Biophysical mechanisms for large-effect mutations in the evolution of steroid hormone receptors

Michael J. Harmsa,b,1, Geeta N. Eickc,d, Devrishi Goswami, Jennifer K. Coluccif, Patrick R. Griffin, and Joseph W. Thortona,d,2

aInstitute of Ecology and Evolution, University of Oregon, Eugene, OR 97403; bDepartment of Molecular Therapeutics, The Scripps Research Institute, Jupiter, FL 33458; cBiochemistry Department, Emory University School of Medicine, Atlanta, GA 30322; and dDepartments of Human Genetics and Ecology and Evolution, The University of Chicago, Chicago, IL 60637

Edited by David Baker, University of Washington, Seattle, WA, and approved May 23, 2013 (received for review March 7, 2013)

The genetic and biophysical mechanisms by which new protein functions evolve is a central question in evolutionary biology, biochemistry, and biophysics. Of particular interest is whether major shifts in protein function are caused by a few mutations of large effect and, if they are, the mechanisms that mediate these changes. Here we combine ancestral protein reconstruction with genetic manipulation and explicit studies of protein structure and dynamics to dissect an ancient and discrete shift in ligand specificity in the steroid receptors, a family of biologically essential hormone-controlled transcription factors. We previously found that the ancestor of the entire steroid receptor family was highly specific for estrogens, but its immediate phylogenetic descendant was sensitive only to progesterogens, and corticosteroids. Here we show that this shift in function was driven primarily by two historical amino acid changes, which caused a ∼70,000-fold change in the ancestral protein’s specificity. These replacements subtly changed the chemistry of two amino acids, but they dramatically reduced estrogen sensitivity by introducing an excess of interaction partners into the receptor/estrogen complex, inducing a frustrated ensemble of suboptimal hydrogen bond networks unique to estrogens. This work shows how the protein’s architecture and dynamics shaped its evolution, amplifying a few biophysically subtle mutations into major shifts in the energetics and function of the protein.

Ancestral reconstruction | molecular evolution | protein evolution | evolutionary biochemistry

A central goal in biochemistry/biophysics is to understand how proteins’ sequences determine their functional specificity. In molecular evolution, a key objective is to reveal the historical processes by which the diverse functions of extant proteins came to be. These goals, pursued separately, have rarely been achieved in full because of the deep interplay between a protein’s history and its physical properties (1–3). A complete explanation of the functional differences among proteins would explicitly identify the historical mutations that caused their functions to diverge, characterize the physical mechanisms that mediated these mutations’ effects, and reveal how the architecture of the protein shaped and was shaped by the evolutionary process. Such studies could help explain key questions in evolution, such as the role and mechanisms of large-effect mutations in phenotypic evolution (4–8), and in biophysics, such as the determinants of ligand specificity (9–11).

Ancestral sequence reconstruction allows the properties of ancient proteins and the effects of historical mutations to be characterized directly (3, 12). It has been used to identify key mutations that led to changes in protein structure and function (13–16), but it has not been used to understand the evolution of proteins as dynamic molecular systems. Scores or hundreds of amino acids participate in a dense network of interactions to determine protein structure, dynamics, and function. The evolution of ligand specificity is a particular challenge because ligand binding may involve multiple protein/ligand conformations (9–11). A key goal for evolutionary biochemistry is therefore to determine how historical mutations shifted specificity by differentially perturbing the ancient energetic landscapes for binding one class of ligands versus another.

An Evolutionary Shift in Hormone Specificity

Steroid receptors (SRs) are hormone-activated transcription factors that mediate the classic effects of gonadal and adrenal steroids on development, reproduction, and physiology (17). Each SR binds its preferred hormone with high affinity and specificity, interacts directly with DNA, and regulates transcription of nearby target genes. Hormone specificity is determined by the ligand-binding domain (LBD), which binds the hormone in a deep hydrophobic pocket and then undergoes a conformational change, causing assembly of a new surface that recruits coactivator proteins, which in turn modify chromatin or otherwise potentiate transcription (18).

