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

Carlos R. Cortez-Romero (1), Jixing Lyu (4), Arvind S. Pillai (2,5), Arthur Laganowsky (4), Joseph W. Thornton (2,3)

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
- Supplementary figures 1 to 13
- 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 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 – involving the same surface patch on interacting subunits, 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 vs homomeric interactions. Our findings suggest that elaborate and specific symmetrical 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 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.

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 (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 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 (1,5,9,15-17). 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. 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 (20-26). But the genetic and structural mechanisms by which those

interactions were acquired long ago are often different from their derived forms in the present (27). 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 (18,19) and in multimers composed of unrelated proteins, which are by definition asymmetrical (26). 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 ((17), 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 to make the tetramer ((29), 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 (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. 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? 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

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

148 Here we isolated the individual contributions of each amino acid changer by introducing them
149 singly into Ancαβ and characterizing their effect on assembly into tetramers using size-exclusion
150 chromatography (SEC) and native mass spectrometry (nMS) (30,31). We found that q40W
151 alone is sufficient to recapitulate the evolution of Hb's tetrameric stoichiometry. Ancαβ forms
152 only dimers in SEC at 100 μM of total protein subunits; by contrast, the mutant Ancαβ$_{q40W}$ is
153 tetrameric, with occupancy of the tetramer similar to that observed in the derived Ancα + Ancβ
154 complex and human Hb (Fig. 1E). We used nMS across a titration series to measure the affinity
155 with which dimers associate into tetramers and found that the tetramerization affinity of
156 Ancαβ$_{q40W}$ (Kd 10 μM) is stronger than that of Ancα + Ancβ (29 μM) and human Hb (34 μM)
157 (Fig. 1F). The conclusion that q40W is sufficient to confer tetramerization is robust to statistical
158 uncertainty about the ancestral sequence: similar experiments using a different reconstruction of
159 Ancαβ that incorporates alternative residues at all ambiguously reconstructed sites yields
160 almost identical results (Fig. S2).

161 The other historical replacement, t37V, is not sufficient to confer tetramerization. Mutant
162 Ancαβ$_{t37V}$ confers no detectable tetramer occupancy by SEC, even at 1 mM (Fig. 1G), and it
163 displays no measurable affinity to form tetramers using nMS (Fig. 1H). When combined with
164 q40W, however, t37V does increase affinity of the dimer-tetramer transition by a factor of 6
165 compared to the effect of q40W alone (Fig. 1D; Fig. S3).

166 In principle, a sequence change could also facilitate tetramerization by increasing affinity of the
167 monomer-to-dimer transition; by increasing the effective concentration of dimers, more
168 tetramers would be produced at a given protein concentration, even if affinity of the dimer-
169 tetramer transition were unchanged. Using nMS, we found that t37V improves the monomer-
170 dimer affinity of Ancαβ by >100-fold (Fig. 1H; Fig. S3). Substitution q40W, in contrast, has no
171 effect on monomer-dimer affinity. These findings are consistent with the structural location of
172 these residues -- t37V contributes to both IF1 and IF2 and q40W to IF2 only -- and they explain
173 why t37V does not confer tetramerization on its own but enhances the impact of q40W.

174 A likely physical mechanism for the effect of q40W is that tryptophan's bulky hydrophobic side
175 chain nestles into a hydrophobic divot on the IF2 surface of the facing subunit, and is further
176 strengthened by a hydrogen bond to 102D (32). To test this hypothesis, we identified alternative
177 amino acid replacements with similar biochemical properties and measured whether they also

178 could have caused Ancαβ to evolve into a tetramer. Like tryptophan, the bulky hydrophobic
179 residues phenylalanine or tyrosine at this position also confer tetramerization, albeit at affinity
180 slightly weaker than q40W but similar to that of Ancα+Ancβ and human Hb. The greater affinity
181 of tryptophan may be due to its longer side chain, which buries more hydrophobic surface area
182 across the interface; the hydrogen bond with 102D could make a small contribution but is not
183 necessary, because phenylalanine confers tetramerization but provides no hydrogen bond
184 donor. Leucine, in contrast, which has a smaller volume and no hydrogen bonding capacity,
185 confers no measurable tetramerization (Fig. 1I). High-affinity homotetramerization could
186 therefore have evolved via any of three different aromatic replacements at site 40.

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

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

198 If this hypothesis is correct, then assembly across IF2 by the derived Hb protein should require
199 assembly across IF1 to multiply the intrinsic affinity of IF2 (Fig. 1A). We tested this prediction by
200 introducing q40W into Ancαβ but doing so under conditions that prevent assembly across IF1.
201 We first compromised dimerization across IF1 genetically by reverting the IF1 surface to the
202 ancestral states of the monomeric ancestor AncMH; these mutations abolish dimer occupancy,
203 leaving a monomers-only population at 20 μM (Fig. 2A). We then introduced q40W into these
204 IF1-ablated mutants and assessed stoichiometry using nMS. As predicted, these proteins do not
205 form any observable dimers or tetramers (detection limit ~1 μM) (Fig. S4). Similar results are
206 found when we used the combination of t37V/q40W to confer association across IF2 or the
207 mutation P127R – which introduces unsatisfied positive charges into IF1 -- to compromise IF1.
208 (Fig. 2B). The IF2 mutations do not compromise heme binding or solubility, because the mutant
209 proteins are purifiable and heme-bound in nMS.

210 We also tested whether assembly across IF2 could have been historically acquired before
211 dimerization across IF1 evolved. We introduced t37V/q40W into the ancestral monomer AncMH
212 – which existed before the evolution of dimerization -- and tested whether dimer assembly
213 across IF2 can be conferred in this background. As predicted, only monomers were observed,
214 with no dimers or higher stoichiometries detected (Fig. 2C). Acquisition of multimerization
215 across IF2 by q40W and by the pair t37V/q40W therefore depends on the prior evolution of
216 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 only 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. 2D). 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). 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 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 (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 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 (17). We therefore hypothesized that heterospecificity of the Hb tetramer is encoded entirely by IF1, such

258 that Ancα and Ancβ specifically heterodimerize across IF1, and these heterodimers then bind to
259 each other via a nonspecific IF2, yielding $\alpha_2\beta_2$ heterotetramers.

260 This hypothesis makes two predictions: 1) IF1 mediates specific assembly of α and β subunits
261 into heterodimers, and 2) this specificity is sufficient to account for the heterospecificity of $\alpha_2\beta_2$
262 heterotetramer. To test the first hypothesis, we characterized the specificity of hetero- vs
263 homodimer assembly by IF1 under two different conditions in which no binding across IF2
264 occurs. First, we diluted a coexpressed mixture of Ancα and Ancβ to concentrations at which
265 dimers rather than tetramers assemble: at 50 μM, only heterodimers and heterotetramers form;
266 at 5 μM, only heterodimers are observed (Fig. 3A). IF2 does not mediate assembly of
267 monomers into dimers in the absence of IF1 (Fig. 2A, 2B), so these heterodimers must be IF1-
268 mediated, indicating that IF1 is heterospecific (Fig. 3A). Second, we expressed Ancα and Ancβ
269 separately and mixed them at equal and moderate concentration; because tetramerization
270 requires co-folding, only IF1 dimers form (33), and these are predominantly heterodimers (Fig.
271 3B, Fig. S5). Finally, we engineered protein Ancβ' – a variant of Ancβ in which all IF2 residues
272 that were substituted between Ancαβ and Ancβ are reverted to the ancestral state, thus
273 abolishing binding across IF2– and found that it also forms predominantly heterodimers when
274 mixed with Ancα (Fig. 3C, Fig. S5). Together, these data indicate that the derived IF1 is specific,
275 preferentially mediating assembly into heterodimers.

276 To test the second prediction – that heterospecificity mediated by IF1 is sufficient to drive
277 specific assembly of $\alpha_2\beta_2$ heterotetramers even if IF2 is nonspecific – we measured the affinities
278 of homomerization and heteromerization across IF1 and used these measurements to predict
279 their effects on tetramer specificity in the absence of any specificity at IF2. Using nMS and
280 Ancβ', we found that IF1's heterodimerization affinity (Kd=0.6 μM) is slightly worse than its
281 homodimerization affinity (0.2 μM), but both are far better than the Ancα homodimer (24 μM)
282 (Fig. 3D, S6, S7, S8, S9). We then predicted occupancy of each stoichiometry as the
283 concentration of Hb subunits changes, given these affinities at IF1 and assuming that IF2 has a
284 dimer-to-tetramer affinity of 30 μM, as measured in Ancα + Ancβ, with no preference for
285 homomeric or heteromeric binding (Fig. 1D). At low concentrations, the system produces almost
286 exclusively IF1-mediated heterodimers. The predominance of heterodimers is attributable to
287 Ancα's weak homodimerization affinity; the excess of unbound Ancα subunits causes Ancβ
288 subunits to preferentially heterodimerize rather than homodimerize at equilibrium, even though
289 Ancβ's homodimerization affinity is slightly stronger than its heterodimerization affinity (Fig. 3D).
290 As protein concentration increases, these dimers begin to assemble with each other across IF2
291 into tetramers. The excess of heterodimers over homodimers means that the vast majority of
292 the tetramers are heterotetramers, even though IF2 itself does not distinguish between subunit
293 types. At physiologically relevant concentrations of 3mM total Hb subunits (34), the population is
294 dominated by $\alpha_2\beta_2$ heterotetramers, with a small fraction of heterodimers and virtually no
295 homotetramers (Fig. 3d; right panel).

296 Taken together, these data establish that the measured specificity of IF1 alone mediates highly
297 specific assembly of Ancα+ Ancβ into heterotetramers, even when IF2 is entirely nonspecific --

298 which our previous experiments suggest is the case – because IF1 is a much stronger interface
299 than IF2. The historical acquisition of heterospecificity across IF1 after the Ancαβ gene
300 duplication is therefore sufficient to account for the evolution of Hb's heterotetrameric
301 architecture.

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

306 The heterospecificity of a pair of dimerizing proteins can be quantified in energetic terms as the
307 difference in the ΔG of binding between the heterodimer and the mean of the two homodimers
308 ($\Delta\Delta G_{spec}$; see methods for calculation). If $\Delta\Delta G_{spec} = 0$, then the fractional occupancy of the
309 heterodimer at saturating and equal concentrations of subunits will be 50%, as will the sum of
310 the homodimers; this is true even if the homodimer ΔGs are very different from each other, as
311 long as the heterodimer ΔG is halfway between them. By contrast, if $\Delta\Delta G_{spec}<0$, then
312 heterodimers will account for the majority of dimers; conversely, if $\Delta\Delta G_{spec}>0$, homodimers
313 together will predominate (Fig. 4A-C). Hetero- or homospecificity thus arises when two paralogs
314 contribute nonadditively to dimerization. Whether or not the system is hetero- or homospecific,
315 the two homodimers will have equal occupancies to each other at saturating conditions:
316 irrespective of the fraction of subunits assembled into heterodimers, the remaining subunits will
317 be at equal concentrations and by definition will be well above the homodimerization Kds (28).

318 We quantified the heterospecificity of Ancα and Ancβ at IF1 by estimating $\Delta\Delta G_{spec}$. We used
319 nMS to measure the homodimer and heterodimer affinities of Ancα and Ancβ', which contains
320 all substitutions that occurred along the Ancβ branch except those that mediate tetramerization
321 across IF2, allowing us to prevent tetramerization and thus isolate the specificity effects at IF1.
322 From these affinities, we calculated the ΔG of binding and the expected fractional occupancy of
323 each dimer at high and equal concentration of subunits. For Ancα+Ancβ', we found that
324 $\Delta\Delta G_{spec}= -1.3$ (in units of kT) and heterodimer occupancy of 82% (Fig. 4D). This represents the
325 total specificity acquired by the two diverging paralogs after the duplication of of Ancαβ, which
326 by definition had no specificity. This specificity was acquired because of evolutionary changes in
327 all three relevant affinities. Relative to the ancestral dimerization affinity of Ancαβ, Ancα's
328 energy of homodimerization became worse ($\Delta\Delta G = 0.9$) while homodimerization by Ancβ'
329 improved substantially ($\Delta\Delta G = -3.7$). The heterodimer affinity improved by $\Delta\Delta G = -2.7$,
330 substantially more than the average of the two homodimers, yielding the observed strong
331 preference for the heterodimer.

