} + \frac{2[\mathrm{B}]^2}{\mathrm{Kd}_2} + \frac{\frac{2[\mathrm{A}]^2[\mathrm{B}]^2}{\mathrm{Kd}_2^2}}{\mathrm{Kd}_4} + \frac{\frac{3[\mathrm{A}][\mathrm{B}]^3}{\mathrm{Kd}_2 * \mathrm{Kd}_3}}{\mathrm{Kd}_4} + \frac{\frac{4[\mathrm{B}]^4}{\mathrm{Kd}_3^2}}{\mathrm{Kd}_4}$$

We used these equations to predict [A] and [B] across a range of $[\mathrm{A}]_{\mathrm{tot}}$ and $[\mathrm{B}]_{\mathrm{tot}}$ values given previously measured equilibrium constants. Predicted [A] and [B] concentrations were used to calculate the concentration of homodimers and heterodimers as described above, and the concentration of tetramers were calculated using the following equations:

$$[\mathrm{BBBB}] = \frac{\frac{[\mathrm{B}]^4}{\mathrm{Kd}_2^2}}{\mathrm{Kd}_4}$$

$$[\mathrm{ABBB}] = \frac{\frac{[\mathrm{A}][\mathrm{B}]^3}{\mathrm{Kd}_2 * \mathrm{Kd}_3}}{\mathrm{Kd}_4}$$

$$[\mathrm{AABB}] = \frac{\frac{[\mathrm{A}]^2[\mathrm{B}]^2}{\mathrm{Kd}_3^2}}{\mathrm{Kd}_4}$$

where $[\mathrm{BBBB}]$ corresponds to the concentration of homotetramer, $[\mathrm{ABBB}]$ is concentration of $\alpha_1\beta_3$ tetramers, and $[\mathrm{AABB}]$ is the concentration of $\alpha_2\beta_2$ heterotetramers.

Homology models. SWISS-Model was used to generate a structural model of the Ancαβ_{q40W} homotetramer using the crystal structure of the human Hbβ homotetramer (PDB 1CMB) as template, which was then refined using Rosetta's Fast Relax protocol, which energetically minimizes the initial structure via small adjustments to the backbone and side chain torsion angles (61). PyMOL V2.1 was used to visualize the proteins and capture images.

IF1-mediated homodimers were generated by the same procedure, except for homodimers of Ancα or Ancα$\beta_{\Delta D}$, for which the homodimer of human Hbα (PDB 3S48) was used as template. IF1-mediated heterodimers were generated by the same procedure but using the heterotetramer of human Hb (PDB 4HHB).

Acknowledgements and funding sources

We thank members of the Thornton and Laganowsky labs for helpful advice and comments. Supported by NIGMS R01GM131128, R35 GM14533601, 1T32GM13978201.

Author Contributions: CRCR: project conception and design, protein expression and purification, chromatography experiments, structural modeling, data analysis, and writing (initial draft and revisions). JL: design, execution, and analysis of mass spectrometry experiments, and writing (revisions). AP: project conception and design, protein expression and purification, chromatography experiments, data analysis, writing (initial draft and revisions). AL: design and analysis of mass spectrometry experiments, writing (revisions). JWT: project conception and design, data analysis, project management, funding acquisition, writing (initial draft and revisions).

Citations

1. D. S. Goodsell, A. J. Olson, Structural symmetry and protein function. *Annual Review of Biophysics and Biomolecular Structure* **29**, 105-153 (2000).
2. J. A. Marsh, S. A. Teichmann, Structure, dynamics, assembly, and evolution of protein complexes. *Annu Rev Biochem* **84**, 551-575 (2015).
3. M. Lynch, The Evolution of multimeric protein assemblages. *Molecular Biology and Evolution* **29**, 1353-1366 (2012).
4. M. Lynch, Evolutionary diversification of the multimeric states of proteins. *Proceedings of the National Academy of Sciences* **110**, E2821-E2828 (2013).
5. A. S. Pillai, G. K. A. Hochberg, J. W. Thornton, Simple mechanisms for the evolution of protein complexity. *Protein Sci* **31**, e4449 (2022).
6. G. K. A. Hochberg *et al.*, Structural principles that enable oligomeric small heat-shock protein paralogs to evolve distinct functions. *Science* **359**, 930-935 (2018).
7. E. J. Capra, B. S. Perchuk, J. M. Skerker, M. T. Laub, Adaptive mutations that prevent crosstalk enable the expansion of paralogous signaling protein families. *Cell* **150**, 222-232 (2012).
8. S. E. Ahnert, J. A. Marsh, H. Hernández, C. V. Robinson, S. A. Teichmann, Principles of assembly reveal a periodic table of protein complexes. *Science* **350**, 245 (2015).
9. H. Garcia-Seisdedos, C. Empereur-Mot, N. Elad, E. D. Levy, Proteins evolve on the edge of supramolecular self-assembly. *Nature* **548**, 244-247 (2017).
10. G. Diss *et al.*, Gene duplication can impart fragility, not robustness, in the yeast protein interaction network. *Science* **355**, 630-634 (2017).
11. S. Jones, J. M. Thornton, Principles of protein-protein interactions. *Proc Natl Acad Sci U S A* **93**, 13-20 (1996).
12. A. Marchant *et al.*, The role of structural pleiotropy and regulatory evolution in the retention of heteromers of paralogs. *Elife* **8** (2019).
13. J. B. Pereira-Leal, E. D. Levy, C. Kamp, S. A. Teichmann, Evolution of protein complexes by duplication of homomeric interactions. *Genome Biol* **8**, R51 (2007).
14. D. Grueninger *et al.*, Designed protein-protein association. *Science* **319**, 206-209 (2008).
15. J. Monod, J. Wyman, J.-P. Changeux, On the nature of allosteric transitions: a plausible model. *J Mol Biol* **12**, 88-118 (1965).
16. A. S. Pillai *et al.*, Origin of complexity in haemoglobin evolution. *Nature* **581**, 480-485 (2020).
17. G. C. Finnigan, V. Hanson-Smith, T. H. Stevens, J. W. Thornton, Evolution of increased complexity in a molecular machine. *Nature* **481**, 360-364 (2012).
18. J. R. EmLaw *et al.*, A single historical substitution drives an increase in acetylcholine receptor complexity. *Proc Natl Acad Sci U S A* **118** (2021).
19. O. Ashenberg, K. Rozen-Gagnon, M. T. Laub, A. E. Keating, Determinants of homodimerization specificity in histidine kinases. *J Mol Biol* **413**, 222-235 (2011).
20. D. Ding *et al.*, Co-evolution of interacting proteins through non-contacting and non-specific mutations. *Nat Ecol Evol* **6**, 590-603 (2022).

