



Published in final edited form as:

Nature. 2014 August 14; 512(7513): 203–207. doi:10.1038/nature13410.

Historical contingency and its biophysical basis in glucocorticoid receptor evolution

Michael J. Harms^{1,3} and Joseph W. Thornton^{2,3}

¹Institute of Molecular Biology & Department of Chemistry, University of Oregon, Eugene, OR 97403 USA

²Departments of Human Genetics and Ecology & Evolution, University of Chicago, Chicago IL 60637 USA

³Institute of Ecology and Evolution, University of Oregon, Eugene, OR 97403 USA

Abstract

Understanding how chance historical events shape evolutionary processes is a central goal of evolutionary biology^{1–7}. Direct insights into the extent and causes of evolutionary contingency have been limited to experimental systems,^{7–9} because it is difficult to know what happened in the deep past and to characterize other paths that evolution could have followed. Here we combine ancestral protein reconstruction, directed evolution, and biophysical analysis to explore alternate “might-have-been” trajectories during the ancient evolution of a novel protein function. We previously found that the evolution of cortisol specificity in the ancestral glucocorticoid receptor (GR) was contingent on permissive substitutions, which had no apparent effect on receptor function but were necessary for GR to tolerate the large-effect mutations that caused the shift in specificity.⁶ Here we show that alternative mutations that could have permitted the historical function-switching substitutions are extremely rare in the ensemble of genotypes accessible to the ancestral GR. In a library of thousands of variants of the ancestral protein, we recovered historical permissive substitutions, but no alternate permissive genotypes. Using biophysical analysis, we found that permissive mutations must satisfy at least three physical requirements—they must stabilize specific local elements of the protein structure, maintain the correct energetic balance between functional conformations, and be compatible with the ancestral and derived structures—thus revealing why permissive mutations are rare. These findings demonstrate that GR evolution depended strongly on improbable, nondeterministic events, and this contingency arose from intrinsic biophysical properties of the protein.

Historians and evolutionary biologists have long wrestled with the idea that historical outcomes may hinge on chance events. How differently would the world have turned out if the Persian cavalry were present at the Battle of Marathon or the KT asteroid missed the earth? In biology, evolutionary trajectories driven solely by the deterministic force of natural

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

Correspondence to: Joe Thornton, joet1@uchicago.edu.

Contributions: MJH and JWT conceived the project, designed the experiments, and wrote the paper. MJH performed the experiments and analyzed the data.

selection will always produce the optimal accessible form, irrespective of chance events^{3,10}. In contrast, when non-deterministic processes such as drift play a strong role, the outcome depends on whatever chance events occur during evolution; if history could be set in motion again from some past starting point, very different results would likely unfold.

Recent studies show that the evolution of some protein functions was contingent on prior “permissive” mutations, which are functionally neutral in isolation but must be present for the function-altering mutations to be tolerated^{6,7,9,11–15}. Permissive mutations cannot be fixed by selection for the derived function and must therefore accumulate stochastically with respect to it. It remains unknown, however, how many permissive mutations could have enabled these evolutionary transitions and therefore whether the dependence on nondeterministic events is strong or weak. If the suite of potential permissive mutations is large, then many different evolutionary paths could enable the function-switching mutations, and the outcome of protein evolution would be only weakly contingent on its specific history. Conversely, if only a few mutations have the potential to permit the realized outcome, the probability that one of these would occur by chance would be very small, and the particular form and function achieved by the evolving protein would be strongly contingent on a low-probability event.

Understanding evolutionary contingency requires measuring the number of potentially permissive mutations and characterizing the factors that determine that number. Because history happened only once, this knowledge has been inaccessible for natural biological systems that evolved in the deep past. We addressed this issue by reconstructing ancestral proteins and subjecting them to directed evolution, a protein engineering strategy to efficiently characterize regions of protein sequence space with respect to some function of interest^{16,17}. We then employed biophysical analyses to explore the mechanistic factors that determined the number of permissive genotypes.

We previously characterized an evolutionary transition in the GR ligand-binding domain (LBD) of bony vertebrates and found that it was contingent on permissive mutations⁶. The LBD serves as an allosterically regulated transcriptional activator: hormone binding causes the “activation-function helix” (AF-H) to pack against the body of the protein, creating a new surface to which coactivator proteins bind and increasing transcription of nearby target genes^{18,19}. Using ancestral protein reconstruction, we previously found that the cortisol-specific GR evolved from a promiscuous ancient receptor (AncGR1) because of seven historical substitutions that are conserved in all extant GRs (Fig 1A, 1B)⁶. Of these, five function-switching substitutions (denoted F) eliminated the response to other hormones by dramatically repositioning a helix (H7) along one side of the binding cavity and establishing new cortisol-specific contacts. Introducing the F substitutions into AncGR1, however, renders the protein non-functional (Fig 1B). The remaining two historical substitutions (P) are permissive: they have no detectable effect on receptor function when introduced into AncGR1, but they allow F to be tolerated, yielding a cortisol-specific receptor (Fig 1B). Contingency is apparent, because selection for cortisol specificity could not deterministically drive acquisition of P, which was required for subsequent evolution of F and the domain's derived structure and function. It is unlikely that the GR passed through a non-functional intermediate containing F without P,²⁰ because the LBD remained

conserved, presumably due to functional constraints, during ~40 million years from the gene duplication event that generated it until the evolution of its new function (see ref²¹).

To understand the strength of contingency, we used directed evolution to estimate the frequency of alternative permissive mutations (P') in a large library of ancestral protein variants. Permissive mutations must fulfill two criteria (Fig 1C): they must rescue the nonfunctional AncGR1+F protein, allowing it to tolerate the F mutations, and they must be compatible with the ancestral sequence and function when introduced into AncGR1. To screen for rescuing mutations that meet the first criterion, we generated a large library of random mutants in AncGR1+F and characterized the resulting distribution of amino acid replacements (Extended Data Fig 1-2). We screened this library with a yeast two-hybrid system that linked growth to the cortisol-dependent interaction of the LBD with its coactivator peptide^{22,23}. We applied a liberal standard of growth to capture all rescuing mutations and verified their effects in both naïve yeast and a mammalian reporter assay (Fig 1C, Extended Data Fig 3). We screened ~12,500 clones, comprising an estimated 1,025 unique single replacements (71% of all accessible neighbors), 1,802 unique double replacements, and 825 higher-order combinations (3,650 total; see Methods, Extended Data Figs 1-2); the remainder were duplicate clones or contained nonsense, frameshift, or zero nonsynonymous mutations. We found no evidence of bias in the library (Extended Data Fig 2; Methods).

This screen identified 12 unique clones that improved AncGR1+F's cortisol sensitivity. These clones carried one, two, or three mutations each, but dissection of the combinations showed that functional effects were due entirely to single mutations that co-occurred with neutral changes (Extended Data Fig 4). In total, we found 10 unique single mutations that completely or partially rescued cortisol sensitivity. Two of these involved historically substituted residues: one was a historical P substitution (n26T, with upper and lower cases denoting derived and ancestral states), and the other reverted one F substitution to its ancestral state (I98f), conferring partial growth in the absence of permissive mutations (Extended Data Fig 3). Of the novel rescuing mutations, three (M222I, M222L, and L231M) improved the cortisol-sensitivity of AncGR1+F by 10-fold or more, an effect as great as historical P (Fig 2A). The remaining five improved cortisol sensitivity by 2- to 3-fold each, comparable to the individual members of P, but much less than the pair together (Fig 2A). To see if pairing any of the small-effect substitutions could recapitulate the effect of P, we generated all twofold combinations of the weak rescuing mutations. Only one pair (Q114L/M197I) affected cortisol sensitivity similarly to the historical set P (Fig 2A). The screen therefore recovered four alternate rescuing combinations – one double and three single mutants – indicating that rescuing mutations are rare, on the order of 4/3,650 or ~0.1%.

To determine whether the rescuing mutations discovered in the screen met the second criterion for permissive mutations – functional compatibility with the ancestral genetic background – we introduced them into AncGR1 and characterized their effects on hormone-dependent activation. Unlike the historical permissive mutations, all four rescuing mutations disrupted the ancestral protein's ligand-regulated transcriptional function. The large-effect rescuing mutations each caused transcriptional activation even in the absence of hormone and promiscuous activation in response to low doses of other steroids, such as progesterone

(Fig 2B), a natural hormone excluded by all known extant and ancestral corticosteroid receptors. The pair Q114L/M197I destroyed AncGR1's transcriptional function entirely, making it unable to activate reporter expression even with high hormone concentrations (Fig 2C).

Permissive mutations are therefore extremely rare. Among ~3,660 unique protein variants (~3,650 in the screened library plus 10 engineered double mutants), zero permissive genotypes were present. One permissive combination, the historical set P, exists in the universe of sequences near AncGR1, so we estimate an upper bound frequency of accessible permissive pathways of $< 1/3,660$ (0.03%). The total frequency is probably far lower, because knowledge of this one permissive pathway was not acquired by sampling. Further, our screen of double mutants was biased towards discovery of rescuing variants, because it included engineered combinations of all single mutations that had a detectable rescuing effect. The universe of possible variants containing two or more replacements is very large, so alternative permissive sets may exist; however, these genotypes would require multiple independent substitutions, and the joint probability of such events would be very low because they cannot be acquired deterministically by selection for the derived function. A permissive mutation might conceivably be subject to selection for some other function; however, unless the selected and derived functions are correlated, the probability that selection would deterministically fix a compound permissive genotype is extremely low. Evolution of the F mutations was therefore strongly contingent on prior low-probability events.

To understand the mechanisms that make permissive mutations both necessary and rare, we characterized the biophysical effects of F, P, and the four sets of rescuing but non-permissive mutations. Permissive mutations are often thought to act via effects on the global stability of folding: function-switching mutations destabilize a protein, making it prone to degradation and aggregation, but permissive mutations increase stability and offset this effect^{13,15,24,25}. Structural considerations suggested that a stability tradeoff might explain the effects of F and P. The F mutations cause a 3 Å shift in the position of H7 relative to H10 and the ligand, disrupting numerous contacts; they also open empty space between the ligand and helix H3 and remove a hydrogen bond from the key loop that connects AF-H to H10^{21,26}. In contrast, the P mutations add favorable interactions—both a new hydrogen bond and improved packing interactions—in the crystal structure and in molecular dynamics (MD) simulations (Extended Data Fig 5). To elucidate the effects of F and P on stability, we measured the midpoint of irreversible thermal denaturation (T_m) of steroid-bound AncGR1 containing each of the historical F and P mutations. As expected, each F mutation except I111Q was destabilizing (Extended Data Fig 6A), and the P mutations were stabilizing (Fig 3A).

