Cell

Evolution of DNA Specificity in a Transcription Factor Family Produced a New Gene Regulatory Module

Alesia N. McKeown,[1,4] Jamie T. Bridgham,[1,4] Dave W. Anderson,[1] Michael N. Murphy,[3] Eric A. Ortlund,[3] and Joseph W. Thornton[1,2,*]
[1]Institute of Ecology and Evolution, University of Oregon, Eugene, OR 97403, USA
[2]Department of Ecology and Evolution and Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA
[3]Department of Biochemistry, Emory University, Atlanta, GA 30322, USA
[4]Co-first author
*Correspondence: joet1@uchicago.edu
http://dx.doi.org/10.1016/j.cell.2014.09.003

SUMMARY

Complex gene regulatory networks require transcription factors (TFs) to bind distinct DNA sequences. To understand how novel TF specificity evolves, we combined phylogenetic, biochemical, and biophysical approaches to interrogate how DNA recognition diversified in the steroid hormone receptor (SR) family. After duplication of the ancestral SR, three mutations in one copy radically weakened binding to the ancestral estrogen response element (ERE) and improved binding to a new set of DNA sequences (steroid response elements, SREs). They did so by establishing unfavorable interactions with ERE and abolishing unfavorable interactions with SRE; also required were numerous permissive substitutions, which nonspecifically improved cooperativity and affinity of DNA binding. Our findings indicate that negative determinants of binding play key roles in TFs' DNA selectivity and—with our prior work on the evolution of SR ligand specificity during the same interval—show how a specific new gene regulatory module evolved without interfering with the integrity of the ancestral module.

INTRODUCTION

Transcription Factor Specificity and the Evolution of Gene Regulatory Networks

Development, homeostasis, and other complex biological functions depend upon the coordinated expression of networks of genes. Thousands of transcription factors (TFs) in eukaryotes play key regulatory roles in these networks because their distinct affinities for DNA binding sites, other proteins, and small molecules allow them to specifically regulate the expression of unique sets of target genes in response to various hormones, kinases, and other upstream molecular stimuli. Most studies of the evolution of gene regulation have focused on how changes in *cis*-regulatory DNA can bring a new target gene under the influence of an existing TF (Carroll, 2008; Wray, 2007) or on changes in protein-protein interactions among TFs (Brayer et al., 2011; Lynch et al., 2011; Baker et al., 2012). TF specificity for DNA can and does evolve (Baker et al., 2011; Sayou et al., 2014), however, and little is known concerning the molecular mechanisms and evolutionary dynamics by which such changes occur. In turn, it remains unclear how distinct gene regulatory modules—defined as a transcription factor, the molecular stimuli that regulate it, and the DNA target sequences it recognizes—emerge during evolution. If TFs are constrained by selection to conserve essential ancestral functions (Stern and Orgogozo, 2009), how can new regulatory modules ever arise? Do specific modules evolve by partitioning the activities of an ancestral TF that is promiscuous in its interactions with DNA targets and molecular stimuli (Sayou et al., 2014) or by acquiring entirely new interactions (Teichmann and Babu, 2004)? What is the genetic architecture of evolutionary transitions in TF specificity, and what kinds of biophysical mechanisms mediate these changes? Answering these questions requires dissecting evolutionary transitions in TFs' capacity to interact specifically with DNA and molecular stimuli. Ancestral protein reconstruction, combined with detailed studies of protein function and biochemistry, has the potential to accomplish this goal (Harms and Thornton, 2010).

The knowledge gap concerning transcription factor evolution mirrors uncertainty about the physical mechanisms that determine TFs' specificity for their DNA targets. DNA recognition is usually thought to be determined by favorable interactions—especially hydrogen bonds but also van der Waals interactions—between a protein and its preferred DNA sequences (Garvie and Wolberger, 2001; Rohs et al., 2010). Supporting this view, structural studies have established that positive interactions are typically present in high-affinity complexes of protein and DNA. Specificity, however, is determined by the distribution of affinities across DNA sequences, and it is unclear whether positive interactions sufficiently explain TFs' capacity to discriminate among targets. In principle, negative interactions that reduce affinity to nontarget binding sites—such as steric clashes or the presence of unpaired polar atoms in a protein-DNA complex—could also contribute to specificity (von Hippel and Berg, 1986). Evaluating the role of negative interactions in determining specificity, however, requires analyzing not only high-affinity