21. D. A. Ghose, K. E. Przydzial, E. M. Mahoney, A. E. Keating, M. T. Laub, Marginal specificity in protein interactions constrains evolution of a paralogous family. *Proc Natl Acad Sci U S A* **120**, e2221163120 (2023).
22. C. J. McClune, A. Alvarez-Buylla, C. A. Voigt, M. T. Laub, Engineering orthogonal signalling pathways
23. C. D. Aakre *et al.*, Evolving new protein-protein interaction specificity through promiscuous intermediates. *Cell* **163**, 594-606 (2015).
24. T. V. Lite *et al.*, Uncovering the basis of protein-protein interaction specificity with a combinatorially complete library. *Elife* **9** (2020).
25. I. Nocedal, M. T. Laub, Ancestral reconstruction of duplicated signaling proteins reveals the evolution of signaling specificity. *eLife* **11**, e77346 (2022).
26. G. K. A. Hochberg, J. W. Thornton, Reconstructing ancient proteins to understand the causes of structure and function. *Annual Review of Biophysics* **46**, 247-269 (2017).
27. A. F. Cisneros, L. Nielly-Thibault, S. Mallik, E. D. Levy, C. R. Landry, Mutational biases favor complexity increases in protein interaction networks after gene duplication. *Mol Syst Biol* **20**, 549-572 (2024).
28. M. F. Perutz, H. Muirhead, J. M. Cox, L. C. G. Goaman, Three-dimensional fourier synthesis of horse oxyhaemoglobin at 2.8 Å resolution: the atomic model. *Nature* **219**, 131-139 (1968).
29. A. C. Leney, A. J. R. Heck, Native mass spectrometry: what is in the name? *Journal of the American Society for Mass Spectrometry* **28**, 5-13 (2016).
30. E. Stellwagen, "Chapter 23 gel filtration" in *Methods in Enzymology,* R. R. Burgess, M. P. Deutscher, Eds. (Academic Press, 2009), **vol. 463**, pp. 373-385.
31. L. Kiger *et al.*, Thermodynamic studies on the equilibrium properties of a series of recombinant betaW37 hemoglobin mutants. *Biochemistry* **37**, 4336-4345 (1998).
32. G. Snyder, B. Sheafor, Red blood ells: centerpiece in the evolution of the vertebrate circulatory system. *Am. Zool.* **39** (1999).
33. A. A. Bogan, K. S. Thorn, Anatomy of hot spots in protein interfaces. *J Mol Biol* **280**, 1-9 (1998).
34. J. Jee, I.-J. L. Byeon, J. M. Louis, A. M. Gronenborn, The point mutation A34F causes dimerization of GB1. *Proteins: Structure, Function, and Bioinformatics* **71**, 1420-1431 (2008).
35. H. Garcia Seisdedos, T. Levin, G. Shapira, S. Freud, E. D. Levy, Mutant libraries reveal negative design shielding proteins from supramolecular self-assembly and relocalization in cells. *Proceedings of the National Academy of Sciences* **119**, e2101117119 (2022).
36. C.-S. Chen *et al.*, How to change the oligomeric state of a circular protein assembly: switch from 11-subunit to 12-subunit TRAP suggests a general mechanism. *PLOS ONE* **6**, e25296 (2011).
37. B. Kuhlman, J. W. O'Neill, D. E. Kim, K. Y. Zhang, D. Baker, Conversion of monomeric protein L to an obligate dimer by computational protein design. *Proc Natl Acad Sci U S A* **98**, 10687-10691 (2001).
38. H. Schweke *et al.*, An atlas of protein homo-oligomerization across domains of life. *Cell* **187**, 999-1010.e1015 (2024).

39. J. A. Marsh *et al.*, Protein complexes are under evolutionary selection to assemble via ordered pathways. *Cell* **153**, 461-470 (2013).
40. E. D. Levy, E. B. Erba, C. V. Robinson, S. A. Teichmann, Assembly reflects evolution of protein complexes. *Nature* **453**, 1262-1265 (2008).
41. R. P. Bahadur, F. Rodier, J. Janin, A dissection of the protein–protein interfaces in icosahedral virus capsids. *Journal of Molecular Biology* **367**, 574-590 (2007).
42. E. T. Powers, D. L. Powers, A perspective on mechanisms of protein tetramer formation. *Biophysical Journal* **85**, 3587-3599 (2003).
43. J. T. Brennecke, B. L. de Groot, Quantifying asymmetry of multimeric proteins. *The Journal of Physical Chemistry A* **122**, 7924-7930 (2018).
44. M. Bonjack-Shterengartz, D. Avnir, The near-symmetry of proteins. *Proteins: Structure, Function, and Bioinformatics* **83**, 722-734 (2015).
45. M. F. Perutz, G. Fermi, B. Luisi, B. Shaanan, R. C. Liddington, Stereochemistry of cooperative mechanisms in hemoglobin. *Accounts of Chemical Research* **20**, 309-321 (1987).
46. C. R. Baker, V. Hanson-Smith, A. D. Johnson, Following gene duplication, paralog interference constrains transcriptional circuit evolution. *Science* **342**, 104-108 (2013).
47. J. T. Bri*f*ham, J. E. Brown, A. Rodríguez-Marí, J. M. Catchen, J. W. Thornton, Evolution of a new Function by degenerative mutation in cephalochordate steroid receptors. *PLOS Genetics* **4**, e1000191 (2008).
48. S. Mallik, D. S. Tawfik, Determining the interaction status and evolutionary fate of duplicated homomeric proteins. *PLOS Computational Biology* **16**, e1008145 (2020).
49. S. Liu et al., Nonnatural protein–protein interaction-pair design by key residues grafting. *Proceedings of the National Academy of Sciences* **104**, 5330-5335 (2007).
50. G. Guntas, C. Purbeck, B. Kuhlman, Engineering a protein–protein interface using a computationally designed library. *Proceedings of the National Academy of Sciences* **107**, 19296-19301 (2010).
51. D. P. Anderson et al., Evolution of an ancient protein function involved in organized multicellularity in animals. *Elife* **5**, e10147 (2016).
52. L. Schulz et al., Evolution of increased complexity and specificity at the dawn of form I Rubiscos. *Science* **378**, 155-160 (2022).
53. F. L. Sendker et al., Emergence of fractal geometries in the evolution of a metabolic enzyme. *Nature* **628**, 894-900 (2024).
54. G. K. A. Hochberg *et al.*, A hydrophobic ratchet entrenches molecular complexes. *Nature* **588**, 503-508 (2020).
55. D. Posada, T. R. Buckley, Model Selection and Model Averaging in Phylogenetics: Advantages of Akaike Information Criterion and Bayesian Approaches Over Likelihood Ratio Tests. *Systematic Biology* **53**, 793-808 (2004).
56. S. Q. Le, O. Gascuel, An improved general amino acid replacement matrix. *Mol Biol Evol* **25**, 1307-1320 (2008).
57. Z. Yang, PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* **13**, 555-556 (1997).
58. J. Yang, Z. Zhang, X. A. Zhang, Q. Luo, A ligation-independent cloning method using nicking DNA endonuclease. *Biotechniques* **49**, 817-821 (2010).

59. C. Natarajan *et al.*, Expression and purification of recombinant hemoglobin in Escherichia coli. *PLoS One* **6**, e20176 (2011).
60. M. T. Marty *et al.*, Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. *Analytical Chemistry* **87**, 4370-4376 (2015).
61. L. G. Nivón, R. Moretti, D. Baker, A Pareto-Optimal Refinement Method for Protein Design Scaffolds. *PLOS ONE* **8**, e59004 (2013).
62. P. J. Carter, G. Winter, A. J. Wilkinson, A. R. Fersht, The use of double mutants to detect structural changes in the active site of the tyrosyl-tRNA synthetase (Bacillus stearothermophilus). *Cell* **38**, 835-840 (1984).
63. A. Horovitz, Double-mutant cycles: a powerful tool for analyzing protein structure and function. *Folding and Design* **1**, R121-R126 (1996).
64. J. A. Wells, Additivity of mutational effects in proteins. *Biochemistry* **29**, 8509-8517 (1990).