Although these data are consistent with the global stability model, several other observations are inconsistent with it. First, the F and P mutations did not affect expression in mammalian cells as measured by western blot (Extended Data Fig 6B), indicating that AncGR1-F is functionally compromised, not subject to degradation or aggregation because of reduced stability. Second, under the global stability model, rescuing mutations should be more frequent than we observed. The global model predicts that any stabilizing mutation should

be permissive^{24,25}, and it is estimated that 1-10% of mutations are stabilizing²⁷; however, only ~0.1% of our library was rescuing, and permissive mutations were even more rare. Third, the global stability model predicts that any rescuing mutation should also be permissive, but we found that several rescuing mutations were deleterious when introduced without the function-switching mutations. Finally, the rescuing mutants all increased the T_m of AncGR1+F more than they did in AncGR1, suggesting a specific epistatic effect rather than generic compensatory mechanism (Fig 3B, Extended Data Table 1). These observations all indicate that permissive mutations must do more than simply increase global stability.

To understand the requirements that permissive mutations must fulfill, we first examined the location of permissive and rescuing mutations in the protein's structure. Under the global stability model, a stabilizing mutation should be permissive irrespective of its location^{24,28}. In contrast, the permissive and rescuing mutations exhibited a striking structural distribution, occurring in two distinct clusters near the F mutations: "pocket" substitutions bordering the ligand cavity, and "AF-H" substitutions at the interface between AF-H and the rest of the protein (Fig 3C). Both the ancestral crystal structures and MD simulations show that the historical P mutations yield new favorable contacts that involve the same structural elements destabilized by F (Extended Data Fig 5). Specifically, Thr26 strengthens a hydrogen bond connecting H3 to the H10/AF-H loop, compensating for the loss of a hydrogen bond in this loop due to F mutation s212. Leu105 improves packing interactions between H3 and H7, apparently compensating for the effects of the other F mutations on the interactions among H3, H7, and the ligand. Similarly, all rescuing mutations we discovered in our screen improve packing interactions involving AF-H or H7 (Figs 4, Extended Data Figs 7-8).

These observations suggest that permissive mutations must stabilize specific local structural elements destabilized by F, rather than generically modulating global stability. To test this hypothesis, we used the structure to identify a potentially stabilizing pair of mutations (E165A and K168E) ~25 Å distant from the ligand pocket and AF-H (Fig 3C). We introduced them into AncGR1+F and found that they raised T_m by 1.4°C; rather than rescuing function, however, they impaired AncGR1+F's cortisol sensitivity by ~10-fold (Fig 3D). These data confirm that increasing global stability is not sufficient to yield a permissive effect and point to a biophysical requirement that limits the number of permissive mutations: they must exert specific local rather than generic global effects on protein stability.

This requirement explains why rescuing mutations were few, but it does not explain why they were functionally incompatible with AncGR1, suggesting that further biophysical requirements limit the number of permissive mutations. To elucidate these requirements, we first examined the mechanisms by which the large-effect rescuing mutations make the ancestral protein super-active. All three increased the stability of both AncGR1 and AncGR1+F (Fig 3A,3B) and are clustered on AF-H, suggesting they exert their effect by disrupting ligand-induced allosteric regulation of this helix's position (Fig 3C), which differentiates inactive and active conformations. For a properly regulated receptor without ligand, the inactive conformation is more stable than the active conformation and thus the dominant species (Fig 4A); binding of hormone stabilizes the active conformation, causing it to become dominant. To test whether the AF-H mutations unconditionally stabilized the active conformation, we performed MD simulations of these mutations in AncGR1 in the

absence of ligand. As predicted, M222I and M222L improve hydrophobic packing between the active position of AF-H and H3 (Fig 4B, 4C), and L231M introduces a new sulfur- π interaction, anchoring AF-H in the active position against H10 (Extended Data Fig 7). Stabilizing the active conformation relative to the inactive conformation is expected to increase the proportion of the protein in the active conformation—explaining why these mutations impart activity in the absence of ligand and make the receptor highly sensitive to formerly weak ligands (Fig 4A). These observations point to a second limiting requirement: permissive mutations must not alter the energetic balance between functional conformations of the protein. That is, they must stabilize the “right” portions of the protein without stabilizing the “wrong” portion. The global stability model does not account for these constraints because GR function depends not only on the stability of folded versus unfolded or misfolded forms but also on the stabilities of active versus inactive conformations in both the presence and absence of ligand.

Finally, we examined why the rescuing pair Q114L/M197I renders the ancestral protein non-functional (Fig 2C). These sites are near the ligand-binding pocket, facing each other on helices H7 and H10 (Fig 4D). In the presence of F, the two residues are slightly offset, and the rescuing states Leu114 and Ile197 improve hydrophobic packing between H7 and H10, explaining their observed positive effect on the derived protein's stability and sensitivity (Extended Data Fig 8). In the AncGR1 structure, however, the shifted position of H7 places these two residues directly across from each other: the large side chains of the rescuing residues clash and destabilize the H7/H10 interaction (Fig 4D). As predicted by this model, the pair of rescuing states increases the T_m of AncGR1+F but lowers that of AncGR1 (Fig 3B). These observations reveal a final requirement: permissive mutations must be compatible with the conformations of both the ancestral and derived proteins.

Evolutionary contingency has usually been discussed in terms of chance external forces, such as random extinction by asteroid impacts or climate change². Our results show that the internal organization of biological systems—in this case, a protein's structure and thermodynamics—can give rise to strong contingency during evolution. The F mutations that triggered GR's functional transition required permissive mutations to stabilize the specific local structural elements F destabilized, without disturbing the energetic balance between the receptor's functional conformations or clashing with ancestral or derived protein structures. Our data indicate that very few mutations can satisfy all these biophysical requirements, making GR's evolution dependent on rare, low-probability historical events.

Our findings point to strong contingency not only in the evolution of GR's specific sequence but also the protein's molecular form—the structural and mechanistic underpinnings that produce its function. GR's cortisol specificity was achieved by a unique repositioning of H7 and reorganization of numerous hormone contacts. If other F-like mutations exist that could produce a form and function similar to the modern GR's, these mutations would reorganize and destabilize the same local elements of the ligand-receptor complex. To be tolerated, these effects would have to be offset by permissive mutations. The permissive mutations, in turn, would be subject to the same biophysical constraints as the historical permissive mutations, because those constraints arise from the functional form itself and the fundamental architecture of the GR LBD. Our experiments establish that very few accessible

genotypes satisfy these constraints. Permissive sequence changes that could enable alternative ways of achieving a similar form and function – even using entirely different mutations -- would therefore be very rare, as well.

If evolutionary history could be replayed from the ancestral starting point, the same kind of permissive substitutions would be unlikely to occur. The transition to GR's present form and function would likely be inaccessible, and different outcomes would almost certainly ensue. Cortisol-specific signaling might evolve by a different mechanism in the GR, or by an entirely different protein, or not at all; in each case, GR -- or the vertebrate endocrine system more generally -- would be substantially different. Because GR is the only ancestral protein for which alternate evolutionary trajectories to historically derived functions have been explored, the generality of our findings is unknown. The specific biophysical constraints, and in turn the degree and nature of contingency, that shape the evolution of other proteins are likely to depend on the particular architecture of each protein and the unique historical mechanisms by which its functions evolved.

Complete Methods Description

Library generation and characterization

To interpret the number and types of clones found in the directed evolution screen, we had to quantify the number and types of clones in the initial library. We characterized three basic aspects of the library: the mutational characteristics of the enzyme, how mutations at the nucleic acid level translated to amino acid substitutions, and the sampling of the library. We generated the mutant library using the Genemorph II Domain Mutagenesis kit (Agilent). This kit uses a dNTP-limited PCR reaction with an error-prone polymerase, which means that altering the amount of template—and thus the number of rounds of PCR prior to running out of dNTPs—alters the number of mutations per clone. We first characterized the relationship between the concentration of LBD template and the number of mutations by generating libraries with template amounts ranging from 2-10 ng/μL, and then sequencing 5-95 clones from each library (Extended Data Fig 1A). We then used the 294 mutations seen in these clones to measure the mutation spectrum of the enzyme. We found close agreement between our measured and the mutation spectrum published in the kit manual (Extended Data Fig 1B).

Given the empirical mutation spectrum of the enzyme and the DNA sequence of the LBD, we could then calculate the expected number of amino acid substitutions at a given mutation rate. These predictions could then be tested using the same libraries we generated to correlate template amount and mutation rate. We found close agreement between the expected and observed amino acid substitutions (Extended Data Fig 1C). For the screen reported in the paper, we used the library highlighted with the box in Extended Data Fig 1C, which had an average mutation rate of 1.04 mutations per clone. We then simulated sampling, with replacement, a library with these characteristics. Given the genetic code, 1,440 single amino acid substitutions were accessible with a single base change from the initial DNA sequence. By counting the number of unique substitutions identified in the simulated screen, we could then estimate what accessible mutations would be observed for a given screen sample size (Extended Data Fig 1D). We also sequenced 95 random clones

taken from an unscreened mutant library and looked for deviation from a Poisson mutation process in the number of unique mutations seen and the number of clones containing one, two, three, and four mutations (Extended Data Fig 2).

Of the 12,500 clones screened (see *Screening pipeline*) we estimate that 3,975 (31.8%) contained no amino acid substitution; 1,975 (15.8%) contained an early stop or frame shift; 3,875 (31.0%) contained a single amino acid substitution; 1,888 (15.1%) contained two substitutions; 600 (4.8%) contained three substitutions; and 188 (1.5%) contained more than three substitutions. In total, 6,551 clones contained one or more substitutions without a frameshift or early stop codon. Because we sampled the library with replacement, the 3,875 single substitutions only sampled 1,025 unique single substitutions.

Analysis of library bias

Mutant libraries generated by error-prone PCR can be biased toward a small number of clones, thus limiting the number of unique clones sampled²⁹. We designed our mutation protocol to minimize this possibility. Bias is minimized by a high concentration of template, fewer rounds of amplification, and mixing replicate reactions^{29,30}. To minimize the effect of population bottlenecks and PCR drift, we started with 1,700 ng of template/reaction, corresponding to $\sim 10^{14}$ molecules. Our initial primer-to-template ratio was 20:1, leading to an exhaustion of primer after 4-5 rounds of amplification. We also diluted the effect of any stochastic PCR drift by performing 12-replicate error-prone PCR reactions and then pooling them for the final library.

To verify that this design successfully limited bias, we looked for evidence that our mutation process deviated from a Poisson-expectation. We sequenced 95 clones from a single error-prone PCR reaction and compared the result to simulated samples of a virtual library generated using a Poisson process. We generated a virtual library using $\lambda = 1.04$ (Extended Data Fig 1A), the empirically derived mutation spectrum of the enzyme (Extended Data Fig 1B), and the sequence of the gene being mutated. We then sampled 95 clones at random from this library and queried the number of times each unique mutation was seen, as well as the number of times each unique clone (combination of mutations) was seen. We repeated this process 1,000,000 times to calculate the expected distributions of the experimental observables in a 95-clone sample.

