PNAS RESEARCH ARTICLE | EVOLUTION

Symmetry facilitated the evolution of heterospecificity and high-order stoichiometry in vertebrate hemoglobin

Carlos R. Cortez-Romero[a], Jixing Lyu[b], Arvind S. Pillai[c,d], Arthur Laganowsky[b], and Joseph W. Thornton[c,e,1]

Affiliations are included on p. 12.

Edited by Michael T. Laub, Massachusetts Institute of Technology, Cambridge, MA; received July 24, 2024; accepted December 4, 2024 by Editorial Board Member Harmit S. Malik

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), an $\alpha_2\beta_2$ heterotetramer that evolved from a homodimeric ancestor after a gene duplication. We show that the mechanisms for this evolutionary transition were 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, binding via the same surface patch on interacting subunits, but rotated 180° relative to each other. 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 and homomeric interactions. Our findings suggest that elaborate and specific symmetrical molecular complexes may often evolve via simple genetic and physical mechanisms.

protein evolution | molecular complexes | multimeric proteins | heterodimers | evolution of specificity

Significance

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. One amino acid change in each protein drove these changes in size and specificity. These 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.

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 (12). 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, 13). 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, 14). 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

Author contributions: C.R.C.-R., J.L., A.S.P., A.L., and J.W.T. designed research; C.R.C.-R., J.L., and A.S.P. performed research; C.R.C.-R., J.L., A.S.P., A.L., and J.W.T. analyzed data; and C.R.C.-R. and J.W.T. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission M.T.L. is a guest editor invited by the Editorial Board.

[1]To whom correspondence may be addressed. Email: joet1@uchicago.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2414756122/-/DCSupplemental.

Published January 23, 2025.

across the interface(s) of an isologous homodimer or heterotetramer, four times in a homotetramer, etc. (Fig. 1*A*). Mutations that weakly affect affinity on their own can therefore confer large effects on the assembly of isologous multimers (1, 5, 9, 15–17). 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 potential 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. 1*A*).

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 (18–23). But the genetic and structural mechanisms by which those interactions were acquired long ago are often different from their derived forms in the present (24). 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 (25, 26) and in multimers composed of unrelated proteins, which are by definition asymmetrical (27). 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 (28), 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 (ref. 17 and Fig. 1*B*), 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 to make the

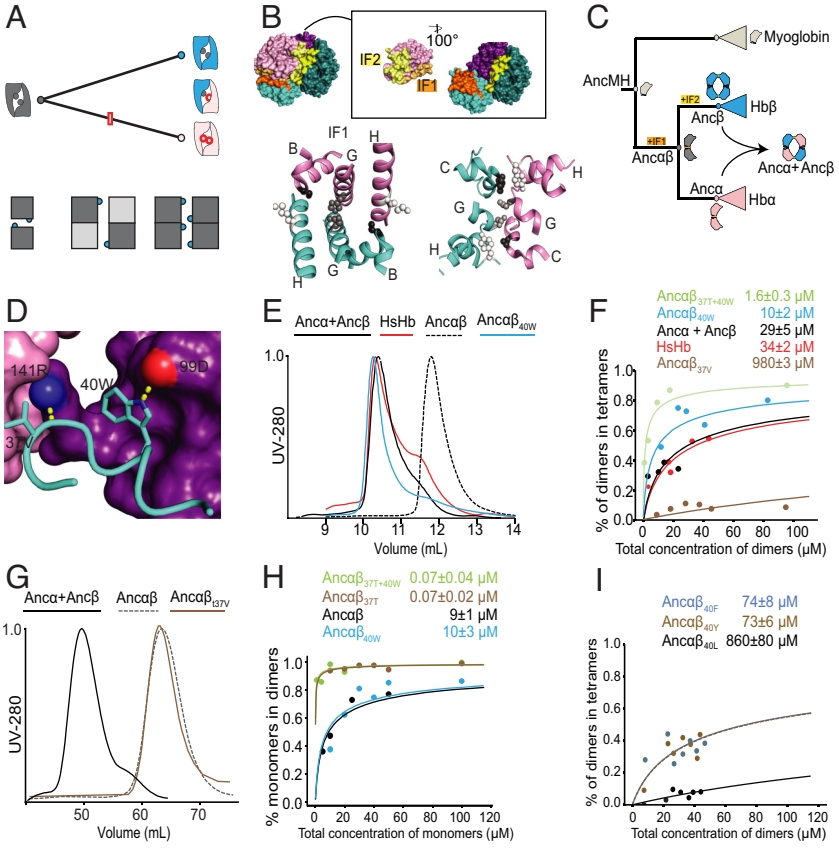


Fig. 1. A single substitution confers tetramerization on an ancestral dimer. (*A*) A substitution in one subunit can potentially affect specificity and stoichiometry in an isologous interface. *Top*: After duplication of an isologous homodimer (gray), a substitution that occurs in one paralog (red box) 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, the side chain of one residue 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 (17). Acquisition of interfaces of IF1 and IF2 is shown (17). (*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*) Effect of historical substitution q40W on stoichiometry, 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 μM. (*F*) 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 95% CI are shown. (*G*) Effect of substitution t37V on stoichiometry, measured by size exclusion stoichiometry. Protein concentration at 1 mM. (*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.

tetramer (ref. 29 and Fig. 1*B*). Hbα and Hbβ descend from a gene duplication deep in the vertebrate lineage (Fig. 1*C*), 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 (17). 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. 1*C*).

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 tetramerization 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? 4) 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 sequence changes that conferred tetramerization after duplication of the ancestral homodimer Ancαβ. We focused on the branch leading from the duplication of Ancαβ to Ancβ (the Hbβ subunit in the last common ancestor of jawed vertebrates), because Ancβ heterotetramerizes with Ancα (the Hbα subunit in the jawed vertebrate ancestor) and, like extant Hbβs, also homotetramerizes with itself. We previously found two amino acid replacements that occurred on the branch which, if introduced together into Ancαβ, are sufficient to confer high-affinity assembly into homotetramers (17). One of these (q40W) is buried in the IF2 interface, whereas the other (t37V) makes contacts across both IF1 and IF2 (Fig. 1*D*, using lower and upper case to denote ancestral and derived amino acids, respectively). W40 is strictly conserved in Hbβ subunits throughout the jawed vertebrates, and V37 is conserved in Hbβ of most taxa (*SI Appendix*, Fig. S1).

Here, we isolated the individual contributions of each amino acid changes by introducing them singly into Ancαβ and characterizing their effect on assembly into tetramers using size-exclusion chromatography (SEC) and native mass spectrometry (nMS) (30, 31). We found that 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 $\text{Ancαβ}_{\text{q40W}}$ is tetrameric, with occupancy of the tetramer similar to that observed in the derived Ancα + Ancβ complex and human Hb (Fig. 1*E*). We used nMS across a titration series to measure the affinity with which dimers associate into tetramers and found that the tetramerization affinity of $\text{Ancαβ}_{\text{q40W}}$ (Kd 10 μM) is stronger than that of Ancα + Ancβ (29 μM) and human Hb (34 μM) (Fig. 1*F* and *SI Appendix*, Fig. S2). The conclusion that q40W is sufficient to confer tetramerization is robust to statistical uncertainty about the ancestral sequence, because similar experiments using a different reconstruction of Ancαβ that incorporates alternative residues at all ambiguously reconstructed sites yield almost identical results (*SI Appendix*, Fig. S3).

