LETTER

doi:10.1038/nature13410

Historical contingency and its biophysical basis in glucocorticoid receptor evolution

Michael J. Harms[1,2] & Joseph W. Thornton[2,3]

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 alternative '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 alternative 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, non-deterministic 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 had been present at the Battle of Marathon or if the KT asteroid had missed the Earth? In biology, evolutionary trajectories driven solely by the deterministic force of natural 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 part, 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 probably 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 non-deterministic 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 mutations[6]. 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 the 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, b)[6]. Of these, five function-switching substitutions (denoted F) eliminated the response to other hormones by repositioning a helix (H7) along one side of the binding cavity and establishing new cortisol-specific contacts. Introducing the F substitutions into AncGR1, however, rendered the protein non-functional (Fig. 1b). The remaining two historical substitutions (P) were permissive: they had no detectable effect on receptor function when introduced into AncGR1, but they allowed F to be tolerated, yielding a cortisol-specific receptor (Fig. 1b). Contingency is apparent, because selection for cortisol specificity could not deterministically drive the acquisition of P, which was required for the subsequent evolution of F and the domain's derived structure and function. It is unlikely that the evolving GR passed through a non-functional intermediate containing F without P[20], because the LBD remained conserved, presumably as a result of functional constraints, for ~40 million years from the gene duplication event that generated it until the evolution of its new function (see ref. 21).

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 fulfil two criteria (Fig. 1c): they must rescue the non-functional 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 of AncGR1+F and characterized the resulting distribution of amino acid replacements (Extended Data Figs 1 and 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

[1]Institute of Molecular Biology and Department of Chemistry & Biochemistry, University of Oregon, Eugene, Oregon 97403, USA. [2]Departments of Human Genetics and Ecology & Evolution, University of Chicago, Chicago, Illinois 60637, USA. [3]Institute of Ecology and Evolution, University of Oregon, Eugene, Oregon 97403, USA.

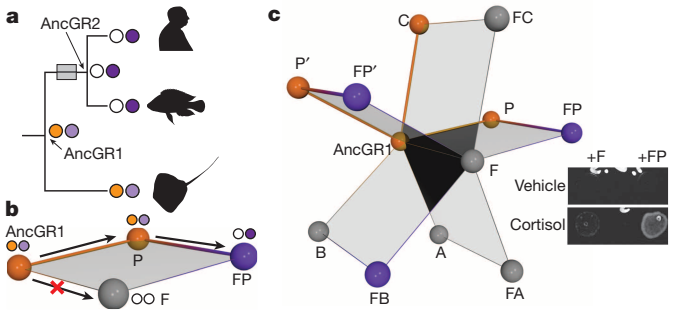


Figure 1 | Searching for alternative permissive mutations in an ancestor of GR. a, Evolution of hormone specificity in vertebrate GRs[6]. Icons indicate taxa (tetrapods, teleosts, elasmobranchs); circles show sensitivity to cortisol (purple) or 11-deoxycorticosterone (orange). The transparent box represents the 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 coloured by primary ligand (orange, 11-deoxycorticosterone; purple, cortisol), or no activation (grey). Thick bars connect functional proteins; thin bars lead to non-functional proteins. Arrows represent evolutionary paths that pass only through functional intermediates. **c**, Historical (P) or alternative permissive (P′) mutations rescue AncGR1+F and are tolerated in the ancestral background. Non-permissive pathways pass through non-functional intermediates (A and B, grey spheres) or fail to rescue F (C). Inset: screening conditions in yeast that identify AncGR1+F variants that confer growth in 1 μM cortisol, compared with vehicle-only control.

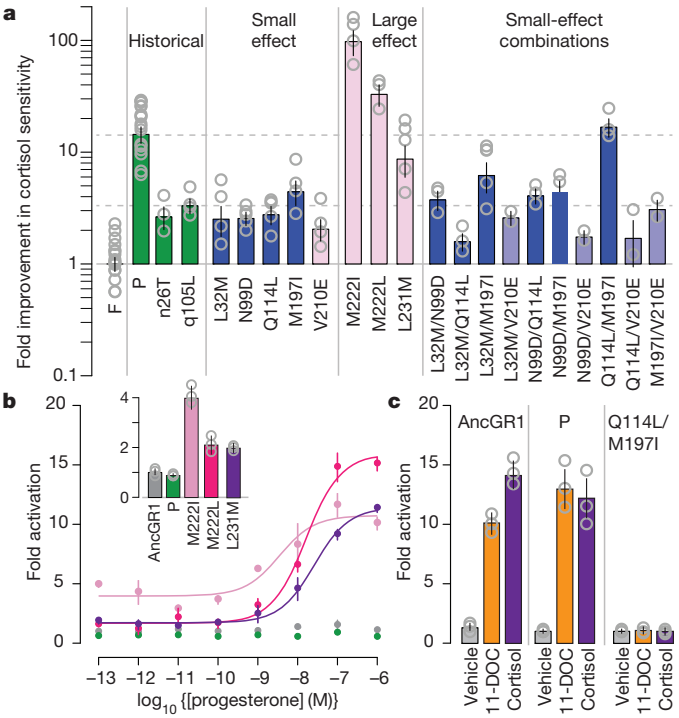


Figure 2 | Rescuing mutations disrupt the ancestral protein's function. a, Effects of rescuing mutations on cortisol sensitivity in AncGR1+F. Sensitivity is defined as the ratio of the mutant to the AncGR1+F concentrations giving half-maximal response (EC_{50}) in a luciferase reporter assay. Results are shown as means and s.e.m. for the number of experimental replicates indicated by grey circles. Green, historical P substitutions, with effect shown by dotted line; rescuing mutations from the screen are coloured 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 (grey), historical P (green) and 3 AF-H mutations (pink shades, corresponding to inset graph). Results are shown as means and s.e.m. for three technical replicates. Inset, fold activation for mutants with no hormone (vehicle only). **c**, Q114L/M197I abolishes activation. The fold activation in 1 μM 11-deoxycorticosterone (11-DOC) or cortisol versus vehicle is shown. Results are means ± s.e.m. for three technical replicates.

in both naive yeast and a mammalian reporter assay (Fig. 1c and Extended Data Fig. 3). We screened ~12,500 clones, comprising an estimated 1,025 unique single replacements (71% of all accessible neighbours), 1,802 unique double replacements and 825 higher-order combinations (3,650 total; see Methods and Extended Data Figs 1 and 2); the remainder were duplicate clones or contained nonsense, frameshift or zero non-synonymous mutations. We found no evidence of bias in the library (Extended Data Fig. 2 and Methods).