We first investigated whether the number of observed clones with zero, one, two, three and four mutations differed from our expectation (Extended Data Fig 2A). We could not reject the null hypothesis ($p = 0.238$). We then investigated whether we saw any of the individual mutations more often than expected (Extended Data Fig 2B). We saw no evidence ($p = 0.242$) that the experimental library differed from the expectation derived from the simulation.

Although these samples were insufficient to reject a Poisson process, we wanted to investigate our power to detect bias in a 95-clone sample of the library. We re-ran the simulations described above, but added a bias towards a particular mutation, capturing the scenario where a mutation occurs early and is then used as a template for subsequent reactions. This bias ranged from 0.0 (clone occurs no more often than by chance) to 1.0

(mutation occurs in every clone). We then estimated the probability of failing to observe a mutation four or more times—our experimental observation—given a particular bias (Extended Data Fig 2C). When bias is present at a level of 0.064 or greater, we expect to observe a single mutation at least 4 times with probability >0.95 . Because this was not observed, we reject the hypothesis that any clone is present with bias > 0.064 .

Even if present and undetected, bias at this level does not change the conclusions described in the manuscript. We estimate that we screened 3,660 unique variants, leading to an estimated permissive mutation frequency of $>0.03\%$. If 6% of the library was redundant, this alters our estimated screen size to 3,440, leaving our estimated permissive mutation frequency unchanged within the precision reported. Further, the maximum bias of 0.06 is a ~ 12 -fold overestimate of the degree of bias in the library because this 95-clone sample was performed on a single PCR reaction. In our experimental protocol, we pooled 12 such reactions, thus diluting any existing bias by a factor of 12. We therefore conclude that the effect of PCR bias on our results is minimal. We designed our screen to minimize bias and see no evidence that our mutation process differs from Poisson-expectation. The maximum bias consistent with our experimental observations does not alter the conclusions made in the paper.

Screening pipeline

A schematic of the screen pipeline is shown in Extended Data Fig 3. The Y2H screen was performed using the GAL4 Two-Hybrid Phagemid Vector Kit (Agilent). We used the YRG-2 cells and pBDGAL4 vector that shipped with the kit, cloning the LBD library into the pBDGAL4 vector. Professor Michael Stallcup generously provided the “Gal4AD-SRC1a.1236-1441” construct²² (pADGAL4-SRC), which contains nuclear receptor box 4 from the C-terminus of the SRC1a protein. Plasmid transformations were done according to the lithium acetate transformation protocol described in the kit. We used synthetic defined minimal media with appropriate amino acid dropout solutions to select for the pADGAL4-SRC plasmid (Leu(-)), the pBDGAL4-LBD plasmid (Trp(-)), or an interaction between the plasmid protein products (His(-)). We sequentially transformed the pADGAL4-SRC and pBDGAL4-LBD constructs.

We then optimized the growth and steroid concentrations to distinguish AncGR1+F and AncGR1+FP. The AncGR1+F genotype used in this study is the AncGR1 maximum likelihood reconstruction plus the five function-switching mutations previously reported (groups X and Y in ref. ⁶), plus five additional historical substitutions from the same interval, which slightly improve activation in the derived state without altering specificity (group W in ref. ²⁶). These W substitutions were included because in yeast the AncGR1+XY +P genotype alone cannot drive hormone-dependent growth in yeast, even at very high cortisol concentrations. With the W substitutions, the screen at 1 μM cortisol distinguishes the poorly growing phenotype of AncGR1+F and the more robust growth of AncGR1+FP (Extended Data Fig 3B).

To perform the screen, the yeast containing the pADGAL4-SRC plasmid were transformed with the AncGR1+F library. Most of the transformation volume was plated onto SD/Leu(-)/Trp(-)/His(-) + 1 μM cortisol plates; however, we took an aliquot to measure transformation

efficiency. To measure efficiency, we plated serial dilutions onto SD/Leu(-)/Trp(-) plates, which allow growth for any yeast that possess both the bait and target plasmids, regardless of interaction between them. We counted the colonies that grew in the serial dilution, and then fit a linear model to the volume vs. counts data to extract the number of cells/ μL . We estimate that we screened $12,485 \pm 220$ clones, which—according to our library statistics—would include $1,025 \pm 18$ and $1,802 \pm 90$ unique single and double mutations, respectively.

We designed our initial screen to minimize the number of false negatives by taking any colony, not matter how small, that grew within 5 days on the plate. The initial screen yielded 232 colonies that grew in 1 μM cortisol (1.9% of the transformants). This liberal initial screen resulted in a large number of false positives. To remove these, we followed this with a secondary screen. We picked each colony into 50 μL of sterile water, then pipetted 3 μL onto 4 Leu(-)/Trp(-)/His(-) plates: EtOH, 0.01, 0.1, and 1.0 μM cortisol. We then looked for colonies that grew better than AncGR1+F. 23 colonies did not grow at all (9.9%), 33 were not better than the AncGR1+F background (14.2%), 11 were constitutive (4.7%) and 165 gave cortisol-dependent growth (71.1%). We then used the “bust-n-grab” protocol³¹ to break open constitutive and cortisol-dependent colonies (176 total), followed colony PCR to amplify and sequence LBDs. When sequenced, only 110 colonies contained amino acid substitutions in the LBD, indicating that many of the yeast had adapted to the selection by some other means than mutations in the LBD. Of these 110, only 47 had unique sequence changes. Because of the extremely high false positive rate, we took the unique clones and back-transformed those preps into a naïve YRG-2 strain containing the pADGAL4-SRC plasmid. We then performed the secondary screen described above a second time on the naïve yeast. 15 clones (31.9%) gave no hormone dependent growth, 1 (2.3%) was constitutive, 18 (40.9%) gave unambiguously hormone dependent growth, and 10 (21.3%) were ambiguous.

To minimize false negatives, we again used liberal criteria, taking all constitutive, cortisol-dependent, and ambiguous clones to the next step. We sub-cloned their LBDs into the pSG5-DBD vector, and measured their sensitivity to cortisol in a transactivation assay. Unlike the Y2H assay, which will activate with any interaction between the LBD and SRC peptide, the functional assay requires a productive interaction and is therefore a more realistic test of LBD function. We assessed the significance of the results using a one-tailed t-test ($p < 0.05$) without multiple-testing correction to avoid type II errors and to maximize the number of clones we examined further. The t-test was performed on four or more independent biological replicates, all of which exhibited similar variances (Extended Data Fig 3F). We measured the change in cortisol sensitivity for all 26 suppressors; however, we found that only 10 clones exhibited significantly improved cortisol sensitivity.

Transactivation assay

The cortisol-dependent transcriptional activity of was assayed using a luciferase reporter system. We cloned the LBDs into the pSG5-DBD vector. 31 amino acids of the GR hinge containing the nuclear localization signal-1 were inserted between the GAL4 DBD and LBD to ensure nuclear localization and conformational independence of the two domains³². CHO-K1 (ATCC #CCL-61) cells were used to measure transactivation. A frozen stock,

created upon receipt of the cells, was used to restart the cell cultures every three months. Cells were grown in 96-well plates and transfected with 1 ng of receptor plasmid, 100 ng of a UAS-driven firefly luciferase reporter (pFRluc), and 0.1 ng of the constitutive phRLtk Renilla luciferase reporter plasmid, using Lipofectamine and Plus Reagent in OPTIMEM (Invitrogen). After 4 h, transfection medium was replaced with phenol-red-free α MEM supplemented with 10% dextran-charcoal stripped FBS (Hyclone). After overnight recovery, cells were incubated in triplicate with the hormone of interest from 10^{-12} to 10^{-5} M for 24 h, and then assayed using Dual-Glo luciferase (Promega). Firefly luciferase activity was normalized by *Renilla* luciferase activity. Dose-response relationships (EC₅₀ and maximum fold increase in activation) were estimated using nonlinear regression in R³³; fold increase in activation was calculated relative to vehicle-only (ethanol) control.

Protein expression/stability measurements

LBDs of interest were cloned into the pETMALc-H10T vector, which allows expression of the LBD as a His-tagged MBP fusion. We expressed the protein in BL21(DE3) Rosetta pLysS cells. Expression was induced during log-phase growth with the addition of 1 mM IPTG, 0.2% glucose, 2% EtOH, and 50 μ M 11-DOC or cortisol. Cells were incubated overnight at 16 °C and then harvested by centrifugation, then frozen at -20 °C. We lysed the cells using B-PER (Thermo Scientific) and then purified the MBP/LBD fusion by nickel affinity chromatography (HisTrap; GE Healthcare). We cleaved the fusion protein overnight using TEV protease in a buffer containing no imidazole. We then ran the cleaved products over the HisTrap column a second time. The LBD interacts non-specifically with the resin and can be eluted with \sim 10 mM imidazole, yielding 99% pure LBD. We added 50 μ M hormone, 0.04% CHAPS and 10% glycerol, then flash froze the protein by dropwise addition to liquid nitrogen.

We were unable to find reversible folding conditions for either temperature or chemical denaturation. We therefore measured the midpoint of irreversible thermal denaturation using a Jasco-815 circular dichroism spectrophotometer with a constant melting rate of 2 °C/min. We followed α helical content via CD at 225 nm. Protein concentration was 0.5 μ M in 25 mM sodium phosphate, 100 mM NaCl, 0.04 % CHAPS, 1 mM TCEP, and 5 μ M cortisol. All melts were done in triplicate (Extended Data Table 1).