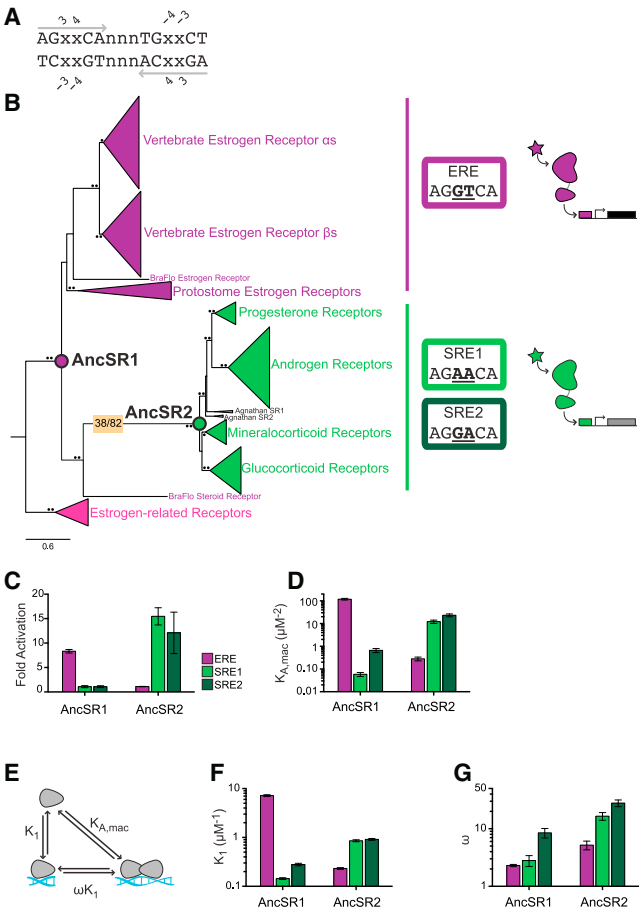


Figure 1. Evolution of Novel Specificity Occurred via a Discrete Shift between AncSR1 and AncSR2

(A) Architecture of SR response elements. All SRs bind to an inverted palindrome of two half-sites (gray arrows) separated by variable bases (n). x indicates sites at which ERE and SREs differ.

(B) SR phylogeny comprises two major clades, which have nonoverlapping specificity for ligands (stars) and REs (boxes). Preferred half-sites for each clade are shown; bases that differ are underlined. Ancestral and extant receptors are colored by RE specificity (purple, ERE; green, SREs; pink, extended monomeric ERE). The orange box indicates evolution of specificity for SREs; number of substitutions on this branch and the total number of DBD residues are indicated. Nodal support is marked by the approximate likelihood ratio statistic (aLRS): unlabeled, aLRS 1 to 10; one solid dot indicates aLRS 10 to 100; two solid dots indicate aLRS > 100. Scale bar is in substitutions per site.

(C) AncSR1 specifically activates reporter gene expression driven by ERE (purple bar) with no activation from SRE1 (light green) or SRE2 (dark green); AncSR2's specificity is distinct. Bar height indicates fold activation relative to vector-only control with SEM of three experimental replicates.

(D) Ancestral binding affinities reflect distinct specificities for ERE versus SREs. Bars heights indicate the macroscopic affinity ($K_{A,mac}$) of binding to palindromic DNA response elements, measured using fluorescence polarization; error bars show SEM of three experimental replicates. Colors as in (C).

(E–G) The components of macroscopic binding affinity—affinity for a half-site (K_1) and cooperativity of binding (ω)—by AncSR1 and AncSR2 were estimated by measuring binding to a half-site and a full palindromic RE and then globally fitting the data to a model containing both parameters. Error bars show SEM of three experimental replicates.

See Figure S1 and Tables S1, S2, and S3.

