Pervasive contingency and entrenchment in a billion years of Hsp90 evolution

Tyler N. Starr[a,1], Julia M. Flynn[b,1], Parul Mishra[b,1,2], Daniel N. A. Bolon[b,3], Joseph W. Thornton[c,3]

[a] Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL 60637, USA

[b] Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605, USA

[c] Departments of Ecology & Evolution and Human Genetics, University of Chicago, Chicago, IL 60637, USA

[1] co-first authors

[2] Present address: Department of Animal Biology, School of Life Sciences, University of Hyderabad, Hyderabad 500046, India

[3] co-senior authors

Editorial correspondence: Joseph W. Thornton, joet1@uchicago.edu, 1-773-834-3423

Classification: Biological Sciences, Evolution

Keywords: epistasis; ancestral protein reconstruction; molecular evolution; protein evolution

Abstract

Interactions among mutations within a protein have the potential to make molecular evolution contingent and irreversible, but the extent to which epistasis actually shaped historical evolutionary trajectories is unclear. We addressed this question by identifying all amino acid substitutions that occurred during the billion-year evolutionary history of the heat shock protein 90 (Hsp90) ATPase domain beginning from a deep eukaryotic ancestor to modern *Saccharomyces cerevisiae* and then precisely measuring their fitness effects when introduced into both extant and reconstructed ancestral Hsp90 proteins. We find a pervasive influence of epistasis: of 98 derived states that evolved during history, most were deleterious at times before they happened, and the vast majority also became subsequently entrenched, with the ancestral state becoming deleterious after its substitution. This epistasis was primarily caused by specific interactions among sites rather than a general permissive or restrictive effect on the protein's tolerance to mutation. Our results show that epistasis continually opens and closes windows of mutational opportunity over evolutionary timescales, producing histories and biological states that reflect the transient internal constraints imposed by a protein's fleeting sequence states.

16 **Significance statement**

17 When mutations within a protein change each other's functional effects—a phenomenon
18 called epistasis—the trajectories available to evolution at any moment in time depend on the
19 specific set of changes that previously occurred in the protein. The extent to which epistasis has
20 shaped historical evolutionary trajectories is unknown. Using a high-precision bulk fitness assay
21 and ancestral protein reconstruction, we measured the fitness effects in ancestral and extant
22 sequences of all historical substitutions that occurred during the billion-year trajectory of an
23 essential protein. We found that most historical substitutions were contingent on prior epistatic
24 substitutions and/or entrenched by subsequent changes. These results establish that epistasis
25 caused widespread, consequential shifts in the site-specific fitness constraints that shaped the
26 protein's historical trajectory.

27

28 **Main text**

29 Epistatic interactions can, in principle, affect the sequence changes that accumulate
30 during evolution. A deleterious mutation's expected fate is to be purged by purifying selection,
31 but it can be fixed if a permissive substitution renders it neutral or beneficial (1-3). Conversely, a
32 neutral mutation – which by definition is initially reversible to the ancestral state without fitness
33 cost – may become entrenched by a subsequent restrictive substitution that renders the ancestral
34 state deleterious (1, 4, 5); reversal of the entrenched mutation would then be unlikely unless the
35 restrictive substitution were itself reversed or another permissive substitution occurred.

36 The extent to which epistasis-induced contingency and entrenchment actually affected
37 protein sequence evolution remains unclear, however, because there is no consensus on the
38 prevalence, effect size, or mechanisms of epistasis among historical substitutions. Deep

39 mutational scans have revealed frequent epistasis among the many possible mutations within
40 proteins (6-10), but how these interactions affect the substitutions that actually occurred during
41 historical evolution is not known. Historical case studies have shown that particular substitutions
42 were contingent (3, 11-13) or became entrenched during evolution (5), but whether these are
43 examples of a general phenomenon is unknown. Computational approaches suggest pervasive
44 contingency and entrenchment among substitutions (1, 4, 14-18), but some of these analyses rely
45 on models of uncertain adequacy (19-21), and their claims have not been experimentally
46 validated. Swapping sequence states among extant orthologs reveals frequent epistasis among
47 substitutions (22), but this "horizontal" approach, unpolarized with respect to time, leaves
48 unresolved whether permissive or restrictive interactions are at play (23). Some experimental
49 studies have systematically examined epistasis among substitutions in an historical context, but
50 most have measured effects on protein function (2, 22) or stability (20, 24), leaving unexamined
51 the prevalence of epistasis with respect to fitness – the phenotype that directly affects
52 evolutionary fate. Others have focused on fitness but used methods that cannot detect effects of
53 relatively small magnitude, which could be both widespread and consequential for evolutionary
54 processes (2, 25).

55 We directly evaluated the roles of contingency and entrenchment on historical sequence
56 evolution by precisely quantifying changes over time in the fitness effects of all substitutions that
57 accumulated during the long-term evolution of heat shock protein 90 (Hsp90) from a deep
58 eukaryotic ancestor to *S. cerevisiae*. Hsp90 is an essential molecular chaperone that facilitates
59 folding and regulation of substrate proteins through an ATP-dependent cycle of conformational
60 changes, modulated by co-chaperone proteins. Orthologs from other fungi, animals, and protists
61 can complement Hsp90 deletion in *S. cerevisiae* (26, 27), indicating that the protein's essential

62 molecular function is conserved over large evolutionary distances. To quantify the context-
63 dependence of historical sequence changes, we used a sensitive deep sequencing-based bulk
64 fitness assay (28) to characterize protein libraries in which each ancestral amino acid is
65 reintroduced into an extant Hsp90 and each derived state is introduced into a reconstructed
66 ancestral Hsp90. We focused our experiments on the N-terminal domain (NTD) of Hsp90, which
67 mediates ATP-dependent conformational changes.
68
69 **Results**
70 ***The historical trajectory of Hsp90 sequence evolution.*** We inferred the maximum
71 likelihood phylogeny of Hsp90 protein sequences from 261 species of Amorphea (the clade
72 comprising Fungi, Metazoa, Amoebozoa, and related lineages (29)), rooted using green algae
73 and plants as an outgroup (Fig. 1a, Fig. S1, Datasets S1, S2, S3). We reconstructed ancestral
74 NTD sequences at all nodes along the trajectory from the common ancestor of Amorphea
75 (ancAmoHsp90) to extant *S. cerevisiae* (ScHsp90) and identified substitutions as differences
76 between the most probable reconstructions at successive nodes (Dataset S4).
77 Along this entire trajectory, substitutions occurred at 72 of the 221 sites in the NTD;
78 because of multiple substitutions, 98 unique ancestral amino acid states existed at these sites at
79 some point in the past and have since been replaced by the ScHsp90 state. The vast majority of
80 these 98 ancestral states are reconstructed with high confidence (posterior probability >0.95) in
81 one or more ancestors along the trajectory (Fig. 1b), and every ancestral sequence has a mean
82 posterior probability across sites of >0.95 (Fig. S2a-c).
83 ***Entrenchment and irreversibility.*** To measure the fitness effects of ancestral amino acids
84 when they are re-introduced into an extant Hsp90, we created a library of ScHsp90 NTD

85 variants, each of which contains one of the 98 ancestral states. We determined the per-generation
86 selection coefficient (*s*) of each mutation to an ancestral state relative to ScHsp90 via bulk
87 competition monitored by deep sequencing (Dataset S5), a technique with highly reproducible
88 results (Fig. S3). Our assay system reduces Hsp90 expression to ~1% of the endogenous level
89 (30), which magnifies the fitness consequences of Hsp90 mutations, enabling us to detect effects
90 of small magnitude.

91 We found that the vast majority of reversions to ancestral states in ScHsp90 are
92 deleterious (Fig. 1c). After experimental noise in fitness measurements is accounted for using a
93 mixture model approach, an estimated 93% of all reversions reduce the fitness of ScHsp90 (95%
94 CI 83–100%; Fig. 1c, Fig. S4). Three other statistical methods that differ in their assumptions
95 yielded estimates that between 54% and 95% of reversions are deleterious (Fig. S5a). Two
96 ancestral states cause very strong fitness defects (s = -0.38 and -0.54), but the typical reversion is
97 only mildly deleterious (median s = -0.010, $P = 1.2 \times 10^{-16}$, Wilcoxon rank sum test; Fig. 1c).
98 This conclusion is robust to excluding ancestral states that are reconstructed with any statistical
99 ambiguity ($P = 4.5 \times 10^{-14}$). The magnitude of each mutation's negative effect on fitness
100 correlates with indicators of site-specific evolutionary, structural, and functional constraint,
101 corroborating the view that they are authentically deleterious (Fig. S6).

102 These results do not imply that reversions can never happen—12 sites did undergo
103 substitution and reversion at some point along the lineage from ancAmoHsp90 to ScHsp90.
104 Rather, our observations indicate that at the current moment in time, the vast majority of
105 ancestral states are selectively inaccessible, irrespective of whether they were available at some
106 moment in the past or might become so in the future (31).

107 ***Intramolecular versus intermolecular epistasis.*** Reversions to ancestral states might be
108 deleterious because the derived states were entrenched by subsequent substitutions within Hsp90
109 (intramolecular epistasis) (1, 4); alternatively, they might be incompatible with derived states at
110 other loci in the *S. cerevisiae* genome (intermolecular epistasis), or the derived states might
111 unconditionally increase fitness. Entrenchment because of intramolecular epistasis predicts that
112 introducing into ScHsp90 sets of deleterious ancestral states that existed together at ancestral
113 nodes should not reduce fitness as drastically as would be predicted from the individual
114 mutations' effects. To test this possibility, we reconstructed complete ancestral NTDs from two
115 ancestral Hsp90s separated by vast time periods on the phylogenetic trajectory (Fig. 2a, Fig. S2)
116 and assayed their relative fitness in *S. cerevisiae* as chimeras with ScHsp90's other domains. This
117 design provides a lower-bound estimate of the extent of intramolecular epistasis, because it does
118 not eliminate interactions between substitutions in the NTD and those in other domains of
119 Hsp90.

120 We found that intramolecular epistasis is the predominant cause of entrenchment. The
121 first reconstruction, ancAscoHsp90, from the ancestor of Ascomycota fungi (estimated age ~450
122 million years (32)), differs from ScHsp90 at 42 NTD sites. If the fitness effects of these ancestral
123 states when combined were the same as when introduced individually, they would confer an
124 expected fitness of 0.65 (95% CI 0.61–0.69; Fig. 2b). When introduced together, however, the
125 actual fitness is 0.99 (Fig. 2d, Fig. S7a), indicating that the current fitness deficit of ancestral
126 states is caused primarily by deleterious interactions within the NTD.

127 The older ancestor, ancAmoHsp90 (estimated age ~1 billion years (33)), differs from
128 ScHsp90 at 60 NTD sites, which would confer an expected fitness of 0.23 (95% CI 0.21–0.26;
129 Fig. 2b), but the actual fitness of the NTD is 0.43 (Fig. S2d, Fig. S7a), again indicating strong

130 epistasis within the NTD. We hypothesized that the remaining fitness deficit caused by the
131 ancAmoHsp90 NTD could be attributed to intramolecular epistasis between the NTD and
132 substitutions in the other Hsp90 domains. We identified a candidate substitution in the protein's
133 middle domain that physically interacts with NTD residues to form the ATP-binding site (34);
134 reverting this substitution to the ancAmorphea state (L378i) together with the ancAmorphea
135 NTD increases fitness to 0.96 (Fig. 2d, Fig. S7a).

136 These findings indicate that virtually all the context-dependent deleterious effects of
137 ancestral states are caused by intramolecular interactions within the NTD and with one other site
138 in the Hsp90 protein. Derived states that emerged along the Hsp90 trajectory have been
139 entrenched by subsequent substitutions within the same protein, which closed the direct path
140 back to the ancestral amino acid without causing major changes in function or fitness (1).

141 ***Contingency and permissive substitutions.*** We next determined whether the derived
142 states that evolved during the protein's history were contingent on prior permissive substitutions.
143 We constructed a library of variants of the ancAmoHsp90 NTD, the deepest ancestor of the
144 trajectory, each of which contains one of the 98 forward mutations to a derived state. We cloned
145 this NTD library into yeast as a chimera with ScHsp90's other domains (with site 378 in its
146 ancestral state) and used our deep sequencing-based bulk fitness assay to measure the selection
147 coefficient of each mutation relative to ancAmoHsp90 (Dataset S6, Fig. S3d,e).

148 We found that most mutations to derived states were selectively unfavorable (Fig. 3a).
149 After accounting for experimental noise in fitness measurements using a mixture model, an
150 estimated 53% of derived states reduce ancAmoHsp90 fitness (95% CI 27–96%), and 32% are
151 neutral (95% CI 0–59%; Fig. S8); three other statistical approaches gave similar results (Fig.
152 S5a). Fifteen percent of the derived states are beneficial in our assay (95% CI 3–57%), which

153 could be because they are unconditionally advantageous or because of epistatic interactions with
154 other loci in *S. cerevisiae* or other regions of ScHsp90. Two derived states had very strong
155 fitness defects, but the typical derived state is weakly deleterious (median s = -0.005, $P = 5.8 \times$
156 10^{-4}, Wilcoxon rank sum test; Fig. 3a).

157 As with the reversions to ancestral states, the effects of individual derived states, as
158 measured in the ancestral background, predict fitness consequences far greater than observed
159 when the derived states are combined in the Hsp90 genotypes that existed historically along the
160 phylogeny (Fig. 2c,d, Fig. S7b). Thus, most derived states would have been deleterious if they
161 had occurred in the ancestral background, but they became accessible following subsequent
162 permissive substitutions that occurred within Hsp90. Taken together, the data from the ancestral
163 and derived libraries indicate that 83% of the amino acid states that occurred along this
164 evolutionary trajectory were contingent on prior permissive substitutions, entrenched by
165 subsequent restrictive substitutions, or both (Fig. 3b).

166 ***Specificity of epistatic interactions.*** Epistatic effects on fitness can emerge from specific
167 genetic interactions between substitutions that directly modify each other's effect on some
168 molecular property, or from nonspecific interactions between substitutions that are additive with
169 respect to bulk molecular properties (e.g. stability (2, 35)) if those properties nonlinearly affect
170 fitness (36-38).

171 To explore which type of epistasis predominates in the long-term evolution of Hsp90, we
172 first investigated the two strongest cases of entrenchment, the strongly deleterious reversions
173 V23f and E7a (upper-case letters indicate the ScHsp90 state and lower-case the ancestral state).
174 We sought candidate restrictive substitutions for each of these large-effect reversions by
175 examining patterns of phylogenetic co-occurrence. Substitution f23V occurred not only along the

176 trajectory from ancAmoHsp90 to ScHsp90 but also in parallel on another fungal lineage; in both
177 cases, candidate epistatic substitution i378L co-occurred on the same branch (Fig. S9a,b). As
178 predicted if i378L entrenched f23V, we found that introducing the ancestral state i378 in
179 ScHsp90 relieves the deleterious effect of the ancestral state f23 (Fig. 4a). These two residues
180 directly interact in the protein's tertiary structure to position a key residue in the ATPase active
181 site (Fig. S9c,d), explaining their specific epistatic interaction.