We recently identified a discrete and biologically important shift in ligand specificity during ancient SR evolution (19). The SR family comprises two major clades. One contains the estrogen receptors (ERs), which bind steroids with an aromatized A-ring and a hydroxyl at C3 on the steroid backbone (Fig. L4). The other clade—the nonaromatized steroid receptors (naSRs)—includes receptors for androgens, progestogens, mineralocorticoids, and glucocorticoids, all of which have a nonaromatized A-ring and a keto or hydroxyl at C3 (SI Appendix, Fig. S1). We used maximum likelihood phylogenetics and >200 extant sequences to infer the LBD sequences of the ancestor of the entire family (AncSR1) and the ancestor of the naSR clade (AncSR2). We synthesized cDNAs for these LBDs, expressed them, and characterized their sensitivity to a broad panel of hormones. AncSR1 responded only to estrogens; AncSR2 was unresponsive to estrogens but sensitive to a broad range of nonaromatized steroids (Fig. 1B). Using a library of steroid hormones, we established that AncSR1’s estrogen specificity is determined primarily by the requirement for an aromatized A-ring, whereas AncSR2 specifically excludes aromatized steroids (19).

Here we identify and characterize the mechanisms by which the shift in SR specificity from estrogens to nonaromatized steroids evolved. By combining ancestral reconstruction with studies of protein structure and dynamics, we show how two historical mutations remodeled the hydrogen bond network between hormone and SR, changing the dynamics of the complex
in a ligand-specific way and radically shifting the receptor’s hormone specificity.

Results and Discussion
Phylogenetic and Structural Analyses to Identify Causal Mutations. To identify candidate historical sequence changes that caused the shift in hormone specificity, we combined phylogenetic and structural modes of inference. On the branch between AncSR1 and AncSR2 LBDs, there were 171 replacements; however, only 22 are conserved in the AncSR1 state in extant ERs and the AncSR2 state in extant naSRs (SI Appendix, Table S2), suggesting these sites are functionally constrained. To further narrow the set of candidate replacements, we examined these diagnostic sites in the crystal structure of AncSR2 and a homology model of AncSR1 (19). Within the ligand cavity, most residues near the ligand’s A-ring are unchanged between the two proteins (SI Appendix, Fig. S2), but two of the diagnostic sites contact the A-ring or its C3 functional group—glu41GLN and leu75MET (Fig. 2B, using upper and lowercase for ancestral and derived states, respectively).

Two Large-Effect Replacements Shifted Hormone Specificity. To test the evolutionary importance of these replacements, we reversed them in AncSR2 to the ancestral state and assayed their effects on hormone sensitivity in a luciferase reporter assay. We quantified selectivity as the ratio of the concentrations at which half-maximal activation is achieved (EC50) for nonprogesterone (norP) and the synthetic estrogen 1,3,5-norprogestatrienelone (NPT); these steroids are identical except that the former is a nonaromatized 3-ketosteroid and NPT is aromatized with a 3-hydroxyl (Fig. 2C).

We found that these two historical replacements are the major causes of the evolutionary shift in ligand specificity. Introducing the ancestral states glu41 and leu75 together increased preference for the estrogen over its nonaromatized analog by a factor of >70,000, transforming AncSR2’s strong preference for norP into a very strong preference for NPT (Fig. 2D and E). Similar effects were observed with other matched pairs of aromatized and nonaromatized steroids. Irrespective of whether they contain a 3-hydroxyl or keto (SI Appendix, Fig. S3). Both replacements make large contributions to the functional shift: reversing GLN41 to the ancestral glu alone moderately reduces sensitivity to norP and dramatically increases sensitivity to NPT, whereas MET75leu increases sensitivity to NPT by about 300-fold. We also introduced GLN41 and MET75 into AncSR1: as predicted, the derived states strongly reduced sensitivity to estrogens and increased sensitivity to nonaromatized A-rings (SI Appendix, Fig. S4).

These data indicate that replacements glu41GLN and leu75-MET were large-effect mutations that drove the evolution of AncSR2’s specific response to nonaromatized steroids. Other historical replacements must have made additional minor contributions, however, because AncSR1-GLN41/MET75 retains weak sensitivity to aromatized steroids (SI Appendix, Fig. S4) and AncSR2-glu41/leu75 retains some sensitivity to nonaromatized steroids (Fig. 2D and E).