This screen identified 12 unique clones that improved AncGR1+F's sensitivity to cortisol. 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 ten unique single mutations that completely or partly 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 tenfold or more, an effect as great as historical P (Fig. 2a). The remaining five mutations improved cortisol sensitivity twofold to threefold each, comparable to the individual members of P, but much less than the pair together (Fig. 2a). To see whether 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 alternative rescuing combinations—one double and three single mutants—indicating that rescuing mutations are rare, on the order of 4 in 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 caused 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 at 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 less than 1 in 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 the 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 through 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 helix 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 (refs 21, 26). In contrast, the P mutations add favourable 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 l111Q 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 analysis (Extended Data Fig. 6b), indicating that AncGR1-F is functionally compromised rather than 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 stabilizing[27]; however, only ~0.1% of our library was rescuing, and permissive mutations were even rarer. 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 a generic compensatory mechanism (Fig. 3b and 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 fulfil, 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 showed that the historical P mutations yielded new favourable contacts involving the same structural elements destabilized by F (Extended Data Fig. 5). Specifically, Thr 26 strengthens a hydrogen bond connecting helix H3 to the H10/AF-H loop, compensating for the loss of a hydrogen bond in this loop as a result of F mutation s212Δ. Leu 105 improves packing interactions between helices H3 and H7, apparently compensating for the effects of the other F mutations on the interactions between H3, H7 and the ligand. Similarly, all rescuing mutations we discovered in our screen improved packing interactions involving AF-H or H7 (Fig. 4 and Extended Data Figs 7 and 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

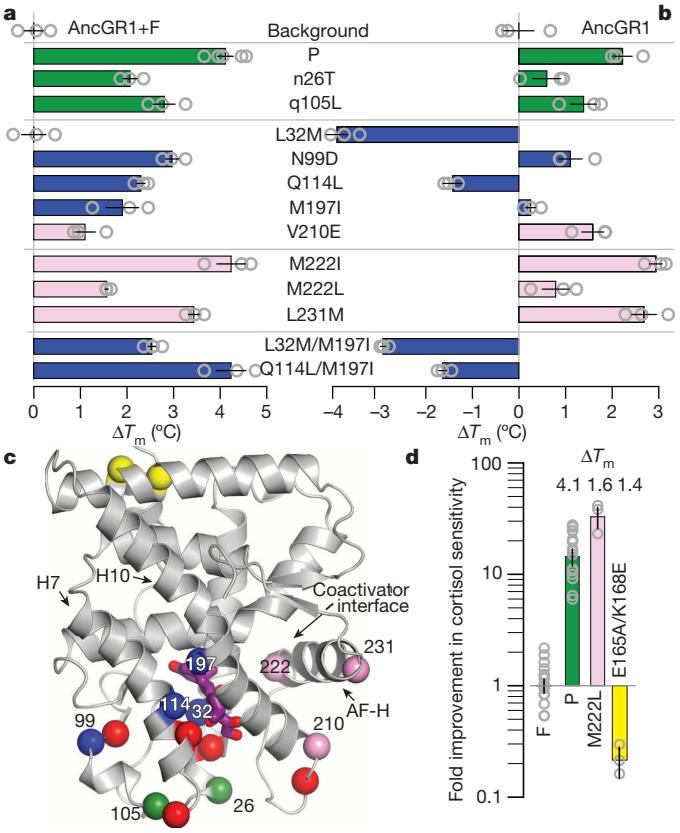


Figure 3 | Permissive mutations must stabilize local structural elements. **a**, **b**, Effect of rescuing mutations on T_m values of AncGR1+F (**a**) and AncGR1 (**b**). Colours correspond to structural position in **c**. **c**, Structural distribution of mutations on AncGR1 (PDB 3RY9). Spheres, Cα 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. ΔT_m values relative to AncGR1+F are shown. Results in **a**, **b** and **d** are shown as means and s.e.m. for the experimental replicates indicated by grey circles.

cortisol sensitivity roughly tenfold (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, b) and are clustered on AF-H, suggesting that 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 improved hydrophobic packing between the active position of AF-H and helix H3 (Fig. 4b, c), and L231M introduced a new sulphur–π interaction, anchoring AF-H in the active position against H10 (Extended

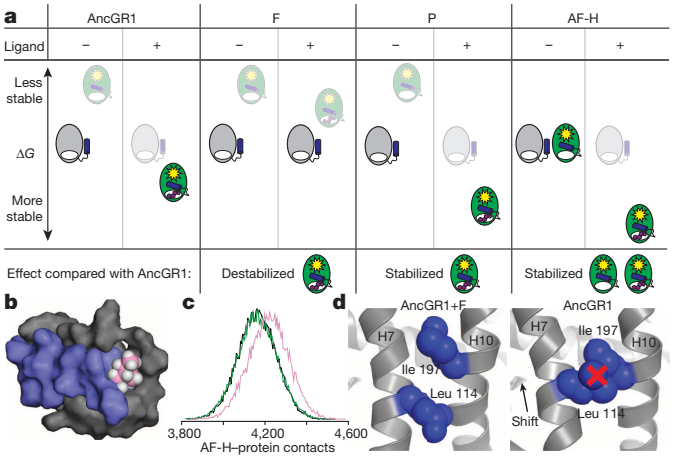


Figure 4 | Biophysical requirements make some rescuing mutations intolerable in the ancestral protein. **a**, A simple thermodynamic model explains why AF-H mutants lead to activity in the absence of hormone. The protein can exist in inactive (grey) or active (green) microstates, which are differentiated by AF-H's position (blue). For each genotype, the relative free energy (ΔG) of active and inactive states is shown with or without hormone. Populated states are opaque; unpopulated states are faded. **b**, Snapshot from MD trajectory of AncGR1+M222I shows tight packing interaction between Ile 222 (pink) and the rest of the protein. Blue, AF-H; grey, surface that AF-H contacts. **c**, Distribution of atom contacts (centre-to-centre distances 3.5 Å or less) between AF-H and the rest of the protein over three replicate MD trajectories for AncGR1+F (black), +P (green) and +M222I (pink). The *y* axis is frequency. **d**, Change in position of H7 with respect to H10 from ancestral to derived GRs changes the effects of mutations Q114L/M197I from incompatible to rescuing (blue spheres). Structures are AncGR2 (left, PDB 3GN8) and AncGR1 (right, PDB 3RY9) with side chains at these sites introduced (spheres).

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 imparted activity in the absence of ligand and made 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 the absence of ligand.

Finally, we examined why the rescuing pair Q114L/M197I rendered 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 Leu 114 and Ile 197 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 change[2]. 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 that 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 in the evolution both of GR's primary sequence and of its molecular form—the structural and mechanistic underpinnings that produce the protein's function. GR's cortisol specificity was achieved by a unique repositioning of H7 and the reorganization of numerous hormone contacts. If other F-like mutations exist that could produce a form and function similar to those of the modern GR, 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 also be very rare.

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 probably be inaccessible, and different outcomes would almost certainly ensue. Cortisol-specific signalling 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 alternative 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.

METHODS SUMMARY

The AncGR1+F mutant library was generated by GeneMorphII-EZClone, using conditions to maximize single and double mutations. We characterized the library's composition by sequencing random clones. For yeast two-hybrid screening, we cloned the LBD library into pBDGAL4 and the human steroid receptor coactivator peptide SRC-1 into pADGAL4 (refs 22, 23). Clones showing any growth in the screen were retransformed into naive yeast and characterized for hormone-dependent growth, then subcloned into pSG5-DBD, transfected into CHO-K1 cells, and assayed using a dual-luciferase reporter. Additional genotypes were generated by Quikchange mutagenesis. Proteins were expressed as His-tagged fusions with maltose-binding protein (MBP), then cleaved and purified to more than 99% purity by sequential affinity chromatography. We followed irreversible thermal denaturation using circular dichroism. For MD simulations, three independent 100-ns trajectories were performed for each genotype, using GROMACS 4.5.5 and the CHARMM27 force field starting from equilibrated crystallographic coordinates with or without *in silico* mutations.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 January; accepted 28 April 2014.
Published online 15 June 2014.