182 In the case of E7a—the other reversion strongly deleterious in ScHsp90—the ancestral
183 state was reacquired in a closely related fungal lineage. We reasoned that the substitutions that
184 entrenched a7E on the lineage leading to ScHsp90 must have themselves reverted or been further
185 modified on the fungal branch in which reversal E7a occurred. We identified two candidates
186 (n13T and a151N) that met these criteria (Fig. S10a,b,c). As predicted, experimentally
187 introducing the ancestral states n13 or a151 into ScHsp90 relieves much of the fitness defect
188 caused by the ancestral state a7, indicating that substitutions n13T and a151N entrenched a7E
189 (Fig. 4b). These three sites are on interacting secondary structural elements that are
190 conformationally rearranged when Hsp90 converts between ADP- and ATP-bound states (Fig.
191 S10d,e).

192 To test whether these modifiers specifically restrict particular substitutions or are general
193 epistatic modifiers, we asked whether the restrictive substitutions that entrenched one
194 substitution also modify the effects of the other (2). As predicted if the interactions among these
195 sets of substitutions are specific, introducing L378i does not ameliorate the fitness defect caused
196 by E7a, and introducing T13n or N151a does not ameliorate the fitness defect caused by V23f
197 (Fig. 4c,d). These data indicate that specific biochemical mechanisms underlie the restrictive
198 interactions for these large-effect examples of epistatic entrenchment.

199 Finally, we investigated whether the epistatic interactions among the set of small-effect
200 substitutions in this trajectory are also specific or the nonspecific result of a threshold-like
201 relationship between fitness and some bulk property such as stability (2, 35). If epistasis is
202 mediated by a nonspecific threshold relationship, mutations that decrease fitness in one
203 background will never be beneficial in another, although they can be neutral if buffered by the
204 threshold (Fig. 5a) (2, 20, 25). In contrast, specific epistatic interactions can switch the sign of a
205 mutation's selection coefficient in different sequence contexts (Fig. 5b) (37). As predicted under
206 specific epistasis, we found that for most differences between ancAmoHsp90 and ScHsp90
207 (65%), the ancestral state confers increased fitness relative to the derived state in the ancestral
208 background but decreases it in the extant background (Fig. 5c).
209 The selection coefficients of mutations are negatively correlated between backgrounds
210 (P=0.009), indicating that the substitutions that became most entrenched in the present also
211 required the strongest permissive effect in the past. This pattern is expected if the structural
212 constraints that determine the selective cost of having a suboptimal state at some site are
213 conserved over time, but the specific states preferred depend on the residues present at other
214 sites. Taken together, these findings indicate that most epistasis during the long-term evolution
215 of the Hsp90 NTD involved specific one-to-one (or few-to-few) interactions among sites, not
216 general effects on the protein's tolerance to mutation.
217
218 **Discussion**
219 ***Relation to prior work.*** We observed widespread and specific epistasis over the course of
220 a billion years of Hsp90 evolution, during which the protein's function, physical architecture,
221 and fitness were conserved. The fraction of historical substitutions that were either contingent on

222 permissive substitutions or entrenched by restrictive substitutions—about 80%—is considerably
223 higher than suggested by previous experimental work (2, 22, 24) and some computational
224 analyses (14), rivaling the highest estimates from computational studies (1, 16, 17). One
225 explanation for the more widespread epistatic interactions in our study may be our method's
226 capacity to detect much smaller growth deficits than have been discernable in previous
227 experimental studies.

228 Another difference from previous research is that we primarily observed specific
229 epistasis, whereas several studies have found a dominant role for nonspecific stability-mediated
230 epistasis, particularly during the short-term evolution of viruses (2, 11, 20). This disparity could
231 be attributable to a difference in selective regime or in time scale: the epistatic constraints caused
232 by specific interactions are expected to be maintained over far longer periods of time than those
233 caused by nonspecific interactions, which are easily replaced by other substitutions because of
234 the many-to-many relationship between permissive and permitted amino acid states (2, 20, 37).
235 The prevalence and type of epistasis may also vary because of differences in proteins' physical
236 architectures. Additional case studies will be necessary to evaluate the causal role of these and
237 other factors in determining the nature of epistatic interactions during evolution.

238 ***Limitations.*** Our strategy has some known limitations, but none are likely to change our
239 major conclusions. For example, we assayed the effect of long-past substitutions in the context of
240 extant yeast cells. Our experiments, however, indicate that there is only very weak epistasis for
241 fitness between historical substitutions within Hsp90 and those at other loci, because the
242 reconstructed ancestral Hsp90 chimeras cause a fitness deficit of only 0.01 to 0.04 when
243 introduced into *S. cerevisiae* cells—much smaller than the sum of intramolecular
244 incompatibilities revealed by introducing the ancestral states individually. Our finding of

245 widespread contingency and entrenchment is therefore not an artifact of incompatibilities
246 between ancestral Hsp90 states and the genotype of present-day *S. cerevisiae* at other loci.
247 A second potential limitation is that the ancestral states we tested were reconstructed
248 phylogenetically, not known empirically. But the vast majority of states were inferred with high
249 statistical confidence, because the Hsp90 NTD is well conserved and we used a densely-sampled
250 alignment. The ambiguity that was present primarily concerned the specific ancestral node at
251 which an inferred ancestral state was present, not whether or not it was ancestral somewhere
252 along the trajectory, which is the key inference for our purposes. Further, even when all states
253 with any degree of statistical uncertainty in the ancestral reconstruction were excluded from the
254 analysis, the remaining data strongly supported our conclusions concerning contingency and
255 entrenchment.
256 Finally, we measured fitness under a particular set of experimental conditions. Our assay
257 system reduces Hsp90 expression to ~1% of the endogenous level (30). Based on previous work
258 quantifying the relationship between Hsp90 function, expression, and growth rate (30), we
259 estimate that the average selection coefficient of -0.01 we observed among contingent or
260 entrenched substitutions corresponds to a fitness deficit of approximately $s = -5 \times 10^{-6}$ under
261 native-like expression levels. Mutations with selection coefficients in this range would likely be
262 subject to purifying selection in large microbial populations (39-41). Our assay also tests fitness
263 under log-phase growth conditions in rich media. A more heterogeneous or demanding
264 environment would likely increase the magnitude of selective effects of Hsp90 mutations,
265 because stress should amplify the fitness consequences of mutations in the proteostasis
266 machinery.

267 ***Implications.*** Our observation that contingency and entrenchment affected the majority
268 of historical substitutions suggests a daisy-chain model by which genetic interactions structured
269 long-term Hsp90 evolution (Fig. 5d). A permissive mutation becomes entrenched and
270 irreversible once a substitution contingent upon it occurs; if the contingent substitution
271 subsequently permits a third substitution, it too becomes entrenched (1, 16).
272 Most of the substitutions along the trajectory from ancAmoHsp90 to ScHsp90 were both
273 contingent and entrenched, suggesting that they occupy an internal position in this daisy chain.
274 Each of these changes closed reverse paths at some sites and opened forward paths at others,
275 which—if taken—would then entrench the previous step. Evolving this way over long periods of
276 time, proteins come to appear exquisitely well-adapted to the conditions of their existence, with
277 most present states superior to past ones. The conditions that make today's states so fit, however,
278 include—or are even dominated by—the transient internal organization of the protein itself.
279
280 **Methods**
281 For additional details, see SI Methods.
282 ***Phylogenetic inference and ancestral reconstruction***
283 We inferred the maximum likelihood (ML) phylogeny for 261 Hsp90 protein sequences
284 from the Amorphea clade, with Viridiplantae as an outgroup, under the LG+Γ+F model in
285 RAxML version 8.1.17 (42). Most probable ancestral NTD sequences were reconstructed on this
286 ML phylogeny using the AAML module of PAML version 4.4 (43). The trajectory of sequence
287 change was enumerated from the amino acid sequence differences between successive ancestral
288 nodes on the lineage from the common ancestor of Amorphea (ancAmoHsp90) to *S. cerevisiae*
289 Hsp82 (ScHsp90). Ancestral states are defined as amino acid states not present in ScHsp90 that

290 occurred in at least one ancestral node on the lineage from ancAmoHsp90 to ScHsp90. Derived
291 states are defined as amino acid states not present in the reconstructed ancAmoHsp90 sequence
292 that occurred in at least one descendent node on the lineage to ScHsp90.

293 ***Bulk growth competitions***

294 The ScHsp90 and ancAmoHsp90 NTD (+L378i, see SI Methods, Fig S2d) protein-coding
295 sequences were expressed together with the other domains from ScHsp90 from the
296 p414ADHΔTer plasmid (30). Individual mutations in each variant library were introduced via
297 PCR. Library genotypes were tagged with short barcodes to simplify sequencing steps during the
298 bulk competition (44); barcodes were associated with variant genotypes via paired-end
299 sequencing on an Illumina MiSeq instrument. Variant libraries were transformed into the
300 DBY288 Hsp90 *S. cerevisiae* shutoff strain (45, 46) and grown for 48 (ScHsp90 library) or 31
301 (ancAmoHsp90 library) hours under selective conditions. Cultures were maintained in log-phase
302 by regular dilution with fresh media to maintain a population size of 10^9 or greater, and samples
303 of ~10^8 cells were collected at regular time points over the course of the bulk competition.
304 Plasmid DNA was isolated from each time point (47), and the frequency of each library genotype
305 at each time point was determined via Illumina sequencing. Bulk competitions were performed
306 in duplicate.

307 ***Determination of selection coefficients***

308 The ratio of the frequency of each variant in the library relative to wildtype
309 (ancAmoHsp90 or ScHsp90) was determined from the number of sequence reads at each time
310 point, and the slope of the logarithm of this ratio versus time (in number of generations) was
311 determined as the raw per-generation selection coefficient (*s*) (48):

312 $$s = \mathrm{d}/\mathrm{dt}\ [\ \ln(\mathrm{n_m}\ /\ \mathrm{n_{wt}})\ ]$$

313 where n_m and n_{wt} are the number of sequence reads of mutant and wildtype, respectively, and
314 time is measured in number of wildtype generations. For genotypes that were not assayed in the
315 bulk competition, selection coefficients were determined from monoculture growth rates (48),
316 measured over 30 hours of growth with periodic dilution to maintain log-phase growth (30).
317 Selection coefficients for both types of measurement were scaled in relative fitness space ($w =$
318 e^s) such that an Hsp90 null allele, which is lethal, has a relative fitness of 0 ($s = -\infty$).

319 ***Estimating the fraction of deleterious mutations***

320 We estimated the fraction of mutations in each library that are deleterious by fitting the
321 distribution of mutant selection coefficients to a mixture model of underlying Gaussian
322 distributions. One of the underlying mixture components was required to have the mean and
323 standard deviation derived from replicate measurements of the wildtype sequences that were
324 present in the library but represented by independent barcodes; remaining mixture components
325 had a freely fit mean and standard deviation, and all components had a freely fit mixture
326 proportion. The optimal number of mixture components was determined via Akaike Information
327 Criterion. The mixture component derived from the wildtype sampling distribution was taken to
328 represent genotypes in the library with fitness indistinguishable from wildtype; mixture
329 components with mean less than zero were taken to reflect deleterious mutations; and mixture
330 components with mean greater than zero were taken to reflect beneficial mutations. For each
331 variant, the posterior probability of being neutral, deleterious, or beneficial was determined from
332 the relative probability density function for mixture components in each category at the selection
333 coefficient measured for that mutation; the total fraction of variants in the library that are
334 deleterious (or beneficial) was determined by summing the posterior probabilities of being
335 deleterious (or beneficial) over all mutants.

336 To determine the robustness of the mixture model's estimates of the fraction of
337 deleterious (or beneficial) mutations, we used three other approaches to analyze the distribution
338 of fitness measurements. The simplest, a nonparametric approach, reports the proportion of
339 mutations with observed selection coefficients < 0; because error in fitness measurements
340 appears to be unbiased (Fig. S5) and the number of deleterious mutations exceeds the number of
341 beneficial mutations, this approach is expected to yield a slightly conservative estimate of the
342 true proportion of deleterious mutations with selection coefficients of any magnitude. The
343 second approach is an empirical Bayes approach that calculates the posterior probability that
344 each mutation is deleterious (or beneficial), given the experimentally observed selection
345 coefficients and measurement error for wildtype and mutant genotypes; these posterior
346 probabilities are summed to yield the estimated proportion of mutations in each fitness category.
347 Third, we constructed 95% confidence intervals for each mutation's selection coefficient given
348 its mean over two replicates and standard error estimated over all mutations and counted the
349 number of mutations with selection coefficients less (or greater) than zero whose confidence
350 intervals do not overlap zero.

351 ***Data and code availability***

352 Processed sequencing data and scripts to reproduce all analyses are available at
353 github.com/JoeThorntonLab/Hsp90_contingency-entrenchment. Tables listing mutants and their
354 selection coefficients are included as Datasets S5 and S6. Raw sequencing data from each bulk
355 competition have been deposited in the NCBI Sequence Read Archive under accession
356 SRP126524. Tables linking barcode variants to their associated Hsp90 genotype are included as
357 Datasets S7 and S8.

358

Acknowledgements We thank members of the Thornton lab for comments on the manuscript and Jeff Boucher for technical assistance. This work was supported by National Institutes of Health R01GM104397 and R01GM121931 (J.W.T.), R01GM112844 (D.N.A.B.), F32GM119205-2 (J.M.F.), T32-GM007183 (T.N.S.), Government of India Department of Biotechnology Ramalingaswami Fellowship (P.M.), and a National Science Foundation Graduate Research Fellowship (T.N.S.)

Author Contributions All authors conceived the project and designed experiments. T.N.S., J.M.F., and P.M. performed experiments and analyzed data. T.N.S. and J.W.T. wrote the paper, with contributions from all authors.

Author Information The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.W.T. (joet1@uchicago.edu) or D.N.A.B (dan.bolon@umassmed.edu).

References

1. Shah P, McCandlish DM, Plotkin JB (2015) Contingency and entrenchment in protein evolution under purifying selection. *P Natl Acad Sci USA* 112(25):E3226–E3235.

2. Gong LI, Suchard MA, Bloom JD (2013) Stability-mediated epistasis constrains the evolution of an influenza protein. *eLife* 2:e00631.

3. Ortlund EA, Bridgham JT, Redinbo MR, Thornton JW (2007) Crystal structure of an ancient protein: evolution by conformational epistasis. *Science* 317(5844):1544–1548.

4. Pollock DD, Thiltgen G, Goldstein RA (2012) Amino acid coevolution induces an evolutionary Stokes shift. *P Natl Acad Sci USA* 109(21):E1352–9.

5. Bridgham JT, Ortlund EA, Thornton JW (2009) An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature* 461(7263):515–519.

6. Olson CA, Wu NC, Sun R (2014) A comprehensive biophysical description of pairwise epistasis throughout an entire protein domain. *Curr Biol* 24(22):2643–2651.

7. Bank C, Hietpas RT, Jensen JD, Bolon DNA (2015) A systematic survey of an intragenic epistatic landscape. *Mol Biol Evol* 32(1):229–238.

8. Podgornaia AI, Laub MT (2015) Pervasive degeneracy and epistasis in a protein-protein interface. *Science* 347(6222):673–677.

9. Melamed D, Young DL, Gamble CE, Miller CR, Fields S (2013) Deep mutational scanning of an RRM domain of the Saccharomyces cerevisiae poly(A)-binding protein. *RNA* 19(11):1537–1551.

10. Sarkisyan KS, et al. (2016) Local fitness landscape of the green fluorescent protein. *Nature* 533(7603):397–401.

11. Bloom JD, Gong LI, Baltimore D (2010) Permissive secondary mutations enable the evolution of influenza oseltamivir resistance. *Science* 328(5983):1272–1275.

12. Natarajan C, et al. (2016) Predictable convergence in hemoglobin function has unpredictable molecular underpinnings. *Science* 354(6310):336–339.