TF/DNA complexes but also poorly bound ones, which are vast in number and difficult to crystallize. We reasoned that, by focusing on a major evolutionary transition in DNA specificity during the history of a family of related TFs, we could gain direct insight into the genetic and biophysical factors that cause differences in DNA recognition (Harms and Thornton, 2013).

Steroid Receptors Coordinate Distinct Gene Regulatory Modules

Steroid hormone receptors (SRs), a family of ligand-activated transcription factors, are a model for the evolution of TF specificity. SRs initiate the cascade of classic transcriptional responses to sex and adrenal steroid hormones in vertebrate physiology, reproduction, development, and behavior (Bentley, 1998). These proteins contain a conserved DNA-binding domain (DBD), which directly binds to DNA sequences in the vicinity of the target genes they regulate. They also contain a conserved ligand-binding domain (LBD), which binds hormonal ligands and then attracts coregulatory proteins, leading to ligand-regulated changes in gene expression (Bain et al., 2007; Beato and Sánchez-Pacheco, 1996; Kumar and Chambon, 1988). Additional poorly conserved N-terminal and hinge domains mediate other activities. All SRs bind as dimers to inverted palindromic DNA sequences consisting of two six-nucleotide half-sites separated by a variable three-nucleotide spacer (Figure 1A; Welboren et al., 2009; So et al., 2007; Lundbäck et al., 1993; Umesono and Evans, 1989; Beato et al., 1989).

There are two phylogenetic classes of SRs in vertebrates, which have distinct specificities for both DNA and hormonal ligands; the two SR classes therefore mediate distinct regulatory modules (Figure 1B). One class, the estrogen receptors (ERs), are activated by steroid hormones with aromatized A-rings (Eick et al., 2012) and bind preferentially to estrogen response elements (ERE, a palindrome of AGGTCA) (Welboren et al., 2009). The other class contains the receptors for the nonaromatized steroid hormones, including androgens, progestogens, glucocorticoids, and mineralocorticoids (AR, PR, GR, and MR; Eick et al., 2012); this class of SR preferentially binds to steroid response elements (SREs), including palindromes of AGAACA (SRE1) or AGGACA (SRE2) (So et al., 2007; Chusacultanachai et al., 1999). The two classes' DNA specificities are distinct—ERs bind poorly to and do not activate SREs, whereas members of the AR/PR/GR/MR group bind poorly to and do not activate ERE (Zilliacus et al., 1992). Although SRs can and do bind variants of these classic sequences (Welboren et al., 2009; So et al., 2007), the classical ERE and SRE sequences are physiologically relevant and have been the subject of extensive biochemical and structural analysis (Beato et al., 1989; Luisi et al., 1991; Zilliacus et al., 1992; Lundbäck et al., 1993; Schwabe et al., 1993).

Understanding the evolution of a TF-mediated regulatory module requires understanding the origin of the TF's interactions with both upstream stimuli and DNA targets. We recently reported on the mechanisms by which the two classes of SRs evolved their distinct specificities for aromatized or nonaromatized hormones (Eick et al., 2012; Harms et al., 2013). Here, we use ancestral protein reconstruction (Harms and Thornton, 2010, 2013; Thornton, 2004) to identify the genetic, biochemical,

and biophysical mechanisms for the evolution of the distinct DNA specificity in the two classes of SRs. The results, together with previous findings on the evolution of SR ligand specificities, allow us to provide a detailed historical and mechanistic account for the evolution of a new regulatory module.

RESULTS

A Discrete Evolutionary Transition in DNA Specificity

To characterize the evolutionary trajectory of DNA recognition in the SRs, we first used ancestral protein reconstruction to infer the amino acid sequences of the DBDs of the ancestral protein from which all SRs descend (AncSR1) and of the ancestor of all ARs, PRs, GRs, and MRs (AncSR2, Figure 1B). Both proteins predate the evolutionary emergence of vertebrates more than 450 million years ago (Eick et al., 2012). We used maximum-likelihood phylogenetics to infer the best-fit evolutionary model and phylogenetic tree for 213 SRs and related nuclear receptors from a wide variety of animal taxa using sequences of both the DBD and LBD (Figure S1 available online). We then inferred the maximum-likelihood amino acid sequences of the DBD and the posterior probability distribution of amino acids at each sequence site at the phylogenetic nodes corresponding to AncSR1 and AncSR2 (Figures S1A and S1B). The vast majority of sites in the two sequences were reconstructed with little or no uncertainty; only three sites in AncSR2 and 12 in AncSR1 were reconstructed ambiguously, defined as having an alternate state with posterior probability > 0.20 (Table S1).