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

In principle, a sequence change could also facilitate tetramerization by increasing affinity of the monomer-to-dimer transition; by increasing the effective concentration of dimers, more tetramers would be produced at a given protein concentration, 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. 1*H*). Substitution q40W, in contrast, has no effect on monomer-dimer affinity. These findings are consistent with the structural location of these residues—t37V contributes to both IF1 and IF2 and q40W to IF2 only—and they explain why t37V does not confer tetramerization on its own but enhances the impact of q40W.

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 is further strengthened by a hydrogen bond to 102D (32). 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 also confer tetramerization, albeit at affinity slightly weaker than q40W but similar to that of Ancα+Ancβ and human Hb. The greater affinity of tryptophan may be due to its longer side chain, which buries more hydrophobic surface area across the interface; the hydrogen bond with 102D could make a small contribution but is not necessary, because phenylalanine confers tetramerization but provides no hydrogen bond donor. Leucine, in contrast, which has a smaller volume and no hydrogen bonding capacity, confers no measurable tetramerization (Fig. 1*I*). High-affinity homotetramerization could therefore have evolved via any of three different aromatic replacements at site 40.

Taken together, these data indicate that replacing the amino acid at a single residue position was sufficient to confer tetramerization during historical Hb evolution, and several alternative replacements at the same site could also have caused the acquisition of this higher-order stoichiometry.

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. 1*A*). 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 20 μM (Fig. 2*A*). 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 ~1 μM) (*SI Appendix*, Fig. S4). Similar results are found when we used the combination of t37V/q40W to confer association across IF2 or the mutation P127R—which introduces unsatisfied positive charges into IF1—to compromise IF1 (Fig. 2*B*). 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 historically acquired before dimerization across IF1 evolved. We introduced t37V/q40W 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. 2*C*). Acquisition of multimerization across IF2 by q40W and by the pair t37V/q40W therefore depends on the prior evolution of dimerization via IF1.

These observations can be explained by a simple model in which the two symmetrical interfaces contribute independently to the energy of binding. A single iteration of IF2 is too weak to confer measurable binding of two monomers into a dimer; however, IF1 is stronger and mediates assembly of dimers. Each IF1-mediated dimer presents two iterations of the IF2 surface patch, doubling the total energy of IF2-mediated assembly of dimers into tetramers. Because of the exponential relationship between energy and occupancy, a weak IF2 can therefore confer high-affinity binding but if IF1 is already present. The affinities that we measured are consistent with this simple model. If the energy of dimer-to-tetramer assembly is twice that of monomer-dimer binding using the same interface, then the Kd of IF2-mediated tetramerization should be the square of the Kd of IF2-mediated dimerization (Fig. 2*D*). 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 ~1 mM. 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 (*Materials and Methods*). This simple additive model therefore explains most—and possibly all—of the difference in affinity conferred when IF2 is doubled in the symmetrical tetramer. The dependence of assembly across IF2 upon the presence of IF1 does not imply any direct physical interaction between the interfaces or any conformational change in one interface caused by binding at the other. 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.

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 form, a single IF2 would have been too weak to mediate multimerization. The fact that q40W's effect depends 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 (17). 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 occurred, 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. Specifically, when all historical substitutions that occurred at the IF2 surface during the postduplication interval are introduced into Ancαβ and this protein is coexpressed

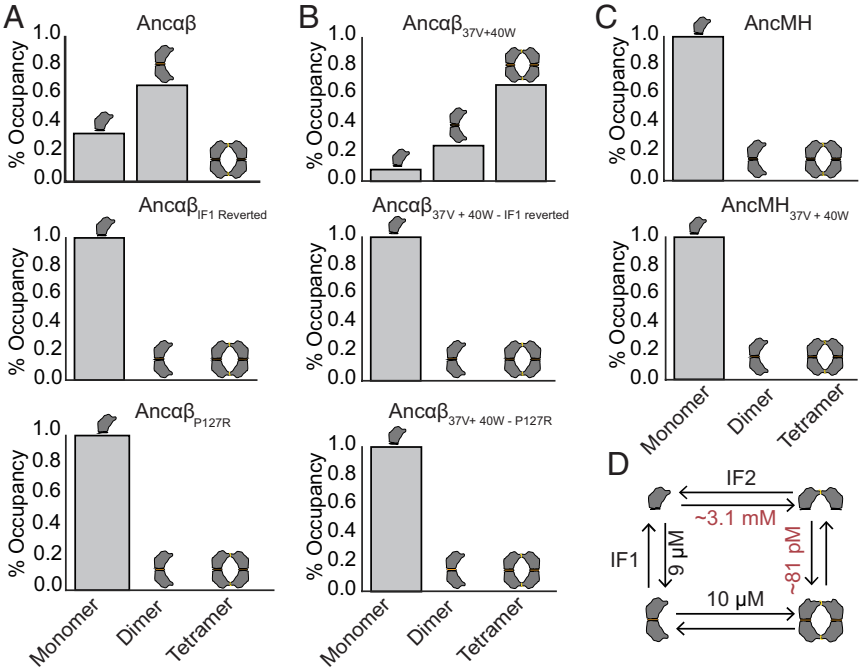


Fig. 2. Multimerization across IF2 requires IF1. (*A*) IF1-mediated dimerization can be compromised by mutations. Relative occupancy of each stoichiometry as measured by native MS at 20 μM total protein is shown for the ancestral dimer Ancαβ (*Top*), Ancαβ$_{IF1\ reverted}$ (*Middle*, a variant of Ancαβ 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 multimerization 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*).

with Ancα, an indiscriminate mixture of homotetramers, $\alpha_1\beta_3$ heterotetramers, and $\alpha_2\beta_2$ heterotetramers is produced (17). 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 conditions that prevent binding across IF2. 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. 3*A*). IF2 does not mediate assembly of monomers into dimers in the absence of IF1 (Fig. 2 *A* and *B*), so these heterodimers must be IF1-mediated, indicating that IF1 is heterospecific (Fig. 3*A*). Second, we expressed Ancα and Ancβ separately and mixed them at equal and moderate concentration; because tetramerization requires cofolding, only IF1 dimers form (33), and these are predominantly heterodimers (Fig. 3*B* and *SI Appendix*, Fig. S5). 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. 3*C* and *SI Appendix*, Fig. S5). Together, these data indicate that the derived IF1 is specific, preferentially mediating assembly into heterodimers.