332 We next sought to isolate the contribution of the evolutionary changes that occurred along each
333 of the two branches. To measure the specificity acquired along the branch leading to Ancα, we
334 measured affinities and calculated $\Delta\Delta G_{spec}$ when Ancα is mixed with the deeper ancestor
335 Ancαβ. This pair of proteins is heterospecific, with $\Delta\Delta G_{spec}= -1.2$ (expected heterodimer
336 occupancy 76%). Changes in the α subunit alone therefore account for >90% of the total
337 $\Delta\Delta G_{spec}$ that was acquired by the entire Ancα+Ancβ' system. This specificity is acquired via a

338 2.6-fold reduction in Ancα's homodimerization affinity compared to the Ancαβ ancestor and a
339 1.8-fold improvement in heterodimer affinity (Fig. 4E; Fig. S7B & D).

340 To isolate the contribution to IF1 specificity of evolutionary changes that occurred along the
341 branch to Ancβ, we measured affinities when Ancβ' is mixed with Ancαβ. This pair of proteins is
342 weakly heterospecific, with $\Delta\Delta G_{spec}$= –0.3 and expected heterodimer occupancy of just 58%.
343 The specificity is weak because both the homodimer and heterodimer affinity improved, but the
344 deviation of the heterodimer from the average of the homodimers is very small (Fig. 4F; Fig.
345 S7A &C).

346 Finally, we assessed whether the evolutionary changes in the Hbα subunit and those in the Hbβ
347 subunit interacted with each other nonindependently. If the changes affect specificity entirely
348 independently, $\Delta\Delta G_{spec}$ should equal the sum of the $\Delta\Delta G_{spec}$ acquired on each of the two
349 branches (–1.2 + –0.3 = –1.5). The observed $\Delta\Delta G_{spec}$ = –1.3, suggesting a very weak negative
350 interaction between changes in the two subunits, which makes the complex slightly less
351 heterospecific than expected if the substitutions were independent (Fig. 4G).

352 Taken together, these data indicate that the IF1 specificity acquired by the derived complex
353 Ancα + Ancβ is primarily attributable to substitutions in the α subunit, with substitutions in the β
354 subunit making a much smaller contribution; nonadditive interactions between the two sets of
355 changes did not contribute to the evolution of heterospecificity. The most important factor was
356 that Ancα became much worse at binding itself than at binding Ancβ. Ancβ, by contrast, became
357 slightly worse at binding Ancα than binding itself (Fig. 4G).

358 **A one-residue deletion was the primary evolutionary cause of heterospecificity.** We next
359 sought to identify the particular historical substitutions in Ancα that conferred this heteromeric
360 specificity on IF1. Only three sequence changes occurred on the branch from Ancαβ to Ancα: a
361 single-residue deletion of a histidine at site 3 (ΔH3), a five-residue deletion in helix D (ΔD), and
362 an amino acid replacement (v140A). ΔH3 is on the protein's N-terminal loop near IF1, and ΔD
363 directly contributes to the interface. Substitution v140A is biochemically conservative and far
364 away from the interface. The deletions are conserved to the present in Hbα subunits throughout
365 the jawed vertebrates, including humans, whereas the amino acid at site 140 varies (Fig. S1).
366 We therefore focused first on the effects of the deletions.

367 To isolate the contribution of each deletion to the evolution of specificity, we introduced each
368 one singly into Ancαβ and measured its effect on affinity and specificity when the mutant protein
369 is mixed with Ancαβ. We found that introducing ΔH3 alone confers substantial specificity,
370 recapitulating >80% of the Ancα's acquired heterospecificity for Ancαβ ($\Delta\Delta G_{spec}$ = –1.0 out of a
371 total $\Delta\Delta G_{spec}$ = –1.2 acquired along this branch) and >75% of the specificity acquired along both
372 branches by the entire Ancα+Ancβ' complex ($\Delta\Delta G_{spec}$ = –1.3, Figs. 5A, C). ΔH3 enhances
373 specificity by improving heterodimer affinity and reducing homodimer affinity, with both Kds very
374 close to those of Ancα (Fig. 5A; Fig. S10A & C).

375 The other deletion, ΔD, removes several residues that directly interact with the other subunit
376 across IF1, but introducing this change into Ancαβ had a much weaker effect on specificity
377 ($\Delta\Delta G_{spec}$ = –0.4, Fig. 5B; Fig. S10B & D). When the contributions of ΔH3 and ΔD to specificity
378 are added together, they slightly exceed the specificity of Ancα (–1.4 rather than –1.3),
379 suggesting the possibility of a very weak negative epistatic interaction between them or a small
380 countervailing effect of the third change v104A. Taken together, these results indicate that ΔH3
381 was a large-effect historical sequence change that accounted for most of the specificity
382 historically acquired by the derived Hb complex.

383 **Structural mechanisms for the gain in specificity.** We next considered the structural
384 mechanisms by which ΔH3 conferred specificity by increasing heterodimer affinity and reducing
385 homodimer affinity. For a mutation to have these opposite effects, it must yield favorable
386 interactions when introduced into one side of the interface (in the heterodimer) but have
387 deleterious effects when introduced twice (in the homodimer). In principle, two kinds of
388 mechanisms could cause these opposite effects. Either 1) the mutated residue could interact
389 directly with the same residue on the other subunit favorably when one is in the derived state
390 but unfavorably when both are, or 2) the symmetry of the interface could be imperfect, such that
391 introducing the mutation on one side of the interface is favorable but introducing it again onto
392 the other side is net-unfavorable. The first scenario does not pertain in this case. Residue H3 is
393 part of the N-terminal loop, which does not participate directly in IF1 but instead packs against
394 helix H, which does contribute to IF1. But neither helix H nor the N-terminal loop contact the
395 same elements in the other subunit across the interface (Fig. 5D). Asymmetry in the interface is
396 therefore the likely of cause ΔH3's differential effects on heterodimer vs. homodimer affinity.

397 To gain insight into the possible nature of this asymmetry and potential mechanisms by which
398 ΔH3 affects specificity, we modeled the structures of the Ancαβ homodimer, the Ancαβ$_{\Delta H2}$
399 homodimer, and the heterodimer of these two proteins. The modeled Ancαβ homodimer
400 contains a subtle asymmetry: on one end of IF1, residue 130H on helix H sits close to 33R on
401 the opposite subunit, which allows a cross-interface hydrogen bond to form; on the other end of
402 the interface, the two residues are slightly further away from each other, leaving their hydrogen-
403 bonding potential unsatisfied when bound (Fig. 5E & F). In the heterodimer, deleting His2 from
404 one subunit repairs this unfavorable interaction. Specifically, the deletion shortens the N-
405 terminal loop and changes its packing interaction against helix H, which causes helix H to slide
406 along the interface by ~1 Å compared to its position in the unmutated Ancαβ homodimer (Figs.
407 5D, 5G). 130H moves closer to 37T on the other subunit, allowing it to form a new hydrogen
408 bond across the interface, and several other interactions across the interface are also
409 enhanced. On the other end of the isologous interactions, the favorable interactions found in the
410 homodimer remain intact. This provides a potential structural explanation for how ΔH3 improves
411 heterodimer affinity (Figs. 5D, G).

412 The modeled Ancαβ$_{\Delta H3}$ homodimer structure is notably asymmetric and suggests a possible
413 mechanism by which introducing ΔH3 into both subunits reduces affinity (Fig. 5H). One side
414 displays the favorable new cross-interface interactions caused by ΔH3 in the heterodimer,

415 including the 130H-37T hydrogen bond. On the other side, however, the effect of the deletion is
416 very different: ΔH3 again causes helix H to slide along the interface, but on this side the
417 movement of 130H breaks the ancestral 130H-33R hydrogen bond, and 37T is also too far
418 away to interact favorably. This leaves the side chains of both 130H and 33R unsatisfied,
419 reducing homodimer affinity. In total, the homodimer of Ancαβ$_{\Delta H3}$ contains three unsatisfied
420 hydrogen-bond donors/acceptors at these sites, whereas only one and two are unsatisfied in the
421 heterodimer and the ancestral homodimer, respectively.

422 These hypothesized mechanisms appear to have persisted over time. The same pattern of
423 interactions are found in the modeled structures of the hetero- and homodimers of Ancα + Ancβ
424 (Fig. S11). High-resolution crystal structures of extant hemoglobin also show notable
425 asymmetries in the multimerization interfaces, which exceed the deviation expected given the
426 resolution of the structures (35). These structures include some of the particular asymmetrical
427 interactions observed in our ancestral models: in the human Hb heterotetramer, 33R hydrogen
428 bonds across IF1 to residue 130, but this interaction is again lacking in the homodimer of human
429 Hbα, leaving 33R unsatisfied, potentially explaining the weak homomeric affinity of Hbα (Fig.
430 S11). At least some of the mechanisms of heterodimer specificity suggested by the structural
431 models of the ancestral proteins are therefore present in the known structures of its present-day
432 descendants. Structural models are prone to error, and the asymmetries we observed are
433 subtle; further research will be required to definitively characterize potential asymmetries in the
434 ancestral multimers.

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

441 To test this hypothesis, we measured the effect on specificity of subsets of changes that
442 occurred along the Ancβ lineage, which the results above show had strong effects on affinity
443 when introduced all together. First, we tested the five substitutions that that occurred at the IF1
444 surface (Fig. 5E & 5F). We introduced these changes into Ancαβ (creating protein Ancαβ$_{IF1}$) and
445 measured affinity and specificity when this protein is mixed with Ancαβ. These substitutions
446 yield a highly heterospecific complex ($\Delta\Delta G_{spec}$ = –2.2, heterodimer occupancy 90%, Fig. 6A; Fig.
447 S12A, C, & E). Unlike the Ancα substitutions, the Ancβ$_{IF1}$ substitutions confer heterospecificity
448 by improving both homodimer and heterodimer affinity, but they improve the latter by more than
449 the former.

450 Because Ancαβ$_{IF1}$ is specific in complex with Ancαβ, we wondered whether it would also be
451 specific with Ancα. We found that this complex is barely heterospecific ($\Delta\Delta G_{spec}$ = –0.1, Fig. 6B),
452 implying that other substitutions on the branch leading to Ancβ but not on the interface must
453 have contributed to the evolution of specificity between Ancαβ$_{IF1}$ and Ancα. We therefore
454 introduced an additional set of five historical substitutions that occurred in Ancβ but one

455 structural layer away from IF1 (see ref. 17). This protein (Ancαβ$_{IF1+Adjacent}$) has strong
456 heterospecificity when mixed with Ancα ($\Delta\Delta G_{spec}$ = –2.0, heterodimer occupancy >85%, Fig. 6D;
457 Fig. S12B, D, & F), because these mutations together improve both heterodimer and
458 homodimer affinity, but with a larger improvement in the heterodimer. It is also moderately
459 heterospecific when mixed with Ancαβ ($\Delta\Delta G_{spec}$ = –0.9).

460 Finally, we tested the effect of the adjacent substitutions on their own and found that they confer
461 specificity when mixed with Ancαβ ($\Delta\Delta G_{spec}$ = –0.9). These mutations impart specificity by
462 causing almost identical changes in homo- and heterodimer affinity. They also confer some
463 heterospecificity when Ancαβ$_{Adjacent}$ is mixed with Ancα ($\Delta\Delta G_{spec}$ = –0.6, Fig. S13A-D).