Western blots

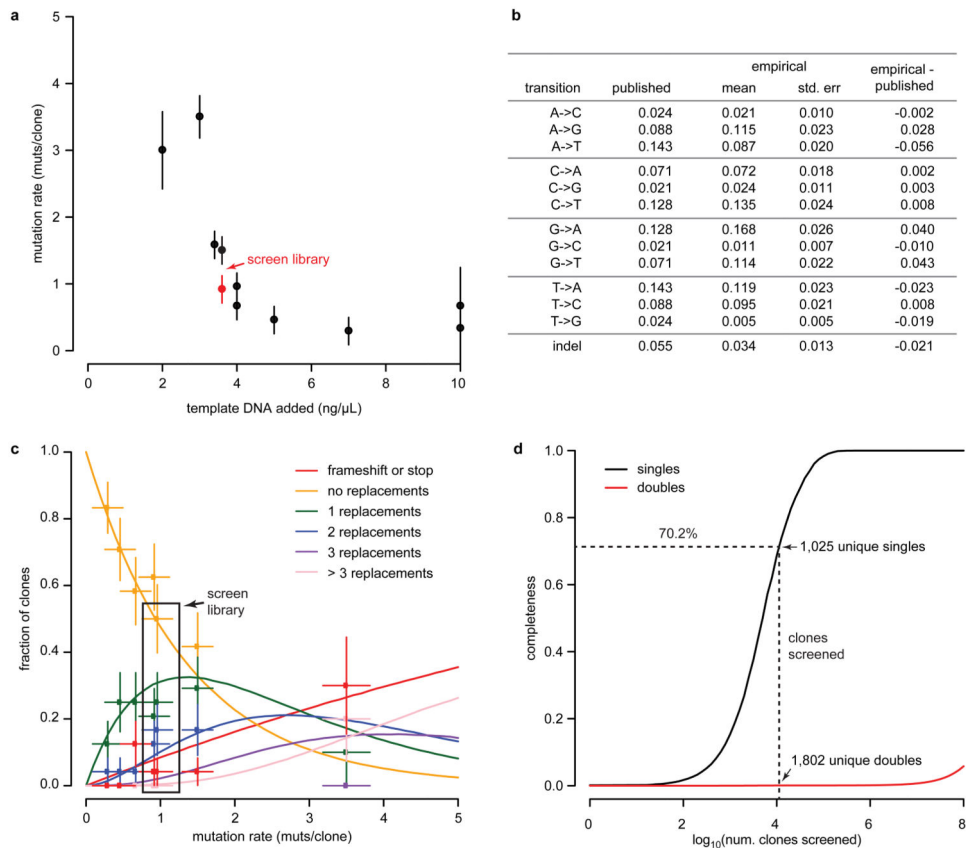
CHO-K1 cells were transiently transfected with plasmids containing the GAL4DBD-LBD fusion constructs used in the steroid sensitivity assays. After 24 hr, soluble proteins were extracted using RIPA buffer (Sigma, St. Louis MO) + 10ul PMSF, 10ul sodium orthovanadate, and 10 ul protease inhibitor. 25 μ g of protein (quantified by Bradford) was loaded onto a 12% Tris SDS-PAGE gel, followed by transfer to a PVDF membrane. Blot was blocked overnight at 4 °C in blocking buffer (50 mM Tris, 150 mM NaCl, 0.02% Tween, 5% dried milk, 2% BSA). Primary antibody was a rabbit polyclonal antibody to the GAL4DBD (Santa Cruz Biotechnology, SC-577); secondary antibody was goat anti-rabbit HRP. Bands were visualized by Luminol (1 min), followed by a 30 minute exposure on an GelDoc system.

Molecular dynamics simulations

Simulations were done in pairs to look for changes correlated with observed mutational effects. Simulations were started from crystal structures of AncGR1/11-deoxycorticosterone (3RY9)²¹ and, as a proxy for AncGR1+F, AncGR2/dexamethasone (3GN8)²⁶. The comparison between AncGR1+F and AncGR2 is useful because of their close functional and phylogenetic relationship^{6,26}. We further showed that the mutations under study have similar functional effects in both AncGR1+F and AncGR2 with P reverted to the ancestral state (AncGR2p; Extended Data Fig 5D). We ran simulations of AncGR2p, AncGR2p+P, AncGR2p+Q114L/M197I, AncGR2p+M222I, and AncGR2p+L231M. Appropriate amino acid replacements were made to the AncGR2 structure using PyMOL (Schrödinger LLC), and the rotamer that minimized steric clashes was chosen visually. Cortisol was placed in the pocket by minimizing the RMSD between the cortisol and dexamethasone structures. Initial cortisol coordinates were downloaded from PubChem (CID 5754). Cortisol structure and electron distributions were calculated using the 6-31G* basis set within the Firefly 7.1.G implementation of GAMESS-US^{34,35}. Partial charges were calculated using the Restrained ElectroStatic Potential (RESP) method using RED-III.5 with multiple ligand orientations³⁶. Topology files were generated using the SwissParam server³⁷.

Molecular dynamics simulations were performed using the CHARMM27 forcefield and TIP3P waters as implemented in GROMACS 4.5.5³⁸⁻⁴⁰. In all simulations, bonds were treated as constraints and fixed using LINCS⁴¹. Electrostatics were treated with the Particle Mesh Ewald model⁴², using an FFT spacing of 12 Å, interpolation order of 4, tolerance of 1e-5, and a Coulomb cutoff of 9 Å. van der Waals forces were treated with a simple cutoff at 9 Å. NaCl counterions were added to neutralize the system at a concentration of 100 mM. Calculations were done at 310K and 1 bar in the NPT ensemble using Nose-Hoover temperature coupling^{43,44} with a τ of 0.1 ps and isotropic Parrinello-Rahman pressure coupling^{45,46} with a τ of 1.0 ps and a compressibility of 4.5e-5 bar⁻¹. Each protein/ligand pair was equilibrated in a periodic water box 20 Å larger than the maximum protein dimension on each axis. The system was energy minimized, velocities were assigned from a Maxwell distribution, and the system was equilibrated for 1 ns with heavy protein atoms fixed. This was followed by a 100 ns equilibration using unrestrained MD. The last frame of this simulation was then used as input for independent, triplicate production MD simulations. New velocities were assigned to each replicate, followed by a second 1 ns position-restrained calculation. Each production run was 100 ns, with the first 1 ns discarded (the protein RMSD reached a plateau within 200 ps). The trajectory time step was 2 fs, with frames recorded every 5 ps. Final analyses were performed on frames taken every 12.5 ps. Analyses were performed using VMD 1.9.1⁴⁷—with its built-in TCL scripting utility—as well as a set of in-house Python and R scripts. For the L231M sulfur-pi analysis, we used geometric criteria of $R < 6 \text{ \AA}$ and $20^\circ < \theta < 60^\circ$ (see Extended Data Fig 7)⁴⁸.

Extended Data



Extended Data Figure 1. Quantitative characterization of the AncGR1+F mutant library

a, Relationship between the amount of DNA in the mutagenesis reaction and the final mutation rate. Each point is an independently generated library, with its mutation rate estimated by sequencing between 5 and 24 clones. Error bar shows the expected standard error for an estimate of the mean of a Poisson distribution with the observed mean given the number of clones sequenced, calculated using the `epicalc` package in the R statistical environment. The library used for the screen is highlighted in red. **b**, Table showing the frequency of each possible nucleotide transition as a proportion of all mutations in the library (empirical) and predicted by the manufacturer (published). The standard error for an estimate of a proportion p given n samples was calculated as $\text{stderr} = \sqrt{p^*(1-p)/n}$. **c**, Fraction of clones in a library containing 0, 1, 2, 3, or more amino acid replacements given varying total mutation rates; points show experimentally measured fractions, and lines show Poisson prediction. Error bars show standard errors, calculated as in panel a (for mutation rate) and panel b (for fraction). Box highlights frequencies of each class in the library used for the screen. **d**, Calculated library coverage for single (black) and double (red) substitutions for the library boxed in panel c. The dashed line shows the screening depth and completeness used in this study.

a Number of clones containing X mutations.

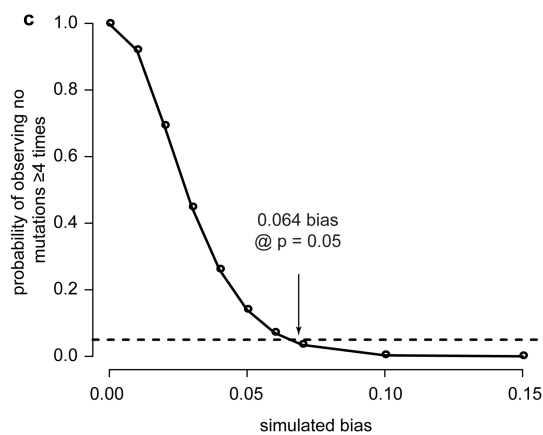
X =	observed	expected
0	27	33.578
1	46	36.920
2	16	18.153
3,4	6	7.939

χ^2 test (two-tailed, 3 degrees of freedom)
p for rejecting Poisson distribution: 0.238

b Number of mutations seen X times.

X =	observed	expected
1	84	93
2	6	3
3	1	0

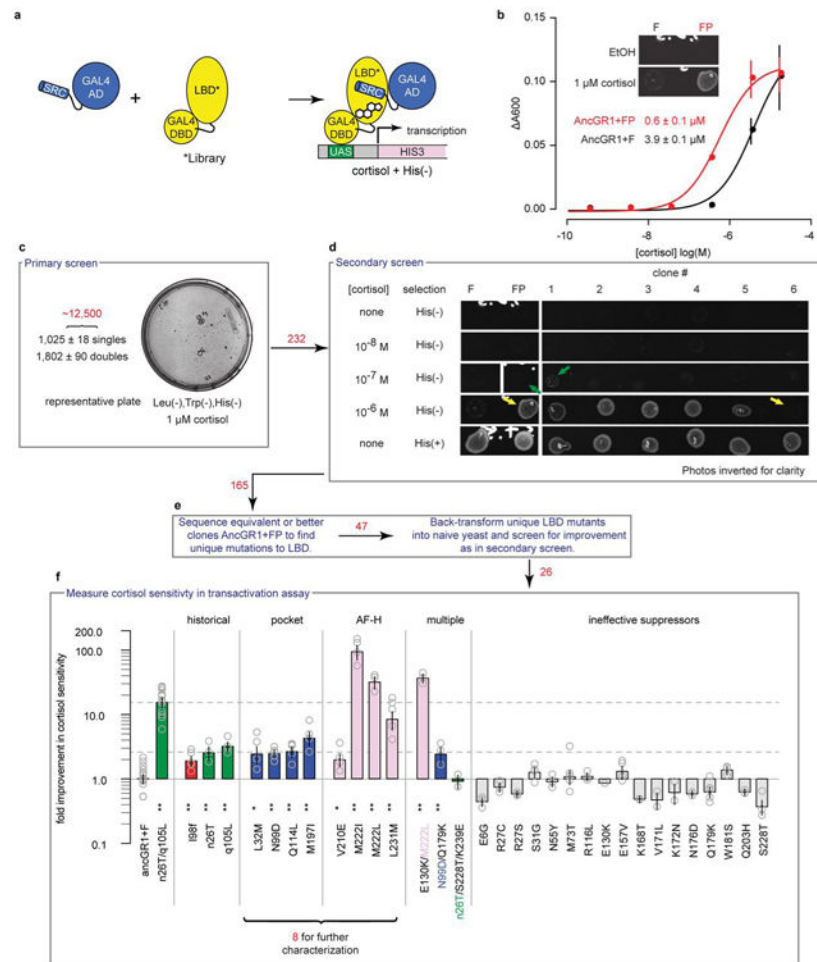
Fisher's exact test:
p for rejecting Poisson distribution: 0.242



Extended Data Figure 2. There is no detectable bias in the initial mutant library