To test the second prediction—that 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 = 0.6 μM) is slightly worse than its homodimerization affinity (0.2 μM), but both are far better than the Ancα homodimer (24 μM) (Fig. 3*D* and *SI Appendix*, Figs. S6–S9). We then predicted occupancy of each stoichiometry as the concentration of Hb subunits changes, given these affinities at IF1 and assuming that IF2 has a dimer-to-tetramer affinity of 30 μM, as measured in Ancα + Ancβ, with no preference for homomeric or heteromeric binding (see Fig. 1 *D* and *F*). At low concentrations, the system produces almost exclusively IF1-mediated heterodimers. The predominance of heterodimers is attributable to Ancα's weak homodimerization affinity; 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. 3*D*). As protein concentration increases, these dimers begin to assemble with each other across IF2 into tetramers. The excess of heterodimers over homodimers means that the vast majority of the tetramers are heterotetramers, even though IF2 itself does not distinguish between subunit types. At physiologically relevant concentrations of 3mM total Hb subunits (34), the population is dominated by $\alpha_2\beta_2$ heterotetramers, with a small fraction of heterodimers and virtually no homotetramers (Fig. 3 *D*, *Right*).

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 characterized

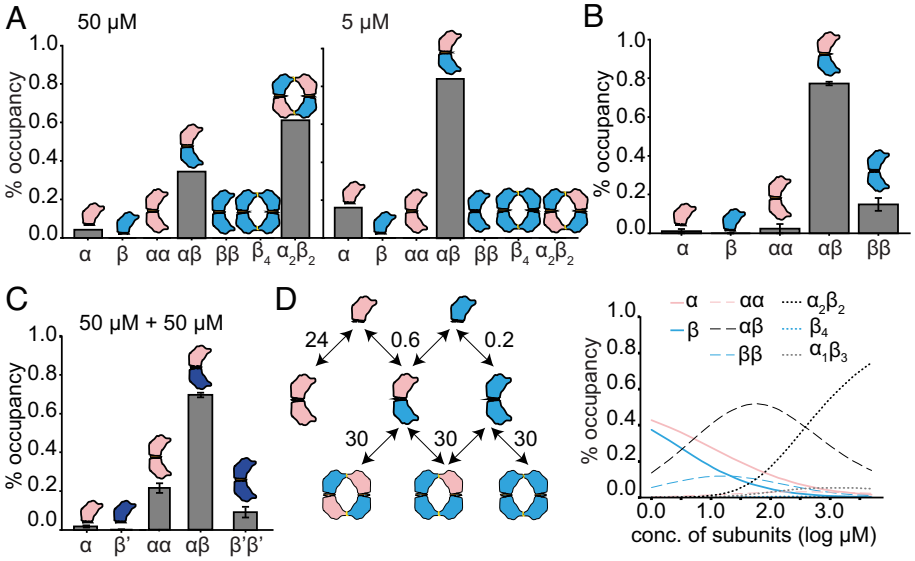


Fig. 3. Heterotetramer specificity is conferred by specificity at IF1. (*A*) Occupancy (as fraction of all Hb subunits) when Ancα +Ancβ are coexpressed, measured by native MS. At 50 μM total protein, heterotetramers and heterodimers predominate (*Left*). At 5 μM (*Right*)—at which assembly occurs only 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 SEM for three replicates. (*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 μM. (*D*) Predicted occupancy of multimeric stoichiometries if IF1 is specific and IF2 is nonspecific. *Left*: binding scheme with experimentally estimated Kds (in mM) for IF1 and IF2-mediated multimerization by Ancα + Ancβ, assuming that all IF2 Kds are equal (for Kds, seeFigs. 1*D* and 4*D*). *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.

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 ($\Delta\Delta G_{spec}$; see *Materials and Methods* for calculation). 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; this is true even if the homodimer ΔGs are very different from each other, as long as the heterodimer ΔG is halfway between them. By contrast, if $\Delta\Delta G_{spec} < 0$, then heterodimers will account for the majority of dimers; conversely, if $\Delta\Delta G_{spec} > 0$, homodimers together will predominate (Fig. 4 *A*–*C*). Hetero- or homospecificity thus arises when two paralogs contribute nonadditively to dimerization. Whether or not the system is hetero- or homospecific, the occupancies of the two homodimers at saturating concentrations will be identical. Irrespective of the fraction of subunits that assemble into heterodimers, the remaining subunits will comprise equal concentrations of the two types; if these concentrations are far above the homodimerization Kds, then virtually all of the remaining subunits will assemble into homodimers, which will therefore have equal occupancies (28).

We quantified the heterospecificity of Ancα and Ancβ at IF1 by estimating $\Delta\Delta G_{spec}$. We used nMS to measure the homodimer and heterodimer affinities of Ancα and Ancβ', which contains all substitutions that occurred along the Ancβ branch except those that mediate tetramerization across IF2, allowing us to prevent tetramerization and thus isolate the specificity effects at IF1. From these affinities, we calculated the ΔG of binding and the expected fractional occupancy of each dimer at high and equal concentration of subunits. For Ancα+Ancβ', we found that $\Delta\Delta G_{spec}$ = –1.3 (in units of kT) and heterodimer occupancy of 82% (Fig. 4*D*). This represents the total specificity acquired by the two diverging paralogs after the duplication 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 of homodimerization became worse (ΔΔG = 0.9) while homodimerization by Ancβ' improved substantially (ΔΔG = –3.7). The heterodimer affinity improved by ΔΔG = –2.7, substantially more than the average of the two homodimers, yielding the observed strong preference for the heterodimer.

We next isolated the contribution 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 deeper ancestor Ancαβ. This pair of proteins is heterospecific, with $\Delta\Delta G_{spec}$ = –1.2 (expected heterodimer occupancy 76%). Changes in the α subunit alone therefore account for >90% of the total $\Delta\Delta G_{spec}$ that was acquired by the entire Ancα + Ancβ' system. This specificity is acquired via a 2.6-fold reduction in Ancα's homodimerization affinity compared to the Ancαβ ancestor and a 1.8-fold improvement in heterodimer affinity (Fig. 4*E* and *SI Appendix*, Fig. S7 *B* and *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 barely heterospecific, with $\Delta\Delta G_{spec}$ = –0.3 and expected heterodimer occupancy of just 58%. The specificity is weak because both the homodimer and heterodimer affinity improved, but the deviation of the heterodimer from the average of the homodimers is very small (Fig. 4*F* and *SI Appendix*, Fig. S7 *A* and *C*).

Finally, we assessed whether the evolutionary changes in the Hbα subunit and those in the Hbβ 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 (–1.2 + –0.3 = –1.5). The observed $\Delta\Delta G_{spec}$ = –1.3, suggesting a very weak negative interaction between changes in the two subunits, which makes the complex slightly less heterospecific than expected if the substitutions were independent (Fig. 4*G*).