Structural Mechanisms for the Shift in Specificity. We next sought to understand how these two mutations, which cause relatively subtle changes in the biochemistry of the side chains, caused such large functional effects. We first compared the AncSR2/progesterone crystal structure with the AncSR1/estradiol structural model (19). Although the basic architecture of the ligand cavity is conserved (Fig. 2F), there are several notable differences in the putative hydrogen bond networks that coordinate the ligands.

First, the structures suggest that glu41GLN increased sensitivity to nonaromatized steroids by establishing a new favorable ligand contact. The ancestral glu41 provides only hydrogen bond acceptors, so it cannot form a direct interaction with the 3-keto acceptor (Fig. 2F). Replacement with GLN41, which substitutes a donor functional group for an acceptor, allows this side chain to hydrogen bond directly to 3-keto ligands (Fig. 2F), explaining why glu41GLN causes a ~100-fold increase in sensitivity to 3-keto ligands (Fig. 2D and E). The two mutations’ most significant effect, however, is to radically reduce activation by aromatized steroids (Fig. 2D and E). This cannot be explained by loss of a favorable contact, because GLN41 does contain an acceptor for the aromatized steroid’s 3-OH group. We hypothesized that specificity is instead a property of the entire hydrogen bond network comprising the ligand’s C3 group, GLN41, MET75, and Arg82 (a conserved residue that interacts with both residue 41 and the C3 group). Replacement glu41GLN adds two additional donors in a location where no apparent rotamers of GLN41 and Arg82 can fulfill all their potential interactions (Fig. 2F). MET75 may exacerbate this effect by adding a new weak hydrogen bond acceptor above the A-ring, further complicating an already overconstrained network (Fig. 2F). We thus predicted that aromatized steroids would be unable to form a single, optimal configuration of the derived network and would therefore fail to stabilize the ligand/receptor interaction.

Changes in the Energetic Landscape of Ligand Binding. To test these hypotheses, we used molecular dynamics (MD) simulations to study the effects of the two key mutations on protein–ligand interactions. We conducted triplicate 50-ns simulations of the atom-scale dynamics of four ligand–receptor complexes: AncSR2 with the ancestral or derived amino acids at sites 41 and 75, each with NPT or norP. Because the protein does not relax to the inactive conformation over this timescale even in the absence of hormone (SI Appendix, Fig. S5), the purpose of this analysis is
Large-effect historical mutations drove the evolution of new ligand for NPT. (in a dose and atoms of diagnostic residues. (thus stabilizing the active conformation of which, along with helix H12, form the coactivator interface (Fig. 4). To verify that this effect arose specifically for aromatized steroids, we also analyzed AncSR2 with 19-nor-4-derivative (Fig. S6). Introducing the derived amino acids causes the amine donor of GLN41 to form a new hydrogen bond to the ligand’s carbonyl (SI Appendix, Fig. S6), supporting the view that this interaction contributes to AncSR2’s increased sensitivity to nonaromatized ligands.

We then tested the hypothesis that replacements 41 and 75 disrupted the A-ring hydrogen bond network in the complex with the estrogen NPT (Fig. 2F). To quantitatively analyze the network’s behavior, we clustered the conformations populated during each simulation based on the state of the A-ring network, determined the frequency of each state and the transition between them, verified that these transitions were at equilibrium, and used the Boltzmann equation to calculate the relative free energies of each state and transition (SI Appendix, Figs. S7 and S8 and Tables S3 and S4). In the complex containing the aromatized steroid and the ancestral residues, a stable network of interactions is formed that connects the ligand’s A-ring with side chains 81, 75, and 82. The charged glu41 side chain stably accepts a hydrogen bond from the 3-hydroxyl, and the hydrophobic leu75 packs against the top of the A-ring (Fig. 3, ref. 20). Just two subtle variants of this network were observed: in one, all of the interactions are direct, whereas in the other network some are mediated by water molecules, which exchange rapidly from bulk solvent through the “mouth” of the hydrophobic pocket (Movie S1). This stable network allows the aromatized ligand to serve as a bridge between helices H3 (via residue 41) and H5 (via residues 75 and 82)—which, along with helix H12, form the coactivator interface—thus stabilizing the active conformation of the receptor.