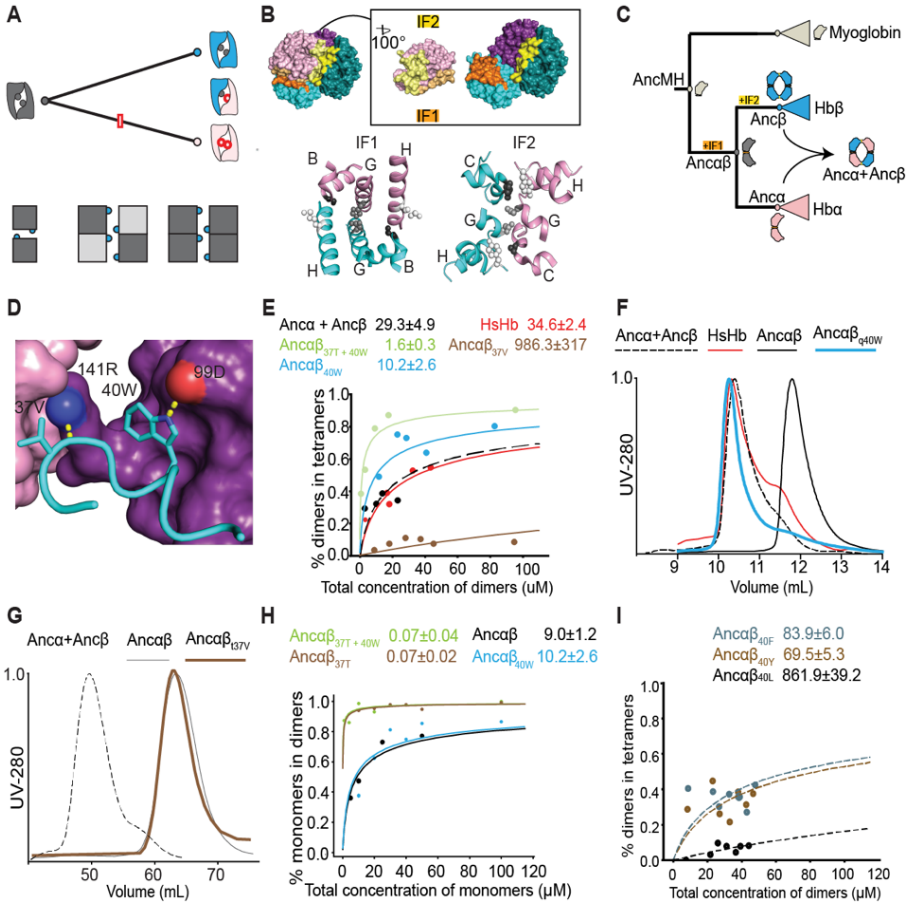


Figure 1. A single substitution confers tetramerization on an ancestral dimer. (A) A substitution in one subunit can potentially affect specificity and stoichiometry in an isologous interface. *Top*: After duplication of an isologous homodimer (gray), a substitution that occurs in one paralog (red box) appears twice in the interface of a homodimer (red circles), once in a heterodimer, and not at all in the other homodimer (blue). *Bottom*: One substitution (blue circle) in an isologous interface appears twice in a homodimer (*left*), twice in a heterotetramer (*middle*), and four times in a homotetramer (*right*), multiplying its effects on affinity. Dark and light gray, paralogous subunits. **(B)** *Top*: Interfaces in the human Hb heterotetramer (PDB 4HHB). Pink, Hbα; blue, Hbβ; α_1 and β_1 are in lighter hues than α_2 and β_2. IF1 surfaces (orange) mediate α_1-β_1 and α_2-β_2 interactions; yellow surfaces (IF2) mediate α_1-β_2 and α_2-β_1 interactions. Only interfaces involving α_1 are shown. Inset, α_1 subunit rotated away from the rest of the tetramer to show IF1 and IF2. *Bottom*: Isology of IF1 and IF2. Helices contributing to each interface are shown and labeled. Balls and sticks: on each helix, one residue's side chain is shown to visualize symmetry. **(C)** Evolution of tetrameric stoichiometry on the phylogeny of Hb and related globins. Icons, oligomeric states determined by experimental characterization of reconstructed ancestral proteins (16). Acquisition of interfaces of IF1 and IF2 is shown (16). **(D)** Key residues V37 and W40 that were substituted in Ancβ. Cyan cartoon helix, β_1 subunit. Pink and violet surfaces, α subunits that interact with β_1 via IF1 and IF2, respectively. Dotted lines to red or blue spheres, hydrogen bonds to oxygen or nitrogen atoms, respectively (PDB 4HHB). **(E)** Dimer-to-tetramer affinity of reconstructed ancestral Hb subunits containing historical substitutions q40W and t37V, measured by native mass spectrometry across a titration series. Points, fraction of dimers that are incorporated into tetramers. Lines, best-fit binding curves. Estimated Kd and SE are shown. **(F,G)** Effect of historical substitutions on stoichiometry, as measured by size exclusion chromatography. The ancestral dimer Ancαβ and the tetramers Ancα+Ancβ and human hemoglobin (HsHb) are shown for comparison. Protein concentration at 100 mM (E) or 1 mM (F). **(H)** Effect of historical substitutions on monomer-dimer affinity measured by native MS. **(I)** Effect on dimer-tetramer affinity of nonhistorical hydrophobic mutations in at residue 40, measured by native MS.

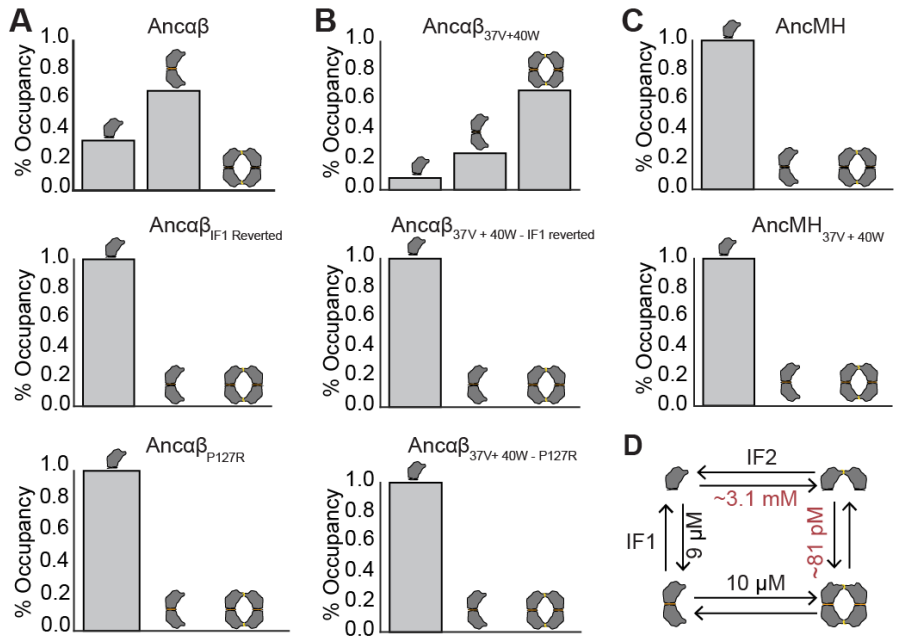


Figure 2. Multimerization across IF2 requires IF1. (A) IF1-mediated dimerization can be compromised by mutations. Relative occupancy of each stoichiometry as measured by native MS at at 20 mM total protein is shown for the ancestral dimer Ancαβ (top), Ancαβ$_{IF1\ reverted}$ (middle, a variant of Ancαβ in which all IF1 residues are reverted to the ancestral state found in AncMH), and Ancαβ-P127R (bottom, in which a mutation known to compromise IF1-mediated dimerization has been introduced). **(B)** Compromising IF1 prevents assembly across IF2. Relative occupancy of Ancαβ$_{40W + 37V}$ with and without mutations that compromise IF1-mediated dimerization. **(C)** AncMH, which does not dimerize across IF1, cannot multimerize across IF2, even when mutations sufficient to confer IF2-mediated mutimerization in Ancαβ are introduced. **(D)** Observed (black) and expected (red) affinities of Ancαβ +q40W interfaces. Expected Kd of a single iteration of IF2 (top) equals the square root of the measured apparent Kd when two iterations are present (bottom). Expected apparent Kd of two iterations of IF1 (right) equals the square of the measured Kd of a single IF1 (left).