1. Monod, J. *Chance and Necessity: An Essay on the Natural Philosophy of Modern Biology* (Vintage Books, 1972).
2. Gould, S. J. *Wonderful Life: The Burgess Shale and the Nature of History* (W. W. Norton & Company, 1990).
3. Losos, J. B., Jackman, T. R., Larson, A., Queiroz, K. & Rodríguez-Schettino, L. Contingency and determinism in replicated adaptive radiations of island lizards. *Science* **279,** 2115–2118 (1998).
4. Morris, S. C. *The Crucible of Creation: The Burgess Shale and the Rise of Animals* (Oxford Univ. Press, 2000).
5. Beatty, J. Replaying life's tape. *J. Phil.* **103,** 336–362 (2006).

6. Ortlund, E. A., Bridgham, J. T., Redinbo, M. R. & Thornton, J. W. Crystal structure of an ancient protein: evolution by conformational epistasis. *Science* **317,** 1544–1548 (2007).
7. Blount, Z. D., Borland, C. Z. & Lenski, R. E. Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **105,** 7899–7906 (2008).
8. Travisano, M., Mongold, J. A., Bennett, A. F. & Lenski, R. E. Experimental tests of the roles of adaptation, chance, and history in evolution. *Science* **267,** 87–90 (1995).
9. Meyer, J. R. *et al.* Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science* **335,** 428–432 (2012).
10. Fisher, R. A. *The Genetical Theory of Natural Selection* (Oxford Univ. Press, 1958).
11. Martin, R. E. *et al.* Chloroquine transport via the malaria parasite's chloroquine resistance transporter. *Science* **325,** 1680–1682 (2009).
12. Field, S. F. & Matz, M. V. Retracing evolution of red fluorescence in GFP-like proteins from faviina corals. *Mol. Biol. Evol.* **27,** 225–233 (2010).
13. Bloom, J. D., Gong, L. I. & Baltimore, D. Permissive secondary mutations enable the evolution of influenza oseltamivir resistance. *Science* **328,** 1272–1275 (2010).
14. Lynch, V. J., May, G. & Wagner, G. P. Regulatory evolution through divergence of a phosphoswitch in the transcription factor CEBPB. *Nature* **480,** 383–386 (2011).
15. Gong, L. I., Suchard, M. A. & Bloom, J. D. Stability-mediated epistasis constrains the evolution of an influenza protein. *eLife* **2,** e00631 (2013).
16. Peisajovich, S. G. & Tawfik, D. S. Protein engineers turned evolutionists. *Nature Methods* **4,** 991–994 (2007).
17. Romero, P. A. & Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nature Rev. Mol. Cell Biol.* **10,** 866–876 (2009).
18. Bledsoe, R. K., Stewart, E. L. & Pearce, K. H. Structure and function of the glucocorticoid receptor ligand binding domain. *Vitamins Hormones* **68,** 49–91 (2004).
19. Moras, D. & Gronemeyer, H. The nuclear receptor ligand-binding domain: structure and function. *Curr. Opin. Cell Biol.* **10,** 384–391 (1998).
20. Smith, J. M. Natural selection and the concept of a protein space. *Nature* **225,** 563–564 (1970).
21. Carroll, S. M., Ortlund, E. A. & Thornton, J. W. Mechanisms for the evolution of a derived function in the ancestral glucocorticoid receptor. *PLoS Genet.* **7,** e1002117 (2011).
22. Ding, X. F. *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.* **12,** 302–313 (1998).
23. Chen, Z., Katzenellenbogen, B. S., Katzenellenbogen, J. A. & Zhao, H. Directed evolution of human estrogen receptor variants with significantly enhanced androgen specificity and affinity. *J. Biol. Chem.* **279,** 33855–33864 (2004).
24. Bershtein, S., Segal, M., Bekerman, R., Tokuriki, N. & Tawfik, D. S. Robustness–epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature* **444,** 929–932 (2006).
25. Bloom, J. D., Labthavikul, S. T., Otey, C. R. & Arnold, F. H. Protein stability promotes evolvability. *Proc. Natl Acad. Sci. USA* **103,** 5869–5874 (2006).
26. Bridgham, J. T., Ortlund, E. A. & Thornton, J. W. An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature* **461,** 515–519 (2009).
27. Tokuriki, N., Stricher, F., Schymkowitz, J., Serrano, L. & Tawfik, D. S. The stability effects of protein mutations appear to be universally distributed. *J. Mol. Biol.* **369,** 1318–1332 (2007).
28. Bloom, J. D., Arnold, F. H. & Wilke, C. O. Breaking proteins with mutations: threads and thresholds in evolution. *Mol. Syst. Biol.* **3,** 76 (2007).

Acknowledgements We thank J. Bridgham, members of the Thornton laboratory, 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 (National Science Foundation (NSF) OCI-0960354). This work was supported by National Institutes of Health (NIH) F32-GM090650 (M.J.H.), NIH R01-GM081592 (J.W.T.) and R01-GM104397 (J.W.T.), NSF IOB-0546906 (J.W.T.) and a Howard Hughes Medical Institute Early Career Scientist award (J.W.T.).

Author Contributions M.J.H. and J.W.T. conceived the project, designed the experiments and wrote the paper. M.J.H. performed the experiments and analysed the data.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.W.T. (joet1@uchicago.edu).

METHODS

Library generation and characterization. We generated the mutant library with the Genemorph II Domain Mutagenesis kit (Agilent). 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. The mutagenesis 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 before 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 to 10 ng μl^{-1}, 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 spectrum 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 by using the same libraries that 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). To maximize the number of single and double mutants, 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 the 'Screening pipeline' section) 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 towards a small number of clones, thus limiting the number of unique clones sampled[29]. 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 per reaction, corresponding to $\sim 10^{14}$ molecules. Our initial primer-to-template ratio was 20:1, leading to an exhaustion of primer after four or five 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 with simulated samples of a virtual library generated using a Poisson process. We generated a virtual library by 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 that each unique mutation was seen, as well as the number of times that 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 in which 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 four 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 diagram of the screen pipeline is shown in Extended Data Fig. 3. The yeast two-hybrid 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. M. Stallcup generously provided the 'Gal4AD-SRC1a.1236-1441' construct[22] (pADGAL4-SRC), which contains nuclear receptor box 4 from the carboxy terminus of the SRC1a protein. Plasmid transformations were performed in accordance with 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 between 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 between the poorly growing phenotype of AncGR1+F and the more robust growth of AncGR1+FP (Extended Data Fig. 3b).

To perform the screen, the yeasts containing the pADGAL4-SRC plasmid were transformed with the AncGR1+F library. Most of the transformation volume was plated onto synthetic defined (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 yeasts 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 fitted a linear model to the volume versus counts data to extract the number of cells per microlitre. We estimate that we screened 12,485 ± 220 clones, which—according to our library statistics—would include 1,025 ± 18 and 1,802 ± 90 unique single and double mutations, respectively.

