Crystal Structure of an Ancient Protein: Evolution by Conformational Epistasis

Eric A. Ortlund,† Jamie T. Bridgham,† Matthew R. Redinbo,† Joseph W. Thornton†

The structural mechanisms by which proteins have evolved new functions are known only indirectly. We report x-ray crystal structures of a resurrected ancestral protein—the ~450 million-year-old precursor of vertebrate glucocorticoid (GR) and mineralocorticoid (MR) receptors. Using structural, phylogenetic, and functional analysis, we identify the specific set of historical mutations that recapitulate the evolution of GR’s hormone specificity from an MR-like ancestor. These substitutions repositioned crucial residues to create new receptor-ligand and intraprotein contacts. Strong epistatic interactions occur because one substitution changes the conformational position of another site. “Permissive” mutations—substitutions of no immediate consequence, which stabilize specific elements of the protein and allow it to tolerate subsequent function-switching changes—played a major role in determining GR’s evolutionary trajectory.

A central goal in molecular evolution is to understand the mechanisms and dynamics by which changes in gene sequence generate shifts in function and therefore phenotype (1, 2). A complete understanding of this process requires analysis of how changes in protein structure mediate the effects of mutations on function. Comparative analyses of extant proteins have provided indirect insights into the diversification of protein structure (3–6), and protein...
engineering studies have elucidated structure-function relations that shape the evolutionary process (7–11). To directly identify the mechanisms by which historical mutations generated new functions, however, it is necessary to compare proteins through evolutionary time.

Here we report the empirical structures of an ancient protein, which we “resurrected” (12) by phylogenetically determining its maximum likelihood sequence from a large database of extant sequences, biochemically synthesizing a gene coding for the inferred ancestral protein, expressing it in cultured cells, and determining the protein’s structure by x-ray crystallography. Specifically, we investigated the mechanistic basis for the functional evolution of the glucocorticoid receptor (GR), a hormone-regulated transcription factor present in all jawed vertebrates (13). GR and its sister gene, the mineralocorticoid receptor (MR), descend from the duplication of a single ancient gene, the ancestral corticoid receptor (AncCR), deep in the vertebrate tree (14). AncCR ligand-binding pocket. Side chains (<4.2 Å from bound ligand) are superimposed from crystal structures of AncCR with aldosterone (green), DOC (orange), and cortisol (purple). Oxygen and nitrogen atoms are red and blue, respectively; dashed lines indicate hydrogen bonds. Arrows show C11, C17, and C18 positions, which differ among the hormones.

Fig. 1. (A) Functional evolution of corticosteroid receptors. Dose-response curves show transcription of a luciferase reporter gene by extant and resurrected ancestral receptors with varying doses (in log M) of aldosterone (green), DOC (orange), and cortisol (purple). Black box indicates evolution of cortisol specificity. The number of sequence changes on each branch is shown (aa, replacement; Δ, deletion). Scale bars, SEM of three replicates. Node dates from the fossil record (19, 20). For complete phylogeny and sequences, see fig. S10 and table S5. (B) Crystal structure of the AncCR LBD with bound aldosterone (green, with red oxygens). Helices are labeled. (C) AncCR’s ligand-binding pocket. Side chains (<4.2 Å from bound ligand) are superimposed from crystal structures of AncCR with aldosterone (green), DOC (orange), and cortisol (purple). Oxygen and nitrogen atoms are red and blue, respectively; dashed lines indicate hydrogen bonds. Arrows show C11, C17, and C18 positions, which differ among the hormones.
plausible alternative states by mutagenesis, but none changed function (fig. S4). GR’s specificity therefore evolved during the interval between these two speciation events, ~420 to 440 Ma (19, 20).

During this interval, there were 36 substitutions and one single-codon deletion (figs. S5 and S6). Four substitutions and the deletion are conserved in one state in all GRs that descend from AncGR2 and in another state in all receptors with the ancestral function. Two of these—S106P and L111Q (21)—were previously identified as increasing cortisol specificity when introduced into AncCR (13). We introduced these substitutions into AncGR1 and found that they recapitulate a large portion of the functional shift from AncGR1 to AncGR2, radically reducing aldosterone and DOC response while maintaining moderate sensitivity to cortisol (fig. 2A); the concentrations required for half-maximal activation (EC50) by aldosterone and DOC increased by 169- and 57-fold, respectively, whereas that for cortisol increased only twofold. A strong epistatic interaction between substitutions was apparent: L111Q alone had little effect on sensitivity to any hormone, but S106P dramatically reduced activation by all ligands. Only the combination switched receptor preference from aldosterone and DOC to cortisol. Introducing these historical substitutions into the human MR yielded a completely nonfunctional receptor, as did reversing them in the human GR (fig. S7). These results emphasize the importance of having the ancestral sequence to reveal the functional impacts of historical substitutions.