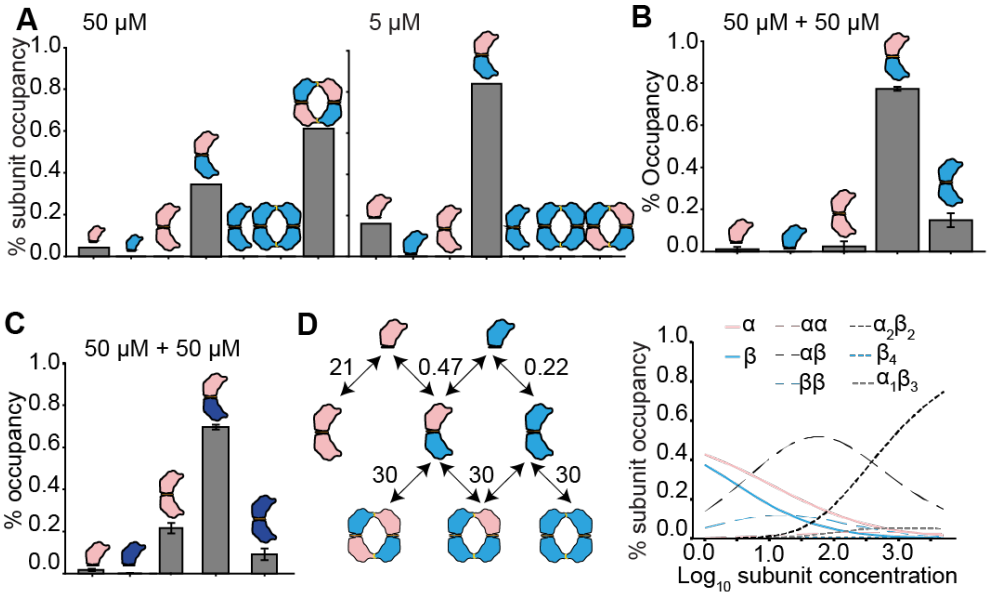


Figure 3. Heterotetramer specificity is conferred by specificity at IF1. (A) Occupancy (as fraction of all Hb subunits) when Ancα +Ancβ are coexpressed, measured by native MS. At 50 uM total protein, heterotetramers and heterodimers predominate (left). At 5 uM (right) – at which assembly occurs only across the high-affinity interface (IF1) -- all dimers are heterodimers. (**B)** Occupancy of subunits in stoichiometries as measured by nMS when Ancα and Ancβ are separately expressed and then mixed at 50 μM each; IF2-mediated tetramer assembly does not occur under these conditions, and dimers are predominantly heterodimers. Error bars represent standard error of measurement. **(C)** Percent occupancy of stoichiometries when Ancα and Ancβ' (Ancβ with all derived IF2 surface residues reverted to the state in Ancαβ) are expressed separately and then mixed at 50 uM. Error bars, SEM over three replicates. **(D)** Predicted occupancy of multimeric stoichiometries if IF1 is specific and IF2 is nonspecific. Left: binding scheme with experimentally estimated Kds (in μM) for IF1 and IF2-mediated multimerization by Ancα + Ancβ, assuming that all IF2 Kds are equal (for Kds, see Fig. 4D and 1D). Right: expected occupancies of each monomer, dimer, and tetramer, given the binding scheme at left. Occupancies are expressed as the fraction of all subunits in each species.

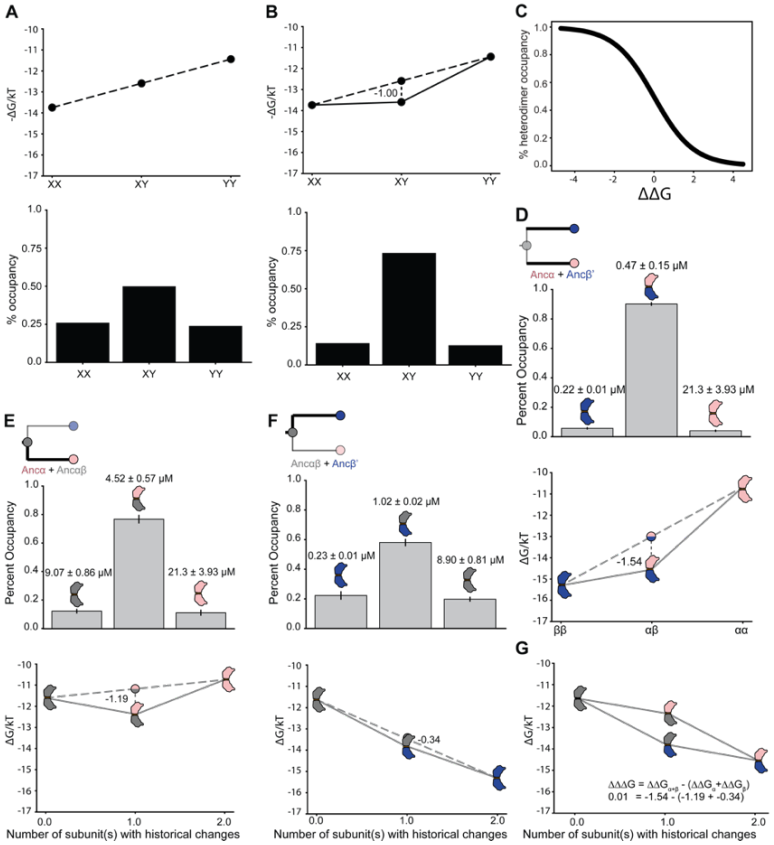
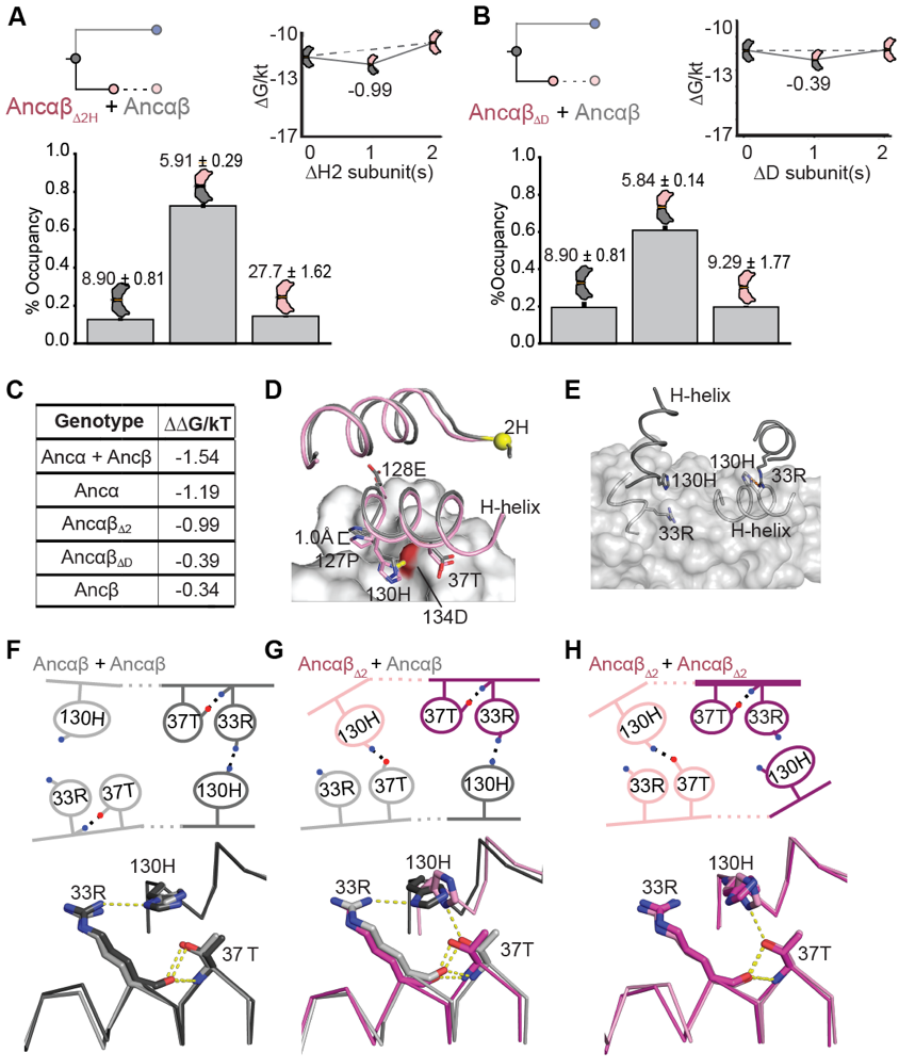