Taken together, these data indicate that the IF1 specificity acquired by the derived complex Ancα + Ancβ is primarily attributable to substitutions in the α subunit; substitutions in the β subunit made a much smaller contribution, and nonadditive interactions between the two sets of changes did not contribute to the evolution of heterospecificity. 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. 4*G*).

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 3 (ΔH3), a five-residue deletion in helix D (ΔD), and an amino acid replacement (v140A). ΔH3 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 conserved to the present in Hbα subunits throughout the jawed vertebrates, including humans, whereas the amino acid at site 140 varies (*SI Appendix*, Fig. S1). 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 ΔH3 alone confers substantial specificity, recapitulating >80% of the Ancα's acquired heterospecificity for Ancαβ ($\Delta\Delta G_{spec}$ = –1.0 out of a total $\Delta\Delta G_{spec}$ = –1.2 acquired along this branch) and >75% of the specificity acquired along both branches by the entire Ancα + Ancβ' complex ($\Delta\Delta G_{spec}$ = –1.3, Fig. 5 *A* and *C*). ΔH3 enhances specificity by improving heterodimer affinity and reducing homodimer affinity, with both Kds very close to those of Ancα (Fig. 5*A* and *SI Appendix*, Fig. S10 *A* and *C*).

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 ($\Delta\Delta G_{spec}$ = –0.4, Fig. 5*B* and *SI Appendix*, Fig. S10 *B* and *D*). When the contributions of ΔH3 and ΔD to specificity are added together, they slightly exceed the specificity of Ancα (–1.4 rather than –1.3), suggesting the possibility of a very weak negative epistatic interaction between them or a small countervailing effect of the third change v104A. Taken together, these results indicate that ΔH3 was a large-effect historical sequence change that caused 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 ΔH3 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). In principle, two kinds of mechanisms could cause these opposite effects. Either 1) the mutated residue could interact 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 could be imperfect, such that introducing the mutation on one side of the interface is favorable but introducing

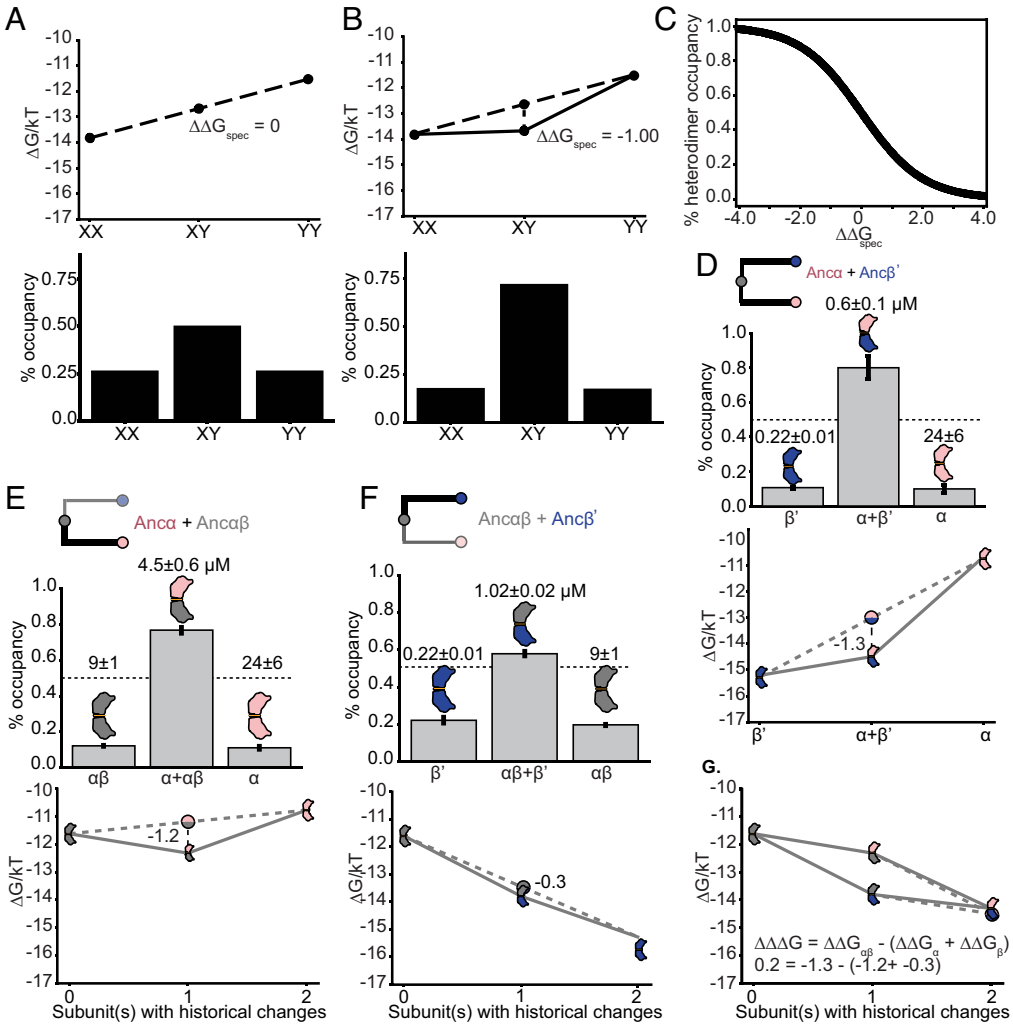
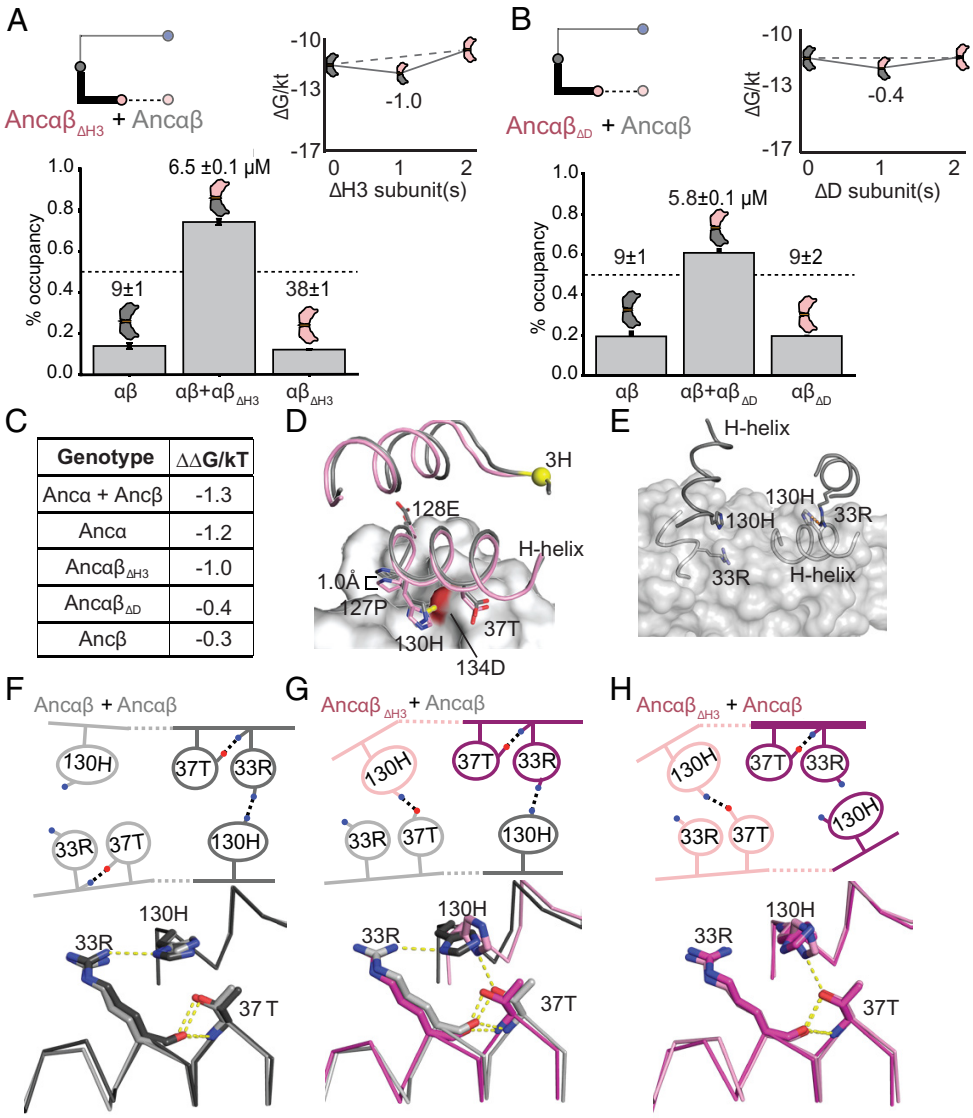