As predicted, introducing the derived residues GLN41 and MET75 in the presence of the estrogen changes this network dramatically. The A-ring system now transitions among a frustrated ensemble of seven distinct and suboptimal states. This behavior reflects a rugged conformational free-energy landscape with multiple basins separated by free-energy barriers (Figs. 3B and 4A). None of these configurations allows satisfaction of the hydrogen bond potential of all polar atoms in the network. In particular, the GLN41 amine and carbonyl groups directly compete for hydrogen bond partners in their vicinity (Fig. 4B). In 44% of sampled states, GLN41’s amine or carbonyl is completely unsatisfied; in the remaining states, these polar groups interact with water that has penetrated through the cavity walls via the internal core of the protein (Figs. 3 and 4C and Movie S2). Internal water penetration occurs because the unsatisfied polar atoms on GLN41 cannot interact with water from the pocket’s mouth; this residue’s interactions with the ligand and MET75 cause its polar groups to face inward, away from bulk solvent (Fig. 3). In the most frequent configuration, a chain of waters runs upward from the ligand pocket through the protein’s core, behind the helices that compose the coactivator interface (Fig. 4C). Helices H3, H5, and H12 separate, disrupting the geometry of the coactivator interface (Fig. 4D). To verify that this effect arose specifically for aromatized steroids, we also analyzed AncSR2 with 19-nor-4-pregnenolone, which has a nonaromatized A-ring but is otherwise identical to NPT. This ligand did not induce frustration, because its hydroxyl faces away from Arg82, leading to a small number of satisfied configurations of the network (SI Appendix, Fig. S9).

Taken together, these observations suggest a mechanistic model that explains the effects of the two key historical replacements on ligand specificity. The derived residues established a new favorable interaction with 3-ketosteroids. They also established a frustrated
The biophysical model

Evolution of Proteins as Complex Physical Systems. The proximate causes of a protein’s specificity are its biophysical properties; the ultimate causes are the evolutionary processes that brought those biophysical mechanisms into being. Our analyses show how combining a phylogenetic approach to history with reductionist mechanistic studies can illuminate both kinds of causes, explaining why

ensemble of suboptimal hydrogen bond networks for aromatized steroids, which specifically excluded estrogens by introducing new interaction partners that cannot be simultaneously satisfied, given the position of the 3-hydroxyl on aromatized steroids.

Experimental Analysis of Changes in Dynamics. To experimentally test this model, we used hydrogen-deuterium exchange (HDX) MS, which quantifies local solvent accessibility and dynamics across a protein by characterizing the rate of deuteration exchange for peptides across the protein (21). We performed HDX-MS on the complexes of AncSR2 and AncSR2-glut41/leu75, each with NPT or norP, and identified regions of each protein in which the two ligands resulted in different rates of deuterium exchange. We then quantified the effects of the two key historical substitutions by comparing the ligand-specific local differences in exchange rates in AncSR2 to those displayed by the ancestralized AncSR2-glut41/leu75 (SI Appendix, Figs. S10 and S11 and Table S5). This approach allowed us to identify regions of the protein where the two mutations specifically caused the NPT:protein complex to undergo increased (or reduced) local motion and/or exposure to solvent.

As predicted by our hypothesis, introducing the derived amino acids increased NPT-specific deuteration exchange in four regions: the C terminus of H3, H6/H7, the loop connecting H10 to H12, and H12 itself (Fig. 4C and SI Appendix, Fig. S10). These regions correspond to the same functionally important locations in which water penetration and disruption of the tertiary structure were observed in the MD simulations (Fig. 4C). These experimental observations corroborate the MD predictions and indicate that during evolution the two key substitutions caused a structural breakdown below the aromatized ligand, where water penetration is occurring, to satisfy the ensemble of suboptimal A-ring hydrogen bond networks in AncSR2. Several regions distant from the ligand-binding site exhibited lower NPT-specific exchange rates when the key amino acids were in the derived states (Fig. 4C); the causes of this effect are unclear.

Arg82 Is Necessary for Ligand-Specificity. The biophysical model derived from the structure (Fig. 2F) and the MD simulations (Fig. 3) includes a key role for the conserved, donor-rich Arg82 in mediating the effects of the two key historical replacements on activation by aromatized estrogens. We therefore predicted that substituting alanine, which does not have excess donors, for Arg82 would attenuate the ligand-selective effect of the historical replacements.