Figure 4. Contribution of historical changes in each subunit to the acquisition of heterospecificity. (A) Theoretical example of affinities and occupancy in a system of dimers with no specificity. *Top*: ΔG of dimerization for homodimers (XX and YY) and heterodimers (XY), in units of kT. In the absence of specificity, ΔG of the heterodimer equals the average of the homodimers (dotted line). *Bottom*: expected fractional occupancies of dimers at 1 mM per subunit and dissociation constants (Kd), given the ΔGs in the top panel. In the absence of specificity, heterodimer occupancy = 50%. **(B)** Example of a system with preference for the heterodimer. ΔΔG (the deviation of the heterodimer ΔG from the average of the homodimers) is shown. *Bottom*: Kd and predicted occupancy of each dimer at 1 mM. **(C)** Relationship between ΔΔG and heteromeric occupancy at 1 mM per subunit, assuming the ΔGs of homodimerization for as shown in panel A. **(D)** Specificity of IF1 dimerization in system of Ancα+Ancβ'. *Top*: expected fractional occupancies at 1 mM, given measured Kds by nMS (shown above each bar). *Bottom*: ΔGs and ΔΔG given measured Kds. **(E)** Specificity of IF1 acquired on the branch leading from Ancαβ to Ancα, shown as occupancy and ΔGs of the Ancαβ + Ancα system. **(F)** Specificity of IF1 acquired on the branch leading from Ancαβ to Ancβ, shown as occupancy and ΔGs of Ancαβ + Ancβ'. **(G)** Interaction effect on specificity when evolutionary changes leading from Ancαβ to Ancα (pink) and Ancβ' (blue) are combined. Homodimer of Ancαβ (gray) and each heterodimer are plotted by their ΔG. The observed ΔΔG of each heterodimer in combination Ancαβ is shown (see panels D-F). If the specificity acquired in the two subunits affects heterodimerization independently, then ΔΔG of Ancα+Ancβ will equal the sum of the ΔΔGs, yielding a parallelogram. The deviation from this expectation is shown.



Genotype	ΔΔG/kT
Ancα + Ancβ	-1.54
Ancα	-1.19
Ancαβ$_{Δ2}$	-0.99
Ancαβ$_{ΔD}$	-0.39
Ancβ	-0.34

Figure 5. Effect of historical sequence changes on specificity. (A) Specificity of Ancα$_{Δ2}$ with Ancαβ, as in Fig. 4. **(B)** Specificity of Ancα$_{ΔD}$ with Ancαβ. **(C)** Gain in specificity caused by various sets of historical mutations, relative to Ancαβ. Ancα+Ancβ, all changes on both post-duplication branches. Ancα, all changes on the branch leading to Ancα. ΔH2 and ΔD, deletions that occurred on the Ancα branches. **(D)** Models of Ancαβ homodimer and Ancαβ$_{Δ2}$ + Ancαβ heterodimer. The N-terminal helix and the portion of IF1 involving helix H is shown. Grey surface, Ancαβ subunit common to both models. Grey cartoon, other Ancαβ subunit in the homodimer; pink cartoon, Ancαβ$_{Δ2}$ subunit in the heterodimer. Yellow, 2H residue deleted in Ancαβ$_{ΔH2}$. Helix H side chains in the interface are shown as sticks. The hydrogen bond in the heterodimer from 130H to 37T (red surface) is shown (dotted line). **(E)** A portion of IF1 in the Ancαβ homodimer model, showing the isologous interactions with imperfect symmetry between 130H and 33R. Orange dashed-line, hydrogen bond. The two subunits are colored different shades of gray. The surface of the light-gray subunit is shown. **(F ,G, H)** Key residues in IF1 with hydrogen bonds that are affected by ΔH2 in the homodimers and heterodimer of Ancαβ and Ancαβ$_{ΔH2}$. Top, cartoon of key contacts. The two iterations of these interactions across the isologous interface are shown, one each in light or dark hue. Blue and red, nitrogen and oxygen atoms, respectively. Dotted lines, hydrogen bonds. The change in position of the H-helix caused by ΔH2 is shown. Bottom, structural alignment of the two iterations of the isologous interface in each dimer. Each dimer structure was duplicated exactly and then aligned to the original by targeting one subunit of the copy to align to the other subunit of the original. Hues correspond to the isologous iterations in the cartoon above.

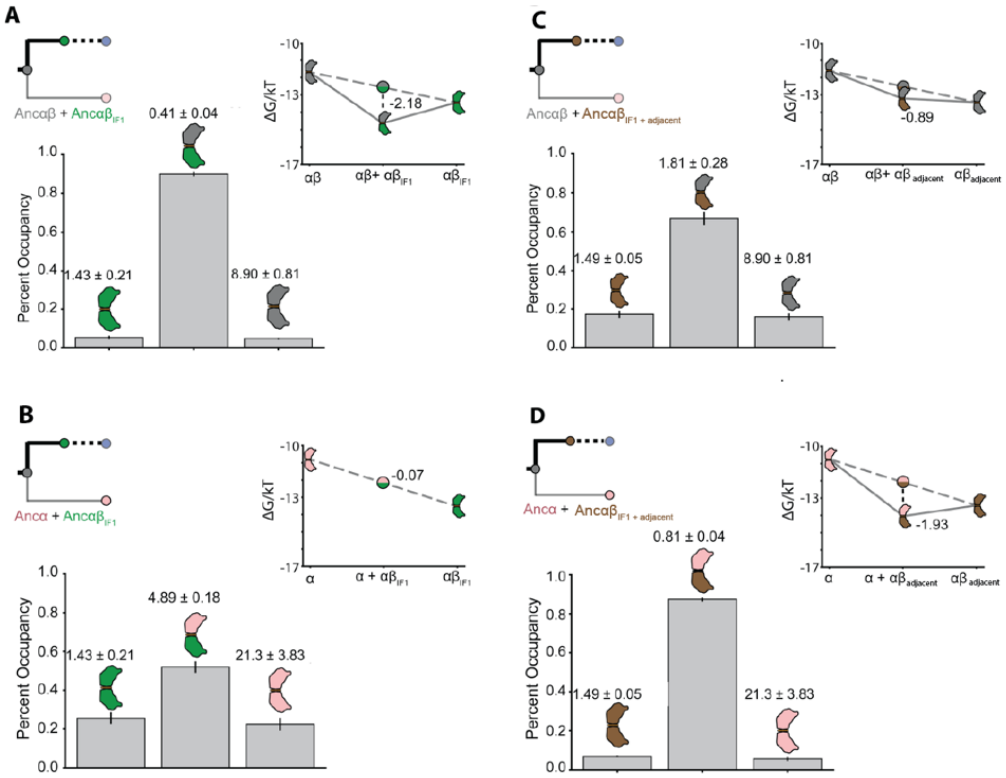


Figure 6. Other subsets of historical substitutions confer heterospecificity on IF1. Affinities measured by nMS, predicted occupancy based on those Kds at 1 mM each subunit, and $\Delta\Delta G_{spec}$ are shown for A) Ancαβ + Ancαβ$_{IF1}$, which contains the five substitutions at the IF1 surface that occurred in the Ancβ lineage; B) Ancαβ + Ancαβ$_{IF1 + adjacent}$, which also includes 4 additional substitutions in Ancβ near but not on the interface; C) Ancα + Ancαβ$_{IF1}$, and D) Ancα + Ancαβ$_{IF1 + adjacent}$.