The distinct specificities of extant SRs could have evolved by partitioning the activities of a promiscuous ancestor among descendants or by a discrete switch from ancestral to derived forms of specificity. To distinguish among these possibilities, we synthesized coding sequences for the inferred ancestral DBDs and characterized their functions and physical properties. We focused on the capacity to bind ERE, SRE1, and SRE2 because these classical REs differ only at two bases in the half-site and are fully distinct in their responses to the two classes of SR (Zilliacus et al., 1992). Using a dual luciferase reporter assay in cultured cells (Figure 1C), we found that AncSR1 had DNA specificity like that of extant ERs, driving strong activation from ERE but exhibiting no expression above background from SREs. AncSR2, in contrast, specifically activated from both SREs but did not activate from ERE. These results are consistent with the strong sequence similarity between AncSR1 and extant ERs and between AncSR2 and the vertebrate ARs, PRs, GRs, and MRs (Figure 1B). They are further corroborated by the pattern of RE specificities across extant members of the SR family tree: because all known descendants of AncSR2 recognize SREs and all other family members and close outgroups bind ERE-like sequences, the most parsimonious expectation by far is SRE specificity by AncSR2 and ERE specificity by AncSR1 (Eick and Thornton, 2011).

Robustness to Uncertainty

To determine whether the inferred functions of AncSR1 and AncSR2 are robust to uncertainty about the ancestral sequences, we synthesized reconstructions of each ancestor that contain every plausible alternate residue. These sequences represent the far edge of the "cloud" of plausible estimates of the true ancestral sequence and are different from the ML sequences at more residues than the expected number of errors in each ML reconstruction (Table S1). These alternative reconstructions therefore provide a conservative test of the robustness of inferences about the ancestral proteins' functions.

We synthesized and assayed these alternate reconstructions and found that the DNA specificities of the alternate reconstructions were nearly identical to those of the ML ancestors (Figure S2A). Moreover, the sequences of extant SRs indicate that none of the plausible alternative residues in AncSR1 or AncSR2 are sufficient to change DNA specificity (Table S2).

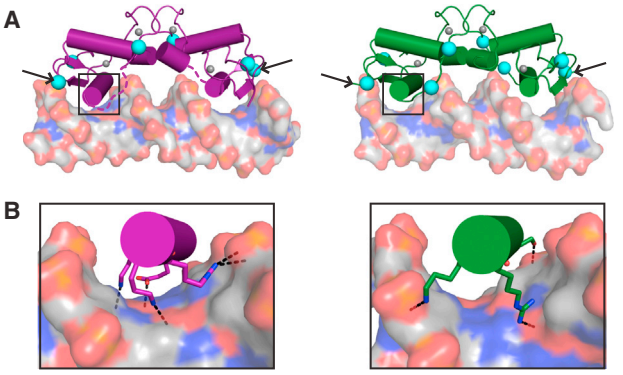
Taken together, these data indicate that the ancestral SR was ERE specific, and recognition of SREs emerged via a discrete change in specificity during the interval between AncSR1 and AncSR2 (Figure 1B). This transition involved a complete loss of activation from the ancestrally preferred ERE and a wholesale gain of novel activation on SREs.

Thermodynamic Basis for Evolution of New DNA Specificity

We next sought to understand the biochemical basis for this ancient change in DNA recognition by expressing and purifying ancestral proteins and characterizing their thermodynamics of binding to DNA. We used fluorescence polarization to determine the macroscopic binding affinity ($K_{A,mac}$) of each ancestral DBD for labeled DNA probes containing palindromic ERE or SREs. The relative affinities followed those in the activation assays, with AncSR1 showing strongly preferential binding to ERE and AncSR2 preferentially binding SREs (Figure 1D and Table S3). These data indicate that the evolutionary transition in the DBD's DNA specificity was due primarily to changes in DNA-binding affinity for the two classes of binding sites (see Bain et al., 2012).