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

470 **DISCUSSION**

471 This work provides a mechanistic history of the evolutionary transition from the ancestral Ancαβ
472 homodimer to the derived Hb heterotetramer (summarized in Fig. S1). Each transition was
473 driven by a very simple genetic mechanism: a single substitution at IF2 conferred high affinity
474 tetramerization, and a single amino acid deletion at IF1 conferred heteromeric specificity. These
475 two key sequence changes have remained conserved in the descendant Hbα and Hbβ subunits
476 of all extant jawed vertebrates (Fig. S1). These transitions were both facilitated by the isologous
477 architecture of Hb's two interfaces, which creates a propensity for mutation to produce high-
478 order complexes and heterospecificity.

479 **Symmetry facilitated evolution of the tetrameric stoichiometry.** We found that
480 tetramerization across IF2 was driven primarily by a single replacement to a bulky hydrophobic
481 amino acid (q40W). In biochemical studies of extant protein interfaces, much of the free energy
482 change in protein-protein binding is attributable to interactions of bulky hydrophobic residues
483 with hydrophobic surface indentations (36), and mutations to bulky hydrophobic amino acids
484 can drive assembly into high-order multimers (9,37-40). Similar substitutions during history may
485 have been driving mechanisms during the evolution not only of Hb but of other molecular
486 complexes, as well.
487
488 The majority of complexes assemble through isologous interfaces (41). It has been suggested
489 that this must imply that isology confers some selective benefit by improving protein function (1).
490 Our results suggest an alternative explanation. If mutations are much more likely to produce
491 isologous complexes than nonisologous ones, then isologous complexes will predominate in
492 nature, even if there is no systematic fitness difference between the two types of multimer. We
493 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. 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.

It has been observed that 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 is attributable to selection: by this hypothesis, selection favors evolutionary intermediates that contain the high-affinity interface, 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 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.

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 (7,13,18,19,21,24-26).

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: 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 (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 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

578 required to yield substantial changes in specificity. The IF1 of Ancα+Ancβ' mediates 80%
579 heterodimer occupancy at equal and saturating concentrations, but its specificity is quite
580 moderate $\Delta\Delta G_{spec} = -1.3$). This difference in binding energy is less than that associated with a
581 typical hydrogen bond or burial of a large hydrophobic residue. The structural differences in
582 physical interactions across the homodimer vs. heterodimer interfaces in our modeled structures
583 could easily yield energetic differences of this magnitude, although the particular form of
584 asymmetry in ancestral hemoglobin complexes remains uncertain. Recent in silico work also
585 found that small differences in ΔG can cause large differences in occupancy between
586 homodimers and heterodimers (28).

587 Why do such subtle differences in energy have such large impacts on specificity? Mutations that
588 cause a modest deviation in binding energies can cause large changes in occupancy because
589 of the nonlinear Boltzmann relationship between these quantities (Fig. 4C). Moreover, specificity
590 is determined by the deviation from additivity in the heterodimer relative to the homodimers, so
591 small differences in the free energy of binding will propagate into even larger changes in
592 specificity. Because of this intrinsic sensitivity, we predict that the evolution of specificity in
593 paralogous complexes with symmetrical interfaces will often be attributable to one or a few
594 genetic changes with relatively small energetic effects and subtle structural mechanisms. That
595 specificity can evolve so easily also implies that paralog interference after gene duplication
596 (49,50) may often be easily resolved through one or a few mutations.

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

615 Taken together, our observations contribute to a growing body of evidence that complex
616 multimeric complexes can evolve through simple genetic mechanisms (5, 15, 37, 39,51-55). In
617 Hb evolution, a single substitution in one of the duplicated genes was sufficient to cause a

618 doubling in stoichiometry from dimer to tetramer, and a single-residue deletion at one interface
619 in the other subunit was sufficient to confer strong preference for the $\alpha_2\beta_2$ heterotetrameric form.
620 Although other substitutions enhanced these effects, and others may have permitted or
621 entrenched them (5,56), our data indicate that discrete evolutionary increases in complexity can
622 occur by very short mutational paths from simpler ancestral forms. The major effects of these
623 small sequence changes was possible because they took place in the context of an isologous
624 complex, and it is likely that its symmetry was slightly imperfect. Many multimers share these
625 structural properties, so we predict that, when other multimeric complexes are studied in detail,
626 simple mechanisms will be found to have driven their historical elaboration.

627 **MATERIALS AND METHODS**
628
629 **Sequence data, alignment, phylogeny, and ancestral sequence reconstruction.** The
630 reconstructed ancestral sequences used here are the same as those reported previously (17).
631 Briefly, 177 amino acid sequences of hemoglobin and related paralogs were collected and
632 aligned. The maximum likelihood (ML) phylogenetic tree was inferred using the AIC best-fit
633 model, LG+G+F (57,58). The phylogeny was rooted using as outgroups neuroglobin and globin
634 X, which are found in both deuterostomes and protostomes and diverged prior to the gene
635 duplications that produced vertebrate myoglobin and the hemoglobin subunits. Ancestral
636 sequence reconstruction was performed using the empirical Bayes method (62), given the
637 alignment, ML phylogeny, ML branch lengths, and ML model parameters. Reconstructed
638 ancestors that were used in this study have been deposited previously in GenBank (IDs
639 MT079112, MT079113, MT079114, MT079115).
640
641 The historical mutations that we introduced into those ancestral proteins are the following. For
642 the set *IF1-reverted*, all sites in IF1 that were substituted on the branch leading to Ancb are
643 reverted to the ancestral state found in Ancab; the mutations introduced are V36t, Y38h, V115a,
644 V119e, H130r, D134e. For the set *IF2-reverted*, all sites that were substituted in IF2 on the
645 branch leading to Ancb are reverted to the ancestral state found in Ancab; the mutations
646 introduced are T37v, W40q, R43t, H100r, E104h. For the set *IF1*, all sites at IF1 that were
647 substituted between Ancab and Ancb are changed to the derived state found in Ancb; the
648 mutations introduced are t37V, k58M, r107K, h130Q, d134Q4. For the set *Adjacent*, five sites
649 adjacent to IF1 that were substituted between Ancab and Ancb are changed to the derived state
650 found in Ancb; the mutations introduced are h47S, s60N, q62K, a96S, h97E. The set
651 *IF1+Adjacent* is the union of the sets *IF1* and *Adjacent*. Deletion DD deletes residues a54, e55,
652 a56, i57, and k58 from Ancab.
653
654 **Recombinant protein expression.** Coding sequences for reconstructed ancestral proteins
655 were optimized for expression in *Escherichia coli* using IDT Codon Optimization and
656 synthesized *de novo* as gBlocks (IDT). Coding sequences were cloned by Gibson assembly into
657 vector pLIC (3) under control of a T7 polymerase promoter. For co-expression of Ancα+Ancβ, a
658 polycistronic operon was constructed under control of a T7 promoter and separated by a spacer
659 containing a stop codon and ribosome binding site, as described in (63).
660
661 BL21 (DE3) *Esherichia coli* cells (New England Biolabs) were heat-shock transformed and
662 plated onto Luria broth (LB) containing 50 ug/mL carbenicillin. For the starter culture, a single
663 colony was inoculated into 50 mL of LB with 1:1000 dilution of working-stock carbenicillin and
664 grown overnight. 5 mL of the starter culture were inoculated into a larger 500-mL terrific broth
665 (TB) mixture containing the appropriate antibiotic concentration. Cells were grown at 37° C and
666 shaken at 225 rpm in an incubator until they reached an optical density at 600 nm of 0.6-0.8.
667
668 For expression of single globin proteins, 100 uM of isopropyl-β-D-1-thiogalactopyranoside
669 (IPTG) and 25 mg/500 mL of hemin were added to each culture. Expression of single proteins in
670 culture were done overnight at 22° C. Cells were collected by centrifugation at 4,000*g* and

671 stored at -80° C until protein purification. Coexpressed proteins were induced using 500 mM
672 IPTG expression with 25 mg/500 mL hemin for 4 hours at 37°C. Cells were collected by
673 centrifugation at 4,000*g*, immediately followed by purification.
674
675 Human hemoglobin was bought commercially (Sigma-Aldridge) and resuspended in PBS.
676
677 We attempted to co-express and purify Ancα$\beta_{\Delta H3}$ in complex with Ancαβ_{40W}, but we were not
678 able to identify conditions at which the two species could be expressed and purified to near-
679 equal concentrations.
680
681 **Protein purification by ion exchange.** All singly expressed proteins (all ancestral globins
682 except Ancα+Ancβ) were purified using ion exchange chromatography. All buffers were vacuum
683 filtered through a 0.2 μM PFTE membrane (Omnipore). After expression, cells were
684 resuspended in 30 mL of 50 mM Tris-Base (pH 6.88). The resuspended cells were placed in a
685 10 mL falcon tube and lysed using a FB505 sonicator (1s on/off for three cycles, each 1 minute).
686 The lysate was saturated with CO, transferred to a 30 mL round bottom tube, and centrifuged at
687 20,000*g* for 60 minutes to separate supernatant from non-soluble cell debris. The supernatant
688 was collected and syringe-filtered using HPX Millex Durapore filters (Millipore) to further remove
689 debris. A HiTrap SP cation exchange (GE) column was attached to an FPLC system (Biorad)
690 and equilibrated in 50 mM Tris-Base (pH 6.88). The lysate was passed over the column. 50 mL
691 of 50 mM Trise-Base (pH 6.88) was run through the SP column to remove weakly bound non-
692 target soluble products. Elution of bound ancestral Hbs was performed with 100-mL gradient of
693 50mM Tris-Base 1 M NaCl (pH 6.88) buffer which was run through the column from 0% to
694 100%. 1.5 mL fractions were captured during the gradient process, all fractions containing red
695 eluant were put into an Amicon ultra-15 tube and concentrated by centrifugation at 4,000g to a
696 final volume of 1 mL. For additional purification, concentrated sample was injected into a HiPrep
697 16/60 Sephacryl S-100 HR size exclusion chromatography (SEC) column. The column was
698 equilibrated in phosphate buffered saline (PBS) at pH 7.4. Purified ancestral globins elute at
699 different volumes depending on the protein's complex stoichiometry: 48-52 for tetramers, 56-60
700 for dimers, and 65-67 for monomers. The purified proteins were concentrated as mentioned
701 above and then flash frozen with liquid nitrogen.
702
703 For experiments in which two proteins were singly expressed and purified and then mixed
704 together, expression and purification of each protein were performed as described above. The
705 concentration of each protein was then quantified using the Hemoglobin Assay Kit (Sigma).
706 Proteins were then mixed together at 50 μM each for nMS. This procedure was performed in
707 triplicate to assess technical error introduced by the quantification and mixing process.
708
709 **Protein purification by zinc affinity chromatography.** Coexpressed proteins Ancα + Ancβ
710 were purified using zinc-affinity chromatography, which was performed using a HisTrap metal
711 affinity column (GE) on a Biorad NGC Quest. Nickle ions were stripped from the column (buffer
712 100 mM EDTA, 100 mM NaCl, 20 mM TRIS, pH 8.0), followed by five column volumes of water.
713 To attach zinc to the column, 0.1 M $ZnSO_4$ was passed over until conductance was stable,
714 approximately 5 column volumes, followed by five column volumes of water. After expression,