We designed our initial screen to minimize the number of false negatives by taking any colony, no 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 four Leu(−)/Trp(−)/His(−) plates: 0.01, 0.1 and 1.0 μM cortisol, and ethanol. 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' protocol[31] to break open constitutive and cortisol-dependent colonies (176 total), followed by colony PCR to amplify and sequence LBDs. When sequenced, only 110 colonies contained amino acid substitutions in the LBD, indicating that many of the yeasts 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 rate of false positives, we took the unique clones and back-transformed those preparations into a naive YRG-2 strain containing the pADGAL4-SRC plasmid. We then performed the secondary screen described above a second time on the naive 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 subcloned their LBDs into the pSG5-DBD vector and measured their sensitivity to cortisol in a transactivation assay. Unlike the yeast two-hybrid 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 by using a one-tailed *t*-test ($P \leq 0.05$) without multiple-testing correction to avoid type II errors and to maximize the number of clones that 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 AncGR variants was assayed with 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 DNA-binding domain (DBD) and LBD to ensure nuclear localization and conformational independence of the two domains[32]. Additional genotypes were generated by Quikchange mutagenesis (Agilent). CHO-K1 (ATCC no. CCL-61) cells were used to measure transactivation. A frozen stock, created on receipt of the cells, was used to restart the cell cultures every 3 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 recovery overnight, cells were incubated in triplicate for 24 h with the hormone of interest from 10^{-12} to 10^{-5} M, and then assayed with Dual-Glo luciferase (Promega). Firefly luciferase activity was normalized by *Renilla* luciferase activity. Dose–response relationships (EC_{50} and maximum fold increase in activation) were estimated by nonlinear regression in R (ref. 33); 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 exponential-phase growth with the addition of 1 mM isopropyl β-D-thiogalactoside, 0.2% glucose, 2% ethanol and 50 μM 11-deoxycorticosterone or cortisol. Cells were incubated overnight at 16 °C and harvested by centrifugation, then frozen at −20 °C. We lysed the cells with B-PER (Thermo Scientific) and then purified the MBP/LBD fusion by nickel-affinity chromatography (HisTrap; GE Healthcare). We cleaved the fusion protein overnight with tobacco etch virus protease in a buffer containing no imidazole. We then ran the cleaved products over the HisTrap column again. The LBD interacts non-specifically with the resin and can be eluted with ~10 mM imidazole, yielding 99% pure LBD. We added 50 μM hormone, 0.04% CHAPS and 10% glycerol, then flash-froze the protein by dropwise addition to liquid nitrogen.

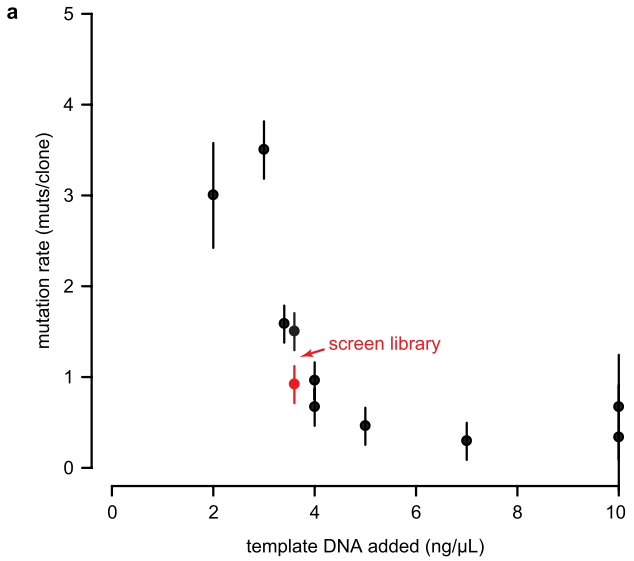
We were unable to find reversible folding conditions for either thermal or chemical denaturation. We therefore measured the midpoint of irreversible thermal denaturation by using a Jasco-815 circular dichroism spectrophotometer with a constant melting rate of 2 °C min^{-1}. We followed α-helical content by means of CD at 225 nm. Protein concentration was 0.5 μM in 25 mM sodium phosphate pH 7.4, 100 mM NaCl, 0.04% CHAPS, 1 mM Tris(2-carboxyethyl)-phosphine (TCEP) and 5 μM cortisol. All melts were performed in triplicate (Extended Data Table 1).

Western blot analyses. CHO-K1 cells were transiently transfected with plasmids containing the GAL4DBD–LBD fusion constructs used in the steroid sensitivity assays. After 24 h, soluble proteins were extracted with RIPA buffer (Sigma) plus 10 μl of phenylmethylsulphonyl fluoride, 10 μl of sodium orthovanadate and 10 μl of protease inhibitor. Protein (25 μg; quantified by Bradford assay) was loaded onto a 12% Tris SDS–PAGE gel, followed by transfer to a poly(vinylidene difluoride) membrane. The blot was blocked overnight at 4 °C in blocking buffer (50 mM Tris, 150 mM NaCl, 0.02% Tween, 5% dried milk, 2% BSA pH 7.6). Primary antibody was a rabbit polyclonal antibody against the GAL4DBD (SC-577; Santa Cruz Biotechnology); secondary antibody was goat anti-rabbit HRP. Bands were revealed with Luminol (1 min), followed by a 30-min exposure on a GelDoc system.

Molecular dynamics simulations. Simulations were conducted in pairs to look for changes correlated with observed mutational effects. Three independent 100-ns trajectories were performed for each genotype. Simulations were started from crystal structures of AncGR1/11-deoxycorticosterone (PDB 3RY9)[21] and, as a proxy for AncGR1+F, AncGR2/dexamethasone (PDB 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 by using PyMOL (Schrödinger LLC), and the rotamer that minimized steric clashes was chosen visually. Cortisol was placed in the pocket by minimizing the root mean squared deviation 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 orientations[36]. Topology files were generated using the SwissParam server[37].