Table a shows the number of clones containing X amino acid replacements in a 95-clone sample of the variant library: “experimental” shows the number of clones in each class observed in the actual library by sequencing, and “expected” shows the number recovered in simulations of samples of clones produced in silico by a Poisson mutation process with the same mutation frequency and spectrum as the experimental library (see methods for details). A χ^2 test (3 degrees of freedom) was used to determine whether the observations deviated from the Poisson expectations. Classes of clones with 3 or 4 replacements were pooled to maintain adequate counts per cell; no observations were made or predicted with more replacements. Table b compares the number of unique amino acid replacements in classes defined by the number of clones X containing that replacement in a 95-clone sample of the experimental and Poisson-simulated libraries. Because of the low expected counts, we employed Fisher's exact test for deviation from the Poisson expectation. **c**, Calculated probability of not observing a replacement in four or more clones out of 95 clones sampled (as occurred in our experimental sample of the library), given variable amounts of bias in the library, where bias ranges from 0.0 (no bias compared to Poisson expectation) to 1.0 (the same replacement is present in every clone). The probability drops below 5% at a bias of

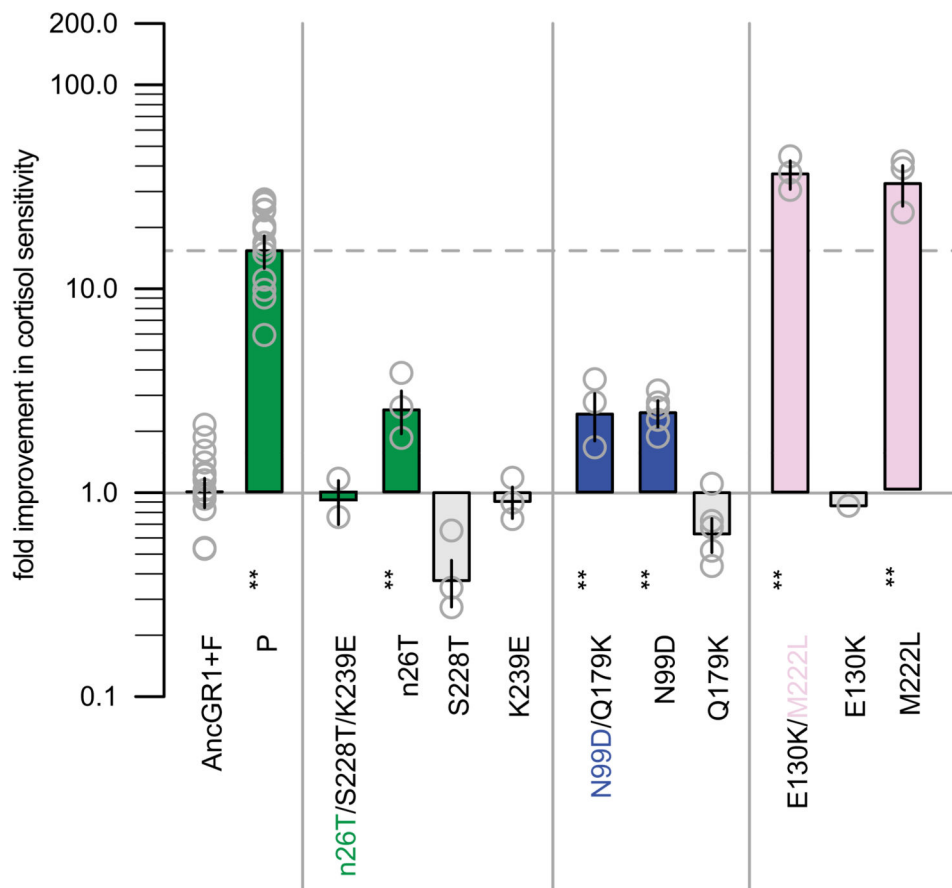
0.064, providing a reasonable upper-bound estimate for the degree of bias in the library given our observations.



Extended Data Figure 3. Experimental library screen pipeline

a, Schematic of the two-hybrid primary screen for cortisol-specific activation of a mutant library of receptor LBDs. Each LBD is fused to the GAL4-DBD and transformed into yeast along with the GAL4-AD activation domain fused to the SRC-1 coactivator peptide (which binds to the active conformation of the LBD) and a selective reporter construct expressing the HIS3 gene, which is required for growth in the absence of histidine. **b**, LBD genotypes with different cortisol sensitivities can be distinguished by their growth in the two-hybrid primary screen. Plot shows OD600 for yeast cultures as a function of cortisol concentration for AncGR1+F (black) and AncGR1+FP (red). Inset shows colonies of AncGR1+F and AncGR1+FP grown on plates with no hormone/vehicle only (top panel, EtOH) or 1 μ M cortisol (bottom panel). Points and error bars are mean and standard error from three technical replicates. This experimental result was reproduced many times with independent cultures. **c-f**, Full screen pipeline. Arrows denote the pipeline, with the number of positive clones recovered at each step shown in red. **c**, Representative plate from the primary screen for mutations that rescue AncGR1+F at 1 μ M cortisol. **d**, Representative clones tested in the secondary screen for dose-responsive growth with increasing cortisol concentration. Each

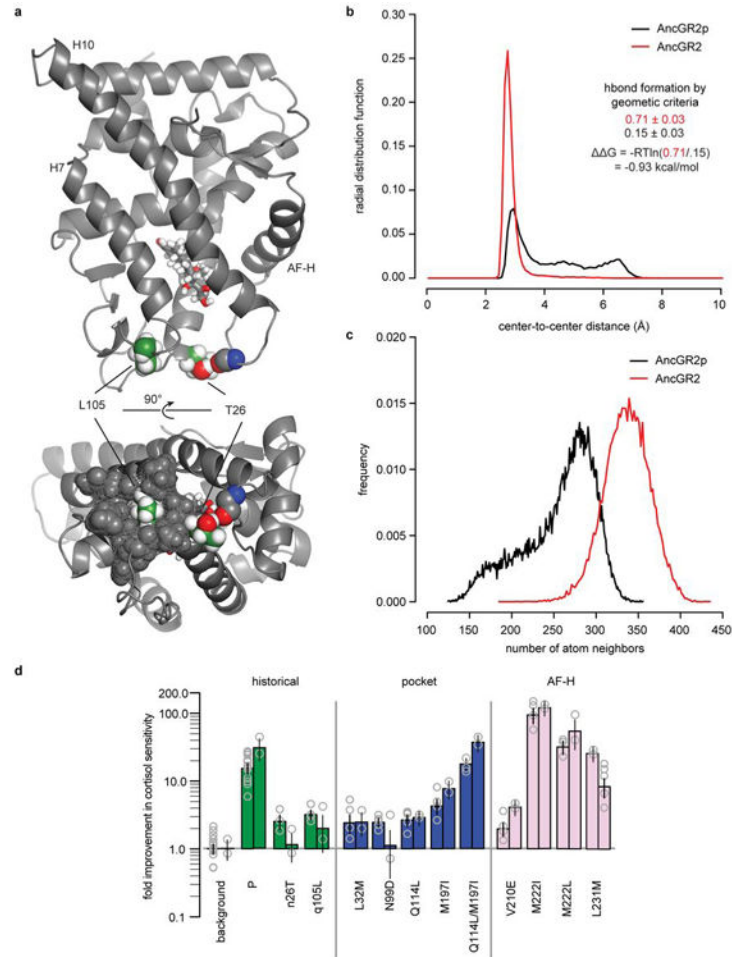
row shows growth of 6 different clones from the primary screen and two reference clones; different rows show growth at increasing cortisol concentrations. The bottom row shows growth with no selection for receptor activity when histidine is supplied. Clone 1 grows better than genotype AncGR1+FP containing historical permissive mutations (green arrows); clone 6 grows worse than AncGR1+FP (yellow arrows). **e**, Two quality control steps were employed after the secondary screen to reduce false positives. **f**, Fold change in cortisol sensitivity measured using a luciferase reporter assay in mammalian cells for the 26 clones identified in the multistage screen. Sensitivity is defined as the ratio of the mutant and wildtype EC_{50} s. Columns and error bars indicate mean and standard error of experimental replicates (gray circles). Historical P substitutions are shown with green bars; reversal of a historical F substitution is in red. Rescuing mutations are colored by their location on the protein structure: near the ligand pocket (blue) or activation function helix AF-H (pink) (see Fig 3C). Mutations that did not improve cortisol sensitivity in this assay are gray. Dots show statistical significance of the difference in fold-activation relative to AncGR1+F (one dot, $p < 0.05$, two dots, $p < 0.01$).



Extended Data Figure 4. Single substitutions explain the sensitivity of clones with multiple substitutions

Bars show fold improvement in cortisol sensitivity for every multi-substitution clone recovered from the library and engineered variants containing the individual substitutions. Columns and error bars indicate mean and standard error of experimental replicates (gray

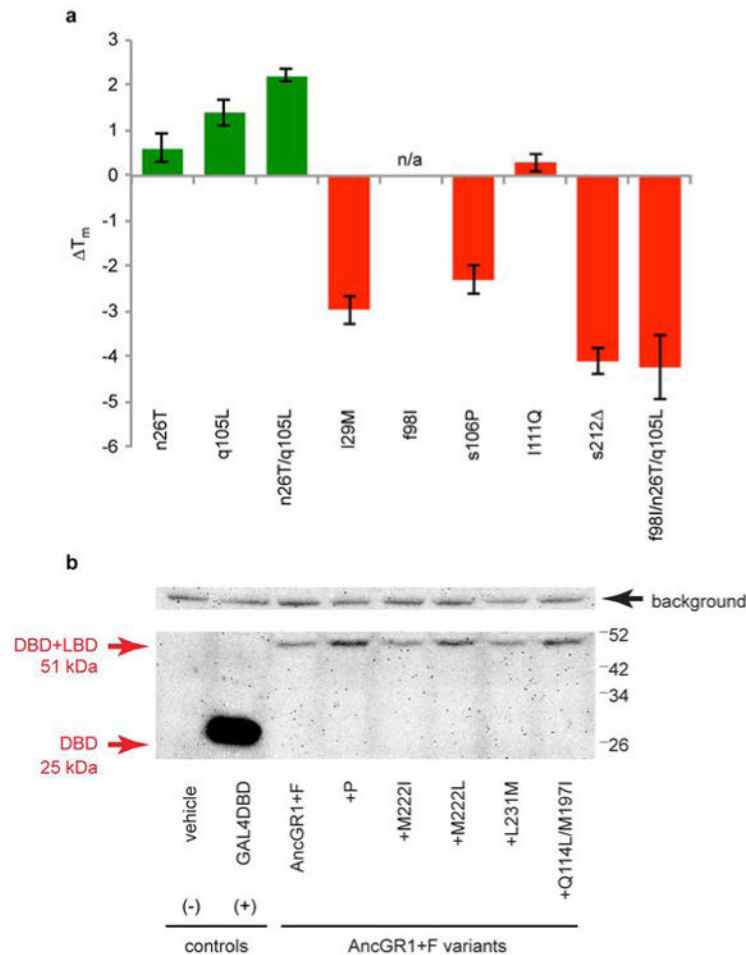
points). Fold improvement is relative to AncGR1+F. Stars indicate result of a one-tailed t-test ($p < 0.05$) assessing the difference between each mutant and AncGR1+F. Colors indicate the class of the clone: historical permissive substitutions (green) and rescuing mutations in the screen that are near the ligand pocket (blue), or activation function helix AF-H (pink).