Fig. 4. Contribution of historical changes in each subunit to the acquisition of heterospecificity. (*A*) Theoretical example of affinities and occupancy in a system of dimers with no specificity. *Top*: ΔG of dimerization for homodimers (XX and YY) and heterodimers (XY), in units of kT. In the absence 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 Kds assessed by nMS (shown above each bar, with 95% CI). *Bottom*: ΔGs and ΔΔG given measured Kds. Dotted line, heterodimer occupancy in the absence of specificity. (*E*) Specificity of IF1 acquired between Ancαβ and Ancα (thick branch), shown as expected occupancies and ΔGs of the Ancαβ + Ancα system. The number of subunits that contain historical changes in each dimer is shown relative to the Ancαβ homodimer. (*F*) Specificity of IF1 acquired on the branch leading from Ancαβ to Ancβ', shown as expected occupancies and ΔGs of Ancαβ + Ancβ'. (*G*) Interaction effect on specificity when evolutionary changes leading from Ancαβ to Ancα (pink) and to 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 interaction effect is the deviation from this expectation (ΔΔΔG).

it again onto the other side is net-unfavorable. The first scenario does not pertain in this case. Residue H3 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. 5*D*). Asymmetry in the interface is therefore the likely cause of ΔH3's differential effects on heterodimer vs. homodimer affinity.

To gain insight into the possible nature of this asymmetry and potential mechanisms by which ΔH3 affects specificity, we modeled the structures of the Ancαβ homodimer, the Ancαβ$_{\Delta H3}$ homodimer, and the heterodimer of these two proteins. The modeled Ancαβ homodimer contains 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. 5 *E* and *F*). 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 (Fig. 5 *D* and *G*). 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 ΔH3 improves heterodimer affinity (Fig. 5 *D* and *G*).

The modeled Ancαβ$_{\Delta H3}$ homodimer structure is notably asymmetric and suggests a possible mechanism by which introducing ΔH3 into both subunits reduces affinity (Fig. 5*H*). One side



Genotype	ΔΔG/kT
Ancα + Ancβ	-1.3
Ancα	-1.2
Ancaβ$_{ΔH3}$	-1.0
Ancaβ$_{ΔD}$	-0.4
Ancβ	-0.3

Fig. 5. Effect of historical sequence changes on specificity. (*A*) Specificity of Ancaβ$_{ΔH3}$ with Ancαβ, represented as in Fig. 4. The branch on which ΔH3 occurred is shown as a thick line, other changes on that branch as a dotted line. (*B*) Specificity of Ancaβ$_{ΔD}$ with Ancαβ. (*C*) Gain in specificity caused by various sets of historical mutations, relative to Ancαβ. Ancα+Ancβ, all changes on both postduplication branches. Ancα, all changes on the branch leading to Ancα. ΔH3 and ΔD, deletions that occurred on the Ancα branches. (*D*) Models of Ancαβ homodimer and Ancaβ$_{ΔH3}$ + Ancαβ heterodimer. The N-terminal helix and the portion of IF1 involving helix H is shown. Gray surface, Ancαβ subunit common to both models. Gray cartoon, other Ancαβ subunit in the homodimer; pink cartoon, Ancaβ$_{ΔH3}$ subunit in the heterodimer. Yellow, 3H residue deleted in Ancaβ$_{ΔH3}$. 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–H*) Key residues in IF1 with hydrogen bonds that are affected by ΔH3 in the homodimers and heterodimer of Ancαβ and Ancaβ$_{ΔH3}$. *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 ΔH3 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.

displays the favorable new cross-interface interactions caused by ΔH3 in the heterodimer, including the 130H-37T hydrogen bond. On the other side, however, the effect of the deletion is very different: ΔH3 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 Ancaβ$_{ΔH3}$ 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.

These hypothesized mechanisms appear to have persisted over time. The same pattern of interactions is found in the modeled structures of the hetero- and homodimers of Ancα + Ancβ (*SI Appendix*, Fig. S11).

High-resolution crystal structures of extant hemoglobin also show notable asymmetries in the multimerization interfaces, which exceed the deviation expected given the resolution of the structures (35). These structures include some of the particular asymmetrical interactions observed in our ancestral models. In the human Hb heterotetramer, 33R hydrogen bonds across IF1 to residue 130, but this interaction is again lacking in the homodimer of human Hbα, leaving 33R unsatisfied, potentially explaining the weak homomeric affinity of Hbα (*SI Appendix*, Fig. S11). 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. Structural models are prone to error, and the asymmetries we observed are subtle; further research will be required to definitively characterize potential asymmetries in the ancestral multimers.

Multiple Sets of Historical 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 several 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 occurred at the IF1 surface. We introduced these changes into Ancαβ (creating protein Ancαβ$_{IF1}$) and measured affinity and specificity when this protein is mixed with Ancαβ. These substitutions yield a highly heterospecific complex ($\Delta\Delta G_{spec}$ = –2.2, heterodimer occupancy 90%, Fig. 6*A* and *SI Appendix*, Fig. S12 *A, C,* and *E*). Unlike the Ancα substitutions, the Ancβ$_{IF1}$ substitutions confer heterospecificity by improving both homodimer and heterodimer affinity, but they improve the latter by more than the former.