To determine the mechanism by which these two substitutions shift function, we compared the structures of AncGR1 and AncGR2, which were generated by homology modeling and energy minimization based on the AncCR and human GR crystal structures, respectively (16). These structures are robust to uncertainty in the reconstruction: Modeling plausible alternate states did not significantly alter backbone conformation, interactions with ligand, or intraprotein interactions. The major structural difference between AncGR1 and AncGR2 involves helix 7 and the loop preceding it, which contain S106P and L111Q and form part of the ligand pocket (fig. 2B and fig. S8). In AncGR1 and AncCR, the loop’s position is stabilized by a hydrogen bond between Ser106 and the backbone carbonyl of Met108. Replacing Ser106 with proline in the derived GRs breaks this bond and introduces a sharp kink into the backbone, which pulls the loop downward, repositioning and partially unwinding helix 7. By destabilizing this crucial region of the receptor, S106P impairs activation by all ligands. The movement of helix 7, however, also dramatically repositions site 111, bringing it close to the ligand. In this conformational background, L111Q generates a hydrogen bond with cortisol’s C17-hydroxyl, stabilizing the receptor-hormone complex. Aldosterone and DOC lack this hydroxyl, so the new bond is cortisol-specific. The net effect of these two substitutions is to destabilize the receptor complex with aldosterone or DOC and restore stability in a cortisol-specific fashion, switching AncGR2’s preference to that hormone. We call this mode of structural evolution conformational epistasis, because one substitution remodels the protein backbone and repositions a second site, changing the functional effect of substitution at the latter.

Although S106P and L111Q (“group X” for convenience) recapitulate the evolutionary switch in preference from aldosterone to cortisol, the receptor retains some sensitivity to MR’s ligands, unlike AncGR2 and extant GRs. We hypothesized that the other three strictly conserved changes that occurred between AncGR1 and AncGR2 (L29M, F98I, and deletion S212Δ) would complete the functional switch. Surprisingly, introducing these “group Y” changes into the AncGR1 and AncGR1 + X backgrounds produced completely nonfunctional receptors that cannot activate transcription, even in the presence of high ligand concentrations (fig. 3A). Additional epistatic substitutions must have modulated the effect of group Y, which provided a permissive background for their evolution that was not yet present in AncGR1.

The AncCR crystal structure allowed us to identify these permissive mutations by analyzing the effects of group Y substitutions (fig. 3B). In all steroid receptors, transcriptional activity depends on the stability of an activation-function helix (AF-H), which is repositioned when the ligand binds, generating the interface for transcriptional coactivators. The stability of this orientation is determined by a network of interactions among three structural elements: the loop preceding AF-H, the ligand, and helix 3 (17). Group Y substitutions compromise activation because they disrupt this network. S212Δ eliminates a hydrogen bond that directly stabilizes the AF-H loop, and L29M on helix 3 creates a steric clash and unfavorable interactions with the D-ring of the hormone. F98I opens up space between helix 3, helix 7, and the ligand; the resulting instability is transmitted indirectly to AF-H, impairing activation by all ligands (fig. 3B). If the protein could tolerate group Y, however, the structures predict that these mutations would enhance cortisol specificity: L29M forms a hydrogen bond with cortisol’s unique C17-hydroxyl, and the additional space created by F98I relieves a steric clash between the repositioned loop and Met108, stabilizing the key interaction between Q111 and the C17-hydroxyl (fig. 3B).

We hypothesized that historical substitutions that added stability to the regions destabilized by group Y might have permitted the evolving protein to tolerate group Y mutations and to complete the GR phenotype. Structural analysis suggested two candidates (group Z): N26T generates a new hydrogen bond between helix 3 and the AF-H loop, and Q105L allows helix 7 to pack more tightly against helix 3, stabilizing the latter and, indirectly, AF-H (fig. 3B). As predicted, introducing group Z into the nonfunctional AncGR1 + X + Y receptor restored transcriptional activity, indicating that Z is permissive for Y (fig. 3A). Further, AncGR1 + X + Y + Z displays a fully GR-like phenotype that is unresponsive to aldosterone and DOC and maintains moderate
cortisol sensitivity. Both N26T and Q105L are required for this effect (table S4). Strong epistasis is again apparent: Adding group Z substitutions in the absence of Y has little or no effect on ligand-activated transcription, presumably because the receptor has not yet been destabilized (Fig. 3A).

Evolutionary trajectories that pass through functional intermediates are more likely than those involving nonfunctional steps (22), so the only historically likely pathways to AncGR2 are those in which the permissive substitutions of group Z and the large-effect mutations of group X occurred before group Y was complete (Fig. 3C).