13. McKeown AN, et al. (2014) Evolution of DNA specificity in a transcription factor family produced a new gene regulatory module. *Cell* 159(1):58–68.

14. Soylemez O, Kondrashov FA (2012) Estimating the rate of irreversibility in protein evolution. *Genome Biol Evol* 4(12):1213–1222.

15. Jordan DM, et al. (2015) Identification of cis-suppression of human disease mutations by comparative genomics. *Nature* 524(7564):225–229.

16. Povolotskaya IS, Kondrashov FA (2010) Sequence space and the ongoing expansion of the protein universe. *Nature* 465(7300):922–926.

17. Breen MS, Kemena C, Vlasov PK, Notredame C, Kondrashov FA (2012) Epistasis as the primary factor in molecular evolution. *Nature* 490(7421):535–538.

18. Goldstein RA, Pollard ST, Shah SD, Pollock DD (2015) Nonadaptive amino acid convergence rates decrease over time. *Mol Biol Evol* 32(6):1373–1381.

19. McCandlish DM, Rajon E, Shah P, Ding Y, Plotkin JB (2013) The role of epistasis in protein evolution. *Nature* 497(7451):E1–2– discussion E2–3.

20. Ashenberg O, Gong LI, Bloom JD (2013) Mutational effects on stability are largely conserved during protein evolution. *P Natl Acad Sci USA* 110(52):21071–21076.

21. Mendes FK, Hahn Y, Hahn MW (2016) Gene tree discordance can generate patterns of diminishing convergence over time. *Mol Biol Evol* 33(12):3299–3307.

22. Lunzer M, Golding GB, Dean AM (2010) Pervasive cryptic epistasis in molecular evolution. *PLOS Genet* 6(10):e1001162.

23. Hochberg GKA, Thornton JW (2017) Reconstructing ancient proteins to understand the causes of structure and function. *Annu Rev Biophys* 46(1):247–269.

24. Risso VA, et al. (2015) Mutational studies on resurrected ancestral proteins reveal conservation of site-specific amino acid preferences throughout evolutionary history. *Mol Biol Evol* 32(2):440–455.

25. Doud MB, Ashenberg O, Bloom JD (2015) Site-specific amino acid preferences are mostly conserved in two closely related protein homologs. *Mol Biol Evol* 32(11):2944–2960.

26. Piper PW, et al. (2003) Yeast is selectively hypersensitised to heat shock protein 90 (Hsp90)-targetting drugs with heterologous expression of the human Hsp90β, a property that can be exploited in screens for new Hsp90 chaperone inhibitors. *Gene* 302(1-2):165–170.

27. Wider D, Péli-Gulli M-P, Briand P-A, Tatu U, Picard D (2009) The complementation of yeast with human or Plasmodium falciparum Hsp90 confers differential inhibitor sensitivities. *Mol Biochem Parasit* 164(2):147–152.

28. Hietpas RT, Jensen JD, Bolon DNA (2011) Experimental illumination of a fitness landscape. *P Natl Acad Sci USA* 108(19):7896–7901.

29. Adl SM, et al. (2012) The revised classification of eukaryotes. *J Eukaryot Microbiol* 59(5):429–514.

30. Jiang L, Mishra P, Hietpas RT, Zeldovich KB, Bolon DNA (2013) Latent effects of Hsp90

mutants revealed at reduced expression levels. *PLOS Genet* 9(6):e1003600.

31. McCandlish DM, Shah P, Plotkin JB (2016) Epistasis and the dynamics of reversion in molecular evolution. *Genetics* 203(3):1335–1351.

32. Taylor JW, Berbee ML (2006) Dating divergences in the Fungal Tree of Life: review and new analyses. *Mycologia* 98(6):838–849.

33. Eme L, Sharpe SC, Brown MW, Roger AJ (2014) On the age of eukaryotes: evaluating evidence from fossils and molecular clocks. *CSH Perspect Biol* 6(8):a016139–a016139.

34. Ali MMU, et al. (2006) Crystal structure of an Hsp90–nucleotide–p23/Sba1 closed chaperone complex. *Nature* 440(7087):1013–1017.

35. Tokuriki N, Tawfik DS (2009) Stability effects of mutations and protein evolvability. *Curr Opin Struc Biol* 19(5):596–604.

36. Harms MJ, Thornton JW (2013) Evolutionary biochemistry: revealing the historical and physical causes of protein properties. *Nat Rev Genet* 14(8):559–571.

37. Starr TN, Thornton JW (2016) Epistasis in protein evolution. *Protein Sci* 25(7):1204–1218.

38. Sailer ZR, Harms MJ (2017) Detecting High-Order Epistasis in Nonlinear Genotype-Phenotype Maps. *Genetics* 205(3):1079–1088.

39. Tsai IJ, Bensasson D, Burt A, Koufopanou V (2008) Population genomics of the wild yeast Saccharomyces paradoxus: Quantifying the life cycle. *P Natl Acad Sci USA* 105(12):4957–4962.

40. Peris D, et al. (2014) Population structure and reticulate evolution of Saccharomyces eubayanusand its lager-brewing hybrids. *Mol Ecol* 23(8):2031–2045.

41. Almeida P, et al. (2015) A population genomics insight into the Mediterranean origins of wine yeast domestication. *Mol Ecol* 24(21):5412–5427.

42. Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.

43. Yang Z, Kumar S, Nei M (1995) A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics* 141(4):1641–1650.

44. Hiatt JB, Patwardhan RP, Turner EH, Lee C, Shendure J (2010) Parallel, tag-directed assembly of locally derived short sequence reads. *Nat Methods* 7(2):119–122.

45. Hietpas RT, Bank C, Jensen JD, Bolon DNA (2013) Shifting fitness landscapes in response to altered environments. *Evolution* 67(12):3512–3522.

46. Mishra P, Flynn JM, Starr TN, Bolon DNA (2016) Systematic mutant analyses elucidate

general and client-specific aspects of Hsp90 function. *Cell Rep* 15(3):588–598.

47. Hietpas R, Roscoe B, Jiang L, Bolon DNA (2012) Fitness analyses of all possible point mutations for regions of genes in yeast. *Nat Protoc* 7(7):1382–1396.

48. Chevin LM (2011) On measuring selection in experimental evolution. *Biol Letters* 7(2):210–213.

49. Pantzartzi CN, Drosopoulou E, Scouras ZG (2013) Assessment and reconstruction of novel HSP90 genes: duplications, gains and losses in fungal and animal lineages. *PLOS ONE* 8(9):e73217.

50. Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22(13):1658–1659.

51. Sievers F, et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:1–6.

52. Le SQ, Gascuel O (2008) An improved general amino acid replacement matrix. *Mol Biol Evol* 25(7):1307–1320.

53. Brown MW, et al. (2013) Phylogenomics demonstrates that breviate flagellates are related to opisthokonts and apusomonads. *P R Soc B* 280(1769):20131755–20131755.

54. Brown MW, Spiegel FW, Silberman JD (2009) Phylogeny of the "forgotten" cellular slime mold, Fonticula alba, reveals a key evolutionary branch within Opisthokonta. *Mol Biol Evol* 26(12):2699–2709.

55. Paps J, Medina-Chacón LA, Marshall W, Suga H, Ruiz-Trillo I (2013) Molecular phylogeny of Unikonts: new insights into the position of Apusomonads and Ancyromonads and the internal relationships of Opisthokonts. *Protist* 164(1):2–12.

56. Kurtzman CP, Robnett CJ (2013) Relationships among genera of the Saccharomycotina(Ascomycota) from multigene phylogenetic analysis of type species. *FEMS Yeast Res* 13(1):23–33.

57. Shen X-X, et al. (2016) Reconstructing the Backbone of the Saccharomycotina Yeast Phylogeny Using Genome-Scale Data. *G3* 6(12):3927–3939.

58. Cunningham CN, Southworth DR, Krukenberg KA, Agard DA (2012) The conserved arginine 380 of Hsp90 is not a catalytic residue, but stabilizes the closed conformation required for ATP hydrolysis. *Protein Sci* 21(8):1162–1171.

59. Sella G, Hirsh AE (2005) The application of statistical physics to evolutionary biology. *P Natl Acad Sci USA* 102(27):9541–9546.

60. Benaglia T, Chauveau D, Hunter D (2009) mixtools: An R package for analyzing finite mixture models. *J Stat Softw* 32(6).

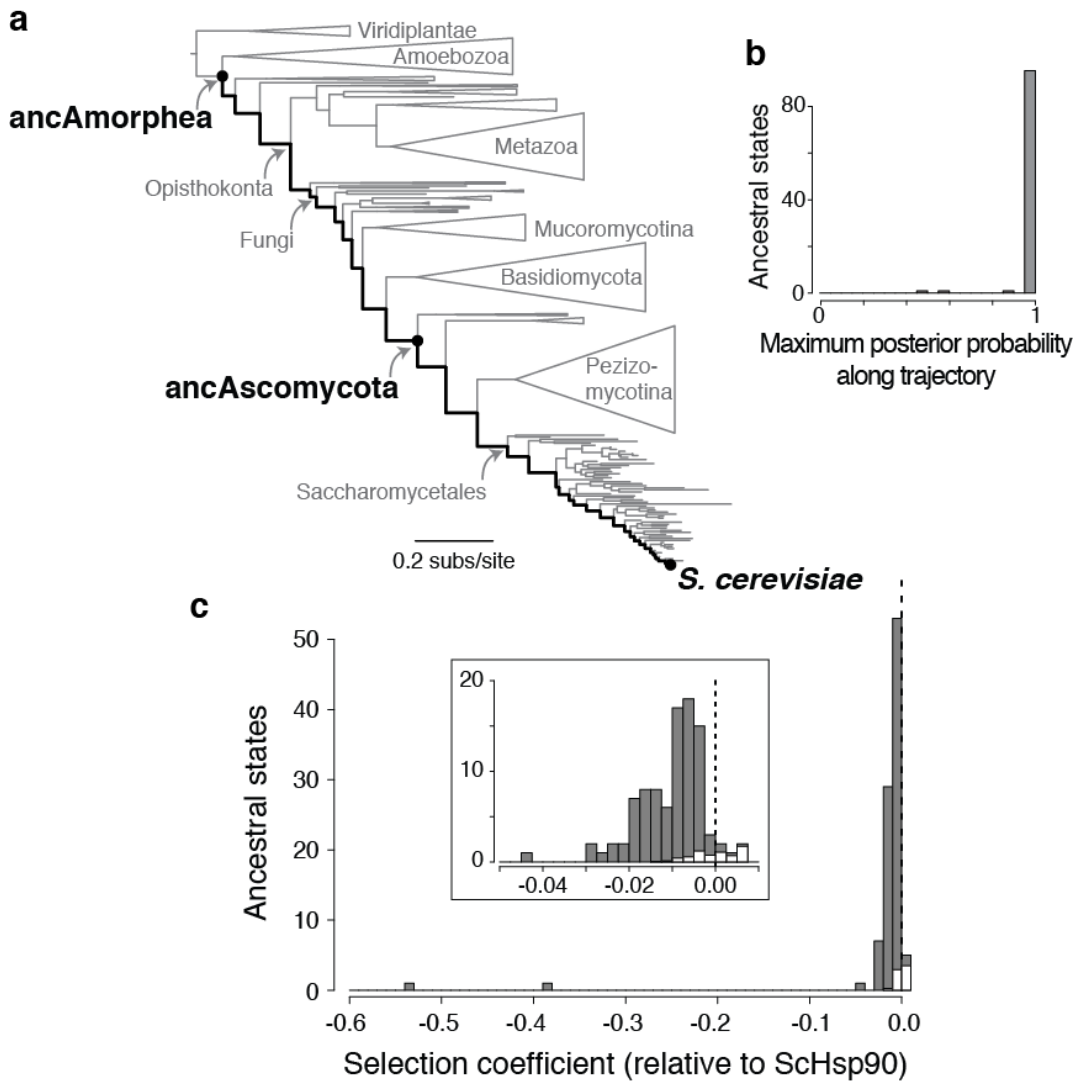


Figure 1. Ancestral states are deleterious in the yeast Hsp90 NTD. a, Maximum likelihood phylogeny of Hsp90 protein sequences from Amorphea. The evolutionary trajectory studied, from the last common ancestor of Amorphea to modern *S. cerevisiae*, is indicated by a dark black line. Major taxonomic groups are labeled in gray. Ancestral and extant genotypes characterized in this study are in black. Complete phylogeny with taxon names is in Fig. S1. **b**, Statistical confidence in ancestral amino acid states. For each of the 98 inferred ancestral states in the NTD, the highest posterior probability of the state at any internal node along the trajectory is shown. **c**, Distribution of selection coefficients of individual ancestral states when introduced into ScHsp90, measured as the logarithm of relative fitness compared to ScHsp90 in a deep-sequencing based bulk competition assay. Dashed line indicates neutrality. Inset, close view of the region near $s = 0$. In each histogram bin, white and grey show the proportion of ancestral states with selection coefficients in that range that are estimated to be neutral or deleterious, respectively, when measurement error is taken into account.

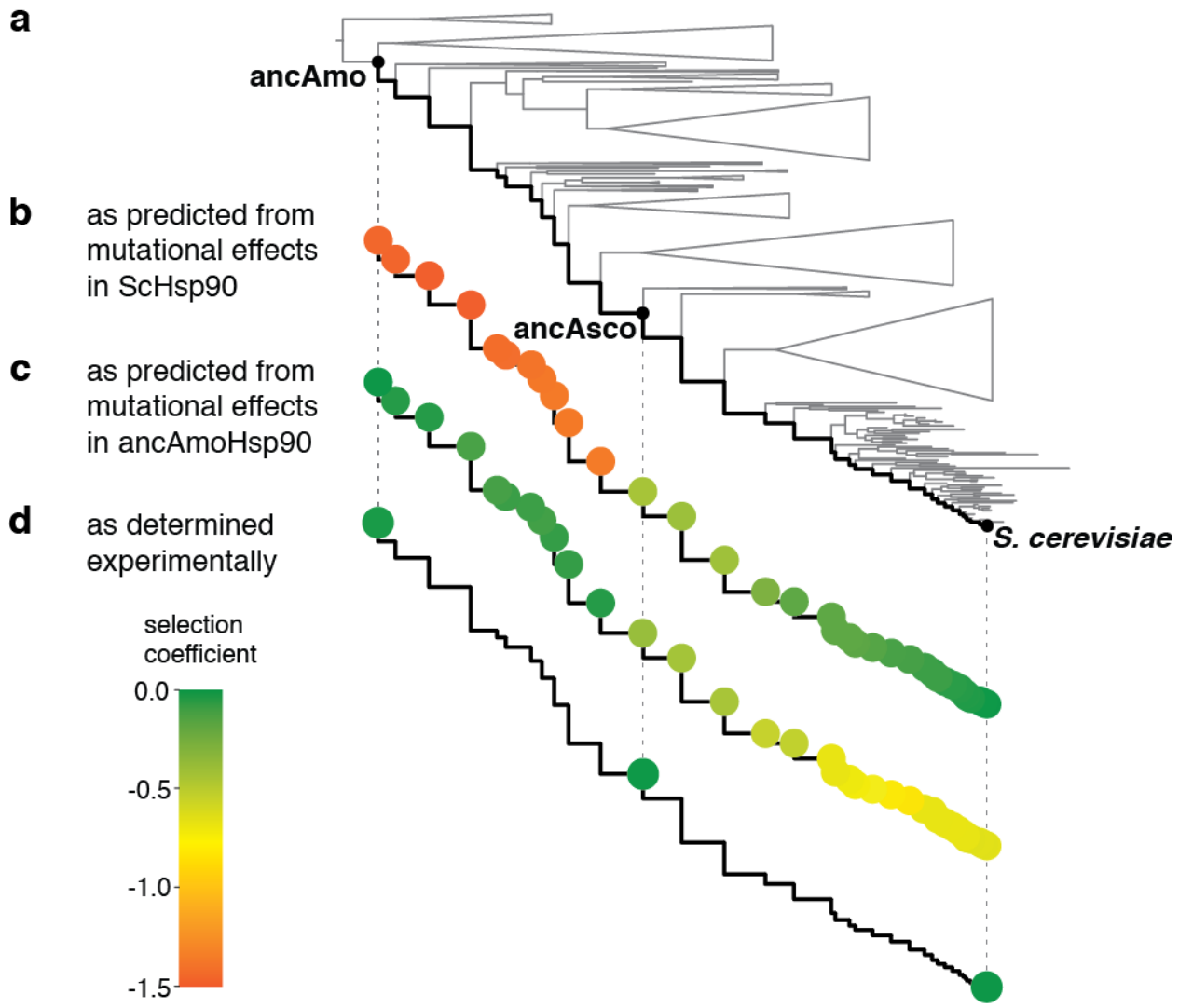


Figure 2. Fitness effects of historical substitutions are modified by intramolecular epistasis. For each node along the trajectory from ancAmoHsp90 to ScHsp90 (black line), the predicted or actual selection coefficient of the entire NTD genotype is represented from green (s = 0) to orange (s = -1.5). **a**, The Hsp90 phylogeny, represented as in Fig. 1a. **b**, The predicted selection coefficient of each ancestral sequence relative to ScHsp90 was calculated as the sum of the selection coefficients of each ancestral state present in that ancestor when measured individually in ScHsp90. **c**, The predicted selection coefficient of each sequence relative to ancAmoHsp90 was calculated as the sum of the selection coefficients of each derived state present at that node when measured individually in ancAmoHsp90. **d**, Experimentally determined selection coefficients for ancAmoHsp90 and ancAscoHsp90 relative to ScHsp90. For selection coefficients of each genotype, see Fig. S7.