Because Ancαβ$_{IF1}$ is specific in complex with Ancαβ, we wondered whether it would also be specific with Ancα. We found that this complex is barely heterospecific ($\Delta\Delta G_{spec}$ = –0.1, Fig. 6*B*), implying that other substitutions on the branch leading to Ancβ but not on the interface must have contributed to the evolution of specificity between Ancαβ$_{IF1}$ 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. 17). This protein (Ancαβ$_{IF1+Adjacent}$) has strong heterospecificity when mixed with Ancα ($\Delta\Delta G_{spec}$ = –2.0, heterodimer occupancy >85%, Fig. 6*D* and *SI Appendix*, Fig. S12 *B, D,* and *F*), because these mutations together improve both heterodimer and homodimer affinity, but with a larger improvement in the heterodimer. It is also moderately heterospecific when mixed with Ancαβ ($\Delta\Delta G_{spec}$ = –0.91, Fig. 6*C*).

Finally, we tested the effect of the adjacent substitutions on their own and found that they confer specificity when mixed with Ancαβ ($\Delta\Delta G_{spec}$ = –0.9). These mutations impart specificity by causing almost identical changes in homo- and heterodimer affinity. They also confer some heterospecificity when Ancαβ$_{Adjacent}$ is mixed with Ancα ($\Delta\Delta G_{spec}$ = –0.6, *SI Appendix*, Fig. S13 *A–D*).

There are therefore several distinct sets of substitutions that occurred during history, any of which would be sufficient to confer heterospecificity on their own (and in various combinations), and which 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 (summarized in *SI Appendix*, Fig. S1). Each aspect of this 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. These two key sequence changes have remained conserved in the descendant Hbα and Hbβ subunits of all extant jawed vertebrates (*SI Appendix*, Fig. S1). These transitions were both facilitated by the isologous architecture of Hb's two interfaces, which creates a propensity for mutation to produce high-order, heterospecific complexes and heterospecificity. Many of the mechanisms involved in Hb evolution are likely to be generally relevant and may explain the evolution of several large-scale patterns in the diversity of molecular complexes.

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), the impact of which was multiplied by the isologous

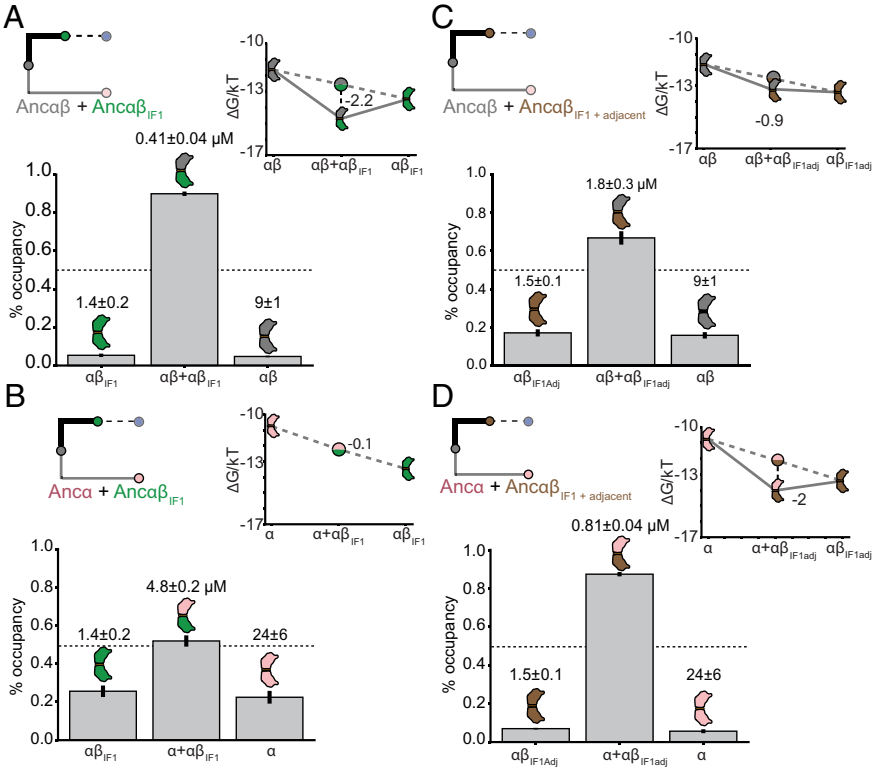


Fig. 6. Other subsets of historical substitutions can confer heterospecificity on IF1. Affinities measured by nMS, predicted occupancy based on those Kds at 1 mM each subunit, and $\Delta\Delta G_{spec}$ are shown for (*A*) Ancαβ + Ancαβ$_{IF1}$, which contains the five substitutions at the IF1 surface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

architecture of the hemoglobin complex. It is plausible that similar mechanisms may have driven stoichiometric increases during the evolution of other isologous complexes. 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 (36), and mutations to bulky hydrophobic amino acids can drive assembly into high-order multimers (9, 37–40).

The majority of complexes assemble through isologous interfaces (41). It has been suggested that this must imply that isology confers some selective benefit by improving protein function (1). Our results point to 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 nonisologous—with q40W interacting with a hydrophobic divot on some other surface of the facing subunit—then this favorable interaction would appear only once, which would be insufficient to confer meaningful tetramer occupancy. By this explanation, isologous complexes are abundant because they are easier to produce by mutation than head-to-tail multimers, not more likely to be fixed by selection.

In high-order multimers, the interface with higher affinity usually evolves before the lower-affinity interface(s) (42–45). Hb evolution displays this pattern, with the stronger interface IF1 evolving before IF2 (17). It has been suggested that this pattern must imply that selection favors evolutionary intermediates that contain the high-affinity interface, perhaps because those containing only the low-affinity interface assemble slowly and/or misassemble into anomalous complexes (42, 43). Our work here suggests a different explanation: It is easier for mutations to generate a new interface that confers a higher stoichiometry if a strong interface is already present, because the affinity of the new interface is multiplied by iteration in an isologous complex. By contrast, low-affinity interfaces do not confer multimerization on their own, so if the low-affinity interface were to evolve first, then the effects of mutations on the second interface would not be multiplied. In complexes with multiple interfaces, the stronger interface tends to be older not because such trajectories improve fitness but because mutation is more likely to build elaborate complexes in this historical order.

One Interface Conferred Specificity on a High-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.

Symmetry Allowed Specificity to Evolve via Changes 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, 13, 19, 22, 23, 25–27).