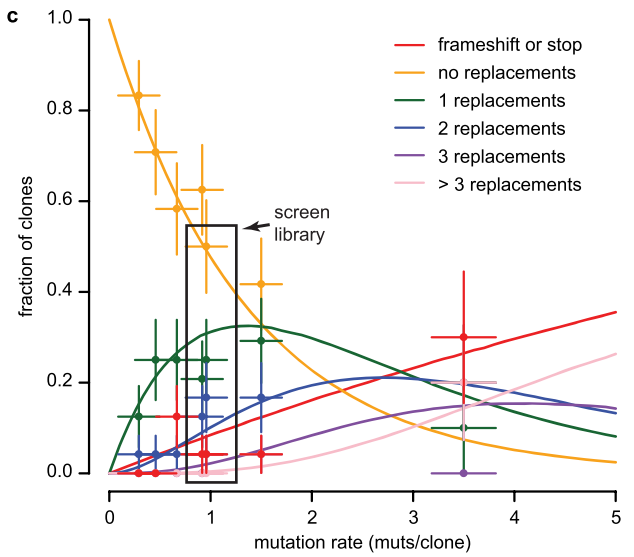
Molecular dynamics simulations were performed using the CHARMM27 force field and TIP3P waters as implemented in GROMACS 4.5.5 (refs 38–40). In all simulations, bonds were treated as constraints and fixed using LINCS[41]. Electrostatics were treated with the Particle Mesh Ewald model[42], using a fast Fourier transform spacing of 12 Å, an interpolation order of 4, a tolerance of 10^{-5} and a Coulomb cut-off 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 performed at 310 K 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 4.5×10^{-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 with 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 root mean squared deviation reached a plateau within 200 ps). The trajectory time step was 2 fs, with frames recorded every 0.5 ps. Final analyses were performed on frames taken every 12.5 ps. Analyses were performed using VMD 1.9.1 (ref. 47)—with its built-in TCL scripting utility—as well as a set of in-house Python and R scripts. For the L231M sulphur–π analysis, we used geometric criteria of $R < 6$ Å and $20° < \theta < 60°$ (see Extended Data Fig. 7)[48].

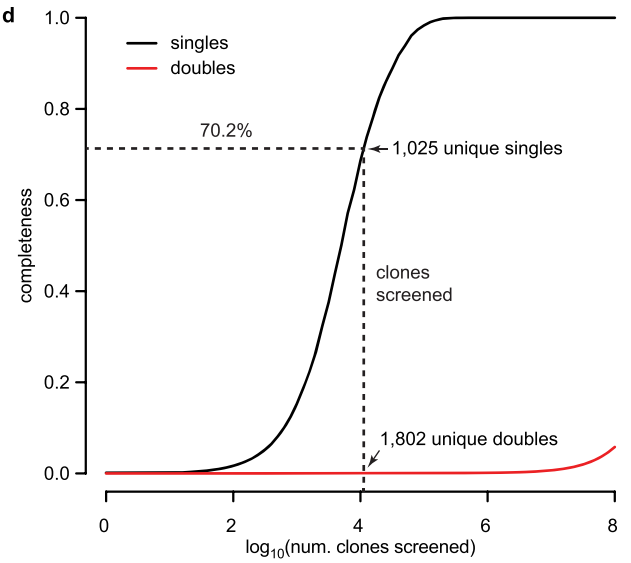
29. Drummond, D. A., Iverson, B. L., Georgiou, G. & Arnold, F. H. Why high-error-rate random mutagenesis libraries are enriched in functional and improved proteins. *J. Mol. Biol.* **350,** 806–816 (2005).
30. Polz, M. F. & Cavanaugh, C. M. Bias in template-to-product ratios in multitemplate PCR. *Appl. Environ. Microbiol.* **64,** 3724–3730 (1998).
31. Harju, S., Fedosyuk, H. & Peterson, K. R. Rapid isolation of yeast genomic DNA: bust n' grab. *BMC Biotechnol.* **4,** 8 (2004).
32. Picard, D. & Yamamoto, K. R. Two signals mediate hormone-dependent nuclear localization of the glucocorticoid receptor. *EMBO J.* **6,** 3333–3340 (1987).
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.* **14,** 1347–1363 (1993).
35. Granovsky, A. A. *Firefly Version 7.1.G* http://classic.chem.msu.su/gran/firefly/index.html.
36. Dupradeau, F.-Y. *et al.* The R.E.D. tools: advances in RESP and ESP charge derivation and force field library building. *Phys. Chem. Chem. Phys.* **12,** 7821–7839 (2010).
37. Zoete, V., Cuendet, M. A., Grosdidier, A. & Michielin, O. SwissParam: a fast force field generation tool for small organic molecules. *J. Comput. Chem.* **32,** 2359–2368 (2011).
38. Brooks, B. R. *et al.* CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4,** 187–217 (1983).
39. Bjelkmar, P., Larsson, P., Cuendet, M. A., 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.* **6,** 459–466 (2010).
40. Spoel, D. V. D. *et al.* GROMACS: fast, flexible, and free. *J. Comput. Chem.* **26,** 1701–1718 (2005).
41. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **18,** 1463–1472 (1997).
42. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: an *N* log(*N*) method for Ewald sums in large systems. *J. Chem. Phys.* **98,** 10089–10092 (1993).
43. Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **52,** 255–268 (1984).
44. Hoover, W. G. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A* **31,** 1695–1697 (1985).
45. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* **52,** 7182–7190 (1981).
46. Nosé, S. & Klein, M. L. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **50,** 1055–1076 (1983).
47. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14,** 33–38 (1996).
48. Valley, C. C. *et al.* The methionine-aromatic motif plays a unique role in stabilizing protein structure. *J. Biol. Chem.* **287,** 34979–34991 (2012).



b

transition	published	empirical mean	empirical std. err	empirical - published
A->C	0.024	0.021	0.010	-0.002
A->G	0.088	0.115	0.023	0.028
A->T	0.143	0.087	0.020	-0.056
C->A	0.071	0.072	0.018	0.002
C->G	0.021	0.024	0.011	0.003
C->T	0.128	0.135	0.024	0.008
G->A	0.128	0.168	0.026	0.040
G->C	0.021	0.011	0.007	-0.010
G->T	0.071	0.114	0.022	0.043
T->A	0.143	0.119	0.023	-0.023
T->C	0.088	0.095	0.021	0.008
T->G	0.024	0.005	0.005	-0.019
indel	0.055	0.034	0.013	-0.021





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. The 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 std. err $= \sqrt{[p(1-p)/n]}$. **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 **a** (for mutation rate) and **b** (for fraction). The 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 **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

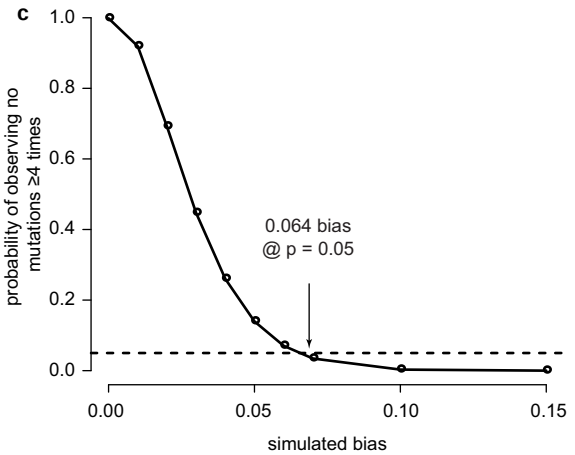
b

Number of mutations seen X times.