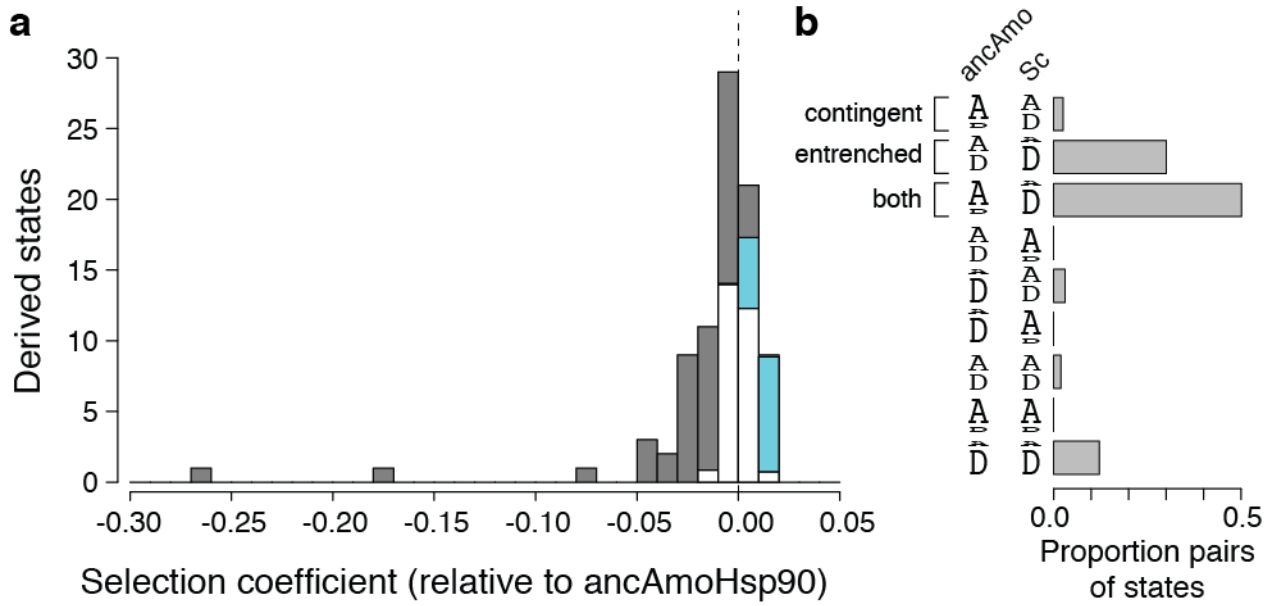


Figure 3. Widespread contingency and entrenchment. a, Distribution of measured selection coefficients of derived NTD states when introduced singly into ancAmoHsp90. Dashed line indicates neutrality. In each histogram bin, white shows the proportion of derived states with selection coefficients in that range that are estimated to be neutral; gray, deleterious; blue, beneficial. **b**, The fraction of pairs of ancestral and derived states that are inferred to be contingent, entrenched or both. Pairs of ancestral and derived states at each site can be classified by the relative fitness of the two states when measured in ancAmoHsp90 or in ScHsp90: ancestral state more fit (A larger than D), derived state more fit (D larger than A), or fitnesses indistinguishable (A and D same size). The fraction of pairs in each category was estimated as the product of the probabilities that each pair of sites is in the relevant selection category (ancestral state with fitness greater than, less than, or indistinguishable from the derived state) in the ScHsp90 and the ancAmoHsp90 backgrounds.

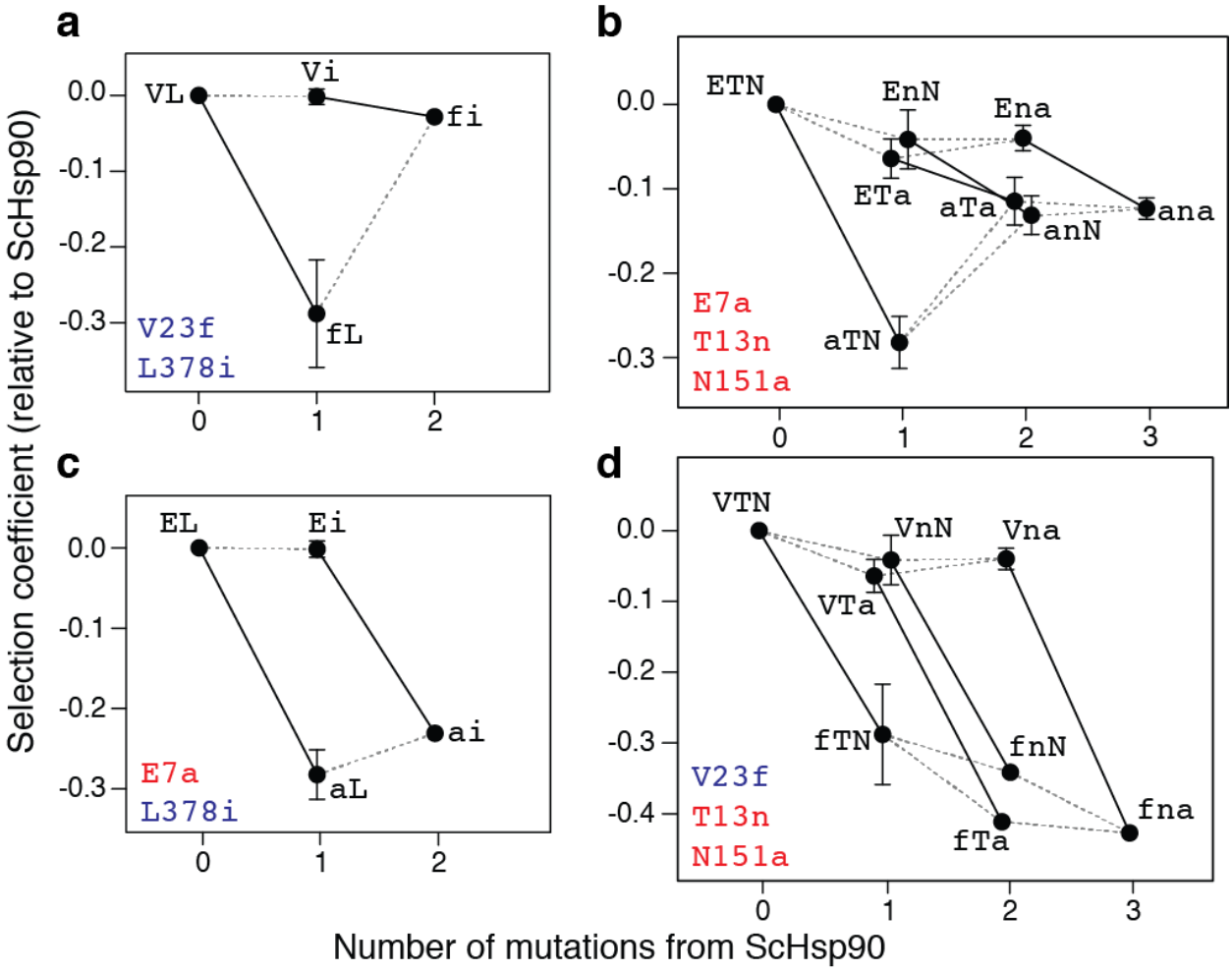


Figure 4. Epistatic interactions are specific. Large-effect deleterious reversions and restrictive substitutions that contributed to their irreversibility. For each single, double, or triple mutant in ScHsp90, the selection coefficient relative to ScHsp90 is shown, as assessed in monoculture growth assays. Lines connect genotypes that differ by a single mutation; solid lines indicate the effect of the large-effect reversions in each background. Error bars, SEM for 2 to 4 replicates (see Methods; absence of error bar indicates one replicate). Data points are labeled by amino acid states: lower case, ancestral state; upper case, derived state. Mutations tested in each cycle are in the bottom-left corner; those in the same color interact specifically with each other. **a**, Deleterious reversion V23f is ameliorated by L378i. **b**, Deleterious reversion E7a is partially ameliorated by N151a or T13n. **c**, L378i does not ameliorate E7a. **d**, N151a and T13n do not ameliorate V23f.

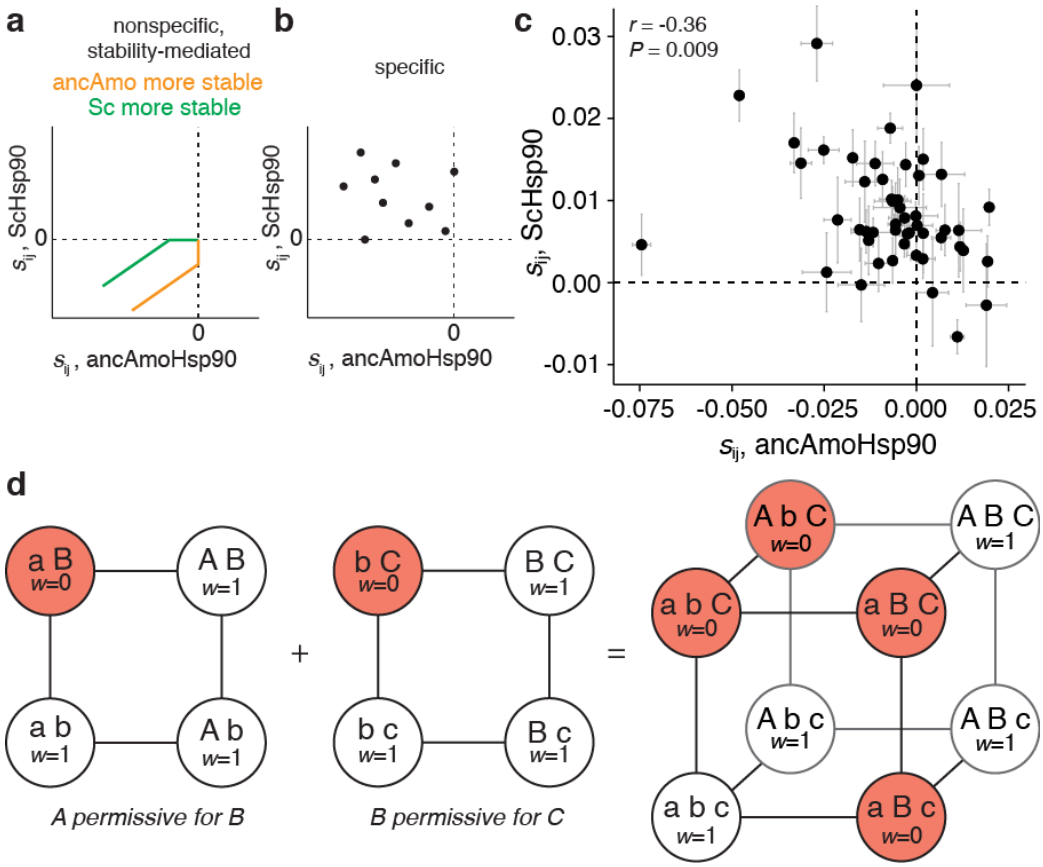


Figure 5. A daisy-chain model of epistasis. a,b, Expected relationship under two models of epistasis between selection coefficients of ancestral-to-derived mutations (s_{ij}) when introduced into ancestral (*x*-axis) or derived (*y*-axis) backgrounds. **a**, Nonspecific epistasis: if genetic interactions are the nonspecific result of a threshold-like, buffering relationship between stability (or another bulk property) and fitness (2, 35), then the effects of strongly deleterious mutations will be positively correlated between the two backgrounds, but weakly deleterious mutations in the less stable background may be neutral in the more stable background (yellow, ancAmoHsp90 more stable; green, ScHsp90 more stable). **b**, Specific epistasis: if interactions reflect specific couplings between sites, then mutations from ancestral to derived states can be deleterious in the ancestral background but beneficial in the derived background (upper left quadrant). **c**, Measured selection coefficients for ancestral-derived state pairs that differ between ancAmoHsp90 and ScHsp90. Dashed lines, $s = 0$. Error bars, SEM from two replicate bulk competition measurements. *r*, Pearson correlation coefficient and associated *P* value. Two additional points that are strongly deleterious outliers in the ScHsp90 or ancAmoHsp90 data are not shown for clarity and are not included in the correlation; the plot including these outliers is shown in Fig. S10c. **d**, Daisy-chain model of specific epistatic interactions. Each square shows the mutant cycle for a pair of substitutions (A and B or B and C; lower-case, ancestral state; upper-case, derived), one of which is permissive for the other. Each circle is a genotype colored by its fitness (*w*): white, neutral; red, deleterious. Edges are single-site amino acid changes. The cube shows the combined mutant cycle for all three substitutions. Permissive substitutions become entrenched when the mutation that was contingent upon it occurs. Substitutions in the middle of the daisy-chain, which require a permissive mutation and are permissive for a subsequent mutation, are both contingent and entrenched.

Supporting Information

SI Methods

373 **Phylogenetic analysis and ancestral reconstruction.** We obtained Hsp90 protein
374 sequences from the Amorphea clade (29) from NCBI, the JGI Fungal Program, the Broad
375 Institute Multicellularity Project, the literature (49), and Iñaki Ruiz-Trello (personal
376 communication). Full identifiers and sources of sequences are listed in Dataset S3. Each protein
377 was used as a query in a BLASTp search against the human proteome to identify and retain
378 Hsp90A orthologs. We used CD-HIT (50) to filter proteins with high sequence similarity. We
379 removed sequences with >67% missing characters and highly diverged, unalignable sequences.
380 Remaining sequences were aligned with Clustal Omega (51). Lineage-specific insertions were
381 removed, as were unalignable linker regions (ScHsp90 sites 1-3, 225-237, 686-701). We added
382 six Hsp90A sequences from Viridiplantae as an outgroup, resulting in a final alignment of 267
383 protein sequences and 680 sites (Dataset S1).