This difference in historical genetic mechanism reflects the opportunities presented by the two different types of multimeric architecture. In nonisologous 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 requires 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 can affect interactions differently when it appears twice in the homodimer versus when it occurs once in the heterodimer. Imperfect asymmetry could facilitate the evolution of specificity in many complexes. Virtually all isologous interfaces contain subtle asymmetries (46). This imperfection arises for two reasons. First, perfect symmetry is entropically unfavorable. Second, 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 (46, 47). Extant human hemoglobin is one of many examples of isologous interfaces in which asymmetry is imperfect (35, 48). Isologous interfaces therefore provide a 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 His3 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' mediates 80% heterodimer occupancy at equal and saturating concentrations, but its specificity is quite moderate ($\Delta\Delta G_{spec} = -1.3$). This difference in binding energy is less than that associated with a typical hydrogen bond or burial of a large hydrophobic residue. The structural differences in physical interactions across the homodimer vs. heterodimer

interfaces in our modeled structures could easily yield energetic differences of this magnitude, although the particular form of asymmetry in ancestral hemoglobin complexes remains uncertain. Recent in silico work also found that small differences in ΔG can cause large differences in occupancy between homodimers and heterodimers (28).

Why do such subtle differences in energy have such large impacts on specificity? Mutations that cause a modest change in binding energy can cause large changes in occupancy because of the nonlinear Boltzmann relationship between these quantities (Fig. 4*C*). Moreover, specificity is determined by the deviation from additivity in the heterodimer relative to the homodimers, so small effects on the free energy of binding of these complexes can propagate into larger changes in specificity. Because of this intrinsic sensitivity, we predict that the evolution of specificity in paralogous complexes with symmetrical interfaces will often be attributable to one or a few genetic changes with relatively small energetic effects and subtle structural mechanisms. That specificity can evolve so easily also implies that paralog interference after gene duplication (49, 50) may often be easily resolved through one or a few mutations.

If specificity can be acquired by small deviations from energetic additivity in either direction, one might expect that homomeric and heteromeric specificity would be equally likely to evolve. But empirical observations suggest that heteromers evolve much more frequently after gene duplication (12, 13). Our findings suggest a plausible explanation for this pattern. The modeled structures of ancestral Hbs suggest that the critical mutation confers specificity because imperfect asymmetry in the interface creates antagonistic pleiotropy between heterodimeric and homodimeric binding. A favorable interaction occurs when the mutation is introduced once in the heteromer, but it fails to produce the same favorable contact and even disrupts a different favorable contact when introduced again on the other side of the interface in the homomer. Heterospecificity will result whenever asymmetry causes antagonistic pleiotropy like this, such that a favorable interaction can be optimized when it is iterated once but not twice. In contrast, homomeric specificity requires a mutation to be even more favorable the second time it is introduced on the other side of an interface. For this to occur, imperfect symmetry must synergistically enhance the interactions caused by the two iterations of the mutation in the homodimer. This scenario seems far less likely than an antagonistic effect, because favorable interactions are constrained in many ways, requiring compatibility of polarity, size, angle, etc. The imperfect symmetry of isologous interfaces may therefore create a mutational propensity that favors the evolution of heteromeric over homomeric specificity.

Our observations contribute to a growing body of evidence that complex multimeric complexes can evolve through simple genetic mechanisms (5, 15, 37, 39, 51–55). In Hb evolution, a single substitution in one of the duplicated genes was sufficient to cause a doubling in stoichiometry from dimer to tetramer, and a single-residue deletion at one interface in the other subunit was sufficient to confer strong preference for the $\alpha_2\beta_2$ heterotetrameric form. Although other substitutions enhanced these effects, and others may have permitted or entrenched them (5, 56), our data indicate that discrete evolutionary increases in complexity can occur by very short mutational paths from simpler ancestral forms. The major effects of these small sequence changes were possible because they took place in the context of an isologous complex, and it is likely that its symmetry was slightly imperfect. Many multimers share these structural properties, so we predict that, when other multimeric complexes are studied in detail, simple mechanisms will be found to have driven their historical elaboration, too.

Materials and Methods

Sequence Data, Alignment, Phylogeny, and Ancestral Sequence Reconstruction. The reconstructed ancestral sequences used here are the same as those reported previously (ref. 17, GenBank IDs MT079112, MT079113, MT079114, MT079115).

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

Recombinant Protein Expression and Purification. Coding sequences for reconstructed ancestral proteins were synthesized de novo and cloned into pLIC by Gibson assembly. For coexpression of Ancα+Ancβ, a polycistronic operon was constructed and separated by a spacer containing a stop codon and ribosome binding site, as described in ref. 57. Plasmids were heat-shock transformed into BL21(DE3) cells, which were then cultured and induced, and globin proteins were purified by ion exchange or zinc affinity chromatography. For experiments in which two proteins were separately expressed and purified and then mixed together at equal concentration prior to nMS, concentration of each protein was quantified using the Hemoglobin Assay Kit (Sigma); this procedure was performed in triplicate to assess technical error introduced by the quantification and mixing process. For details, see *SI Appendix*.

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 a 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 1 mL sample and monitored by absorbance at 280 nm.

Native Mass Spectrometry. Protein samples were buffer exchanged into 200 mM 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, 10,000 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 (30). 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 (2.65×10^{-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 follows: source DC offset of 15 V, injection flatapole DC to 13 V, interflatapole 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 8,750 resolution, microscans set to 1 and averaging set to 100. Mass spectra were deconvoluted using Unidec (58).

Calculating Multimerization Affinity, Occupancy, and ΔΔG of Specificity. For equations and explanation for calculating these metrics from nMS experimental data, see *SI Appendix, Material and Methods*. Occupancy calculations assume each dimer at physiologically relevant concentrations (1 mM total globin subunits).

We estimated the minimum Kd of assembly across IF2 by $Anc\alpha\beta_{37V+40W;\ IF1\ removed}$, 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

$$Kd_{min} = \frac{\left(20 \times 10^{-6}\right)^2 M}{1 \times 10^{-6} M} = 400\ \mu\text{M}.$$

Homology Models. SWISS-Model was used to generate a structural model of the Anc$\alpha\beta_{q40W}$ 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 (59). PyMOLV2.1 was used to visualize the proteins and capture images.

IF1-mediated homodimers were generated by the same procedure, except for homodimers of Ancα or Anc$\alpha\beta_{\Delta D}$, 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). For PyMol PSE file containing these models, see Supplemental Structures S2.

Data, Materials, and Software Availability. Raw Data, Code, Hb specificity and tetramerization data have been deposited in Figshare (60).