We identified permissive substitutions in AncGR1 (green) and AncGR2 (yellow). Group X (S106P, L111Q) and Z (N26T and Q105L) are present; the XYZ combination yields complete cortisol-specificity. The 95% confidence interval for each EC50 is in parentheses. Dash, no activation. (B) Structural prediction of permissive substitutions. Models of AncGR1 (green) and AncGR2 (yellow) are shown with cortisol. Group X and Y substitutions (circles and rectangles) yield new interactions with the C17-hydroxyl of cortisol (purple) but destabilize receptor regions required for activation. Group Z (underlined) imparts additional stability to the destabilized regions. (C) Restricted evolutionary paths through sequence space. The corners of the cube represent states for residue sets X, Y, and Z. Edges represent pathways from the ancestral sequence (AncGR1) to the cortisol-specific combination (+XYZ). Filled circles at vertices show sensitivity to aldosterone (green), DOC (orange), and cortisol (purple); empty circles, no activation. Red octagons, paths through nonfunctional intermediates; arrows, paths through functional intermediates with no change (white) or switched ligand preference (green).

Our discovery of permissive substitutions in the AncGR1-AncGR2 interval suggested that other permissive mutations might have evolved even earlier. We used the structures to predict whether any of the 25 substitutions between AncCR and AncGR1 (fig. S5) might be required for the receptor to tolerate the substitutions that later yielded GR function. Only one was predicted to be important: Y27R, which is conserved in all GRs, stabilizes helix 3 and the ligand pocket by forming a cation-π interaction with Tyr17 (Fig. 4A). When we reversed Y27R in the GR-like AncGR1 + X + Y + Z, activation by all ligands was indeed abolished (Fig. 4B). In contrast, introducing Y27R into AncCR (Fig. 4B) or AncGR1 (fig. S9) had negligible effect on the receptor’s response to any hormone. By conferring increased stability on a crucial part of the receptor, Y27R created a permissive sequence environment for substitutions that, millions of years later, remodeled the protein and yielded a new function.

These results shed light on long-standing issues in evolutionary genetics. One classic question is whether adaptation proceeds by mutations of large or small effect (23). Our findings are consistent with a model of adaptation in which large-effect mutations move a protein from one sequence optimum
to the region of a different function, which smaller-effect substitutions then fine-tune (24, 25); permissive substitutions of small immediate effect, however, precede this process. The intrinsic difficulty of identifying mutations of small effect creates an ascertainment bias in favor of large-effect mutations; the ancestral structures allowed us isolate key combinations of small-effect substitutions from a large set of historical possibilities.

A second contentious issue is whether epistasis makes evolutionary histories contingent on chance events (26, 27). We found several examples of strong epistasis, where substitutions that have very weak effects in isolation are required for the protein to tolerate subsequent mutations that yield a new function. Such permissive mutations create “ridges” connecting functional sequence combinations and narrow the range of selectively accessible pathways, making evolution more predictable (28). Whether a ridge is followed, however, may not be a deterministic outcome. If there are few potentially permissive substitutions and these are nearly neutral, then whether they will occur is largely a matter of chance. If the historical “tape of life” could be played again (29), the required permissive changes might not happen, and a ridge leading to a new function could become an evolutionary road not taken.

Our results provide insights into the structural mechanisms of epistasis and the historical evolution of new functions. GR’s functional specificity evolved by substitutions that destabilized the receptor structure with all hormones but restored a lost molecular interaction (30). Whether a ridge is followed, however, may not be a deterministic outcome. If there are few potentially permissive substitutions and these are nearly neutral, then whether they will occur is largely a matter of chance. If the historical “tape of life” could be played again (29), the required permissive changes might not happen, and a ridge leading to a new function could become an evolutionary road not taken.

Proteins containing membrane attack complex/perforin (MACPF) domains play important roles in vertebrate immunity, embryonic development, and neural-cell migration. In vertebrates, the ninth component of complement and perforin form oligomeric pores that lyse bacteria and kill virus-infected cells, respectively. However, the mechanism of MACPF function is unknown. We determined the crystal structure of a bacterial MACPF protein, Plu-MACPF from Photorhabdus luminescens, to 2.0-angstrom resolution. The MACPF domain reveals structural similarity with pore-forming cholesterol-dependent cytolysins (CDCs) from Gram-positive bacteria. This suggests that lytic MACPF proteins may use a CDC-like mechanism to form pores and disrupt cell membranes. Sequence similarity between bacterial and vertebrate MACPF domains suggests that the fold of the CDCs, a family of proteins important for bacterial pathogenesis, is probably used by vertebrates for defense against infection.