Extended Data Figure 5. Molecular dynamics simulations reveal stabilization mechanisms of historical P mutations

a, Snapshot of trajectory from AncGR2 simulation showing the hydrogen bond from atom OG1 of derived state Thr26 to Val214-O and packing of Leu105 against the protein. **b**, Historical substitution n26T allows formation of a new hydrogen bond. Radial distribution function of the distance to Val214-O from ancestral residue Asn26 (atom ND2) in simulation of AncGR2p (black) and from derived residue Thr26 (atom OG1) in simulation of AncGR2 (red). Numbers show fraction of time hydrogen bond was formed over each simulation using a 3.0 Å, 30° geometric criterion. The change in hydrogen bond frequency was used to calculate G_{hbond} , the favorable effect of this historical substitution on hydrogen bond energy at 310K. **c**, Historical substitution q105L improves packing interactions. Histogram of van der Waal's contacts (3.5 Å cutoff) between residue 105 and other protein atoms for ancestral state Gln105 in the AncGR2p simulations (black) and

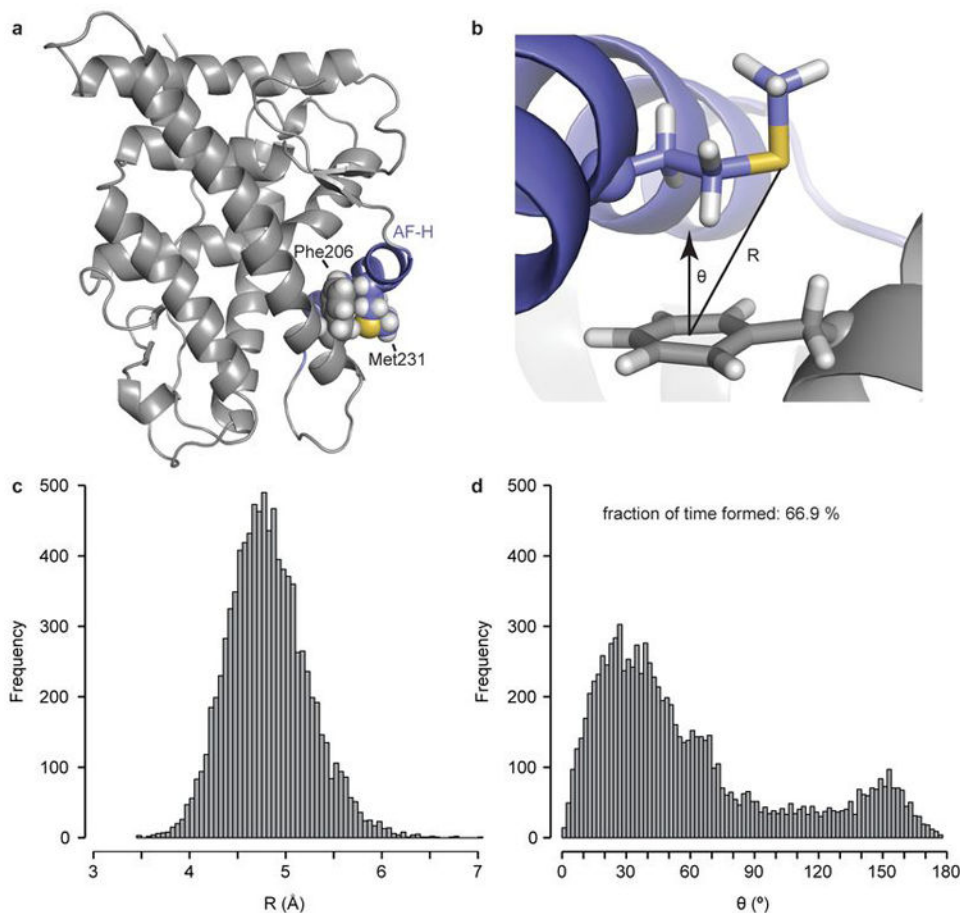
derived state Leu105 in the AncGR2 simulations (red). **d**, Mutations have the same functional effects in the AncGR1+F and AncGR2p (AncGR2/N26t/Q105l) backgrounds, allowing interpretation of experiments in AncGR1+F using MD simulations starting from the AncGR2 crystal structure. Paired bars are changes in cortisol sensitivity for each mutation measured in the AncGR1+F (left) or AncGR2p (right) background. Columns and error bars indicate mean and standard error of experimental replicates (gray points). There was no statistically significant ($p < 0.05$) difference in the effect of each mutation introduced in either the AncGR1+F and AncGR2p backgrounds, as assessed by a two-tailed t-test. No multiple testing correction was performed to minimize type II errors.



Extended Data Figure 6. “F” and “P” mutations have opposite effects on melting temperature, but do not affect expression

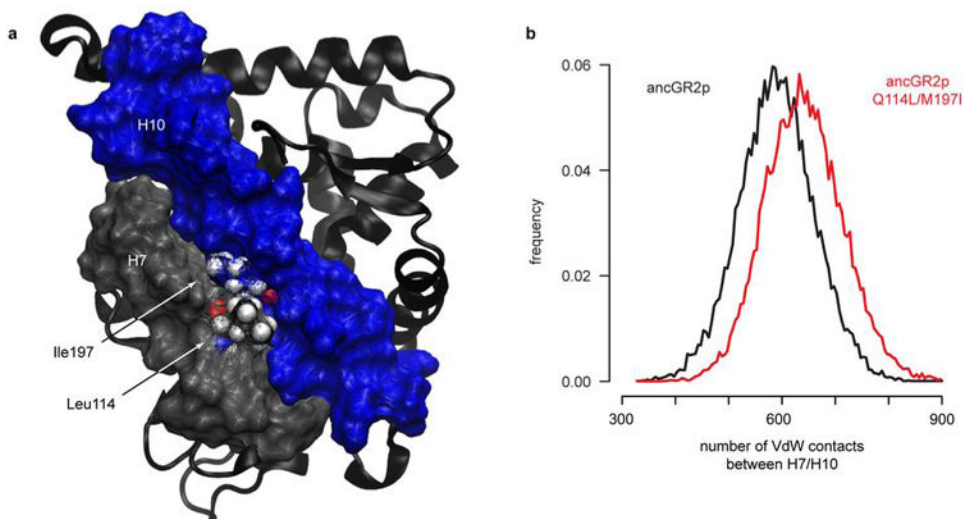
a, Change in T_m induced by “F” and “P” mutations in the AncGR1 background. Colors indicate P (green) or F (red) substitutions. Bars indicate mean change in T_m for triplicate measurements; error bars are standard error. We were unable to express and purify soluble AncGR1/f98I (n/a); comparing n26T/q105L to n26T/q105L/f98I shows that this substitution has a very strong destabilizing effect. **b**, Rescuing mutations do not alter LBD expression in AncGR1+F background. Figure shows a western blot of soluble proteins extracted from CHO-K1 cells, visualized using a polyclonal GAL4DBD antibody. Expression is similar for

all constructs. The small amount of variation does not correlate with sensitivity or fold activation; for example, the non-functional protein AncGR1+F exhibits expression comparable to the highly active AncGR1+F+M222I and AncGR1+F+L231M proteins. Molecular weights (determined by standard marker) are indicated on the right. Red arrows highlight the expected molecular weights of the GAL4DBD and GAL4DBD-LBD fusion protein products. The background band (top) is a high molecular weight, cross-reactive protein that indicates a similar global protein expression level across samples.



Extended Data Figure 7. In MD simulations, rescuing mutation Met231 forms a sulfur- π interaction with Phe206

a, Snapshot from an MD simulation showing the location of the Met231-Phe206 stack at the C-terminal end of the AF-H (slate). **b**, Alternate view of the same snapshot, showing the relative orientation of Met231 and Phe206 as sticks. θ is defined as the angle between A (the vector normal to the Phe plane, extending from its centroid) and B (the vector connecting the Phe centroid to the Met231 sulfur). The distance R is the length of vector B. **c**, Distribution of observed R over 3 independent 100 ns trajectories (9,200 snapshots total). **d**, Distribution of observed θ over the same trajectories. The percentage at the top shows the fraction of time in which the interaction is formed by simple geometric criteria ($R < 6 \text{ \AA}$ and $20^\circ < \theta < 60^\circ$).



Extended Data Figure 8. In MD simulations, rescuing pair Q114L/M197I improves packing between H7 and H10

a, A representative snapshot from the trajectory of AncGR2p+Q114L/M197I shows the favorable interaction of derived states Leu114 and Ile197 (spheres). Helices 7 (gray) and 10 (blue) are shown as solvent accessible surfaces. **b**, A histogram of all van der Waals contacts (3.5 Å cutoff) between H7 and H10 for trajectories of AncGR2p (black) and AncGR2p+Q114L/M197I (red).

Extended Data Table 1
Raw irreversible melting temperatures of mutants of AncGR1 and AncGR1+F

Table a shows melting temperatures of various purified mutants of AncGR1+F (°C). Mean and standard error, and number of replicate melts “n” are shown. **b**, Melting temperatures of various mutants of AncGR1.

a

construct	T_m			T_m		
	mean	stderr	n	mean	stderr	n
AncGR1+F	50.6	0.2	3	0.0	0.3	
n26T	52.7	0.1	3	2.1	0.2	
q105L	53.5	0.2	3	2.8	0.3	
n26T/q105L	54.8	0.2	5	4.1	0.3	
L32M	50.7	0.3	3	0.0	0.3	
N99D	53.6	0.1	3	3.0	0.3	
Q114L	53.0	0.1	3	2.3	0.2	
M197I	52.6	0.4	3	1.9	0.4	
V210E	51.8	0.2	3	1.1	0.3	
M222I	54.9	0.3	3	4.2	0.4	
M222L	52.2	0.0	3	1.6	0.2	

a

construct	T_m			T_m		
	mean	stderr	n	mean	stderr	n
L231M	54.1	0.1	3	3.4	0.2	
L32M/M197I	53.2	0.1	3	2.5	0.2	
Q114L/M197I	54.9	0.3	3	4.3	0.4	
E165A/K168E	52.1	0.1	3	1.4	0.2	

b

construct	T_m			T_m		
	mean	stderr	n	mean	stderr	n
AncGR1	63.0	0.3	3	0.0	0.4	
n26T	63.5	0.1	3	0.5	0.3	
q105L	64.4	0.6	3	1.4	0.7	
n26T/q105L	65.3	0.3	3	2.3	0.4	
L32M	59.7	0.2	3	-3.3	0.4	
N99D	64.6	0.2	3	1.6	0.4	
Q114L	61.7	0.5	3	-1.3	0.6	
M197I	63.5	0.3	3	0.5	0.4	
V210E	64.8	0.5	3	1.8	0.6	
M222I	66.2	0.2	3	3.2	0.4	
M222L	63.2	0.5	3	0.2	0.6	
L231M	65.7	0.1	3	2.7	0.3	
L32M/M197I	60.0	0.3	3	-3.0	0.4	
Q114L/M197I	61.8	0.1	3	-1.2	0.3	
I29M	60.0	0.3	3	-3.0	0.5	
s106P	60.5	0.3	3	-2.5	0.4	
I111Q	63.3	0.1	3	0.3	0.3	
s212	59.0	0.3	3	-4.0	0.4	
f98I/n26T/q105L	61.3	0.7	3	-4.0	0.8	

Acknowledgments

We thank Jamie Bridgham, members of the Thornton lab, and B. Buckley McAllister for technical assistance and fruitful discussions. We thank M. Stallcup for sharing plasmids and the University of Oregon ACISS cluster for computing resources (NSF OCI-0960354). This work was supported by NIH F32-GM090650 (MJH), NIH R01-GM081592 (JT) and R01-GM104397 (JT), NSF IOB-0546906 (JT), and a Howard Hughes Medical Institute Early Career Scientist award (JT).