384 We inferred the maximum likelihood (ML) phylogeny (Dataset S2) given our alignment
385 and the LG model (52) with gamma-distributed among-site rate variation (4 categories) and ML
386 estimates of amino acid frequencies, which was the best-fit model as judged by AIC. The
387 phylogeny was inferred using RAxML version 8.1.17 (42). The ML phylogeny reproduces
388 accepted relationships between major taxonomic lineages (29, 53-57). Most probable ancestral
389 sequences (Dataset S4) were reconstructed on the maximum likelihood phylogeny using the
390 AAML module of PAML version 4.4 (43) given the alignment, ML phylogeny, and LG+Γ
391 model. The trajectory of sequence change was enumerated from the amino acid sequence
392 differences between successive ancestral nodes on the lineage from the common ancestor of
393 Amoebozoa + Opisthokonta (ancAmorphea) to *S. cerevisiae* Hsp82 (ScHsp90, Uniprot P02829).

394 Coding sequences for the most probable ancestral amino acid sequences of the Hsp90 N-
395 terminal domain (NTD) from ancAmorphea (ancAmoHsp90) and the common ancestor of
396 Ascomycota yeast (ancAscoHsp90) were synthesized by IDT (Dataset S9). These sequences
397 were cloned as chimeras with the ScHsp90 middle and C-terminal domains and intervening
398 linkers via Gibson Assembly. AncAmoHsp90 also carries an additional reversion to the
399 ancAmorphea state at site 378 in the middle domain (Fig. S2d), which is part of a loop that
400 extends down and interacts with ATP and the NTD (34, 58).
401 **Generating mutant libraries**. ScHsp90 and ancAmoHsp90 gene constructs were
402 expressed from the p414ADHΔTer plasmid (30). The ScHsp90 library consists of variants of the
403 ScHsp90 NTD, each containing one mutation to an ancestral amino acid state. The
404 ancAmoHsp90 library consists of variants of the ancAmoHsp90 NTD, each containing one
405 mutation to a derived state. Two sets of PCR primers were designed for each mutation, to
406 amplify Hsp90 NTD fragments N-terminal and C-terminal to the mutation of interest; primers
407 introduce the mutation of interest and generate a 25-bp overlap between fragments, as well as 20-
408 bp overlaps between each fragment and the destination vector for gene re-assembly (Dataset S9).
409 PCR was conducted with Pfu Turbo polymerase (Agilent) for 15 amplification cycles. The
410 resulting PCR fragments were stitched together with a 10-cycle assembly PCR, pooled, and
411 combined via Gibson Assembly (NEB) with a linearized p414ADHΔTer Hsp90 destination
412 vector excised of the NTD.
413 **Barcode labeling of library genotypes.** Following construction of the plasmid libraries,
414 each variant in the library was tagged with a unique barcode to simplify sequencing steps during
415 bulk competition (44). A pool of DNA constructs containing a randomized 18 base-pair barcode
416 sequence (N18) and Illumina sequencing primer annealing regions (IDT; Dataset S9) was cloned

417 200 nucleotides downstream from the hsp90 stop codon via restriction digestion, ligation, and
418 transformation into chemically-competent *E. coli*. Cultures with different amounts of the
419 transformation reaction were grown overnight and the colony forming units in each culture were
420 assessed by plating a small fraction. We isolated DNA from the transformation that contained
421 approximately 10-20 fold more colony-forming units than mutants, with the goal that each
422 mutant would be represented by 10-20 unique barcodes.

423 To associate barcodes with Hsp90 mutant alleles, we conducted paired end sequencing of
424 each library using primers that read the N18 barcode in the first read and the Hsp90 NTD in the
425 other (Dataset S9). To generate short DNA fragments from the plasmid library that would be
426 efficiently sequenced, we excised the gene region between the NTD and the N18 barcode via
427 restriction digest, followed by blunt ending with T4 DNA polymerase (NEB) and plasmid
428 ligation at a low concentration (3 ng/μL) that favors circularization over bi-molecular ligations.
429 The resulting DNA was re-linearized by restriction digest, and Illumina adapter sequences were
430 added via an 11-cycle PCR (Dataset S9). The resulting PCR products were sequenced using an
431 Illumina MiSeq instrument with asymmetric reads of 50 bases and 250 bases for Read1 and
432 Read2 respectively. After filtering low quality reads (Phred scores < 10), the data were organized
433 by barcode sequence. For each barcode that was read more than 3 times, we generated a
434 consensus sequence of the N-domain indicating the mutation that it contained (Datasets S7, S8).

435 **Bulk growth competitions.** For bulk fitness assessments, we transformed *S. cerevisiae*
436 with the ScHsp90 library along with wildtype ScHsp90 and a no-insert control; we also
437 transformed *S. cerevisiae* with the ancAmoHsp90 library along with wildtype ScHsp90, wildtype
438 ancAmoHsp90, and a no-insert control. Concentrations of plasmids were adjusted to yield a
439 2:6:1 molar ratio of wildtype: no-insert control: average variant in the library. Plasmid libraries

440 and corresponding controls were transformed into the DBY288 Hsp90 shutoff strain (45, 46),
441 resulting in ~150,000 unique yeast transformants representing 50-fold sampling for the average
442 barcode. Following recovery, transformed cells were washed 5 times in SRGal-W (synthetic 1%
443 raffinose and 1% galactose lacking tryptophan) media to remove extracellular DNA, and then
444 transferred to plasmid selection media SRGal-W and grown at 30°C for 48 hours with repeated
445 dilution to maintain the cells in log phase of growth. To select for function of the plasmid-borne
446 Hsp90 allele, cells were shifted to shutoff conditions by centrifugation, washing and re-
447 suspension in 200 mL SD-W (synthetic 2% dextrose lacking tryptophan) media and ampicillin
448 (50µg/mL), and growth at 30°C 225 rpm. Following a 16-hour growth period required to shut off
449 expression of the wildtype chromosomal Hsp90, we collected samples of ~10^8 cells at 8 or more
450 time points over the course of 48 (ScHsp90 library) or 31 (ancAmoHsp90 library) hours and
451 stored them at -80°C. Cultures were maintained in log phase by regular dilution with fresh
452 media, maintaining a population size of 10^9 or greater throughout the bulk competition. Bulk
453 competitions of each library were conducted in duplicate from independent transformations.
454 **DNA preparation and sequencing**. We collected plasmid DNA from each bulk
455 competition time point as previously reported (47). Purified plasmid was linearized with AscI.
456 Barcodes were amplified by 18 cycles of PCR using Phusion polymerase and primers that add
457 Illumina adapter sequences, as well as an 8-bp identifier used to distinguish among libraries and
458 time points (Dataset S9). Identifiers were designed so that each differed by more than two bases
459 from all others to avoid misattributions due to sequencing errors. PCR products were purified
460 two times over silica columns (Zymo research), and quantified using the KAPA SYBR FAST
461 qPCR Master Mix (Kapa Biosystems) on a Bio-Rad CFX machine. Samples were pooled and

462 sequenced on an Illumina NextSeq (ancAmoHsp90 library) or HiSeq 2000 (ScHsp90 library)
463 instrument in single-end 100 bp mode.

464 **Analysis of bulk competition sequencing data**. Illumina sequence reads were filtered
465 for Phred scores >20, strict matching of the sequence of the intervening bases to the template,
466 and strict matching of the N18 barcode and experimental identifier to those that were expected in
467 the given library. Reads that passed these filters were parsed based on the identifier sequence.
468 For each identifier, the data was condensed by generating a count of each unique N18 read. The
469 unique N18 count file was then used to identify the frequency of each mutant using the variant-
470 barcode association table. For each variant in the library, the counts of each associated barcode
471 were summed to generate a cumulative count for that mutant.

472 **Determination of selection coefficient.** The ratio of the frequency of each variant in the
473 library relative to wildtype (ancAmoHsp90 or ScHsp90) was determined at each time point, and
474 the slope of the logarithm of this ratio versus time (in number of generations) was determined as
475 the raw per-generation selection coefficient (*s*) (48):

476 $$s = \mathrm{d}/\mathrm{dt}\ [\ \ln(\mathrm{n_m}\ /\ \mathrm{n_{wt}})\ ]$$

477 where n_m and n_{wt} are the number of sequence reads of mutant and wildtype, respectively, and
478 time is measured in number of wildtype generations. No-insert plasmid selection coefficients
479 were determined from the first three time points because their counts drop rapidly over time.
480 Mutants with selection coefficients within three standard deviations of the mean of no-insert
481 variants were considered null-like and also analyzed based on the first three time points. For all
482 other variants, selection coefficients were determined from all time points. Final selection
483 coefficients for each variant were scaled in relative fitness space ($w = e^s$) such that the Hsp90
484 null allele, which is lethal, has a relative fitness of 0 ($s = -\infty$). This definition of relative fitness,

485 unlike that which defines $w = 1 - s$, has the advantage of making selection coefficients additive
486 and reversible (the selection coefficient of mutation from state *i* to *j* is the opposite of the
487 selection coefficient of that from *j* to *i*) (59).

488 **Generation of individual mutants and monoculture analysis of yeast growth.** To
489 measure the relative fitness of ancAscoHsp90, mutations missed in the bulk libraries, and
490 genotypes in mutant cycles that we sought to test in combination for epistatic interactions, we
491 assayed growth rate in monoculture and related this to fitness, which assumes the relative rate of
492 growth of two genotypes is the same in isolation as in direct competition (48). The growth rate of
493 individually cloned mutants was estimated over 30 hours of growth with periodic dilution to
494 maintain log-phase growth, as per Jiang et al. (30). Growth rates were determined as the slope of
495 the linear model relating the log-transformed dilution-corrected cell density to time. The growth
496 rate was converted to an estimate of the selection coefficient by taking the difference in growth
497 rate (Malthusian parameter) between mutant and wildtype and multiplying this by the wildtype
498 generation time (48), then rescaling selection coefficients in relative fitness space such that a null
499 mutant has relative fitness 0 ($s = -\infty$).

500 Individual mutants of ancAmoHsp90 and ScHsp90 were generated in the p414ADHΔTer
501 background by Quikchange site-directed mutagenesis (Dataset S9), confirmed by Sanger
502 sequencing. Mutations that were generated and assayed in ancAmoHsp90 (with number of
503 replicate measurements in parentheses) include: S49A (n=1), T137I (n=1), V147I (n=1), I158V
504 (n=1), R160L (n=1), G164N (n=1), E165P (n=1), L167I (n=1), K172I (n=1), L193I (n=1), and
505 V194I (n=1). Mutations generated and assayed in ScHsp90 include: T5S (n=3), E7A (n=4),
506 T13N (n=3), V23F (n=2), N151A (n=3), L378I (n=2), double mutants E7A/T13N (n=3),
507 E7A/N151A (n=3), T13N/N151A (n=3), V23F/T13N (n=1), V23F/N151A (n=1), E7A/L378I

508 (n=1), V23F/L378I (n=1) and triple mutants E7A/T13N/N151A (n=2) and V23F/T13N/N151A
509 (n=1).

510 **Robustness of results to statistical uncertainty and technical variables.** The
511 conclusion that the typical ancestral state is deleterious in ScHsp90 is robust to the exclusion of
512 20 ancestral states that have posterior probability < 1.0 at all ancestral nodes along the trajectory
513 ($P = 4.5 \times 10^{-14}$, Wilcoxon rank sum test with continuity correction). The mutation to one
514 ancestral state was missed in the bulk competition: its selection coefficient was inferred
515 separately via monoculture, and including it in the analysis still leads to the conclusion that the
516 typical ancestral state is deleterious ($P = 7.8 \times 10^{-17}$, Wilcoxon rank sum test with continuity
517 correction).

518 The conclusion that the average derived state is deleterious in ancAmoHsp90 is retained
519 when we include only the 32 mutations for which the ancAmoHsp90 state is inferred with a
520 posterior probability of 1.0 and the derived state is inferred with posterior probability 1.0 in at
521 least one node along the trajectory ($P = 1.1 \times 10^{-4}$, Wilcoxon rank sum test with continuity
522 correction). The conclusion is also robust if we include selection coefficients as determined
523 separately via monoculture for mutations to 11 derived states that were missed in the bulk
524 competition ($P = 5.4 \times 10^{-4}$, Wilcoxon rank sum test with continuity correction).

525 We assessed relative fitness for six genotypes (ScHsp90+E7a, ScHsp90+V23f,
526 ScHsp90+N151a, ScHsp90+T13n, ancAmoHsp90, and ancAmoHsp90+i378L) both by
527 monoculture and by bulk competition. These two measures are well correlated (Pearson R^2 =
528 0.95), although the magnitude of a fitness effect is smaller when measured by monoculture
529 growth assays (Fig. S3f), perhaps because of differences in experimental conditions for bulk
530 versus monoculture growth, such as the type of growth vessel and culture volume (and

531 consequential aeration). The only conclusion involving a comparison between these two kinds of
532 measurements is that ancAscoHsp90 (measured via monoculture) is more fit than would be
533 predicted from the sum of selection coefficients of its component states (measured via bulk
534 competition) (Fig. 2, Fig. S7). We therefore used the observed linear relationship between the
535 two types of fitness assays to transform ancAscoHsp90's fitness as measured by monoculture
536 (0.991); the expected fitness of ancAscoHsp90 in a bulk competition is 0.986, still much larger
537 than the predicted fitness of 0.65 in the absence of epistasis.

538 **Expected versus observed fitness.** To identify epistasis between candidate interacting
539 sites (e.g. Fig. 4a-d) or among the broader set of substitutions (e.g. Fig. 2), we compared the
540 observed fitness of genotypes with multiple mutations to that expected in the absence of
541 epistasis. In the absence of epistatic interactions, selection coefficients combine additively (59).
542 We therefore calculated the expected selection coefficient of a genotype as the sum of selection
543 coefficients of its component mutations as measured independently in a reference background
544 (ancAmoHsp90 or ScHsp90). The standard error of a predicted fitness given the sum of selection
545 coefficients was calculated as the square root of the sum of squared standard errors of the
546 individual selection coefficient estimates, as determined from the duplicate bulk competition
547 measurements. Epistasis was implicated if the observed fitness of a genotype differed from that
548 predicted from the sum of its corresponding single-mutant selection coefficients.