715 cells were resuspended in a 50 mL lysis buffer (20 mM Tris, 150 mM Nacl, 10% glycerol (v/v),
716 1mM BME, 0.05% Tween-20, and 1 Roche Protease EDTA-free inhibitor tablet, pH 7.40),
717 sonicated as described above, and the lysate passed through the prepared column. To remove
718 non-specifically bound protein, the column was washed with 50 mL of lysis buffer. Bound protein
719 was then eluted across a gradient of imidazole concentrations (0 to 500 mM) in a total of 100
720 mL elution buffer (20 mM Tris, 150 mM NaCl, 500 mM imidazole, 10% glycerol, and 1 mM BME,
721 pH 7.4). 1 mL fractions were collected. The fraction corresponding to the second peak of UV
722 absorbance at 280 nm has a visible red color and was collected and concentrated as described
723 above. The concentrated solution was injected into a Biorad ENrich 650 10 x 300 columns for
724 additional purification and eluted in PBS buffer.
725
726 **Size exclusion chromatography assay.** For protein concentrations from 0 to 500 μM, size
727 exclusion chromatography was performed using a Superdex 75 increase 10/300 GL column
728 (GE) equilibrated in PBS, then injected with 250 μL of sample using a 2 mL injection loop on an
729 Biorad NGC Quest FPLC and monitored by absorbance at 280 nm. For proteins at
730 concentration 1 mM, a HiPrep 16/60 Sephacryl S-100 HR was equilibrated in PBS using an
731 AKTAprime FPLC, then injected with 1mL sample and monitored by absorbance at 280 nm.
732
733 **Native Mass Spectrometry.** Protein samples were buffer exchanged into 200mM ammonium
734 acetate using either a centrifugal buffer exchange device (Micro Bio-Spin P-6 Gel, Bio-Rad) or a
735 dialysis device (Slide-A-Lyzer MINI Dialysis Unit, 10000 MWCO, Thermo) prior to native MS
736 experiments. Samples were loaded into gold-coated glass capillaries made in-house and
737 introduced to Synapt G1 HDMS instrument (Waters corporation) equipped with a 32k RF
738 generator (30). The instrument was set to a source pressure of 5.47 mbar, capillary voltage of
739 1.75 kV, sampling cone voltage of 20 V, extractor cone voltage of 5.0 V, trap collision voltage of
740 10 V, collision gas (Argon) flow rate of 2 mL/min (2.65 x 10 -2 mbar), and T-wave settings
741 (velocity/height) for trap, IMS and transfer of 100 ms -1 /0.2 V, 300 ms -1 /16.0 V, and 100 ms -1
742 /10.0 V, respectively. The source temperature (70 °C) and trap bias (30 V) were optimized. Part
743 of the native MS experiments were conducted by Thermo Scientific Exactive Plus Orbitrap with
744 Extended Mass Range (EMR) with tuning as follow: source DC offset of 15 V, injection flatapole
745 DC to 13 V, inter flatapole lens to 5, bent flatapole DC to 4, transfer multipole DC to 3 and C
746 trap entrance lens to 0, trapping gas pressure to 5.0 with the CE to 10, spray voltage to 1.50 kV,
747 capillary temperature to 100 °C, maximum inject time to 100 ms. Mass spectra were acquired
748 with a setting of 8750 resolution, microscans set to 1 and averaging set to 100. Mass spectra
749 were deconvoluted using Unidec (65).
750
751 **Calculating multimerization affinity of homdimers.** To estimate Kd of the monomer-to-
752 homodimer transition of singly expressed proteins, we performed nMS at variable protein
753 concentrations (P_{tot}). The occupancy of each oligomeric state at each concentration was
754 calculated as the proportion of all globin subunits in that state, based on the summed areas
755 under the corresponding peaks in the native MS spectrum. The fraction of subunits assembled
756 into dimers (*Fd*) includes dimers and tetramers and is defined as
757
758 $$Fd = \frac{2x_d+4x_t}{(x_m + 2x_d+4x_t)},$$

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:

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

As an alternative model of homodimerization, we also used a version of the Hill-Langmuir equation:

$$F_d = \frac{P_{tot}}{P_{tot} + K_d}$$

The Hill-Langmuir model, which is typically used for ligand-receptor binding, does not account for depletion of free monomeric subunits as homodimerization takes place and is therefore not a valid model in this case; we used it solely to determine the robustness of the estimated Kd to the specific binding equation used.

Calculating multimerization affinity of homtetramers. To estimate the Kd of the homodimer–homotetramer 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:

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

Parameters were estimated using the curvefit script in the Scipy package (66). The 95% confidence interval on the K_d was estimated as 1.96 times the estimated standard error.

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 μM and the other was added at variable concentration (1 to 50 μM). 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 <100μM. 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 μM. The minimum detection limit for dimers in the nMS assay is 1 μM. Kd is

defined as $Kd = \frac{[M]^2}{[D]}$, where $[M]$ and $[D]$ are the equilibrium 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:

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

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

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

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 (67-69).

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 (70). 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). For PyMol PSE file containing these models, see Supplementary file.

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

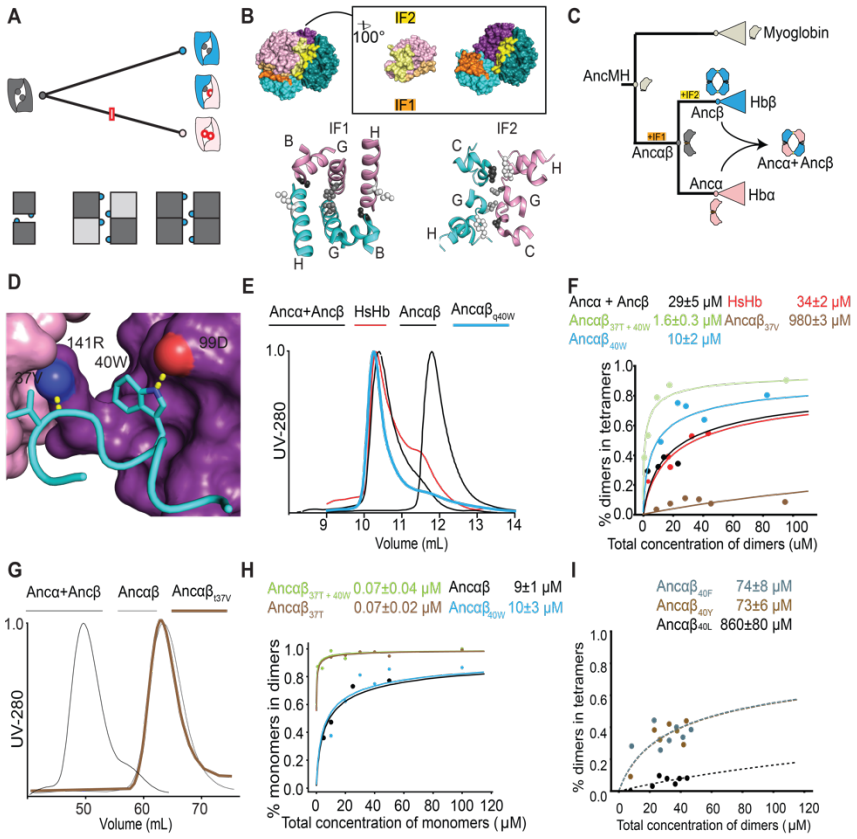
1. D. S. Goodsell, A. J. Olson, Structural symmetry and protein function. *Annual Review of Biophysics and Biomolecular Structure* **29**, 105-153 (2000).
2. J. A. Marsh, S. A. Teichmann, Structure, dynamics, assembly, and evolution of protein complexes. *Annu Rev Biochem* **84**, 551-575 (2015).
3. M. Lynch, The Evolution of multimeric protein assemblages. *Molecular Biology and Evolution* **29**, 1353-13 (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. S. Mallik, D. S. Tawfik, Determining the interaction status and evolutionary fate of duplicated homomeric proteins. *PLOS Computational Biology* **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** (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. 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).
19. 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).
20. 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).
21. D. Ding *et al.*, Co-evolution of interacting proteins through non-contacting and non-specific mutations. *Nat Ecol Evol* **6**, 590-603 (2022).

22. 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).
23. C. J. McClune, A. Alvarez-Buylla, C. A. Voigt, M. T. Laub, Engineering orthogonal signalling pathways
24. C. D. Aakre *et al.*, Evolving new protein-protein interaction specificity through promiscuous intermediates. *Cell* **163**, 594-606 (2015).
25. T. V. Lite *et al.*, Uncovering the basis of protein-protein interaction specificity with a combinatorially complete library. *Elife* **9** (2020).
26. I. Nocedal, M. T. Laub, Ancestral reconstruction of duplicated signaling proteins reveals the evolution of signaling specificity. *eLife* **11**, e77346 (2022).
27. 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).
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? *Journal of the American Society for Mass Spectrometry* **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. *Journal of the American Society for Mass Spectrometry* **18**, 8-16 (2007).
34. G. Snyder, B. Sheafor, Red blood ells: centerpiece in the evolution of the vertebrate circulatory system. *Am. Zool.* **39** (1999).
35. G. Fermi, M. F. Perutz, B. Shaanan, R. Fourme, The crystal structure of human deoxyhaemoglobin at 1.74 Å resolution. *Journal of Molecular Biology* **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: Structure, Function, and Bioinformatics* **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. *Proceedings of the National Academy of Sciences* **119**, e2101117119 (2022).

990 39. C.-S. Chen *et al.*, How to change the oligomeric state of a circular protein assembly:
991 switch from 11-subunit to 12-subunit TRAP suggests a general mechanism. *PLOS ONE*
992 **6**, e25296 (2011).
993 40. B. Kuhlman, J. W. O'Neill, D. E. Kim, K. Y. Zhang, D. Baker, Conversion of monomeric
994 protein L to an obligate dimer by computational protein design. *Proc Natl Acad Sci U S A*
995 **98**, 10687-10691 (2001).
996 41. H. Schweke *et al.*, An atlas of protein homo-oligomerization across domains of life. *Cell*
997 **187**, 999-1010.e1015 (2024).
998 42. J. A. Marsh *et al.*, Protein complexes are under evolutionary selection to assemble via
999 ordered pathways. *Cell* **153**, 461-470 (2013).
1000 43. E. D. Levy, E. B. Erba, C. V. Robinson, S. A. Teichmann, Assembly reflects evolution of
1001 protein complexes. *Nature* **453**, 1262-1265 (2008).
1002 44. R. P. Bahadur, F. Rodier, J. Janin, A dissection of the protein–protein interfaces in
1003 icosahedral virus capsids. *Journal of Molecular Biology* **367**, 574-590 (2007).
1004 45. E. T. Powers, D. L. Powers, A perspective on mechanisms of protein tetramer formation.
1005 *Biophysical Journal* **85**, 3587-3599 (2003).
1006 46. J. T. Brennecke, B. L. de Groot, Quantifying asymmetry of multimeric proteins. *The*
1007 *Journal of Physical Chemistry A* **122**, 7924-7930 (2018).
1008 47. M. Bonjack-Shterengartz, D. Avnir, The near-symmetry of proteins. *Proteins: Structure,*
1009 *Function, and Bioinformatics* **83**, 722-734 (2015).
1010 48. M. F. Perutz, G. Fermi, B. Luisi, B. Shaanan, R. C. Liddington, Stereochemistry of
1011 cooperative mechanisms in hemoglobin. *Accounts of Chemical Research* **20**, 309-321
1012 (1987).
1013 49. C. R. Baker, V. Hanson-Smith, A. D. Johnson, Following gene duplication, paralog
1014 interference constrains transcriptional circuit evolution. *Science* **342**, 104-108 (2013).
1015 50. J. T. Brigham, J. E. Brown, A. Rodríguez-Marí, J. M. Catchen, J. W. Thornton, Evolution
1016 of a new Function by degenerative mutation in cephalochordate steroid receptors. *PLOS*
1017 *Genetics* **4**, e1000191 (2008).
1018 51. S. Liu et al., Nonnatural protein–protein interaction-pair design by key residues grafting.
1019 *Proceedings of the National Academy of Sciences* **104**, 5330-5335 (2007).
1020 52. G. Guntas, C. Purbeck, B. Kuhlman, Engineering a protein–protein interface using a
1021 computationally designed library. *Proceedings of the National Academy of Sciences* **107**,
1022 19296-19301 (2010).
1023 53. D. P. Anderson et al., Evolution of an ancient protein function involved in organized
1024 multicellularity in animals. *Elife* **5**, e10147 (2016).
1025 54. L. Schulz et al., Evolution of increased complexity and specificity at the dawn of form I
1026 Rubiscos. *Science* **378**, 155-160 (2022).
1027 55. F. L. Sendker et al., Emergence of fractal geometries in the evolution of a metabolic
1028 enzyme. *Nature* **628**, 894-900 (2024).
1029 56. G. K. A. Hochberg *et al.*, A hydrophobic ratchet entrenches molecular complexes.
1030 *Nature* **588**, 503-508 (2020).
1031 57. D. Posada, T. R. Buckley, Model Selection and Model Averaging in Phylogenetics:
1032 Advantages of Akaike Information Criterion and Bayesian Approaches Over Likelihood
1033 Ratio Tests. *Systematic Biology* **53**, 793-808 (2004).