SUPPLEMENTARY FIGURES

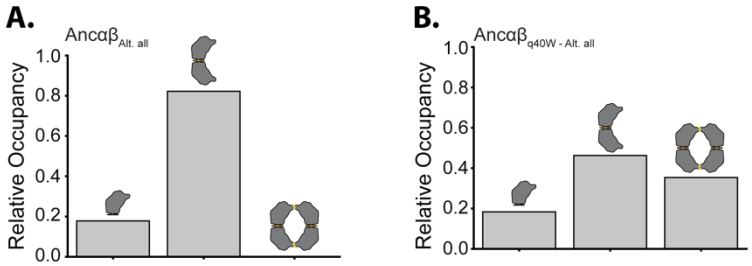
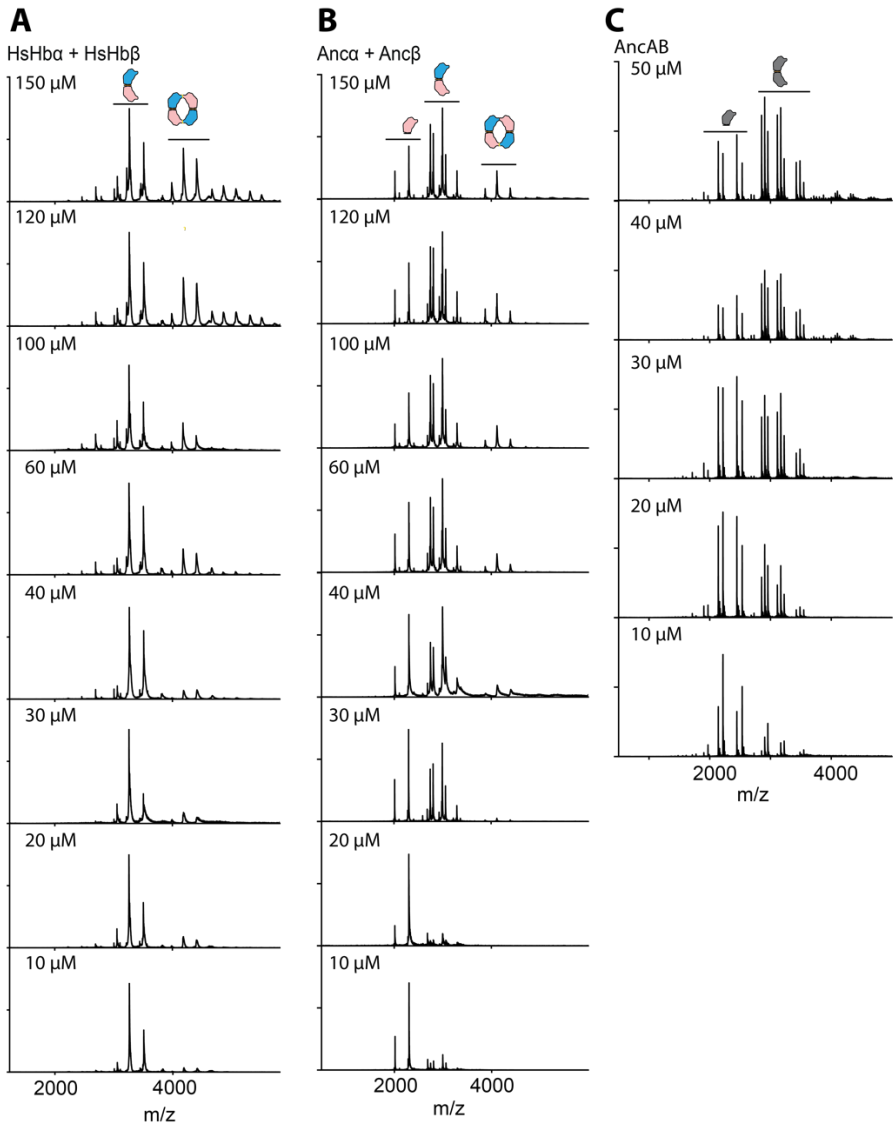


Fig. S1. Effect of q40W tetramerization is robust to statistical uncertainty. (A) Relative occupancy of monomer, dimer, and tetramer of Ancαβ$_{Alt. all}$, an alternative reconstruction of Ancαβ that contains the second most likely state at all ambiguously reconstructed sites, measured at 20 μM total protein using native MS. (B) Relative occupancy Ancαβ$_{Alt. all}$ with substitution q40W.

1103



1104

1105 **Fig. S2. Native mass spectrometry spectra.** nMS spectra across a concentration series is
1106 shown for A) human Hb, B) Anca + Ancb, and (C) Ancab. Peaks corresponding to monomers,
1107 dimers, and tetramer are labeled.

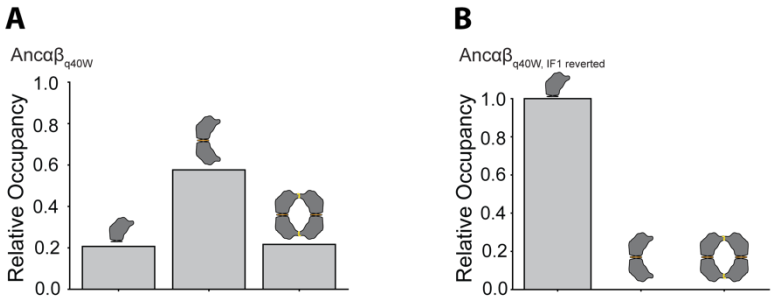


Fig. S3. The effect of q40W on tetramerization depends on IF1. (A) Relative occupancy of Ancαβ$_{q40W}$, measured by native MS at 20 μM total protein. (B) Relative occupancy of Ancαβ$_{q40W_IF1\text{-}reverted}$, which contains mutation q40W, as well as reversions to the ancestral state found in AncMH of all residues that were substituted between AncMH and Ancαβ .

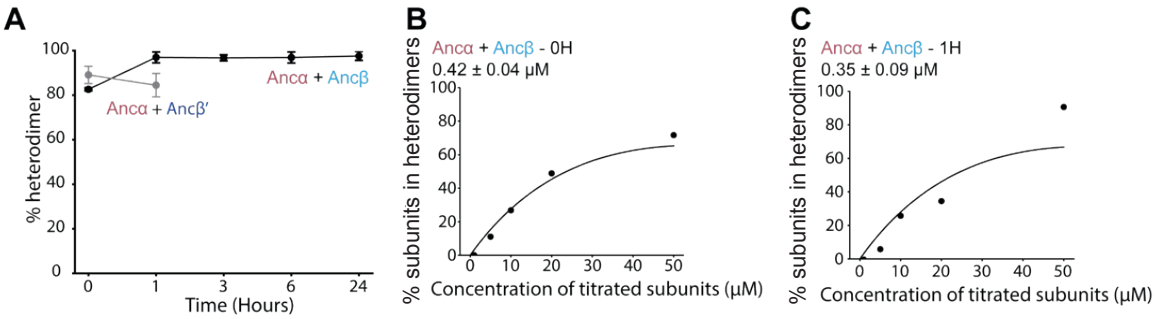


Fig. S4. Heterodimer occupancy of Ancα and Ancβ is near equilibrium after mixing. (A) The percent of all dimers that are heterodimers, measured by nMS when proteins are mixed at 50 μM each and allowed to incubate for 0, 1, 3, 6, or 24 hours. Black line and points, Ancα + Ancβ (which only dimerize when expressed separately and then mixed). Grey line and points, Ancα + Ancβ' (Ancβ in which IF2 surface substitutions are reverted to their ancestral state in Ancαβ, thus preventing tetramerization). Each dot shows the mean of three replicates; error bars, standard error of measurement. (B) Affinity of monomer-to-heterodimer assembly measured by nMS immediately upon mixing of Ancα and Ancβ. Ancα was kept constant at 50 μM, while the concentration of Ancβ varied. Points, fraction of all subunits in the mixture that are incorporated into heterodimers. Line, best-fit binding curve. Estimated Kd and 95% confidence interval are shown. (C) Estimated heterodimerization affinity measured as in panel B, but 1 hour after mixing.

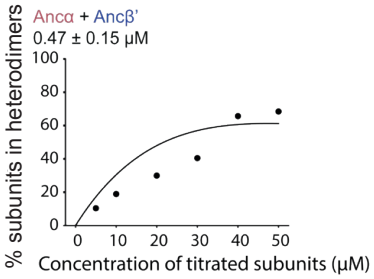


Fig. S5. Heterodimerization by Ancα+Ancβ'. Monomer-to-heterodimer assembly measured by nMS. Ancα was kept constant at 50 μM while Ancβ' was at variable concentration. Points, fraction of all subunits in the mixture that are incorporated into heterodimers at each concentration. Line, best-fit binding curve. Estimated Kd and 95% confidence interval are shown.