549 **Estimating the fraction of deleterious mutations.** We sought to determine the fraction
550 of mutations in each dataset that are deleterious using a modeling approach that incorporates
551 measurement error and which does not require individual mutations to be classified as
552 deleterious, neutral, or beneficial. We used the mixtools package (60) in R to estimate mixture
553 models of underlying Gaussian distributions that best fit the observed distributions of mutant

554 selection coefficients in each library. First, we fit a single Gaussian distribution to the measured
555 selection coefficients of replicate wildtype sequences that were present in the library but
556 represented by independent barcodes. We then required one of the Gaussian distributions in each
557 mixture model to have a mean and standard deviation fixed to that of the wildtype
558 measurements, with a freely estimated mixture proportion. The other Gaussian components in
559 each mixture model had a freely fit mean, standard deviation, and mixture proportion. Mixture
560 models were fit to all non-outlier selection coefficients, because the presence of strongly
561 deleterious selection coefficients ($s < -0.04$), which are unambiguously deleterious, interfered
562 with model convergence. We assessed mixture models with a variable number of mixture
563 components ($k = 2$ to 6 for the ancAmoHsp90 library and 2 to 5 for the ScHsp90 library, because
564 the 6-component model would not converge), and obtained the maximum likelihood estimate of
565 each component's mean, standard deviation, and mixture proportion via an expectation-
566 maximization algorithm as implemented in mixtools. We compared the models built for each k
567 using AIC. For ScHsp90, the 3-component mixture model was favored by AIC (Fig. S4a). For
568 ancAmoHsp90, the 2-component and 5-component mixture models had virtually
569 indistinguishable AIC (Fig. S8a), but the 2-component mixture model had a visually suboptimal
570 fit (Fig. S8c,d) and attributed a larger proportion of mutations as belonging to a deleterious
571 sampling distribution (0.78 versus 0.53 for the 5-component mixture model), so we selected the
572 more conservative and visually superior 5-component mixture model.
573 The mixture component derived from the wildtype sampling distribution was taken to
574 represent genotypes in the library with fitness indistinguishable from wildtype; mixture
575 components with mean < 0 were taken to reflect deleterious variants; and mixture components
576 with mean > 0 were taken to reflect beneficial variants. For each variant, the posterior probability

577 of being deleterious, neutral, or beneficial was determined from the relative probability density
578 function for mixture components in each category at the selection coefficient measured for that
579 mutation; for variants with $s < -0.04$ that were excluded from model inference, the posterior
580 probability of being deleterious was 1. The total fraction of variants in the library that are
581 deleterious (or beneficial) was determined by summing the posterior probabilities of being
582 deleterious (or beneficial) over all mutants. To generate the representations in Figures 1c and 3a,
583 posterior probabilities were summed separately for the set of measurements that fall within each
584 histogram bin. We also report the estimate of the fraction deleterious (or beneficial) by summing
585 the mixture proportion parameters for mixture components centered below (or above) zero (Figs.
586 S4a, S8a).

587 Uncertainty in the estimated fraction of mutations that are deleterious or beneficial was
588 determined via a bootstrapping procedure. For each of 10,000 bootstrap replicates, measured
589 selection coefficients from the bulk competition were resampled with replacement. Mixture
590 models with fixed k were fit to each bootstrap sample, and the estimated fractions of mutations in
591 deleterious, neutral, or beneficial sampling distributions were determined as above.

592 To estimate the probability that a pair of states exhibit contingency and/or entrenchment,
593 we calculated the joint posterior probability as the product of the probabilities that each pair of
594 sites is in the relevant selection category (ancestral state with fitness greater than, less than, or
595 indistinguishable from the derived state) in the ScHsp90 and the ancAmoHsp90 backgrounds.
596 For sites that substituted from the ancAmoHsp90 state i to the ScHsp90 state j ($i \rightarrow j$, n = 35), i is
597 the ancestral state and j the derived state for measurements in both backgrounds. For sites that
598 substituted from the ancAmoHsp90 state i to an intermediate state j before substituting back to i
599 in ScHsp90 ($i \rightarrow j \rightarrow i$, n=12), then i is the ancestral state and j derived in ancAmoHsp90 assay,

600 and j is the ancestral state and i derived in ScHsp90. For sites that substituted from the
601 ancAmoHsp90 state i to an intermediate state j that was further modified to k in ScHsp90
602 ($i \rightarrow j \rightarrow k$, n=25), two comparisons were made: in the first, i was ancestral and k was derived for
603 measurements in both backgrounds, while in the second comparison, i was ancestral and j
604 derived in ancAmoHsp90, and j ancestral and k derived in ScHsp90.

605 In addition to the mixture model approach presented above, we report three independent
606 methods for estimating the fraction of mutations in each dataset that are deleterious (or
607 beneficial) (Fig. S5a). The simplest estimate of the fraction of mutations in each distribution that
608 are deleterious (or beneficial) is the fraction of observed selection coefficients (s_{obs}) that are less
609 than (or greater than) zero. This counting approach assumes that, at some magnitude, all
610 mutations have a true $s > 0$ or $s < 0$. This method would be unbiased if experimental errors are
611 random and if the number of truly beneficial and truly deleterious mutations is equal. In our data,
612 experimental errors are unbiased with respect to s_{obs} (Fig. S5b,c), but there appear to be more
613 deleterious than beneficial mutations. As a result, measurement error is likely to cause the
614 number of mutations with true $s < 0$ and $s_{ob\ s} > 0$ to exceed the number with true $s > 0$ and $s_{obs} < 0$;
615 this approach is therefore expected to underestimate the fraction of mutations with true $s < 0$.

616 Second, we used an empirical Bayes approach. For each mutation, we compute the
617 posterior probability that it is non-neutral by comparing the likelihoods of two hypotheses: the
618 null hypothesis, that a variant is neutral and therefore $s \sim N(0, SEM_{wt})$ where SEM_{wt} is the
619 standard deviation of the sampling distribution of repeated wildtype measurements present in the
620 corresponding experiment (Figs. S4b, S8b); and the alternative hypothesis, that a variant is non-
621 neutral and therefore $s \sim N(s_{\text{obs}}, SEM_{mut})$, where SEM_{mut} is calculated as an estimated SEM from

622 all duplicate bulk fitness measurements, which makes the assumption that all variants have the

623 same experimental error (Fig. S5b,c). SEM_{mut} was calculated as:

624 $$SEM_{mut} = \sqrt{\frac{\sum_{i=1}^{N}(s_i - \overline{s_i})^2}{N-1}}$$

625 where s_i is a measured selection coefficient of a mutant in a single replicate, $\overline{s_i}$ is the

626 corresponding mean selection coefficient for that mutant as calculated from both replicates, and

627 N is the total number of observations from both replicates. The posterior probability that a

628 variant is non-neutral is calculated from the relative likelihoods of the two hypotheses, with a

629 uniform prior on the two hypotheses:

630 $$P(\text{non-neutral}) = \frac{P(s_{obs}|s \sim N(s_{obs}, SEM_{mut})}{P(s_{obs}|s \sim N(s_{obs}, SEM_{mut}) + P(s_{obs}|s \sim N(0, SEM_{wt})}$$

631 If a variant has $s_{obs} > 0$, then P(non-neutral) corresponds to a probability that a mutation is

632 beneficial; if a variant has $s_{obs} < 0$, then P(non-neutral) corresponds to a probability that a mutant

633 is deleterious. Like the counting approach, this empirical Bates approach will call some truly

634 deleterious mutations beneficial, because there is a greater density of deleterious than beneficial

635 mutations in the distributions.

636 Last, we constructed a 95% confidence interval (CI) for each mutation given its mean

637 selection coefficient and the estimated SEM_{mut} described above. We then counted the fraction of

638 mutations whose 95% CI excludes zero. This yields a conservative estimate for our parameter of

639 interest, the total fraction of mutations that are deleterious (or beneficial), as it is designed to

640 indicate whether any particular mutation is deleterious (or beneficial), not to estimate the

641 proportion (which does not depend on unequivocally classifying any one individual mutation as

642 neutral or not). Nonetheless, we report this value as a conservative lower bound on the estimate

643 of non-neutral mutations.

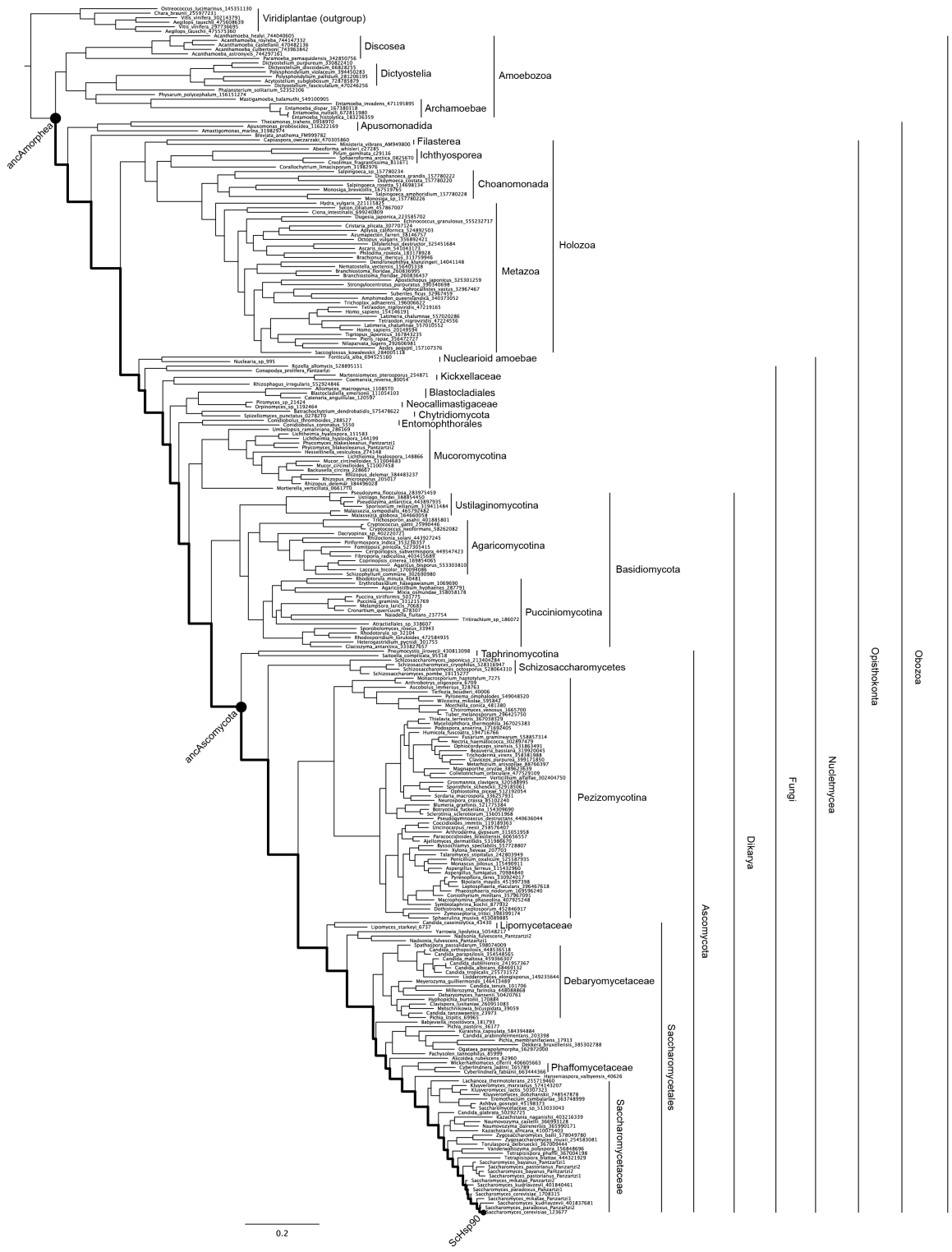


Figure S1. Hsp90 phylogeny. The maximum likelihood phylogeny of 267 Hsp90 protein sequences, with major taxonomic groups labeled. Taxon names indicate genus, species, and an accession number or sequence identifier; complete sequence identification information is given in Dataset S3. Nodes characterized in this study are shown as black dots; the trajectory studied is shown as a thick black line.

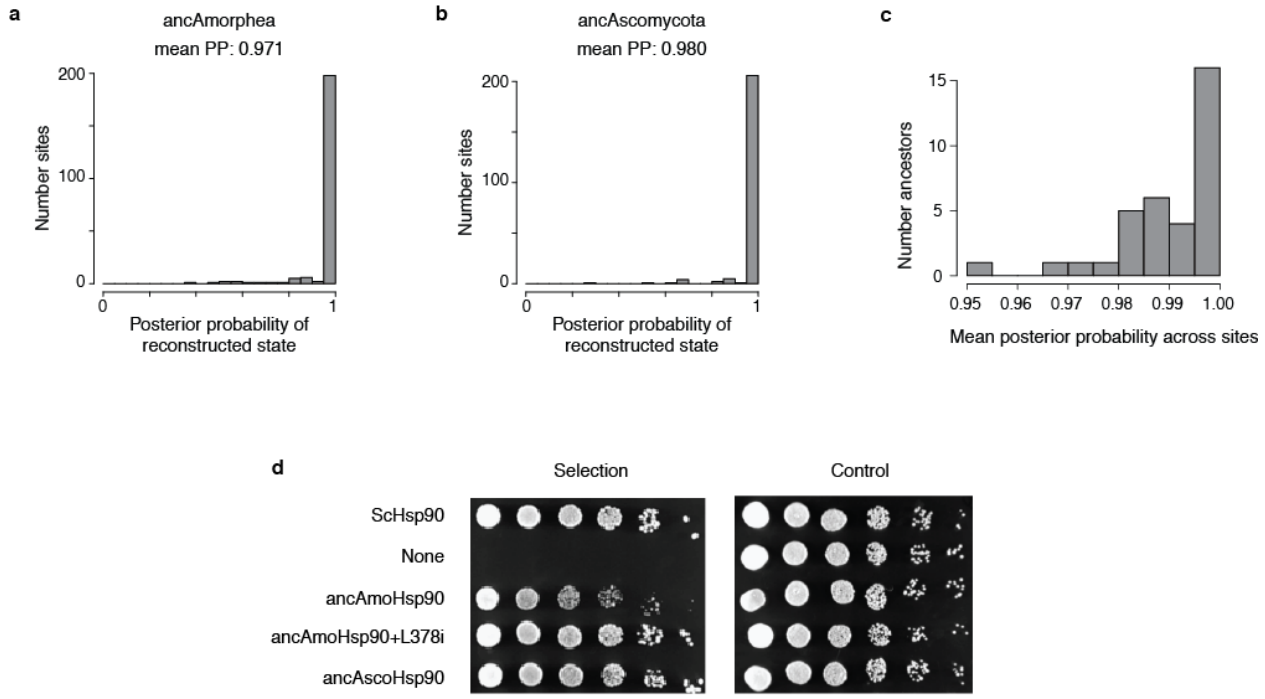


Figure S2. Ancestral Hsp90 sequences have high statistical support and complement yeast growth. a,**b**, For the ancestral NTD sequences reconstructed in this study, the distribution of posterior probability of ancestral states across NTD sites is shown as a histogram. The mean posterior probability of the most probable state across sites (mean PP) is shown for each ancestor. **c**, The distribution of mean PP for reconstructed ancestral sequences along the trajectory from ancAmoHsp90 to ScHsp90. **d**, Growth of *S. cerevisiae* Hsp90 shutoff strains complemented with ancestral Hsp90 NTD variants. Spots from left to right are 5-fold serial dilutions. Control plates represent conditions in which the native ScHsp90 allele is expressed. Under selection conditions, the native ScHsp90 allele is turned off, and growth can only persist when a complementary Hsp90 allele is provided. The ancAmoHsp90 NTD expressed as a chimera with the Sc middle and C-terminal domains exhibits a slight growth defect; this is rescued by adding an additional reversion to the ancAmoHsp90 state in the middle domain (L378i), which occurs on a middle domain loop that extends down and interacts directly with the N-terminal domain and contributes to the NTD ATP-binding pocket. We subsequently refer to ancAmoHsp90+L378i as ancAmoHsp90.