1034 58. S. Q. Le, O. Gascuel, An improved general amino acid replacement matrix. *Mol Biol Evol*
1035 **25**, 1307-1320 (2008).
1036 59. C. Natarajan *et al.*, Expression and purification of recombinant hemoglobin in
1037 Escherichia coli. *PLoS One* **6**, e20176 (2011).
1038 60. M. T. Marty *et al.*, Bayesian deconvolution of mass and ion mobility spectra: from binary
1039 interactions to polydisperse ensembles. *Analytical Chemistry* **87**, 4370-4376 (2015).
1040 61. L. G. Nivón, R. Moretti, D. Baker, A Pareto-Optimal Refinement Method for Protein
1041 Design Scaffolds. *PLOS ONE* **8**, e59004 (2013).
1042 62. Z. Yang, PAML: a program package for phylogenetic analysis by maximum likelihood.
1043 *Comput Appl Biosci* **13**, 555-556 (1997).
1044 63. J. Yang, Z. Zhang, X. A. Zhang, Q. Luo, A ligation-independent cloning method using
1045 nicking DNA endonuclease. *Biotechniques* **49**, 817-821 (2010).
1046 64. C. Natarajan *et al.*, Expression and purification of recombinant hemoglobin in
1047 Escherichia coli. *PLoS One* **6**, e20176 (2011).
1048 65. M. T. Marty *et al.*, Bayesian deconvolution of mass and ion mobility spectra: from binary
1049 interactions to polydisperse ensembles. *Analytical Chemistry* **87**, 4370-4376 (2015).
1050 66. P. Virtanen *et al.*, SciPy 1.0: fundamental algorithms for scientific computing in Python.
1051 *Nature Methods* **17**, 261-272 (2020).
1052 67. P. J. Carter, G. Winter, A. J. Wilkinson, A. R. Fersht, The use of double mutants to
1053 detect structural changes in the active site of the tyrosyl-tRNA synthetase (Bacillus
1054 stearothermophilus). *Cell* **38**, 835-840 (1984).
1055 68. A. Horovitz, Double-mutant cycles: a powerful tool for analyzing protein structure and
1056 function. *Folding and Design* **1**, R121-R126 (1996).
1057 69. J. A. Wells, Additivity of mutational effects in proteins. *Biochemistry* **29**, 8509-8517
1058 (1990).
1059 70. L. G. Nivón, R. Moretti, D. Baker, A Pareto-Optimal Refinement Method for Protein
1060 Design Scaffolds. *PLOS ONE* **8**, e59004 (2013).
1061

1062 **FIGURES**

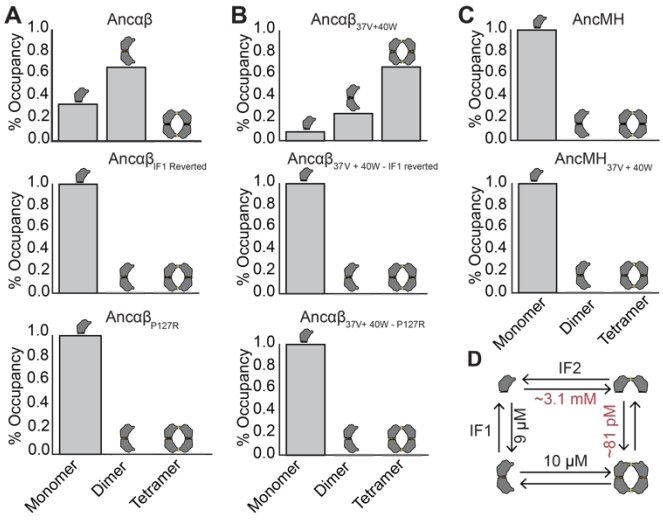


1063

1064 **Figure 1. A single substitution confers tetramerization on an ancestral dimer. (A)** A
1065 substitution in one subunit can potentially affect specificity and stoichiometry in an isologous
1066 interface. *Top*: After duplication of an isologous homodimer (gray), a substitution that occurs in
1067 one paralog (red box) appears twice in the interface of a homodimer (red circles), once in a
1068 heterodimer, and not at all in the other homodimer (blue). *Bottom*: One substitution (blue circle)
1069 in an isologous interface appears twice in a homodimer (*left*), twice in a heterotetramer (*middle*),
1070 and four times in a homotetramer (*right*), multiplying its effects on affinity. Dark and light gray,
1071 paralogous subunits. **(B)** *Top*: Interfaces in the human Hb heterotetramer (PDB 4HHB). Pink,
1072 Hbα; blue, Hbβ; α_1 and β_1 are in lighter hues than α_2 and β_2. IF1 surfaces (orange) mediate α_1-
1073 β_1 and α_2-β_2 interactions; yellow surfaces (IF2) mediate α_1-β_2 and α_2-β_1 interactions. Only
1074 interfaces involving α_1 are shown. Inset, α_1 subunit rotated away from the rest of the tetramer to
1075 show IF1 and IF2. *Bottom*: Isology of IF1 and IF2. Helices contributing to each interface are
1076 shown and labeled. Balls and sticks: on each helix, one residue's side chain is shown to
1077 visualize symmetry. **(C)** Evolution of tetrameric stoichiometry on the phylogeny of Hb and
1078 related globins. Icons, oligomeric states determined by experimental characterization of
1079 reconstructed ancestral proteins (16). Acquisition of interfaces of IF1 and IF2 is shown (16). **(D)**
1080 Key residues V37 and W40 that were substituted in Ancb. Cyan cartoon helix, b_1 subunit. Pink
1081 and violet surfaces, a subunits that interact with b_1 via IF1 and IF2, respectively. Dotted lines to
1082 red or blue spheres, hydrogen bonds to oxygen or nitrogen atoms, respectively (PDB 4HHB).
1083 **(E,G)** Effect of historical substitutions on stoichiometry, as measured by size exclusion

1084 chromatography. The ancestral dimer Ancαβ and the tetramers Ancα+Ancβ and human
1085 hemoglobin (HsHb) are shown for comparison. Protein concentration at 100 mM (E) or 1 mM
1086 (G). **(F)** Dimer-to-tetramer affinity of reconstructed ancestral Hb subunits containing historical
1087 substitutions q40W and t37V, measured by native mass spectrometry across a titration series.
1088 Points, fraction of dimers that are incorporated into tetramers. Lines, best-fit binding curves.
1089 Estimated Kd and 95% confidence interval are shown. **(H)** Effect of historical substitutions on
1090 monomer-dimer affinity measured by native MS. **(I)** Effect on dimer-tetramer affinity of
1091 nonhistorical hydrophobic mutations in at residue 40, measured by native MS.

1092



1093

1094 **Figure 2. Multimerization across IF2 requires IF1.** (A) IF1-mediated dimerization can be
1095 compromised by mutations. Relative occupancy of each stoichiometry as measured by native
1096 MS at at 20 mM total protein is shown for the ancestral dimer Ancαβ (top), Ancαβ$_{\text{IF1 reverted}}$
1097 (middle, a variant of Ancαβ in which all IF1 residues are reverted to the ancestral state found in
1098 AncMH), and Ancαβ-P127R (bottom, in which a mutation known to compromise IF1-mediated
1099 dimerization has been introduced). **(B)** Compromising IF1 prevents assembly across IF2.
1100 Relative occupancy of Ancαβ$_{40W + 37V}$ with and without mutations that compromise IF1-mediated
1101 dimerization. **(C)** AncMH, which does not dimerize across IF1, cannot multimerize across IF2,
1102 even when mutations sufficient to confer IF2-mediated mutimerization in Ancαβ are introduced.
1103 **(D)** Observed (black) and expected (red) affinities of Ancαβ +q40W interfaces. Expected Kd of a
1104 single iteration of IF2 (top) equals the square root of the measured apparent Kd when two
1105 iterations are present (bottom). Expected apparent Kd of two iterations of IF1 (right) equals the
1106 square of the measured Kd of a single IF1 (left).

1107

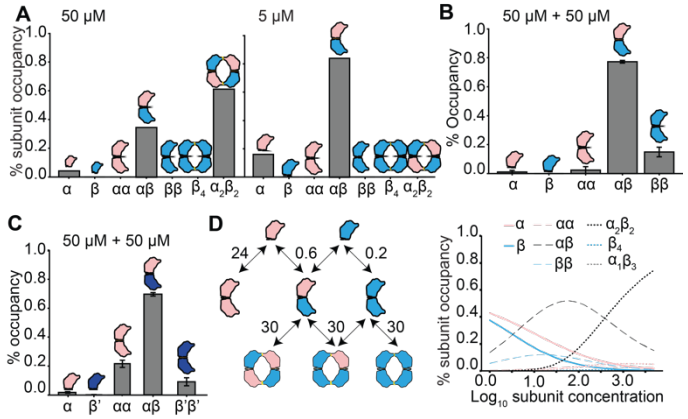


Figure 3. Heterotetramer specificity is conferred by specificity at IF1. (A) Occupancy (as fraction of all Hb subunits) when Ancα +Ancβ are coexpressed, measured by native MS. At 50 uM total protein, heterotetramers and heterodimers predominate (left). At 5 uM (right) – at which assembly occurs only 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 mM 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 uM. **(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, 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.