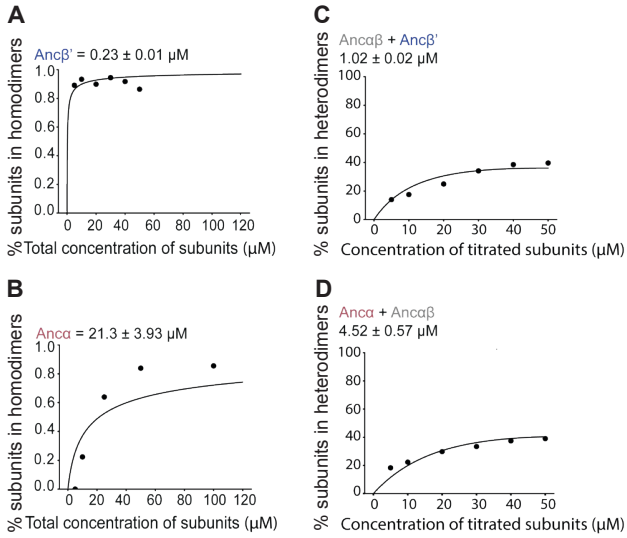


Fig. S6. Dimerization by Ancα and Ancβ' (A-B) Homodimerization by Ancβ' (panel A) and by Ancα (B). measured by nMS across a titration series. Each point shows the fraction of subunits incorporated into dimers as the concentration of protein varied. Best-fit binding curve, Kd, and 95% confidence interval are shown. (C-D) Heterodimerization by mixtures of Ancαβ+ Ancβ (C) and Ancαβ+Ancα and Ancα+Ancαβ (D). Each point shows the fraction of all subunits incorporated into heterodimers. In each case, one protein was held constant at 50 mM while the other was varied.

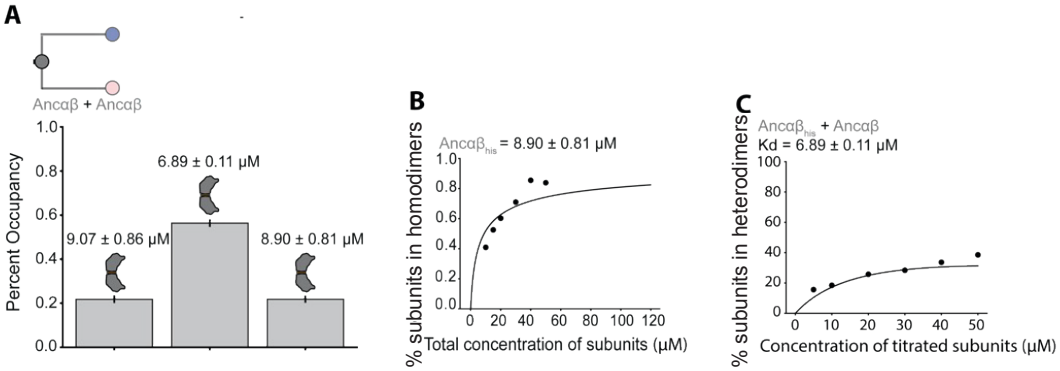


Fig. S7. Dimerization affinities and occupancy of Ancαβ. (A) Expected fractional occupancies of homodimer and heterodimers when Ancαβ is mixed at equal concentrations with Ancαβ$_{his}$ (500 mM each), given the measured dimerization affinities (shown above each column, with 95% confidence interval). Ancαβ$_{his}$ is Ancαβ with an N-terminal polyhistidine tag, which allows the masses of the three kinds of dimer to be distinguished. (B-C) Homodimerization by Ancαβ$_{his}$ and heterodimerization by affinity of Ancαβ+ Ancαβ$_{his}$, measured and represented as in Fig. S5.

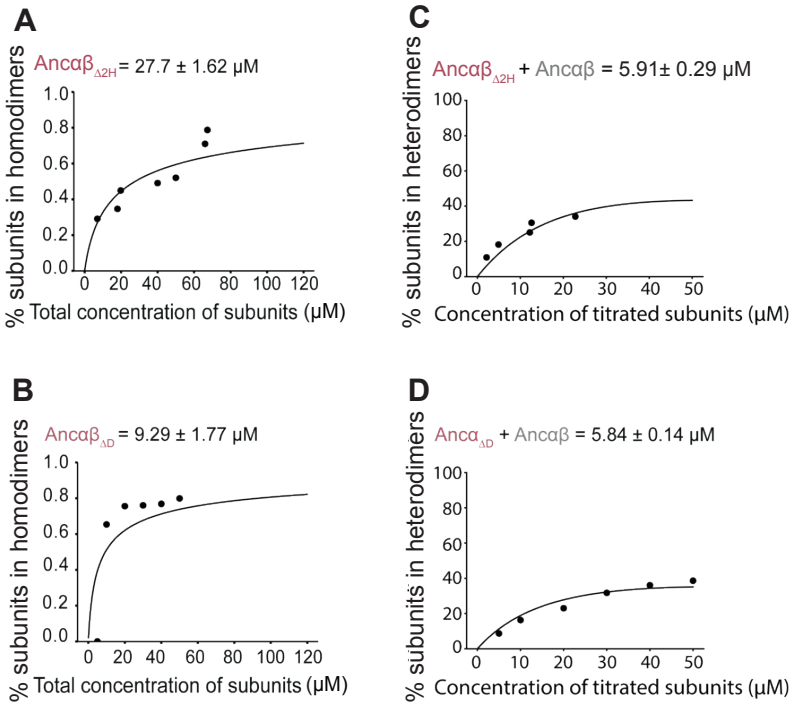


Fig. S8. Effect of historical deletions on dimerization. (A-B) Homodimerization and (C-D) Heterodimerization by mixtures, measured and represented as in Fig. S5.