To test this hypothesis, we introduced R82A into AncSR2 and characterized the effect of the glut41 and leu75 mutations on ligand-dependent reporter expression in this background. As predicted, the 70,000-fold effect of the two key mutations on hormone preference when Arg82 is present is virtually abolished: in the Ala82 background, glut41GLN/leu75MET triggers a mere fourfold shift in preference (Fig. 4E). This result corroborates the hypothesis that an excess of hydrogen bond donors near the A-ring mediated the two key mutations’ effect on estrogen sensitivity. It also demonstrates that a conserved residue within the interaction network amplified the effects of the historical replacements to yield a dramatic shift in ligand selectivity.

Evolution of Proteins as Complex Physical Systems. The proximate causes of a protein’s specificity are its biophysical properties; the ultimate causes are the evolutionary processes that brought those biophysical mechanisms into being. Our analyses show how combining a phylogenetic approach to history with reductionist mechanistic studies can illuminate both kinds of causes, explaining why
proteins have their current sequences, functions, and architectures, and revealing how they got that way.

Our findings shed light on the classic evolutionary debate concerning the size distribution of mutational effects during evolution (4–8) and reveal the underlying causes of that distribution during the evolution of a biologically important function. Two historical replacements were sufficient to drive a huge shift in SR ligand recognition, and the protein’s biophysical architecture made these dramatic effects possible. Despite having no apparent effect on protein structure and minor effects on the local biochemical properties of the ligand pocket, the two key replacements fundamentally altered the energetic landscape of ligand binding by introducing new polar atoms into the interaction network. The result was to favor binding of a new ligand while producing an unstable frustrated hydrogen-bond network when the ancestral ligand was present.

Our observations underscore that proteins evolve as complex systems with astronomical degrees of freedom and nonlinear relationships between sequence, biophysical properties, and function (22). A protein–ligand complex samples a huge number of conformational microstates; just one or a few mutations may cause states with different conformations and couplings to become energetically favorable (23). As a result, subtle changes in the biochemistry of a few amino acids can be magnified into large perturbations in the protein’s biophysical behavior, leading to major evolutionary shifts in function. A protein’s biophysical architecture thus acts as more than a negative constraint on the freedom of protein sequences to evolve (24); it also enables relatively small steps in sequence space to produce dramatic evolutionary shifts in a protein’s biochemical behavior and biological functions.

**Methods**

Ancestral LBD sequences were inferred using likelihood-based phylogenetics, synthesized, subcloned into vector pSG5-DBD, mutated with Stratagene Quikchange, cotransfected into Chinese hamster ovary (CHO-K1) cells with a pFR-LUC reporter and phRLtk normalization plasmid, and assayed using dual luciferase assays. Molecular dynamics simulations were performed using GROMACS and the GROMOS96 53a6 force field with simple point charge parameters for proteins. For each ligand–receptor complex, three independent trajectories were initiated, beginning with the AncSR2-progesterone crystal structure coordinates (Protein Data Bank ID code 4FN9), followed by energy minimization, equilibration with protein atoms fixed (except for mutated residues), and 50 ns of unrestrained MD. The initial 10 ns—a point well after the rmsd of ligand and protein backbone atoms reached a plateau—were excluded as burn-in. To generate free energy landscapes, we assigned each frame to a conformation based on the rotamer of residue 41 and the presence/absence of hydrogen bonds among the ligand and residues 41, 75, and 82; we calculated the frequencies of each conformation across trajectories and the transition frequencies between them; verified that these transitions were at equilibrium; and then used the Boltzmann equation to estimate the free energies of the states and barriers between them. Solution-phase amide HDX was carried out using electrospray ionization directly coupled to a high-resolution MS. Spectra were collected at six time points between 10 and 3,600 s. Deuterium incorporation kinetics were extracted using nonlinear regression. To determine the effect of the historical mutations on ligand-specific exchange properties, we compared the exchange rates for peptides shared across all protein backgrounds and ligands. For detailed methods, see_SI Appendix._

**ACKNOWLEDGMENTS.** We thank J. Bridgham and J.W.T. laboratory members for comments. This work was supported by National Institutes of Health Grants R21GM081592 and F32-GM090650, National Science Foundation Grants IOB-0546906 and DEB-0516530, and the Howard Hughes Medical Institute.