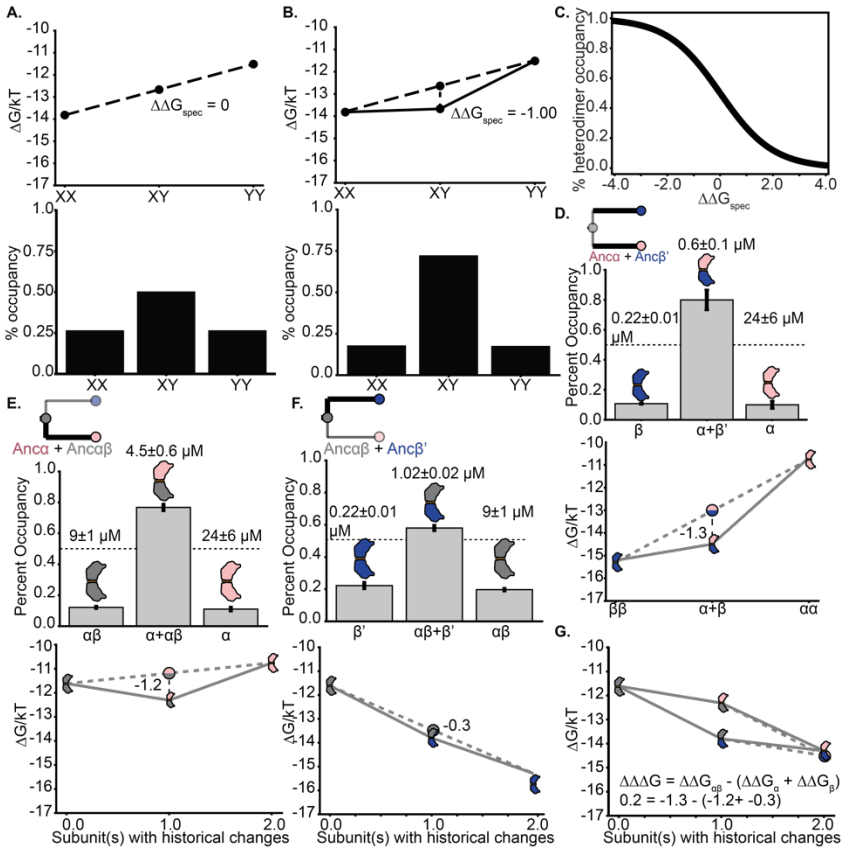
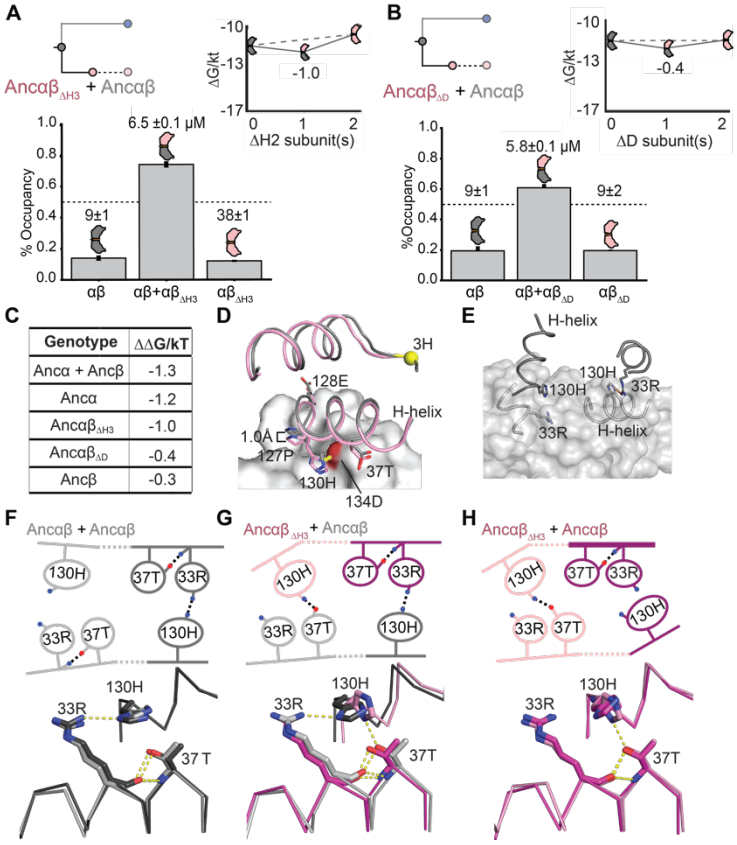


Figure 4. Contribution of historical changes in each subunit to the acquisition of heterospecificity. (A) Theoretical example of affinities and occupancy in a system of dimers with no specificity. *Top*: ΔG of dimerization for homodimers (XX and YY) and heterodimers (XY), in units of kT. In the absence 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 Kds assessed by nMS (shown above each bar, with 95% confidence interval). *Bottom*: ΔGs and ΔΔG given measured Kds. Dotted line, expected occupancies in the absence of specificity. **(E)** Specificity of IF1 acquired on the branch leading from Ancαβ to Ancα, shown as occupancy 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 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	-1.3
Ancα	-1.2
Ancαβ$_{\Delta H3}$	-1.0
Ancαβ$_{\Delta D}$	-0.4
Ancβ	-0.3

Figure 5. Effect of historical sequence changes on specificity. (A) Specificity of Ancαβ$_{\Delta H3}$ with Ancαβ, as in Fig. 4. **(B)** Specificity of Ancαβ$_{\Delta D}$ 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$_{\Delta H3}$ + 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$_{\Delta H3}$ subunit in the heterodimer. Yellow, 3H residue deleted in Ancαβ$_{\Delta H3}$. 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 ΔH3 in the homodimers and heterodimer of Ancαβ and Ancαβ$_{\Delta H3}$. 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.

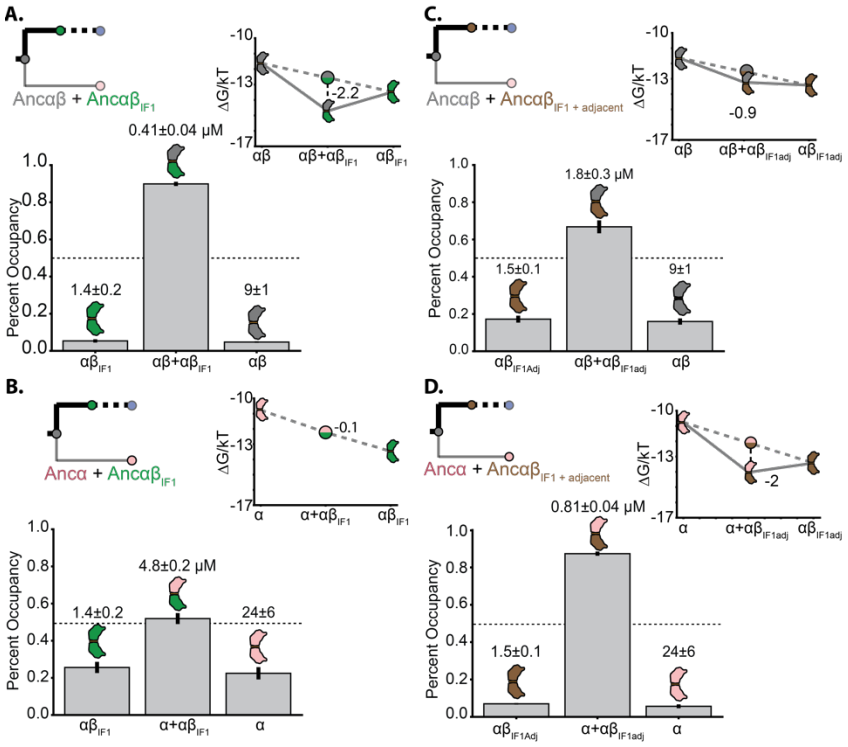


Figure 6. Other subsets of historical substitutions confer heterospecificity on IF1. Affinities measured by nMS, predicted occupancy based on those Kds at 1 mM each subunit, and $\Delta\Delta G_{spec}$ are shown for A) Ancαβ + Ancαβ$_{IF1}$, which contains the five substitutions at the IF1 surface that 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

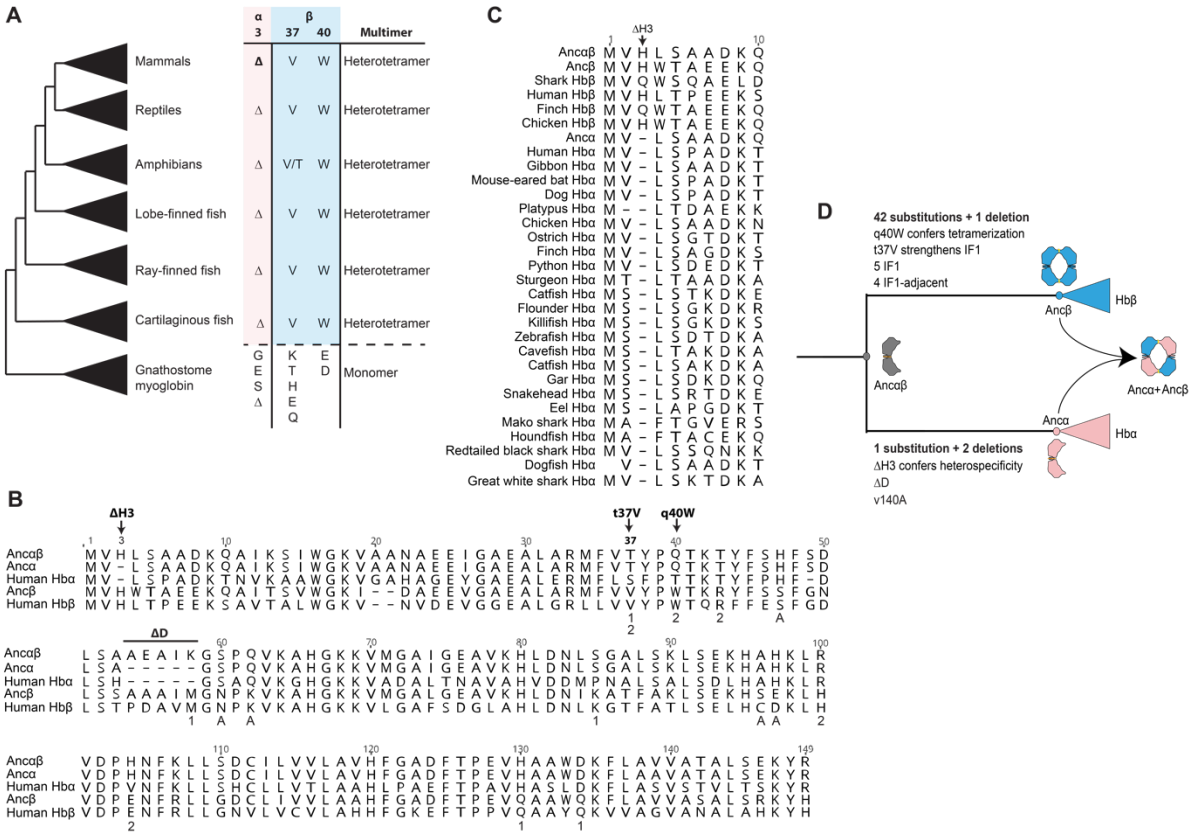
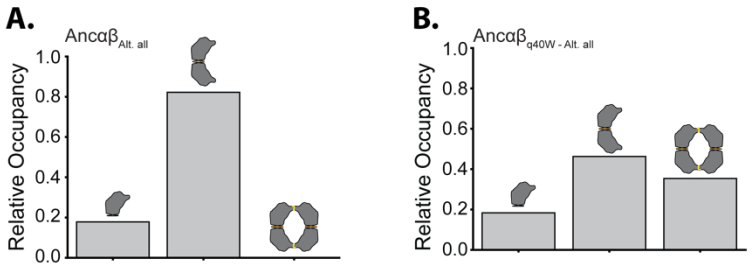