References

1. Monod, J. *Chance and Necessity: An Essay on the Natural Philosophy of Modern Biology*. Vintage Books; 1972.
2. Gould, SJ. *Wonderful Life: The Burgess Shale and the Nature of History*. W. W. Norton & Company; 1990.

3. Losos JB, Jackman TR, Larson A, Queiroz K, Rodríguez-Schettino L. Contingency and Determinism in Replicated Adaptive Radiations of Island Lizards. *Science*. 1998; 279:2115–2118. [PubMed: 9516114]
4. Morris, SC. *The Crucible of Creation: The Burgess Shale and the Rise of Animals*. Oxford University Press; USA: 2000.
5. Beatty J. Replaying Life's Tape. *The Journal of Philosophy*. 2006; 103:336–362.
6. Ortlund EA, Bridgham JT, Redinbo MR, Thornton JW. Crystal Structure of an Ancient Protein: Evolution by Conformational Epistasis. *Science*. 2007; 317:1544–1548. [PubMed: 17702911]
7. Blount ZD, Borland CZ, Lenski RE. Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *PNAS*. 2008; 105:7899–7906. [PubMed: 18524956]
8. Travisano M, Mongold JA, Bennett AF, Lenski RE. Experimental tests of the roles of adaptation, chance, and history in evolution. *Science*. 1995; 267:87–90. [PubMed: 7809610]
9. Meyer JR, et al. Repeatability and Contingency in the Evolution of a Key Innovation in Phage Lambda. *Science*. 2012; 335:428–432. [PubMed: 22282803]
10. Fisher, RA. *The Genetical Theory of Natural Selection*. Oxford University Press; 1958.
11. Martin RE, et al. Chloroquine Transport Via the Malaria Parasite's Chloroquine Resistance Transporter. *Science*. 2009; 325:1680–1682. [PubMed: 19779197]
12. Field SF, Matz MV. Retracing Evolution of Red Fluorescence in GFP-Like Proteins from Faviina Corals. *Mol Biol Evol*. 2010; 27:225–233. [PubMed: 19793832]
13. Bloom JD, Gong LI, Baltimore D. Permissive Secondary Mutations Enable the Evolution of Influenza Oseltamivir Resistance. *Science*. 2010; 328:1272–1275. [PubMed: 20522774]
14. Lynch VJ, May G, Wagner GP. Regulatory evolution through divergence of a phosphoswitch in the transcription factor CEBPB. *Nature*. 2011; 480:383–386. [PubMed: 22080951]
15. Gong LI, Suchard MA, Bloom JD. Stability-mediated epistasis constrains the evolution of an influenza protein. *eLife*. 2013; 2
16. Peisajovich SG, Tawfik DS. Protein engineers turned evolutionists. *Nat Meth*. 2007; 4:991–994.
17. Romero PA, Arnold FH. Exploring protein fitness landscapes by directed evolution. *Nat Rev Mol Cell Biol*. 2009; 10:866–876. [PubMed: 19935669]
18. Bledsoe, RK.; Stewart, EL.; Pearce, KH. *Nuclear Receptor Coregulators*. Vol. 68. Academic Press; 2004. p. 49-91.
19. Moras D, Gronemeyer H. The nuclear receptor ligand-binding domain: structure and function. *Current Opinion in Cell Biology*. 1998; 10:384–391. [PubMed: 9640540]
20. Smith JM. Natural Selection and the Concept of a Protein Space. *Nature*. 1970; 225:563–564. [PubMed: 5411867]
21. Carroll SM, Ortlund EA, Thornton JW. Mechanisms for the Evolution of a Derived Function in the Ancestral Glucocorticoid Receptor. *PLoS Genet*. 2011; 7:e1002117. [PubMed: 21698144]
22. Ding XF, et al. Nuclear Receptor-Binding Sites of Coactivators Glucocorticoid Receptor Interacting Protein 1 (GRIP1) and Steroid Receptor Coactivator 1 (SRC-1): Multiple Motifs with Different Binding Specificities. *Mol Endocrinol*. 1998; 12:302–313. [PubMed: 9482670]
23. Chen Z, Katzenellenbogen BS, Katzenellenbogen JA, Zhao H. Directed Evolution of Human Estrogen Receptor Variants with Significantly Enhanced Androgen Specificity and Affinity. *J Biol Chem*. 2004; 279:33855–33864. [PubMed: 15159406]
24. Bershtein S, Segal M, Bekerman R, Tokuriki N, Tawfik DS. Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature*. 2006; 444:929–932. [PubMed: 17122770]
25. Bloom JD, Labthavikul ST, Otey CR, Arnold FH. Protein stability promotes evolvability. *Proc Natl Acad Sci USA*. 2006; 103:5869–5874. [PubMed: 16581913]
26. Bridgham JT, Ortlund EA, Thornton JW. An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature*. 2009; 461:515–519. [PubMed: 19779450]
27. Tokuriki N, Stricher F, Schymkowitz J, Serrano L, Tawfik DS. The Stability Effects of Protein Mutations Appear to be Universally Distributed. *Journal of Molecular Biology*. 2007; 369:1318–1332. [PubMed: 17482644]
28. Bloom JD, Arnold FH, Wilke CO. Breaking proteins with mutations: threads and thresholds in evolution. *Mol Syst Biol*. 2007; 3

29. Drummond DA, Iverson BL, Georgiou G, Arnold FH. Why High-error-rate Random Mutagenesis Libraries are Enriched in Functional and Improved Proteins. *Journal of Molecular Biology*. 2005; 350:806–816. [PubMed: 15939434]
30. Polz MF, Cavanaugh CM. Bias in Template-to-Product Ratios in Multitemplate PCR. *Appl Environ Microbiol*. 1998; 64:3724–3730. [PubMed: 9758791]
31. Harju S, Fedosyuk H, Peterson KR. Rapid isolation of yeast genomic DNA: Bust n' Grab. *BMC Biotechnol*. 2004; 4:8. [PubMed: 15102338]
32. Picard D, Yamamoto KR. Two signals mediate hormone-dependent nuclear localization of the glucocorticoid receptor. *EMBO J*. 1987; 6:3333–3340. [PubMed: 3123217]
33. R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; 2011.
34. Schimdt M, et al. General atomic and molecular electronic structure system (Gamess). *J Comput Chem*. 1993; 14:1347–1363.
35. Granovsky, AA. Firefly Version 7.1.G. <http://classic.chem.msu.su/gran/firefly/index.html>
36. Dupradeau FY, et al. The R.E.D. tools: advances in RESP and ESP charge derivation and force field library building. *Phys Chem Chem Phys*. 2010; 12:7821–7839. [PubMed: 20574571]
37. Zoete V, Cuendet MA, Grosdidier A, Michielin O. SwissParam: A fast force field generation tool for small organic molecules. *Journal of Computational Chemistry*. 2011; 32:2359–2368. [PubMed: 21541964]
38. Brooks BR, et al. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*. 1983; 4:187–217.
39. Bjelkmar P, Larsson P, Cuendet MA, Hess B, Lindahl E. Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models. *J Chem Theory Comput*. 2010; 6:459–466. [PubMed: 26617301]
40. Spoel DVD, et al. GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry*. 2005; 26:1701–1718. [PubMed: 16211538]
41. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A linear constraint solver for molecular simulations. *J Comput Chem*. 1997; 18:1463–1472.
42. Darden T, York D, Pedersen L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J Chem Phys*. 1993; 98:10089.
43. Nosé S. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics*. 1984; 52:255–268.
44. Hoover WG. Canonical dynamics: Equilibrium phase-space distributions. *Phys Rev A*. 1985; 31:1695–1697. [PubMed: 9895674]
45. Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*. 1981; 52:7182–7190.
46. Nosé S, Klein ML. Constant pressure molecular dynamics for molecular systems. *Molecular Physics*. 1983; 50:1055–1076.
47. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*. 1996; 14:33–38. [PubMed: 8744570]
48. Valley CC, et al. The Methionine-aromatic motif plays a unique role in stabilizing protein structure. *J Biol Chem*. 201210.1074/jbc.M112.374504

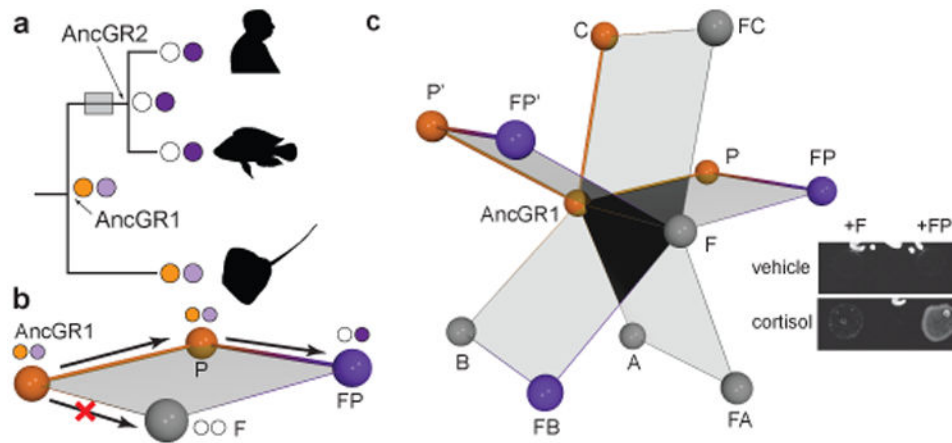


Fig 1. Searching for alternate permissive mutations in an ancestor of GR

A) Evolution of hormone specificity in vertebrate GRs⁶. Icons indicate taxa (tetrapods, teleosts, elasmobranchs); circles show sensitivity to cortisol (purple) or 11-deoxycorticosterone (11-DOC, orange). Transparent box, evolution of new function. B) Seven historical substitutions recapitulate the shift in specificity. Two permissive mutations (P), which have no effect on specificity when introduced alone, allow AncGR1 to tolerate five function-switching mutations (F)⁶. Spheres are colored by primary ligand (11-DOC, orange; cortisol, purple), or no activation (gray). Thick bars connect functional proteins; thin bars lead to non-functional proteins. Arrows, evolutionary paths that pass only through functional intermediates. C) Historical (P) or alternative permissive mutations (P') rescue AncGR1+F and are tolerated in the ancestral background. Non-permissive pathways pass through non-functional intermediates (A, B, gray spheres) or fail to rescue F (C). Inset shows screening conditions in yeast that identify AncGR1+F variants that confer growth in 1 μ M cortisol compared to vehicle-only control.