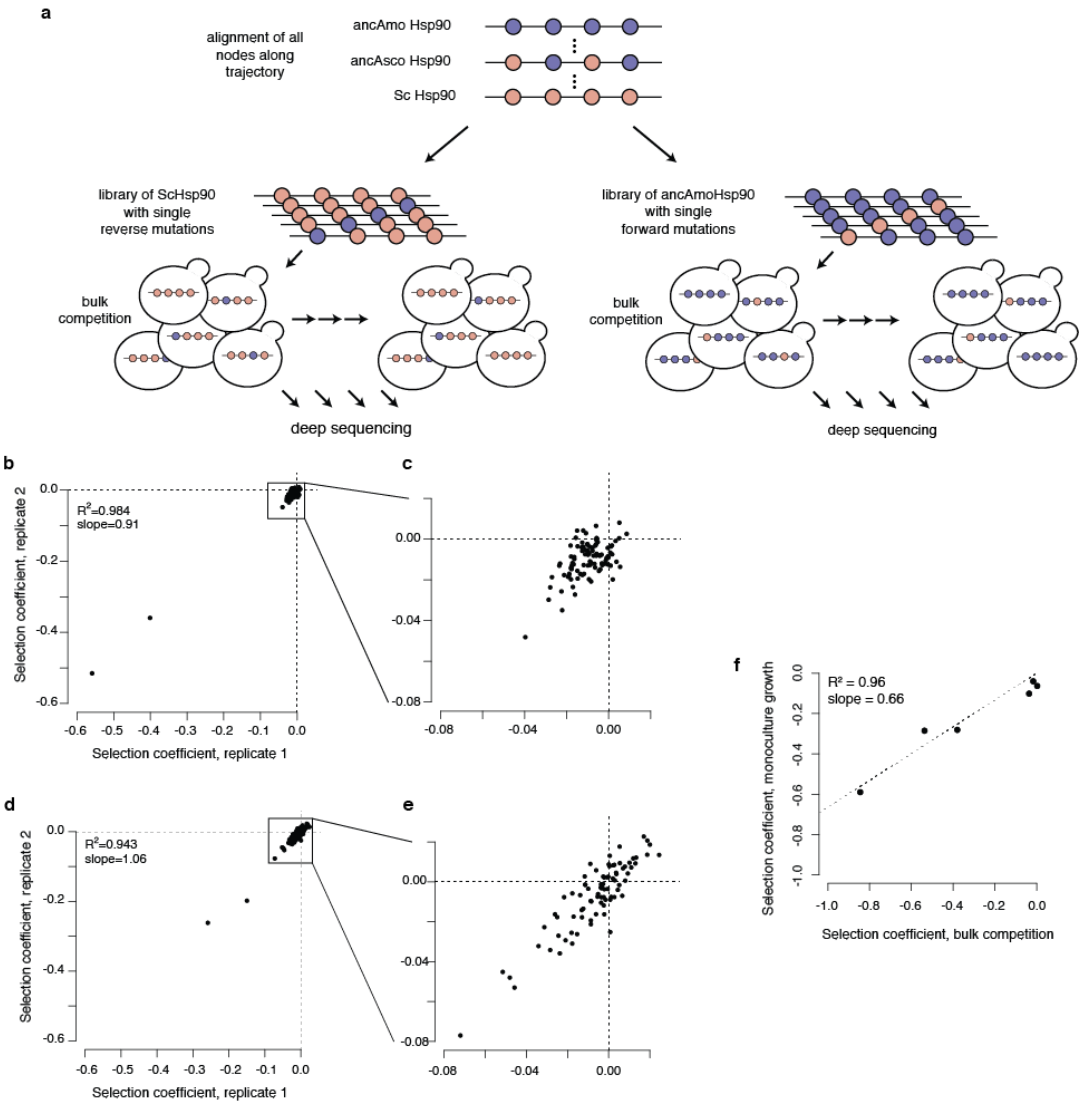


Figure S3. Experimental scheme and reproducibility. a, Experimental scheme for testing the fitness effects of individual mutations to ancestral states in ScHsp90 (left) or individual mutations to derived states in ancAmoHsp90 (right). An alignment of all ancestors along the focal trajectory was constructed to identify the trajectory of Hsp90 NTD sequence change from ancAmoHsp90 to ScHsp90. In each background, a library was constructed consisting of the wildtype sequence and all individual mutations to ancestral or derived states. This library was transformed into yeast, which grew through a bulk competition. The frequency of each genotype at each time point was determined by deep sequencing, allowing us to calculate a selection coefficient for each mutation relative to the respective wildtype sequence. **b**, Reproducibility in selection coefficient estimates for replicate bulk competitions of the ScHsp90 library. R^2, Pearson coefficient of determination. **c**, For visual clarity, zoomed in representation of the boxed region in (**b**). **d**, Reproducibility in selection coefficient estimates for replicate bulk competitions of the ancAmoHsp90 library. R^2, Pearson coefficient of determination. **e**, For visual clarity, zoomed in representation of the boxed region in (**d**). **f**, Correlation in fitness as measured via bulk competition or monoculture growth assay. R^2, Pearson coefficient of determination. The line was forced to go through (0, 0); when freely fit, the intercept term was not significantly different from zero.

a

Number components	log-likelihood	Number parameters	AIC	Estimated proportion deleterious	Estimated proportion beneficial
2	334.4	3	-662.7	1.000	0.000
3	341.0	6	-669.9	0.929	0.000
4	341.8	9	-665.6	0.932	0.000
5	343.2	12	-662.3	0.957	0.043

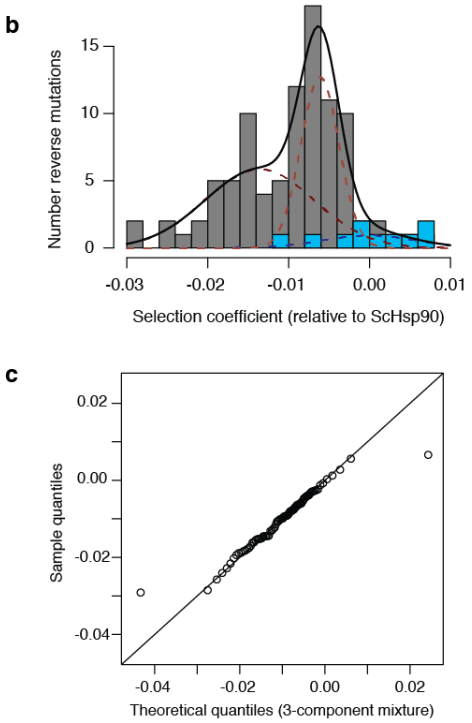


Figure S4. Estimating the proportion of mutations to ancestral states that are deleterious with a mixture model. a, Observed selection coefficients of reversions were fit to mixture models containing a variable number of Gaussian distributions; in each case, one distribution is fixed to have the mean and standard deviation of the sampling distribution of independent wildtype ScHsp90 sequences present in the library, the mixture proportion of which is a free parameter; each additional mixture component has a free mean, standard deviation, and mixture proportion. The empirical data were best fit by a 3-component mixture model, as assessed by AIC. Estimated proportion deleterious (beneficial) comes from summing the mixture proportions of components centered below (above) zero. **b**, The best-fit mixture model. Gray bars, observed distribution of selection coefficients of ancestral reversions; blue bars, distribution of observed selection coefficients of wildtype ScHsp90 sequences present in the library. Black line, best-fit mixture model; red dashed lines, individual mixture components centered below zero; blue dashed line, wildtype mixture component. The area under the curve for each mixture component corresponds to the proportion it contributes to the overall mixture model. **c**, Quantile-quantile plot showing the quality of fit of the 3-component mixture model (*x*-axis) to the empirical distribution of selection coefficients of ancestral reversions (*y*-axis). The mixture model assigns more extreme selection coefficients to the tails than is observed in the empirical distribution, but provides a reasonable fit along the bulk of the distribution.

a

Method	Proportion deleterious, Sc	Proportion neutral, Sc	Proportion beneficial, Sc	Proportion deleterious, ancAmo	Proportion neutral, ancAmo	Proportion beneficial, ancAmo
Mixture model	0.93	0.07	0.00	0.53	0.32	0.15
Count	0.95	0.00	0.05	0.66	0.00	0.34
Empirical Bayes	0.78	0.18	0.03	0.54	0.19	0.26
Confidence Interval	0.54	0.45	0.01	0.48	0.25	0.26

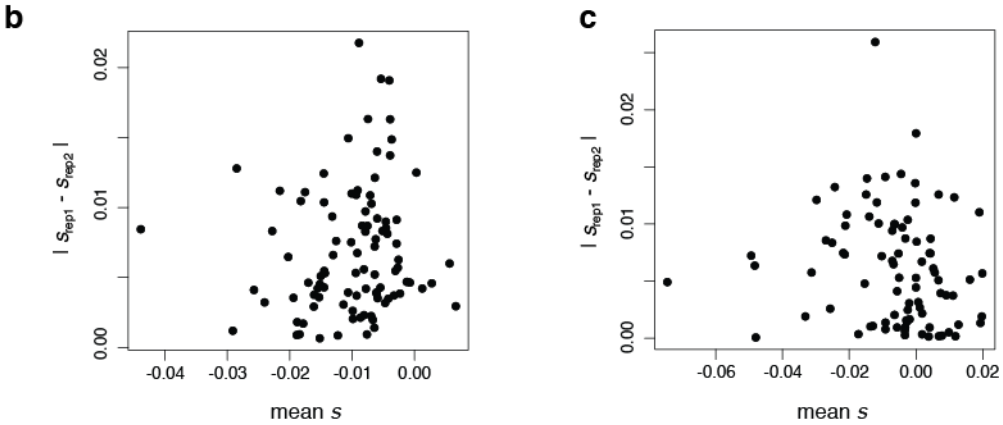
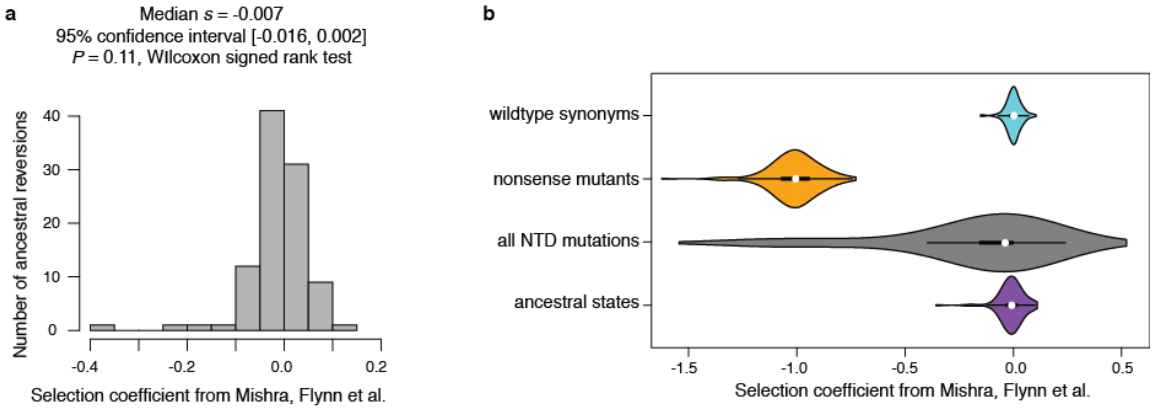


Figure S5. Alternate approaches for estimating the proportion of mutations that are deleterious. a, The estimated proportion of mutations that are deleterious, neutral, or beneficial in each background, as determined by each of four statistical methods. See SI Methods for descriptions of each method. **b,c**, Experimental errors are unbiased with respect to the observed selection coefficient. For the ScHsp90 (**b**) and ancAmoHsp90 (**c**) backgrounds, the absolute difference in s as determined in each replicate is shown versus their mean. In each background, there is no significant linear relationship between experimental error and s_{obs} (P = 0.27 and 0.24, respectively, Pearson's correlation).



c

	Spearman's rank correlation, $s > -0.1$		Spearman's rank correlation, all mutations	
Properties of sites	rho	*P*-value	rho	*P*-value
Average effect of all mutations at position (Mishra, Flynn et al.)	0.267	0.009	0.308	0.002
Relative evolutionary rate	0.294	0.004	0.326	0.001
Relative solvent accessibility (PDB 2CG9)	0.331	0.002	0.313	0.003
Distance to ATP gamma-phosphate	0.380	0.0003	0.395	0.0001
Properties of mutations				
BLOSUM62 score	0.048	0.65	0.077	0.45
Difference in hydrophobicity	-0.209	0.04	-0.178	0.08
Absolute difference in hydrophobicity	-0.121	0.24	-0.136	0.18
Difference in side chain volume	0.134	0.19	0.129	0.21
Absolute difference in side chain volume	-0.042	0.68	-0.072	0.48
Evolutionary distance from ScHsp90 to most recent ancestor with ancestral state	0.020	0.84	0.005	0.96

Figure S6. Ancestral states are deleterious in yeast Hsp90. a, The signature of deleterious ancestral states is present in the independent but lower-resolution dataset of Mishra, Flynn et al. (46). For each mutation to an ancestral state, the selection coefficient as determined by Mishra, Flynn et al. is shown. The median selection coefficient is -0.007, close to that estimated in the current study; however, this median selection coefficient is not significantly different than zero ($P = 0.11$). Because Mishra, Flynn et al. tested a much larger panel of mutations (all single mutations across the entire NTD), experimental variability of estimated selection coefficients was much larger, possibly explaining the lack of significance of this result in this dataset. **b**, Violin plots show the distribution of mutant effects in the dataset of Mishra, Flynn et al. (46). Ancestral states are less detrimental than the average random mutation in the NTD ($P = 3.5\times10^{-9}$, Wilcoxon rank sum test with continuity correction). **c**, Reversions exhibit properties typical of genuinely deleterious mutations. For various properties of sites at which we measured the fitness of ancestral variants (top) or properties of the specific amino acids mutated (bottom), we asked whether there was a significant correlation between the property and the selection coefficients of mutations via Spearman's rank correlation. Ancestral states tend to be more deleterious at positions that are less robust to any mutation, evolve more slowly, are less solvent accessible, and are closer to the gamma-phosphate of bound ATP. These properties are not completely independent; for example, there is a significant positive correlation between relative solvent accessibility and distance to ATP gamma-phosphate. Biochemical properties particular to the amino acid states in each mutation are generally not significantly correlated with the selective effect. Furthermore, we see no evidence for older states being more entrenched, as has been observed by others (1, 4, 14).