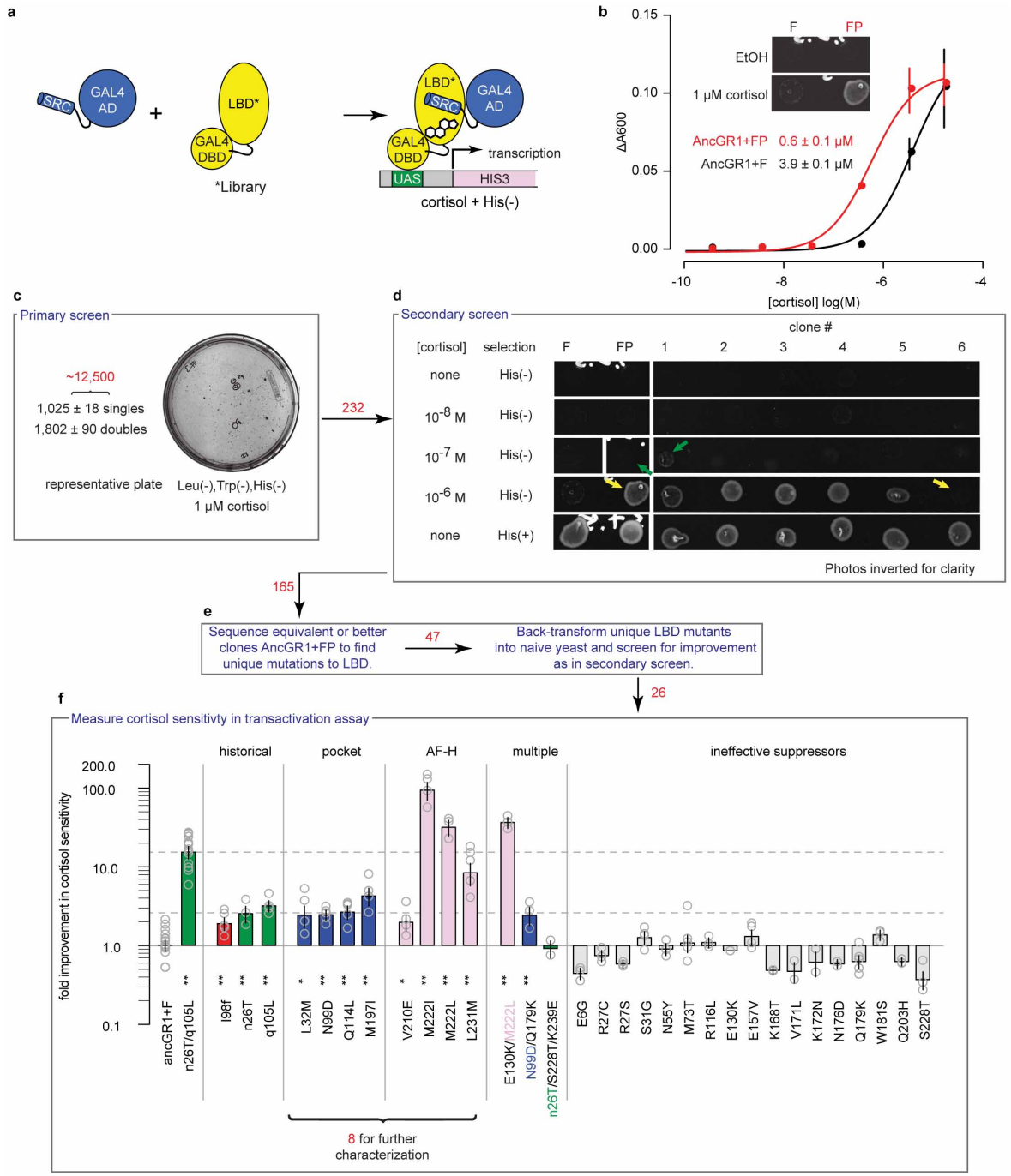
X =	observed	expected
1	84	93
2	6	3
3	1	0

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

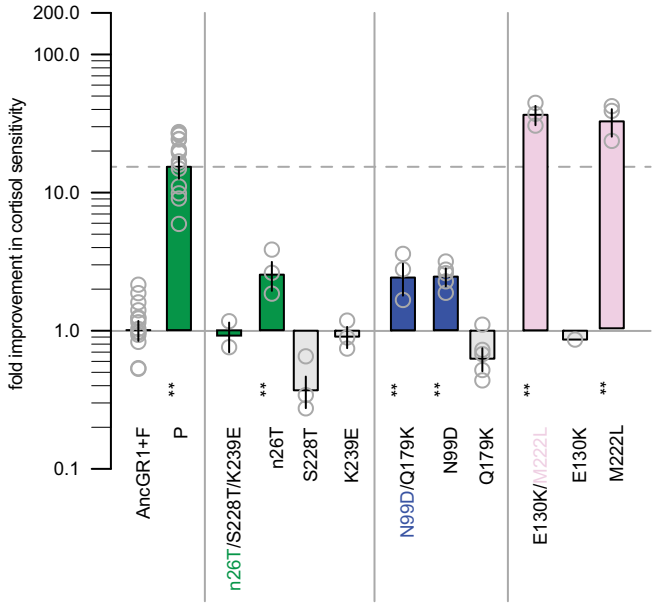
c



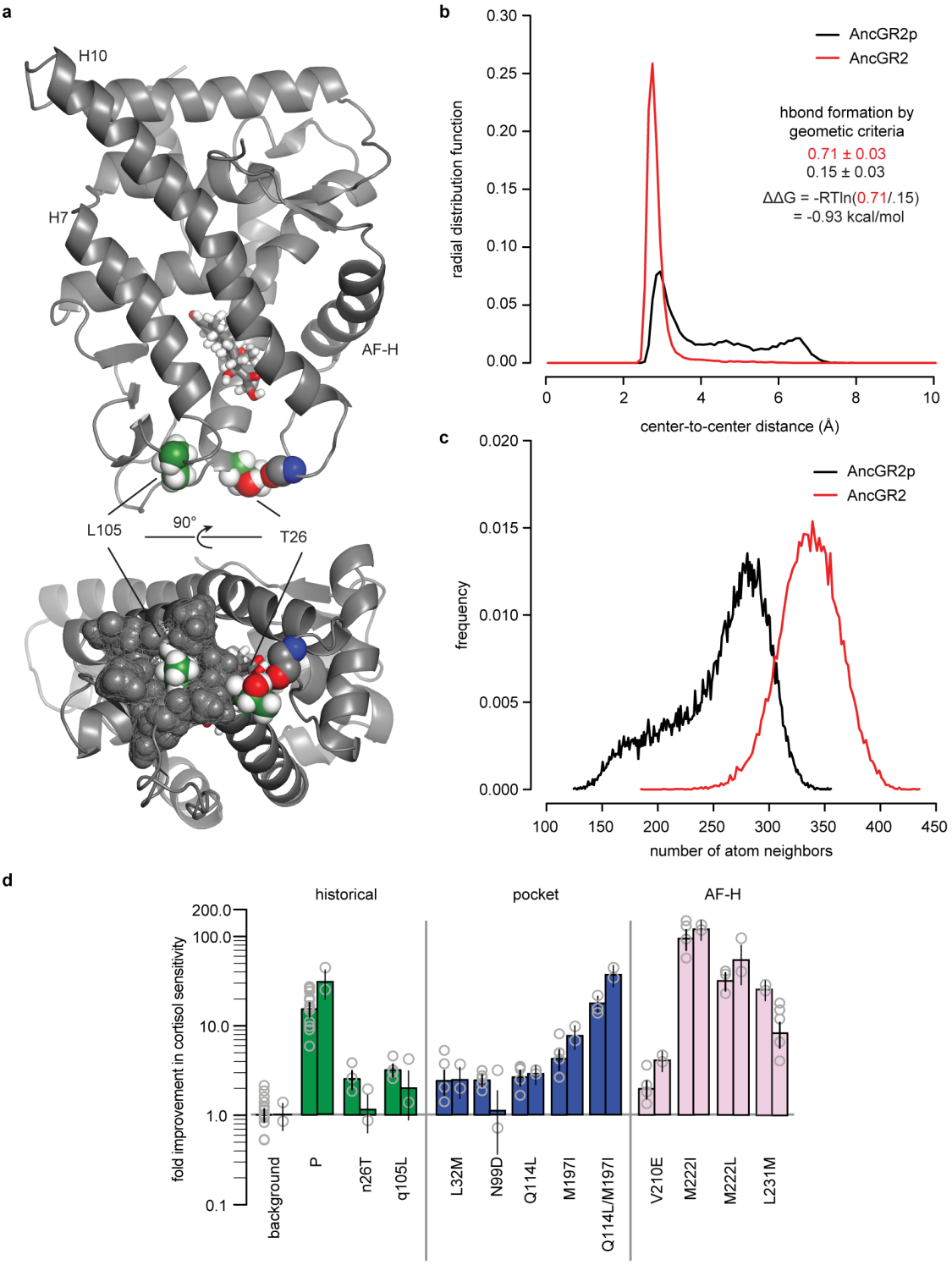
Extended Data Figure 2 | There is no detectable bias in the initial mutant library. **a**, The number of clones containing X amino acid replacements in a 95-clone sample of the variant library: 'experimental' shows the number of clones in each class observed in the actual library by sequencing, and 'expected' 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 (three degrees of freedom) was used to determine whether the observations deviated from the Poisson expectations. Classes of clones with three or four replacements were pooled to maintain adequate counts per cell; no observations were made or predicted with more replacements. **b**, Comparison of 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 with 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.