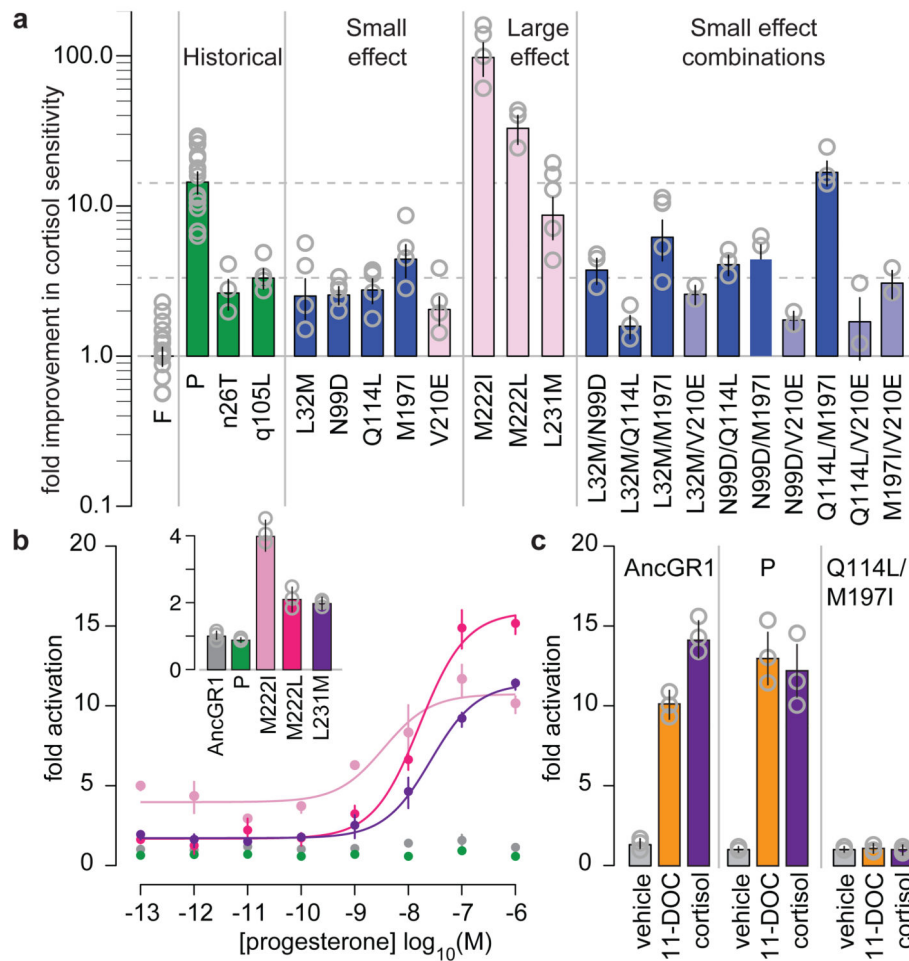


Fig 2. Rescuing mutations disrupt the ancestral protein's function

A) Effects of rescuing mutations on cortisol sensitivity in AncGR1+F. Sensitivity is defined as the ratio of the mutant to AncGR1+F EC_{50} s in a luciferase reporter assay. Columns and error bars show mean and SE of experimental replicates (gray points). Green, historical P substitutions, with effect shown by dotted line; rescuing mutations from the screen are colored by their structural location (see Fig 3C). B) Rescuing AF-H mutations disrupt AncGR1 regulation. Fold reporter activation with progesterone over vector-only control is shown for AncGR1 (gray), historical P (green), and 3 AF-H mutations (pink shades, corresponding to inset graph). Points and error bars show mean and SE for 3 technical replicates. Inset, fold activation for mutants with no hormone (vehicle only). C) Q114L/M197I abolishes activation by AncGR1. Bars, fold activation in 1 μ M of DOC or cortisol vs. vehicle.

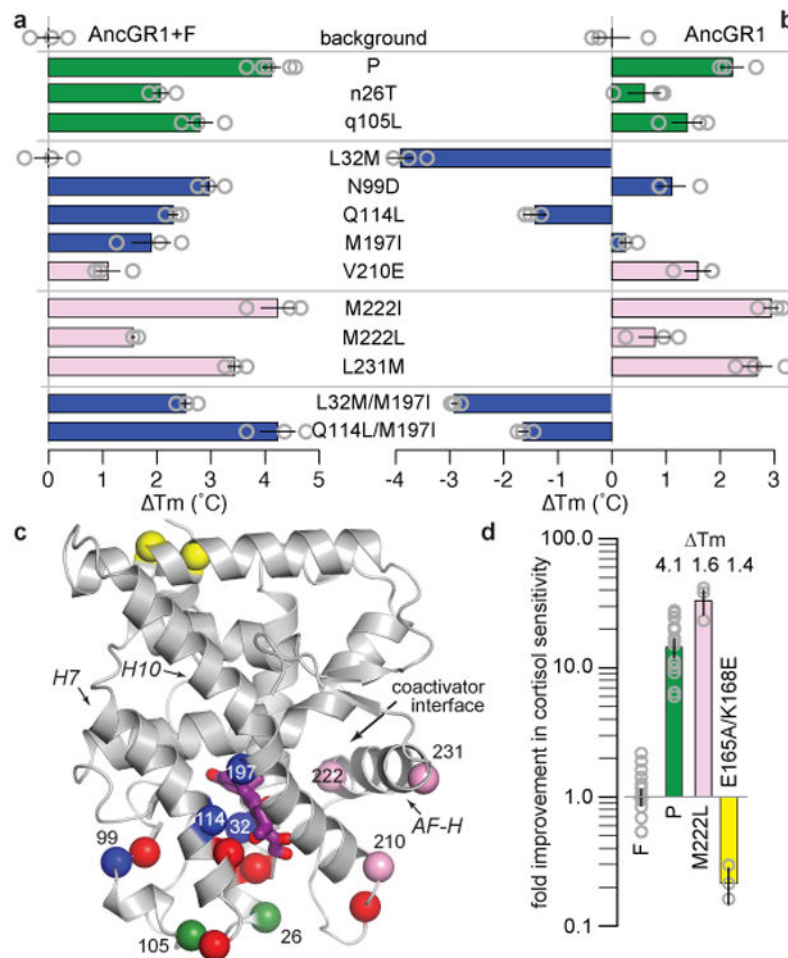


Fig 3. Permissive mutations must stabilize local structural elements

A,B) Effect of rescuing mutations on T_m s of AncGR1+F (A) and AncGR1 (B). Columns and error bars show mean and SE of experimental replicates (grey circles). Colors correspond to structural position in panel C. (C) Structural distribution of mutations on AncGR1 (3RY9). Spheres, Ca atoms. Red, historical F substitutions; green, historical P; blue, rescuing ligand-pocket mutations; pink, rescuing AF-H mutations; yellow, distant mutations that stabilize but do not rescue. Purple sticks show cortisol; helices are indicated. (D) Change in cortisol sensitivity caused by E165A/K168E in AncGR1+F (yellow bar). Effects of P and M222L are shown for comparison. T_m s relative to AncGR1+F are shown.

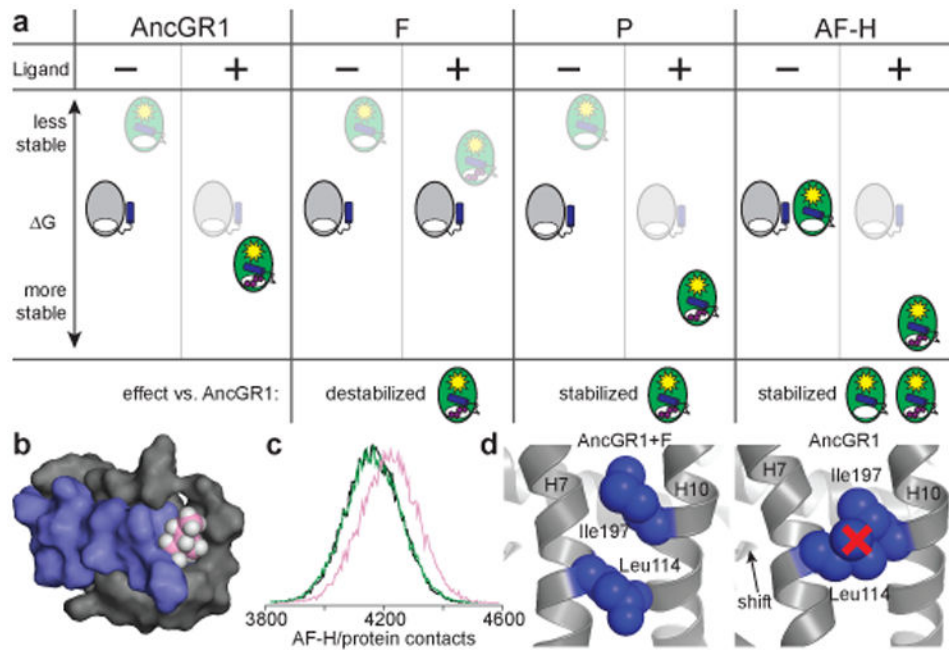


Fig 4. Biophysical requirements make some rescuing mutations intolerable in the ancestral protein

A) A simple thermodynamic model explains why AF-H mutants lead to activity in the absence of hormone. The protein can exist in inactive (grey) or active (green) microstates, which are differentiated by AF-H's position (blue). For each genotype, the relative free energy of active and inactive states is shown with or without hormone. Populated states are opaque, unpopulated states faded. B) Snapshot from MD trajectory of AncGR1+M222I shows tight packing interaction between Ile222 (pink) and the rest of the protein. Blue, AF-H; gray, surface that AF-H contacts. C) Distribution of atom contacts (center-to-center distances $\sim 3.5\text{\AA}$) between AF-H and the rest of the protein over 3 replicate MD trajectories for AncGR1+F (black), +P (green), and +M222I (pink). Y-axis is frequency. D) The change in position of H7 vis-à-vis H10 from ancestral to derived GRs changes the effects of mutations Q114L/M197I from incompatible to rescuing (blue spheres). Structures are AncGR2 (left, 3GN8) and AncGR1 (right, 3RY9) with side chains at these sites introduced (spheres).