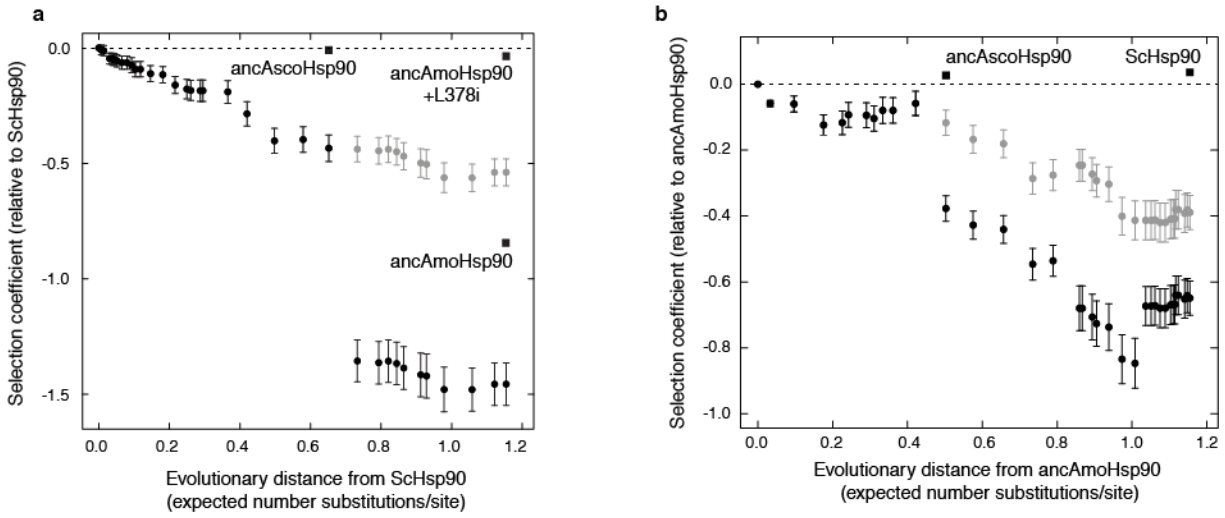
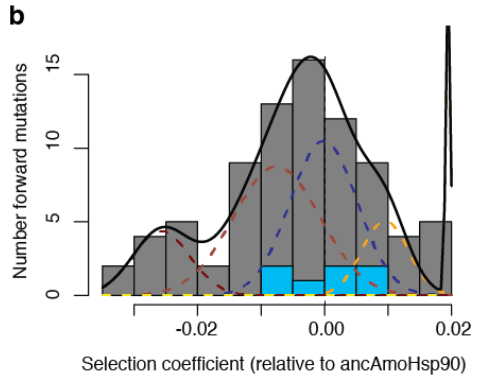


Figure S7. Fitness effects of historical substitutions are modified by intramolecular epistasis. Each black circle represents an ancestral protein along the trajectory from ancAmoHsp90 to ScHsp90. Position along the *x*-axis shows the evolutionary distance that separates it from ScHsp90 (**a**) or ancAmoHsp90 (**b**); *y*-axis position shows the predicted selection coefficient assuming no epistasis relative to ScHsp90 (**a**) or ancAmoHsp90 (**b**). Predicted selection coefficients were calculated as the sum of individual selection coefficients for all sequence differences present in its sequence as measured in ScHsp90 (**a**) or ancAmoHsp90 (**b**). Error bars show the 95% confidence interval for the predicted value, calculated by propagating the standard errors of individual site-specific selection coefficient measurements. Light gray dots show the same data, but excluding the effects of the two strongly deleterious outliers in each library. Labeled squares indicate experimentally determined selection coefficients for complete genotypes: ancAscoHsp90, ancestral Ascomycota (fitness determined via monoculture growth); ancAmoHsp90, ancestral Amorphea (fitness determined via bulk competition); ancAmoHsp90+L378i, ancAmoHsp90 with a candidate epistatic substitution in the Middle Domain also reverted to its ancAmorphea state (fitness determined via bulk competition). Dashed line, $s = 0$.

a

Number components	log-likelihood	Number parameters	AIC	Estimated proportion deleterious	Estimated proportion beneficial
2	243.9	3	-481.7	0.766	0.000
3	245.5	6	-479.0	0.910	0.000
4	246.2	9	-474.4	0.376	0.133
5	252.7	12	-481.5	0.493	0.162
6	252.7	15	-475.5	0.493	0.162



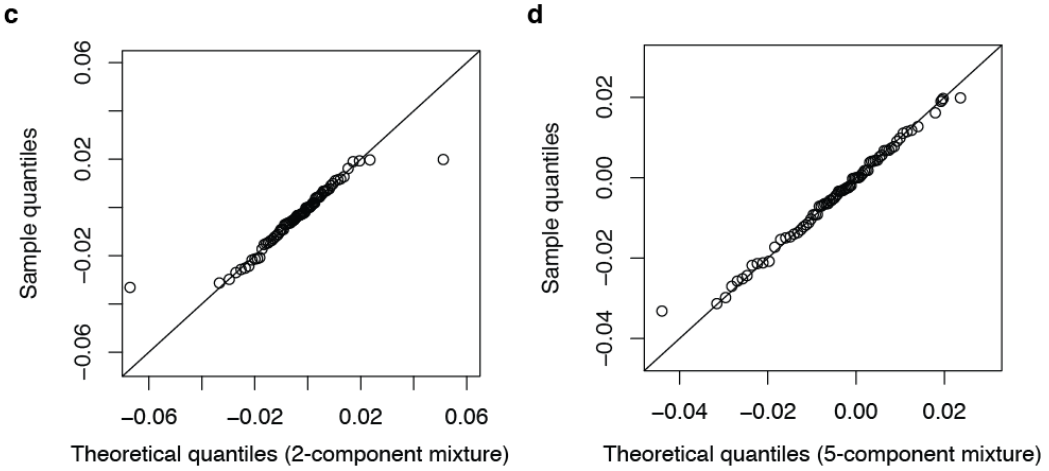


Figure S8. Estimating the proportion of mutations to derived states that are deleterious with a mixture model. a, The distribution of selection coefficients of mutations to derived states was fit by mixture models containing a variable number of Gaussian distributions; in each case, one distribution is fixed to have the mean and standard deviation of the sampling distribution of independent wildtype ancAmoHsp90 alleles in the library, the mixture proportion of which is a free parameter; each additional mixture component has a free mean, standard deviation, and mixture proportion. The empirical data were best fit by a 2-component mixture model, as judged by AIC, with a 5-component mixture being almost equally well fit; the 5-component mixture resulted in a more conservative estimate of the proportion of mutations that were deleterious than the 2-component mixture, and so was chosen despite the AIC difference of 0.2. Estimated proportion deleterious (beneficial) comes from summing the mixture proportions of components centered below (above) zero. **b**, The fit of the 5-component mixture model. Gray bars, distribution of selection coefficients of mutations to derived states; blue bars, distribution of selection coefficients of independent ancAmoHsp90 alleles present in the library. Black line, five-component mixture model. Red dashed lines, individual mixture components centered below zero; blue dashed line, wildtype mixture component; yellow dashed lines, individual mixture components centered above zero; relative integrated areas of mixture components correspond to the relative proportions they contribute to the overall mixture model. **c**,**d**, Quantile-quantile plot showing the quality of fit of the 2-component (**c**), or 5-component (**d**), mixture models (*x*-axis) to the empirical distribution of selection coefficients of mutations to derived states (*y*-axis).

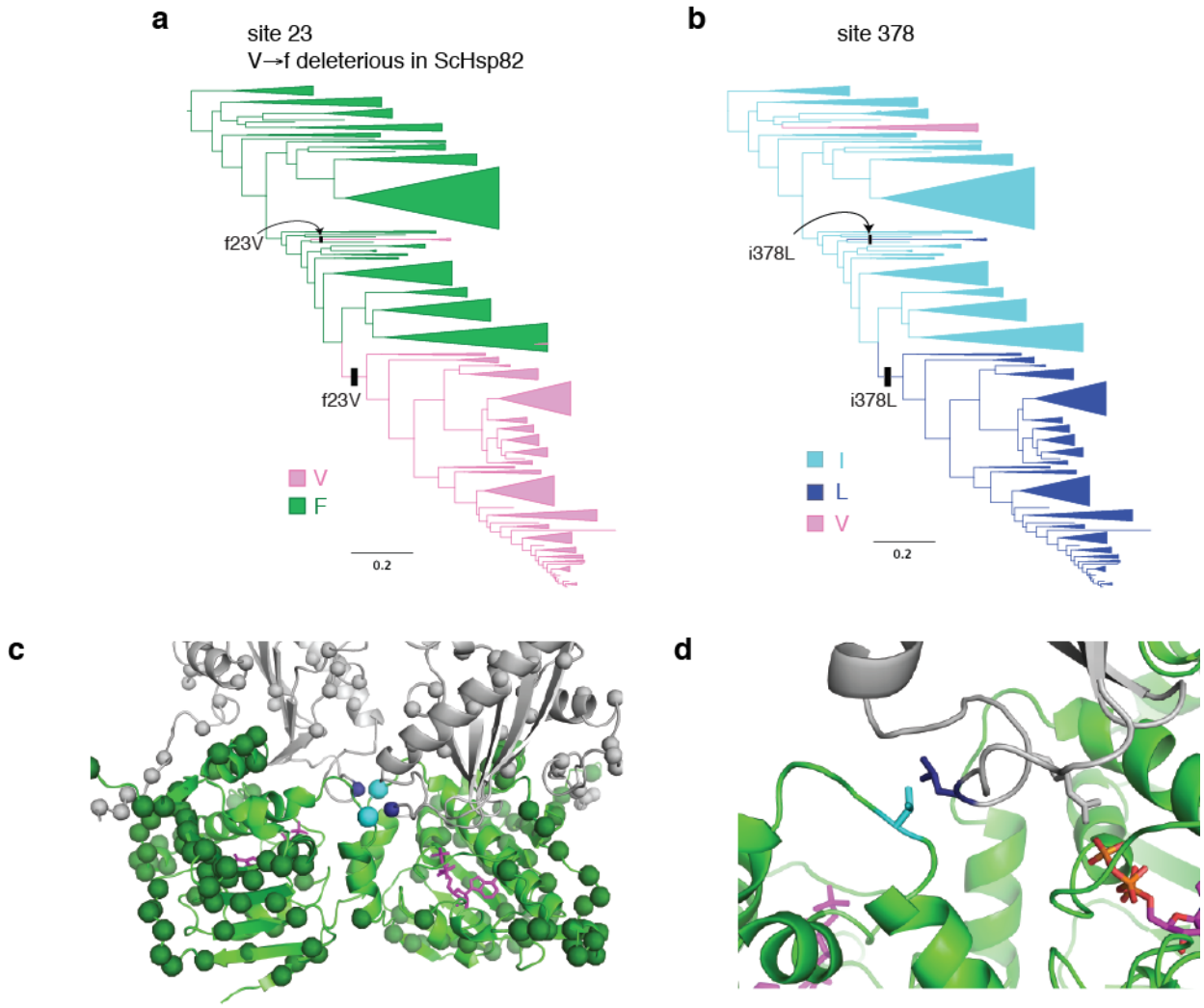


Figure S9. The deleterious V23f reversion is ameliorated by L378i. a,**b** Character state patterns at sites 23 (**a**) and 378 (**b**). On the lineage to ScHsp90, f23V co-occurred with i378L before the common ancestor of Ascomycota. The same two substitutions also co-occur on an independent lineage on this phylogeny (Kickxellaceae fungi), and in the distantly related Rhodophyta red algae (not shown). **c**, The locations of sites 23 and 378 on the ATP-bound Hsp90 dimer structure (PDB 2CG9). Cyan spheres, site 23; dark blue; site 378; dark green, other variable NTD sites; gray, other variable middle and C-terminal domain sites. Magenta sticks, ATP. **d**, Zoomed view of sites 23 and 378. These side chains are in direct structural contact, and may be important for the positioning of the middle domain loop that bears R380 (gray sticks), which forms a salt bridge with the ATP gamma-phosphate and is critical for ATP binding and hydrolysis (34, 58).

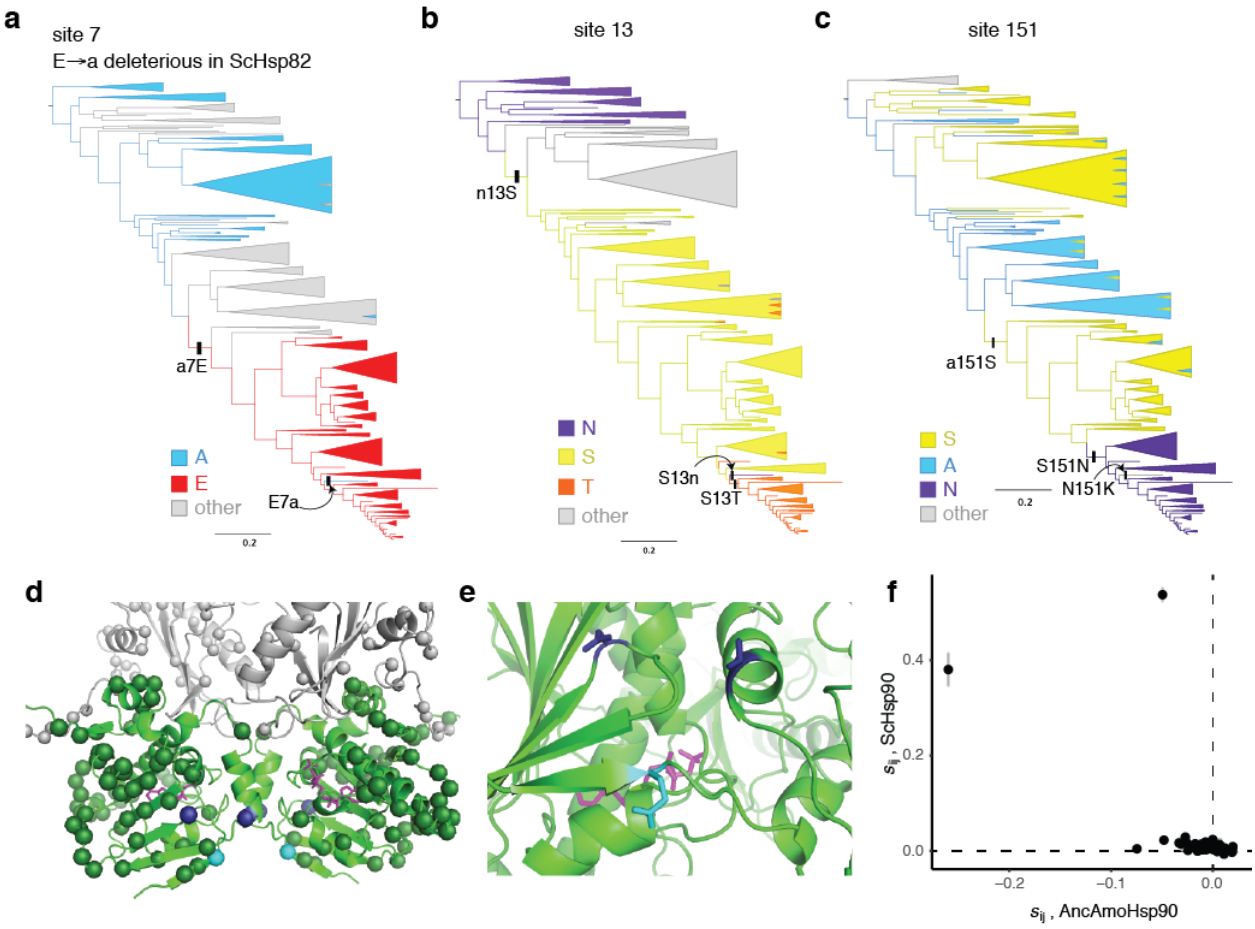


Figure S10. The deleterious E7a reversion is partially ameliorated by N151a or T13n. a,**b**,**c**, Character state patterns at sites 7 (**a**), 13 (**b**) and 151 (**c**). On the trajectory to ScHsp90, a7E occurred before the common ancestor of Ascomycota, then later reverted in the lineage leading to *Ascoidea rubescens* (arrow)*;* on this latter lineage, site 13 also reverted to the ancestral state asparagine, and site 151 substituted to a third state lysine. **d**, The locations of sites 7, 13, and 151 on the ATP-bound Hsp90 structure (2CG9), represented as in Fig. S9c. Cyan spheres, site 7; dark blue, sites 13 and 151. **e**, Zoomed in view of sites 7, 13, and 151. These side chains are not in direct physical contact; however, site 7 is on a beta strand that undergoes extensive conformational movement when Hsp90 converts between ADP- and ATP-bound states. **f**, The same plot as Fig. 5c is shown, including the two strongly outliers V23f and E7a. See Fig. 5c legend for details.