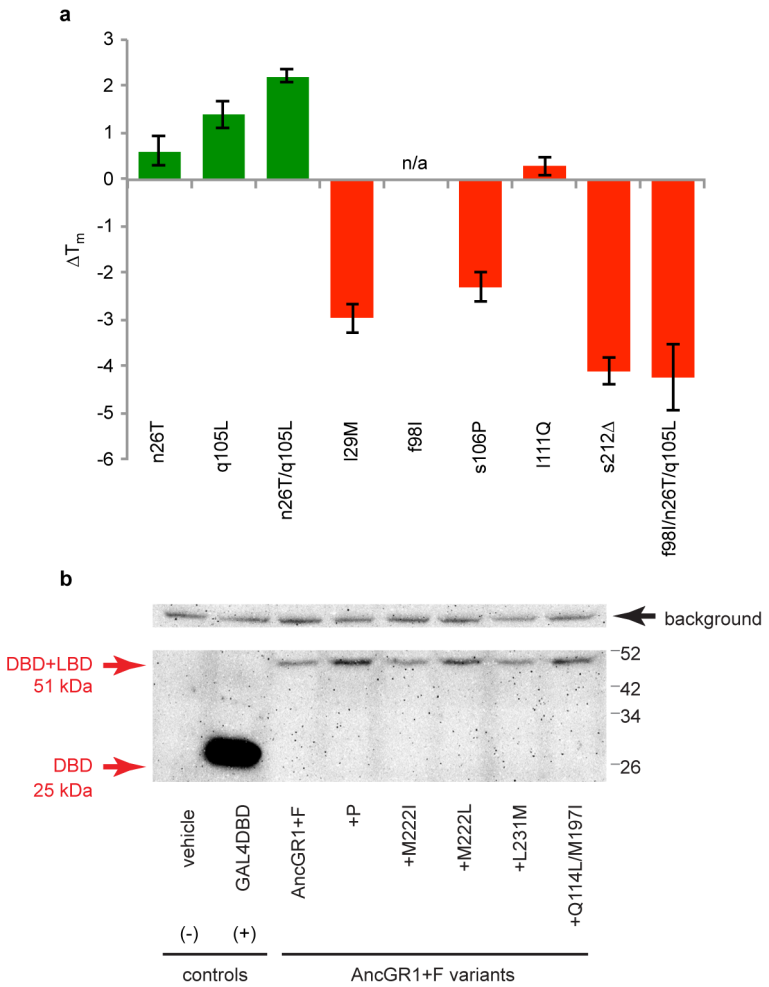
Extended Data Figure 3 | Experimental library screen pipeline. **a**, Diagram of the two-hybrid primary screen for cortisol-specific activation of a mutant library of receptor LBDs. Each LBD is fused to the GAL4-DBD and transformed into yeast along with the GAL4-AD activation domain fused to the SRC-1 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. The plot shows D_{600} for yeast cultures as a function of cortisol concentration for AncGR1+F (black) and AncGR1+FP (red). Inset: colonies of AncGR1+F and AncGR1+FP grown on plates with no hormone/vehicle only (top panel, ethanol (EtOH)) or 1 μM 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 μM cortisol. **d**, Representative clones tested in the secondary screen for dose-responsive growth with increasing cortisol concentration. Each row shows the growth of six 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 decrease false positives. **f**, Fold change in cortisol sensitivity measured with 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 wild-type EC_{50} values. Columns and error bars indicate the mean and standard error of experimental replicates (grey circles). Historical P substitutions are shown with green bars; reversal of a historical F substitution is in red. Rescuing mutations are coloured 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 grey. Dots show statistical significance of the difference in fold activation relative to AncGR1+F (one dot, $P < 0.05$; two dots, $P < 0.01$).