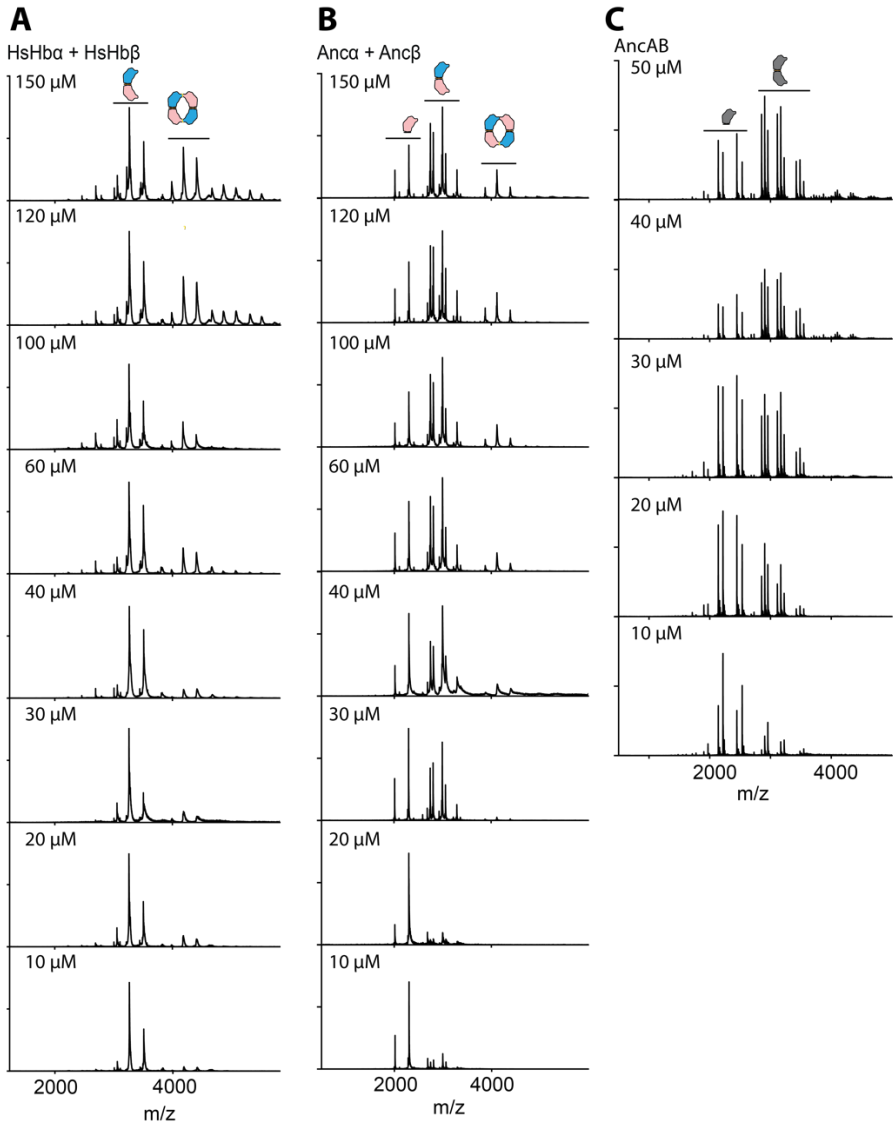
Fig. S1. Key sequence changes are conserved in extant hemoglobin subunits. A) Multimerization state and residues at key sequence sites in extant Hb subunits from major jawed vertebrate taxa. Myoglobin is the closest paralogous protein in the globin family. For complete alignment, see https://tinyurl.com/yc6kvdb8. **B**) Alignment of extant human and reconstructed ancestral Hb subunits is shown. Historical sequence changes that confer heterospecificity (ΔH3 in Hbα) and tetramerization (q40W in Hbβ) are labeled and shown with arrows. Additional substitutions that occurred on the branch leading to Ancβ and assayed in this paper are also labeled, including changes on the surface of IF1 (1), IF2 (2), and at sites adjacent to IF1 (A). **C)** Alignment of N-terminal portion of Hb subunits from species representative of major vertebrate taxa. The deletion of residue 3 in alpha subunits is marked. **D**) Summary of key sequence changes that occurred after the duplication of Ancαβ. Multimerization states of reconstructed ancestral proteins is shown.

1193



1194
1195
1196 **Fig. S2. Effect of q40W tetramerization is robust to statistical uncertainty**. (A) Relative
1197 occupancy of monomer, dimer, and tetramer of Ancαβ$_{Alt.\ all}$, an alternative reconstruction of
1198 Ancαβ that contains the second most likely state at all ambiguously reconstructed sites,
1199 measured at 20 μM total protein using native MS. (B) Relative occupancy Ancαβ$_{Alt.\ all}$ with
1200 substitution q40W.
1201

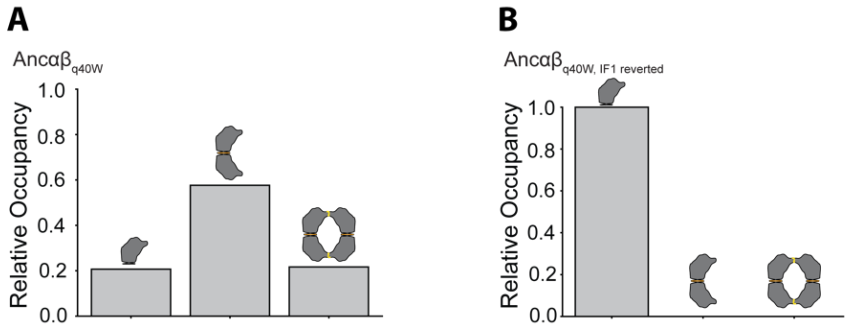
1202



1203

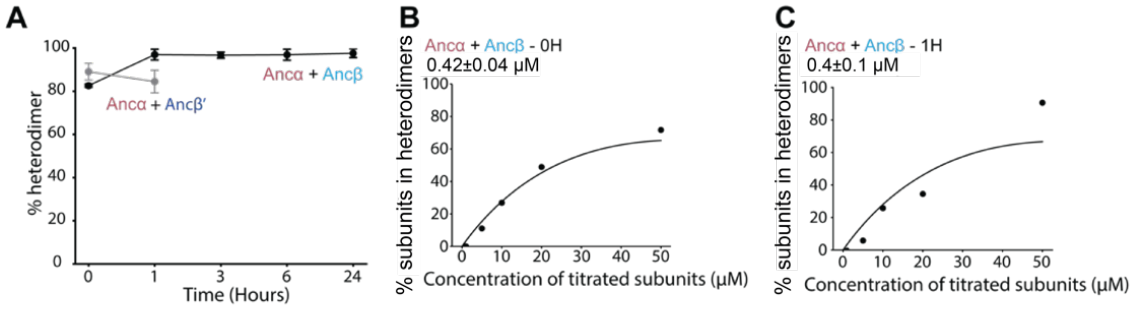
1204 **Fig. S3. Native mass spectrometry spectra.** nMS spectra across a concentration series is
1205 shown for A) human Hb, B) Ancα + Ancβ, and (C) Ancαβ. Peaks corresponding to monomers,
1206 dimers, and tetramer are labeled.

1207



1208
1209
1210
1211 **Fig. S4. The effect of q40W on tetramerization depends on IF1.** (A) Relative occupancy of
1212 Ancαβ$_{q40W}$, measured by native MS at 20 μM total protein. (B) Relative occupancy of
1213 Ancαβ$_{q40W_IF1\text{-}reverted}$, which contains mutation q40W, as well as reversions to the ancestral state
1214 found in AncMH of all residues that were substituted between AncMH and Ancαβ .
1215

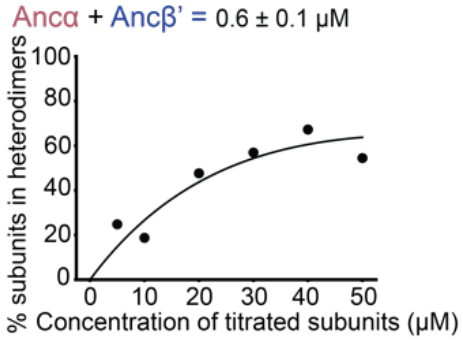
1216



1217
1218

1219 **Fig. S5. Heterodimer occupancy of Ancα and Ancβ is near equilibrium after mixing.** (A)
1220 The percent of all dimers that are heterodimers, measured by nMS when proteins are mixed at
1221 50 μM each and allowed to incubate for 0, 1, 3, 6, or 24 hours. Black line and points, Ancα +
1222 Ancβ (which only dimerize when expressed separately and then mixed). Grey line and points,
1223 Ancα + Ancβ' (Ancβ in which IF2 surface substitutions are reverted to their ancestral state in
1224 Ancαβ, thus preventing tetramerization). Each dot shows the mean of three replicates; error
1225 bars, SEM. (B) Affinity of monomer-to-heterodimer assembly measured by nMS immediately
1226 upon mixing of Ancα and Ancβ. Ancα was kept constant at 50 μM, while the concentration of
1227 Ancβ varied. Points, fraction of all subunits in the mixture that are incorporated into
1228 heterodimers. Line, best-fit binding curve. Estimated Kd and 95% confidence interval are
1229 shown. (C) Estimated heterodimerization affinity measured as in panel B, but 1 hour after
1230 mixing.
1231

1232



1233
1234
1235 **Fig. S6. Heterodimerization by Ancα+Ancβ'.** Monomer-to-heterodimer assembly measured
1236 by nMS. Ancα was kept constant at 50 μM while Ancβ' was at variable concentration. Points,
1237 fraction of all subunits in the mixture that are incorporated into heterodimers at each
1238 concentration. Line, best-fit binding curve. Estimated Kd and 95% confidence interval are
1239 shown.
1240

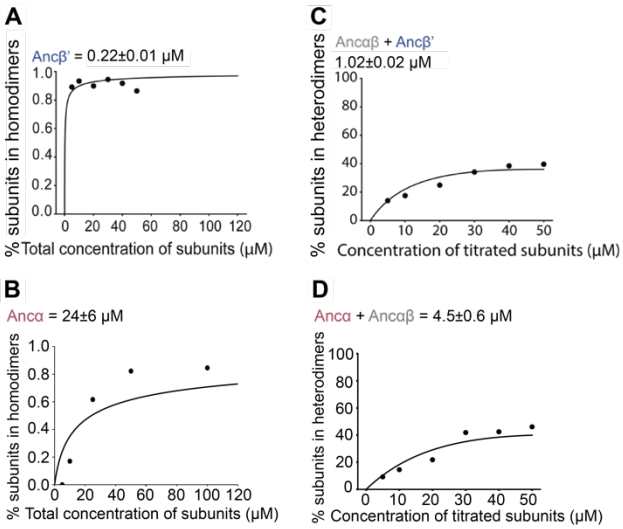
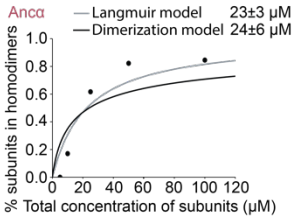


Fig. S7. Dimerization by Ancα and Ancβ' (A-B) Homodimerization by Ancβ' (panel A) and by Ancα (B). measured by nMS across a titration series. Each point shows the fraction of subunits incorporated into dimers as the concentration of protein varied. Best-fit binding curve, Kd, and 95% confidence interval 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 μM while the other was varied.

1250



1251
1252

1253 **Fig. S8. Estimated affinity of the monomer-dimer transition by the Ancα homodimer is**
1254 **robust to the binding model used.** We fit two different binding equations (curves) to the
1255 stoichiometries of **Ancα** measured using nMS. The Langmuir-Hill model (black line) assumes
1256 that the concentration of free monomeric subunits is not depleted by dimerization. The
1257 dimerization model used throughout the rest of this paper accounts for this depletion (described
1258 in Materials and Methods). The estimated Kds (shown with their 95% confidence intervals) are
1259 statistically indistinguishable.

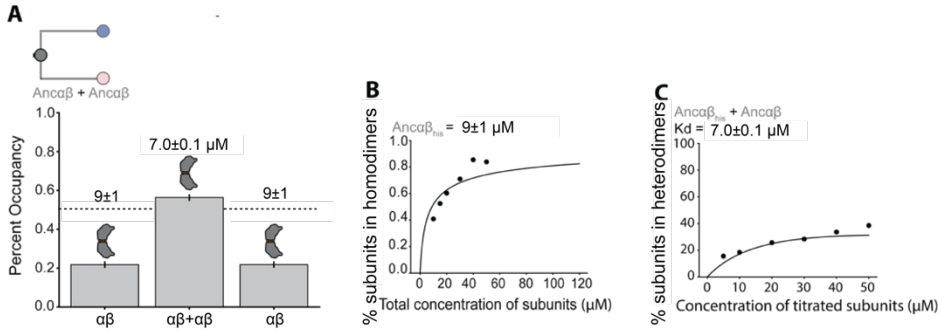


Fig. S9. Dimerization affinities and occupancy of Ancαβ. (A) Expected fractional occupancies of homodimer and heterodimers when Ancαβ is mixed at equal concentrations with Ancαβ$_{his}$ (500 μM each), given the measured dimerization affinities (shown above each column, with 95% confidence interval). 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.