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

Author affiliations: [a]Department of Cell and Molecular Biology, University of Chicago, Chicago, IL 60637; [b]Department of Chemistry, Texas A&M University, College Station, TX 77843; [c]Department of Ecology and Evolution, University of Chicago, Chicago, IL 60637; [d]Institute of Protein Design, University of Washington, Seattle, WA 98195; and [e]Department of Human Genetics, University of Chicago, Chicago, IL 60637

1. D. S. Goodsell, A. J. Olson, Structural symmetry and protein function. *Annu. Rev. Biophy. Biomol. Struct.* **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. *Mol. Biol. Evolut.* **29**, 1353–1366 (2012).
4. M. Lynch, Evolutionary diversification of the multimeric states of proteins. *Proc. Natl. Acad. Sci. U.S.A.* **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. S. Mallik, D. S. Tawfik, Determining the interaction status and evolutionary fate of duplicated homomeric proteins. *PLoS Comput. Biol.* **16**, e1008145 (2020).
13. A. Marchant *et al.*, The role of structural pleiotropy and regulatory evolution in the retention of heteromers of paralogs. *Elife* **8**, e46754(2019).
14. 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).
15. D. Grueninger *et al.*, Designed protein-protein association. *Science* **319**, 206–209 (2008).
16. J. Monod, J. Wyman, J.-P. Changeux, On the nature of allosteric transitions: A plausible model. *J. Mol. Biol.* **12**, 88–118 (1965).
17. A. S. Pillai *et al.*, Origin of complexity in haemoglobin evolution. *Nature* **581**, 480–485 (2020).
18. 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).
19. D. Ding *et al.*, Co-evolution of interacting proteins through non-contacting and non-specific mutations. *Nat. Ecol. Evol.* **6**, 590–603 (2022).
20. 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).
21. C. J. McClune, A. Alvarez-Buylla, C. A. Voigt, M. T. Laub, Engineering orthogonal signalling pathways reveals the sparse occupancy of sequence space. *Nature* **574**, 702–706 (2019).
22. C. D. Aakre *et al.*, Evolving new protein-protein interaction specificity through promiscuous intermediates. *Cell* **163**, 594–606 (2015).
23. T. V. Lite *et al.*, Uncovering the basis of protein-protein interaction specificity with a combinatorially complete library. *Elife* **9**, e60924 (2020).
24. G. K. A. Hochberg, J. W. Thornton, Reconstructing ancient proteins to understand the causes of structure and function. *Annu. Rev. Biophys.* **46**, 247–269 (2017).
25. 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).
26. J. R. Emlaw *et al.*, A single historical substitution drives an increase in acetylcholine receptor complexity. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2018731118 (2021).
27. I. Nocedal, M. T. Laub, Ancestral reconstruction of duplicated signaling proteins reveals the evolution of signaling specificity. *Elife* **11**, e77346 (2022).
28. 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).
29. 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).
30. A. C. Leney, A. J. R. Heck, Native mass spectrometry: What is in the name? *J. Am. Soc. Mass Spectrom.* **28**, 5–13 (2016).
31. 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.
32. L. Kiger *et al.*, Thermodynamic studies on the equilibrium properties of a series of recombinant betaW37 hemoglobin mutants. *Biochemistry* **37**, 4336–4345 (1998).
33. B. L. Boys, L. Konermann, Folding and assembly of hemoglobin monitored by electrospray mass spectrometry using an on-line dialysis system. *J. Am. Soc. Mass Spectrom.* **18**, 8–16 (2007).
34. G. Snyder, B. Sheafor, Red blood cells: Centerpiece in the evolution of the vertebrate circulatory system. *Am. Zool.* **39**, 189–198 (1999).
35. G. Fermi, M. F. Perutz, B. Shaanan, R. Fourme, The crystal structure of human deoxyhaemoglobin at 1.74 Å resolution. *J. Mol. Biol.* **175**, 159–174 (1984).
36. A. A. Bogan, K. S. Thorn, Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **280**, 1–9 (1998).
37. J. Jee, I.-J.L. Byeon, J. M. Louis, A. M. Gronenborn, The point mutation A34F causes dimerization of GB1. *Proteins Struct. Funct. Bioinf.* **71**, 1420–1431 (2008).
38. 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. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2101117119 (2022).
39. 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).
40. 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).
41. H. Schweke *et al.*, An atlas of protein homo-oligomerization across domains of life. *Cell* **187**, 999–1010.e15 (2024).
42. J. A. Marsh *et al.*, Protein complexes are under evolutionary selection to assemble via ordered pathways. *Cell* **153**, 461–470 (2013).
43. E. D. Levy, E. B. Erba, C. V. Robinson, S. A. Teichmann, Assembly reflects evolution of protein complexes. *Nature* **453**, 1262–1265 (2008).
44. R. P. Bahadur, F. Rodier, J. Janin, A dissection of the protein–protein interfaces in icosahedral virus capsids. *J. Mol. Biol.* **367**, 574–590 (2007).
45. E. T. Powers, D. L. Powers, A perspective on mechanisms of protein tetramer formation. *Biophys. J.* **85**, 3587–3599 (2003).
46. J. T. Brennecke, B. L. de Groot, Quantifying asymmetry of multimeric proteins. *J. Phys. Chem. A* **122**, 7924–7930 (2018).
47. M. Bonjack-Shterengartz, D. Avnir, The near-symmetry of proteins. *Proteins Struct. Funct. Bioinf.* **83**, 722–734 (2015).
48. M. F. Perutz, G. Fermi, B. Luisi, B. Shaanan, R. C. Liddington, Stereochemistry of cooperative mechanisms in hemoglobin. *Acc. Chem. Res.* **20**, 309–321 (1987).
49. C. R. Baker, V. Hanson-Smith, A. D. Johnson, Following gene duplication, paralog interference constrains transcriptional circuit evolution. *Science* **342**, 104–108 (2013).
50. J. T. Brigham, 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 Genet.* **4**, e1000191 (2008).
51. S. Liu *et al.*, Nonnatural protein–protein interaction-pair design by key residues grafting. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 5330–5335 (2007).
52. G. Guntas, C. Purbeck, B. Kuhlman, Engineering a protein–protein interface using a computationally designed library. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 19296–19301 (2010).
53. D. P. Anderson *et al.*, Evolution of an ancient protein function involved in organized multicellularity in animals. *Elife* **5**, e10147 (2016).
54. L. Schulz *et al.*, Evolution of increased complexity and specificity at the dawn of form I Rubiscos. *Science* **378**, 155–160 (2022).
55. F. L. Sendker *et al.*, Emergence of fractal geometries in the evolution of a metabolic enzyme. *Nature* **628**, 894–900 (2024).
56. G. K. A. Hochberg *et al.*, A hydrophobic ratchet entrenches molecular complexes. *Nature* **588**, 503–508 (2020).
57. C. Natarajan *et al.*, Expression and purification of recombinant hemoglobin in Escherichia coli. *PLoS One* **6**, e20176 (2011).
58. M. T. Marty *et al.*, Bayesian deconvolution of mass and ion mobility spectra: From binary interactions to polydisperse ensembles. *Anal. Chem.* **87**, 4370–4376 (2015).
59. L. G. Nivón, R. Moretti, D. Baker, A Pareto-optimal refinement method for protein design scaffolds. *PLoS One* **8**, e59004 (2013).
60. C. R. Cortez-Romero, J. Lyu, A. S. Pillai, A. Laganowsky, J. W. Thornton, Hb specificity and tetramerization. Figshare data repository. https://figshare.com/projects/Hb_specificity_and_tetramerization/214114. Deposited 11 December 2024.