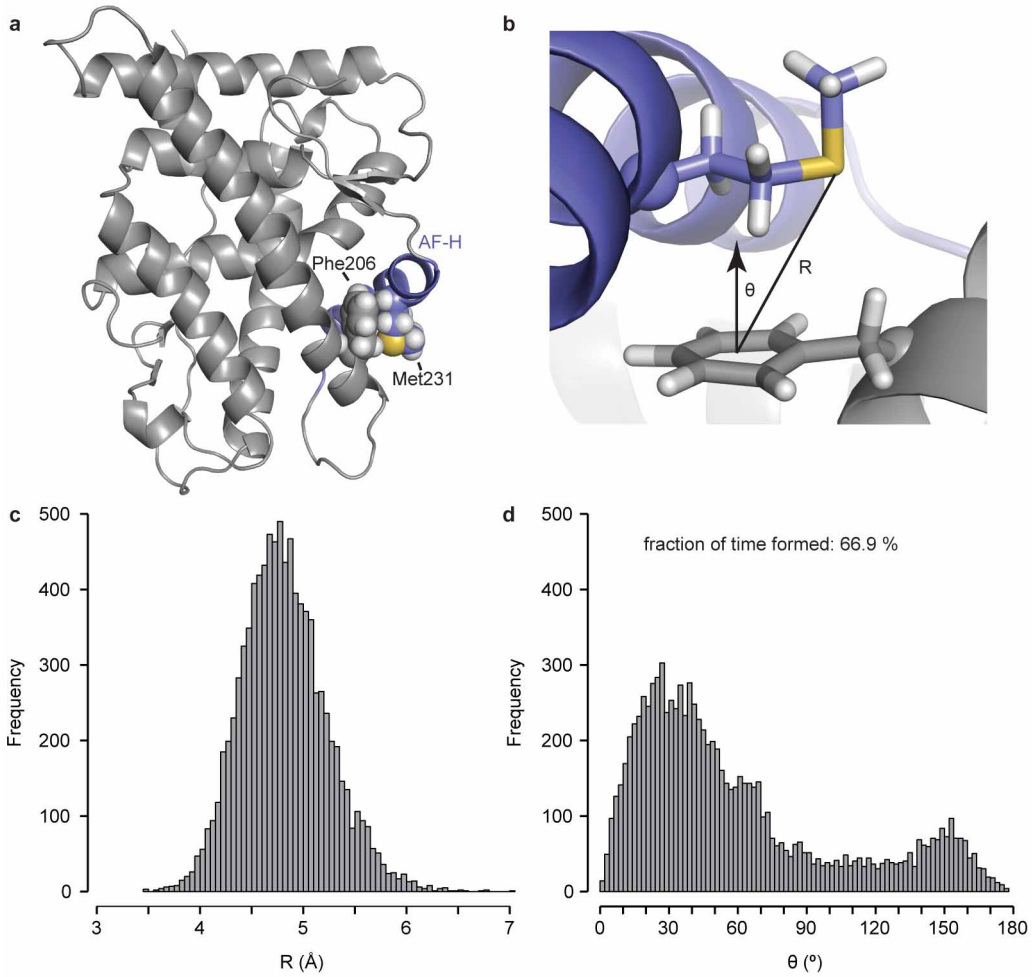
Extended Data Figure 4 | Single substitutions explain the sensitivity of clones with multiple substitutions. Bars show fold improvement in cortisol sensitivity for every multi-substitution clone recovered from the library and engineered variants containing the individual substitutions. Columns and error bars indicate the mean and standard error of experimental replicates (grey points). Fold improvement is relative to AncGR1+F. Stars indicate the result of a one-tailed *t*-test ($P < 0.01$) assessing the difference between each mutant and AncGR1+F. Colours 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).



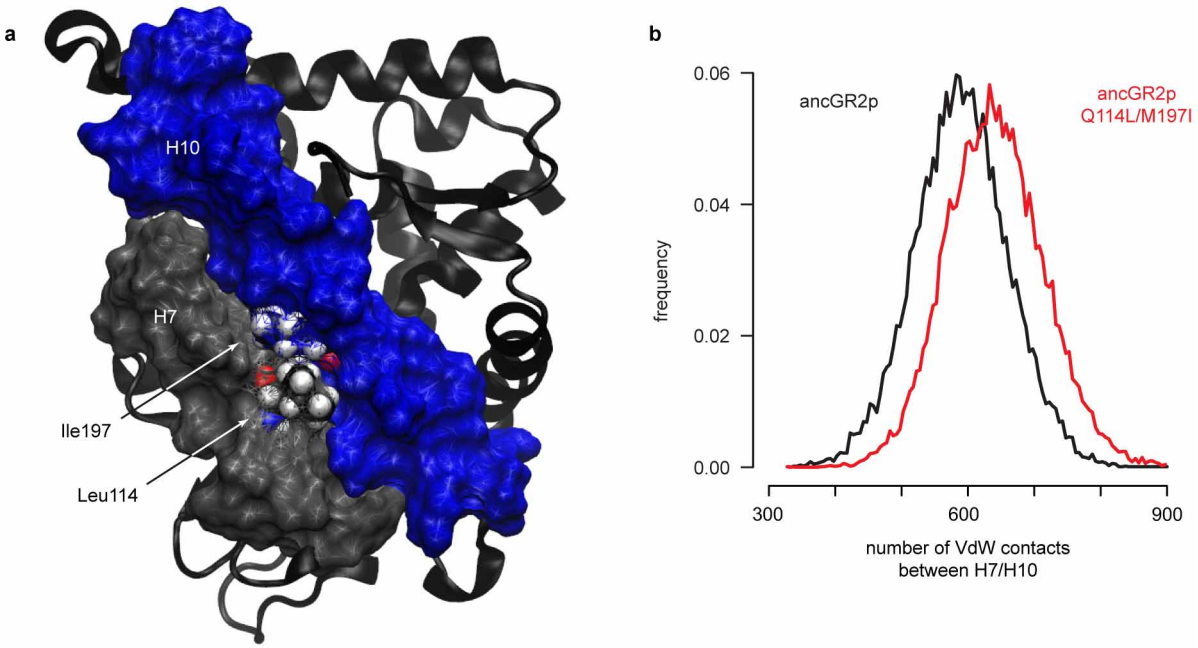
Extended Data Figure 5 | Molecular dynamics simulations reveal stabilization mechanisms of historical P mutations. **a**, Snapshot of trajectory from AncGR2 simulation showing the hydrogen bond from atom OG1 of derived state Thr 26 to Val 214-O and packing of Leu 105 against the protein. **b**, Historical substitution n26T allows the formation of a new hydrogen bond. Radial distribution function of the distance to Val 214-O from ancestral residue Asn 26 (atom ND2) in simulation of AncGR2p (black) and from derived residue Thr 26 (atom OG1) in simulation of AncGR2 (red). Numbers show the fraction of time that a 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 $\Delta\Delta G_{Hbond}$, the favourable effect of this historical substitution on hydrogen bond energy at 310 K. **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 Gln 105 in the AncGR2p simulations (black) and derived state Leu 105 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 the mean and standard error of experimental replicates (grey points). There was no statistically significant ($P < 0.05$) difference in the effect of each mutation introduced in either the AncGR1+F or 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. Colours indicate P (green) or F (red) substitutions. Bars indicate the 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 with 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, revealed with 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 those of the highly active AncGR1+F+M222I and AncGR1+F+L231M proteins. Molecular masses (determined by standard marker) are indicated on the right. Red arrows highlight the expected molecular masses of the GAL4DBD and GAL4DBD–LBD fusion protein products. The background band (top) is a high-molecular-mass 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 sulphur–π interaction with Phe-206. **a**, Snapshot from an MD simulation showing the location of the Met 231–Phe 206 stack at the C-terminal end of the AF-H (slate). **b**, Alternative view of the same snapshot, showing the relative orientation of Met 231 and Phe 206 as sticks. θ is defined as the angle between A (the vector normal to the Phe plane, extending from its centroid) and $\boldsymbol{B}$ (the vector connecting the Phe centroid to the Met 231 sulphur). The distance R is the length of vector B. **c**, Distribution of observed R over three independent 100-ns trajectories (9,200 snapshots in all). **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 ($R < 6$ Å 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 favourable interaction of derived states Leu 114 and Ile 197 (spheres). Helices 7 (grey) 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

a

	T_m			ΔT_m	
construct	mean	stderr	n	mean	stderr
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
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

	T_m			ΔT_m	
construct	mean	stderr	n	mean	stderr
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
s212Δ	59.0	0.3	3	-4.0	0.4
f98I/n26T/q105L	61.3	0.7	3	-4.0	0.8

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