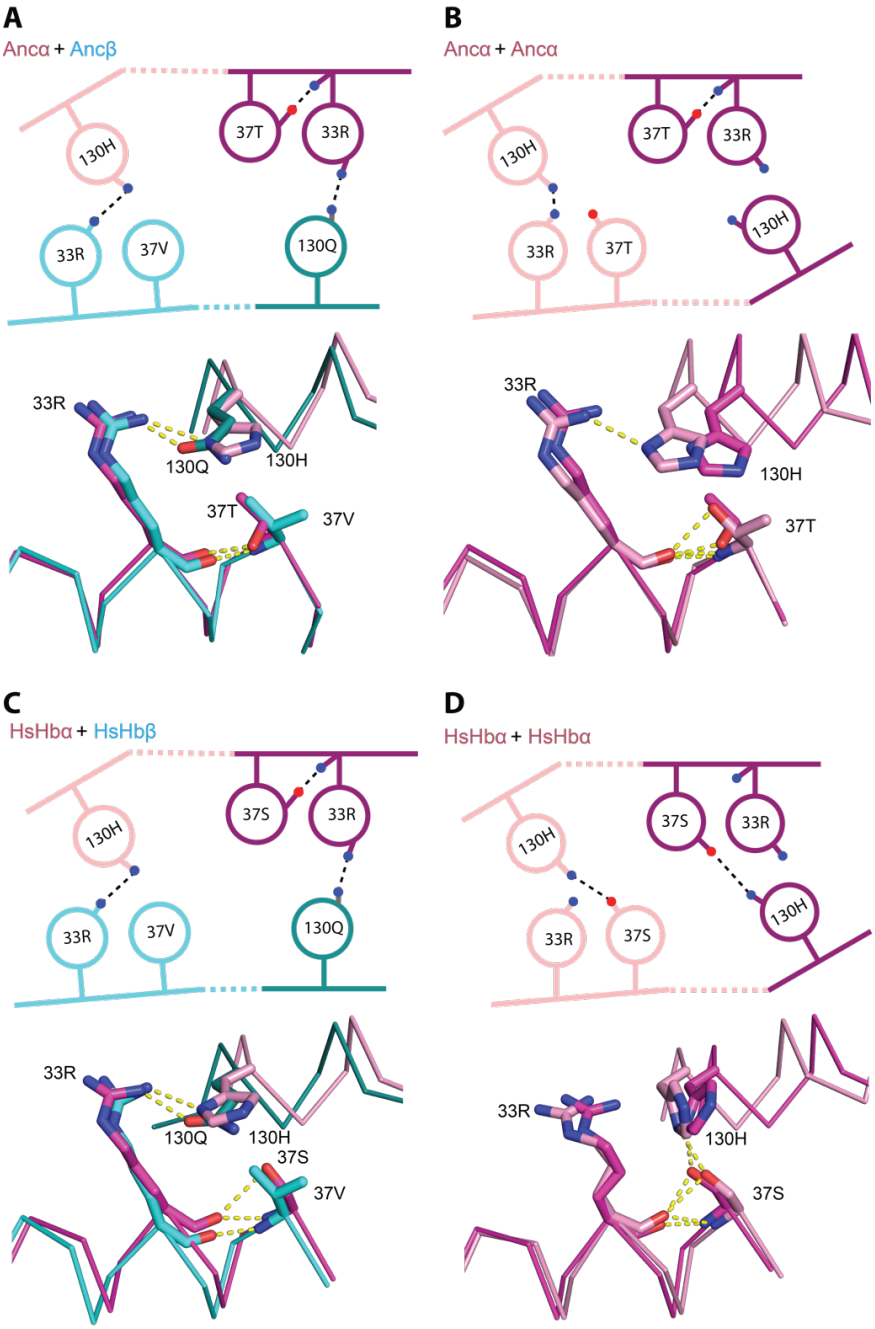


Fig. S9. Nonadditive interactions that contribute to specificity are conserved in derived Hb complexes. In the modeled homodimers and heterodimers of Ancα+Ancβ (panels A, B) and X-ray crystal structure of human hemoglobin (PDB 4HHB and 3S48), the figure shows the key IF1 residues with nonadditive interactions in Ancαβ+Ancαβ$_{\Delta H2}$ (see Fig. 5G for comparison). Top, cartoon of key contacts. The two iterations of these interactions across the isologous interface are shown, one each in light or dark hue. Blue and red, nitrogen and oxygen atoms, respectively. Dotted lines, hydrogen bonds. Bottom, structural alignment of the two iterations of the isologous interface in each dimer. Each dimer structure was duplicated exactly and then aligned to the original by targeting one subunit of the copy to align to the other subunit of the original. Hues correspond to the isologous iterations in the cartoon above

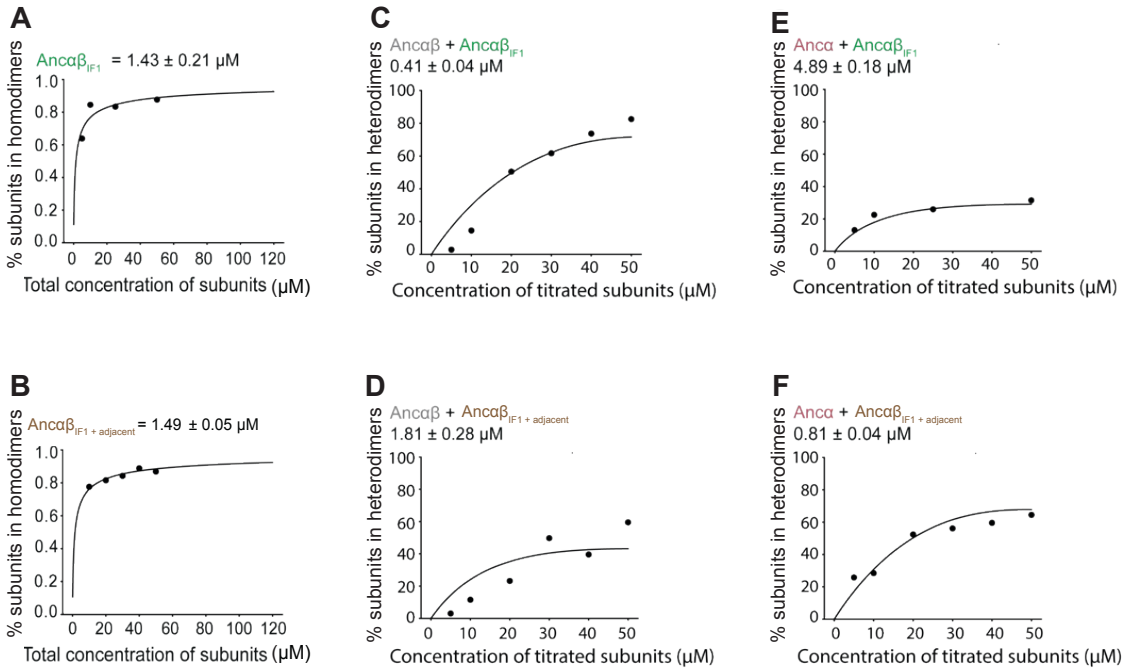


Fig. S10. Homodimerization by Ancαβ$_{IF1}$ and Ancαβ$_{IF1 + Adjacent}$ (A,B) and heterodimerization by those proteins when mixed with Ancαβ (C,D) or Ancα (E,F). Measurements and representation as in Fig. S5.

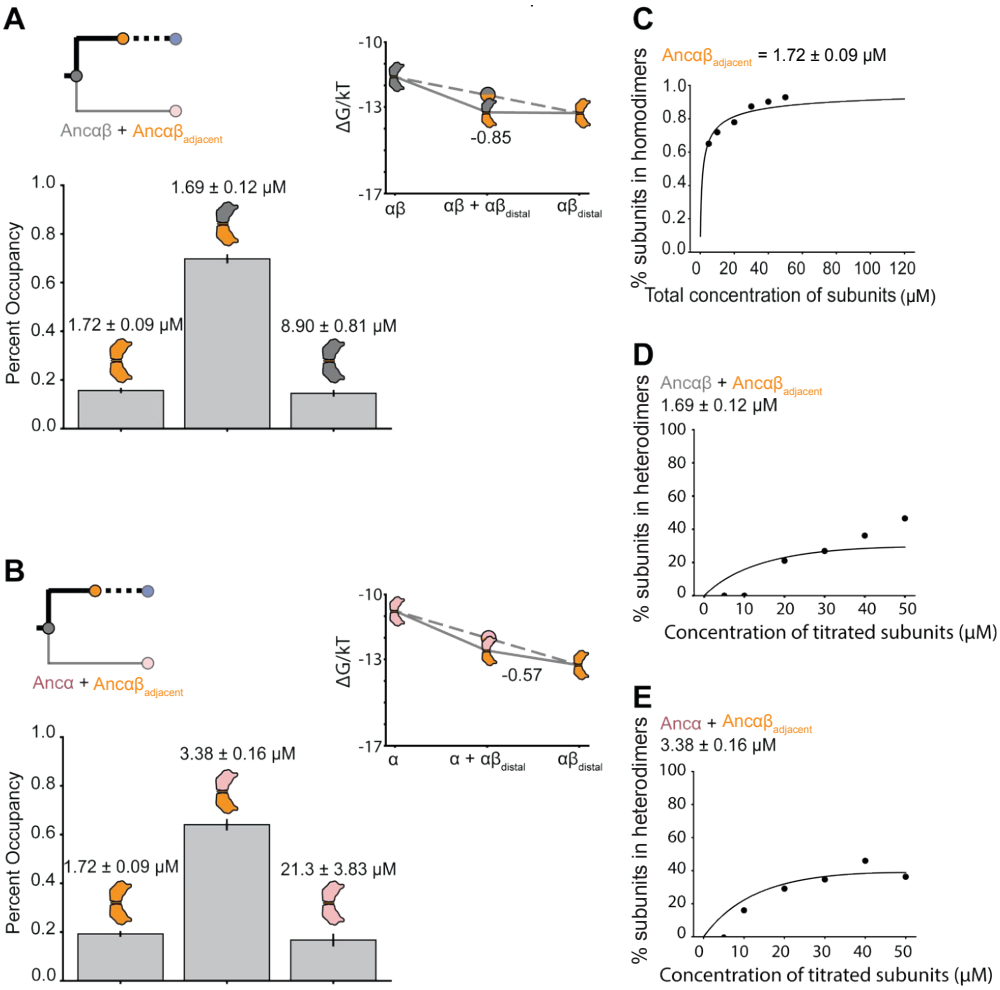


Fig. S11. Dimerization affinity and occupancies for Ancαβ$_{Adjacent}$. Expected fractional occupancies of homodimer and heterodimers when Ancαβ$_{Adjacent}$ Is mixed with Ancαβ (A) or Ancα (B), each at (500 mM), given the measured dimerization affinities (shown above each column, with 95% confidence interval). Inset, ΔG of each dimerization (measured in units of kT), with ΔG_{spec} of the heterodimer shown. (C,D,E) Measurement of binding affinities, measured and represented as in Fig. S5.