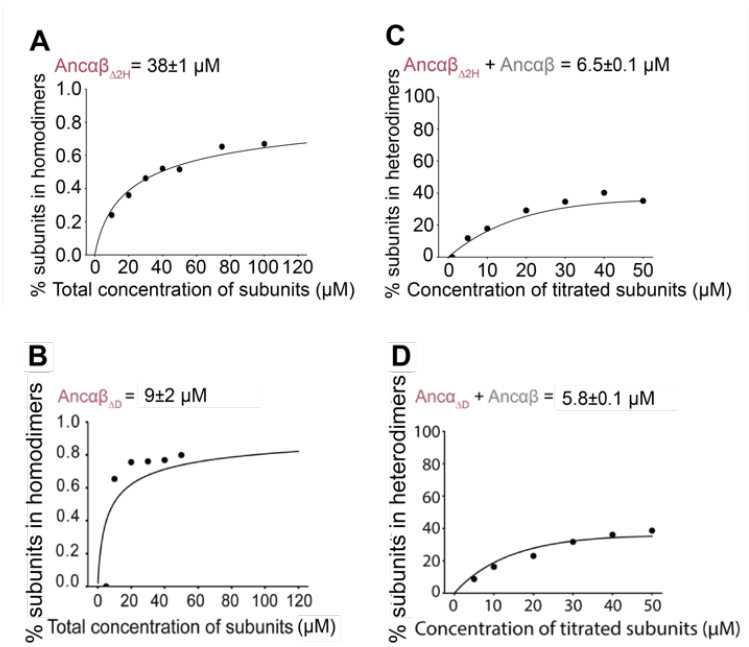


Fig. S10. Effect of historical deletions on dimerization. (A-B) Homodimerization and (C-D) Heterodimerization by mixtures, measured and represented as in Fig. S5.

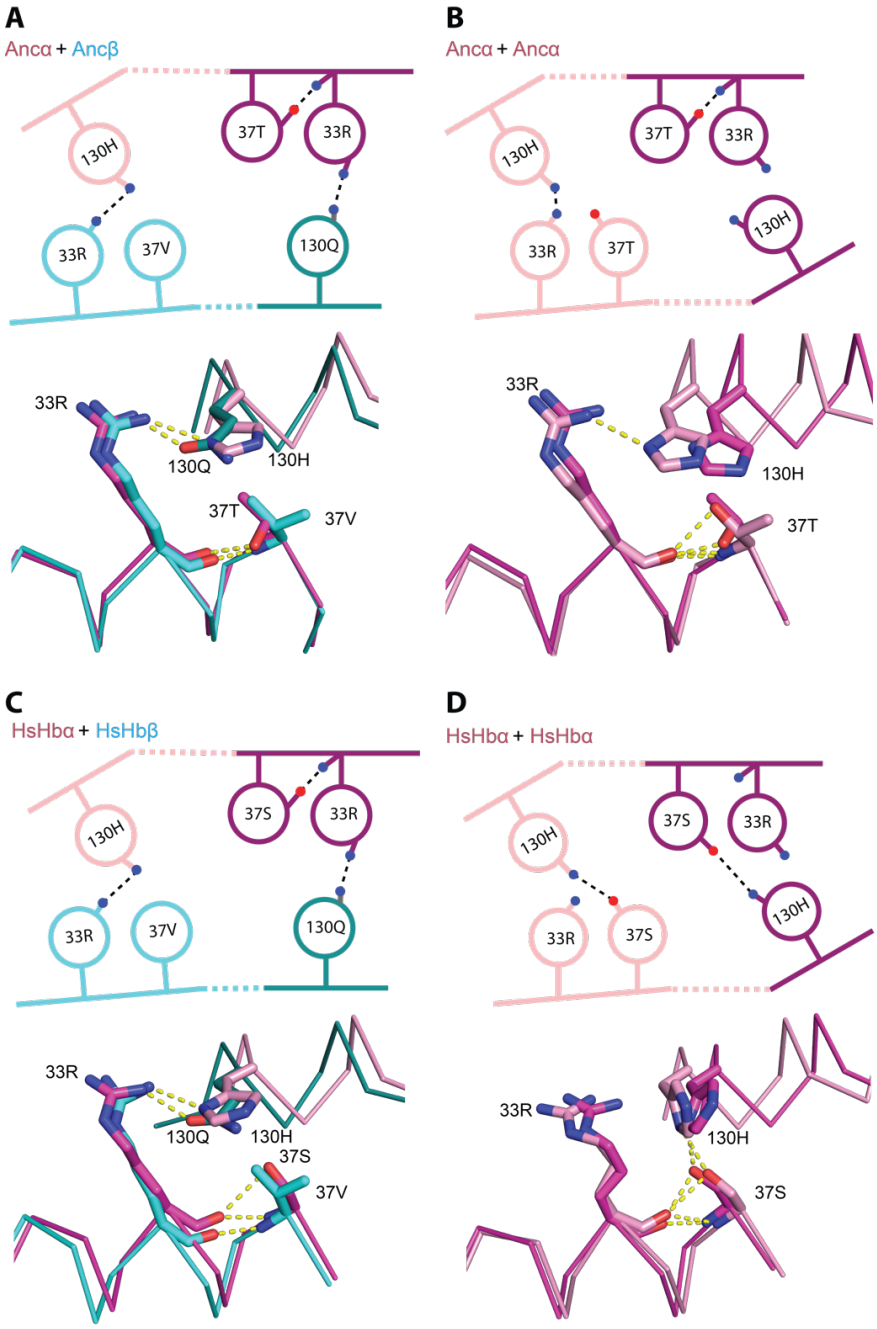


Fig. S11. Nonadditive interactions that contribute to specificity are conserved in derived Hb complexes. In the modeled homodimers and heterodimers of Ancα+Ancβ (panels A, B) and X-ray crystal structure of human hemoglobin (PDB 4HHB and 3S48), the figure shows the key IF1 residues with nonadditive interactions in Ancαβ+Ancαβ$_{\Delta H3}$ (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

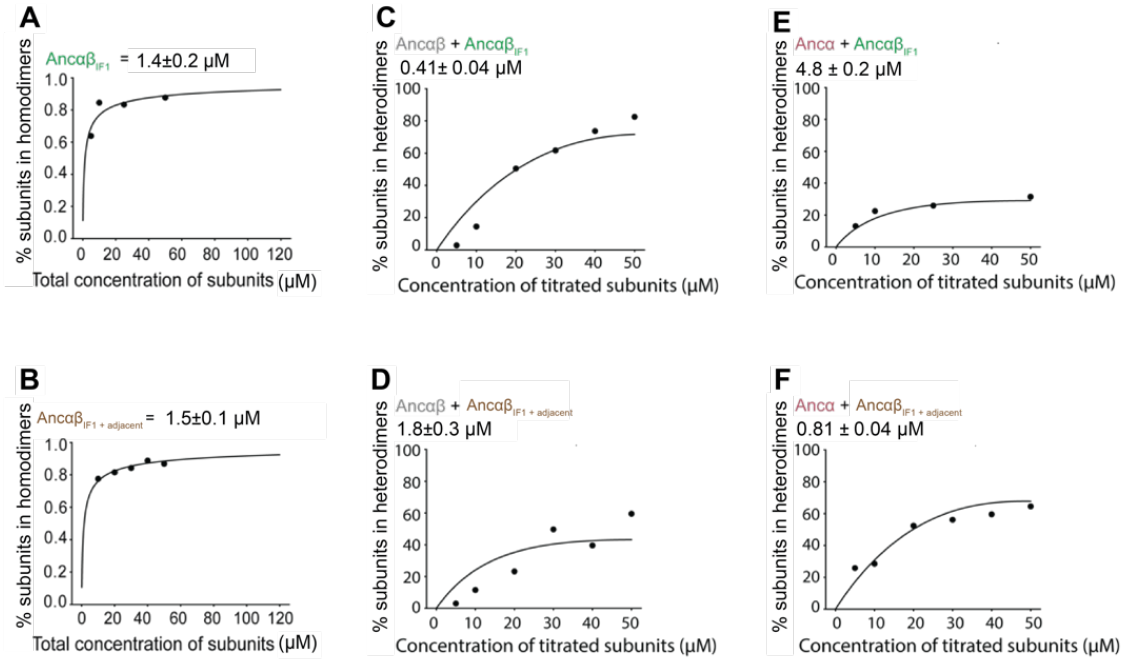


Fig. S12. Homodimerization by Ancαβ$_{IF1}$ and Ancαβ$_{IF1 + Adjacent}$ (A,B) and heterodimerization by those proteins when mixed with Ancαβ (C,D) or Ancα (E,F). Measurements and representation as in Fig. S5.

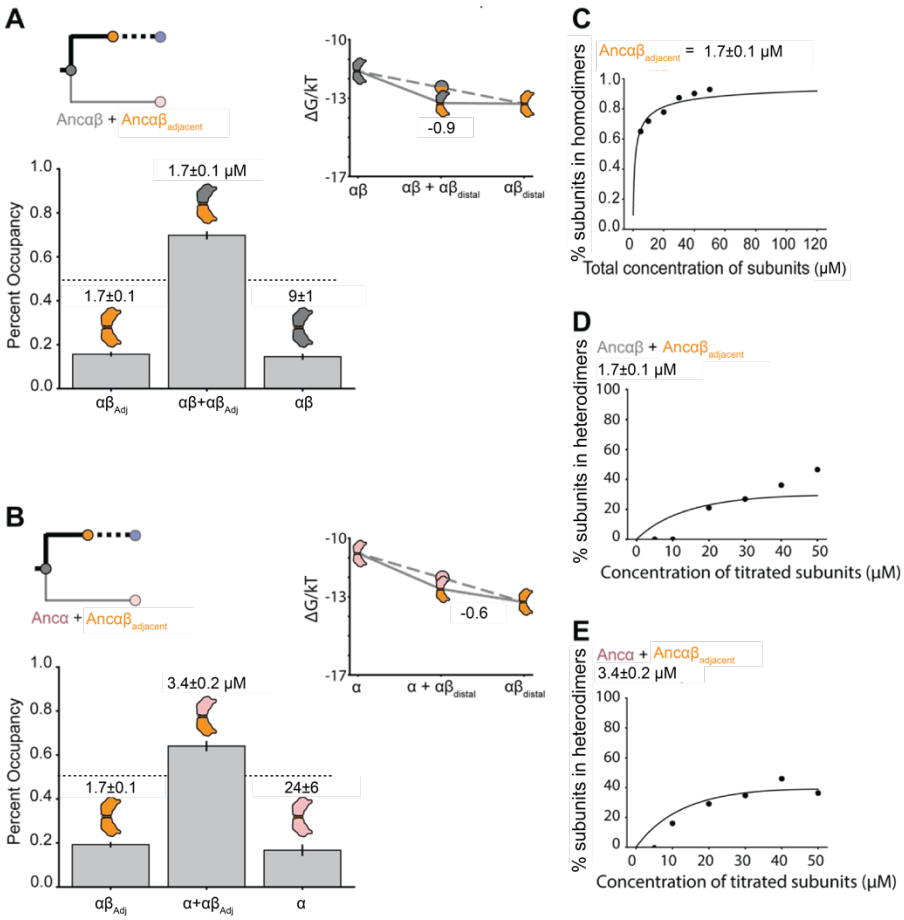


Fig. S13. Dimerization affinity and occupancies for Ancαβ$_{Adjacent}$. Expected fractional occupancies of homodimer and heterodimers when Ancαβ$_{Adjacent}$ Is mixed with Ancαβ (A) or Ancα (B), each at (500 μM), given the measured dimerization affinities (shown above each column, with 95% confidence 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.