The macroscopic affinity of an SR dimer for a palindromic DNA sequence is determined by two components: the half-site binding affinity (K_1) of each monomer for its half-site and the binding cooperativity (ω) between half-sites, defined as the fold excess of the macroscopic affinity beyond that expected if each monomer binds independently (Figure 1E; Härd et al., 1990). To estimate these parameters, we performed separate fluorescence polarization binding experiments with half-site and palindromic DNA constructs and globally fit the parameters of a two-monomer cooperative binding model to these data.

We found that AncSR1 binds ERE with high half-site affinity and low cooperativity. In contrast, AncSR2 displays much lower half-site affinity but greater cooperativity (Figures 1F and 1G and Table S3). AncSR2's novel RE specificity therefore evolved through a trade-off in the energetic mechanisms of binding: the protein's direct interactions with DNA became weaker as its specificity changed, but this effect was offset by an increase in cooperativity of binding. As a result, the derived DBD retained macroscopic DNA binding affinity for its favored targets similar to that of its ancestor but for a new family of DNA sequences. These ancient changes in binding energetics persist to the present: human ERs, like AncSR1, bind DNA with high half-site affinity and low cooperativity, whereas human GR, like AncSR2, displays considerable cooperativity but lower half-site affinity (Alroy and Freedman, 1992; Härd et al., 1990).



Buried Surface Area (Å^2)

	Chain A::DNA	Chain B::DNA	Chain A::Chain B
AncSR1	617	610	401
AncSR2	405 (388)	343 (324)	565 (565)

Figure 2. Structures of Ancestral Proteins Give Insight into the Molecular Determinants of Specificity

(A) X-ray crystal structures of AncSR1 bound to ERE (left); AncSR2 bound to SRE1 (right). Cartoon shows protein dimers; surface shows DNA. Black arrow, beginning of unresolved C-terminal tail. Dotted line, unresolved loop in AncSR1 near dimerization interface. Cyan spheres indicate sites of permissive substitutions. Gray spheres indicate zinc atoms.

(B) Enlarged view of recognition helix in the DNA major groove (black box in A). Sticks indicate side chains of RH residues making polar contacts with DNA. Dotted lines indicate hydrogen bonds and salt bridges from protein to DNA.

(C) Buried solvent-inaccessible surfaces in Å^2 at the protein-DNA and protein-protein interfaces in the crystal structures for each DBD monomer (chains A and B). Parentheses indicate calculations when residues unresolved in the AncSR1 crystal structure are excluded.

See Table S4.

Atomic Structures of Ancestral DBDs

To identify the causes of these evolutionary changes in DNA binding and recognition, we determined the crystal structures of AncSR1-DBD bound to ERE and of AncSR2-DBD bound to SRE1 at 1.5 and 2.7 Å, respectively (Figure 2 and Table S4). Although their sequences are only 54% identical, AncSR1 and AncSR2 have very similar conformations (RMSD for protein backbone atoms = 0.82 Å). Each monomer buries a recognition helix (RH) in the DNA major groove of one half-site and makes additional contacts to the DNA backbone; the monomers contact each other via a dimerization surface composed of an extended loop coordinated by a zinc atom (Schwabe and Rhodes, 1991; Schwabe et al., 1993; Luisi et al., 1991).

Despite these general similarities, there are several differences between the AncSR1 and AncSR2 structures. First, AncSR1's RH makes more hydrogen bonds to DNA than AncSR2 does (Figure 2B). Second, the loop that connects the RH to the dimerization surface is disordered in AncSR1 but adopts a resolved structure in AncSR2. Third, AncSR1 buries ~60% more of its surface area at the DNA interface than AncSR2 does, but AncSR2 buries ~40% more surface in its dimerization interface than AncSR1 (Figure 2C). These differences are consistent with AncSR1's greater affinity for DNA half-sites and AncSR2's greater cooperativity of dimeric binding.

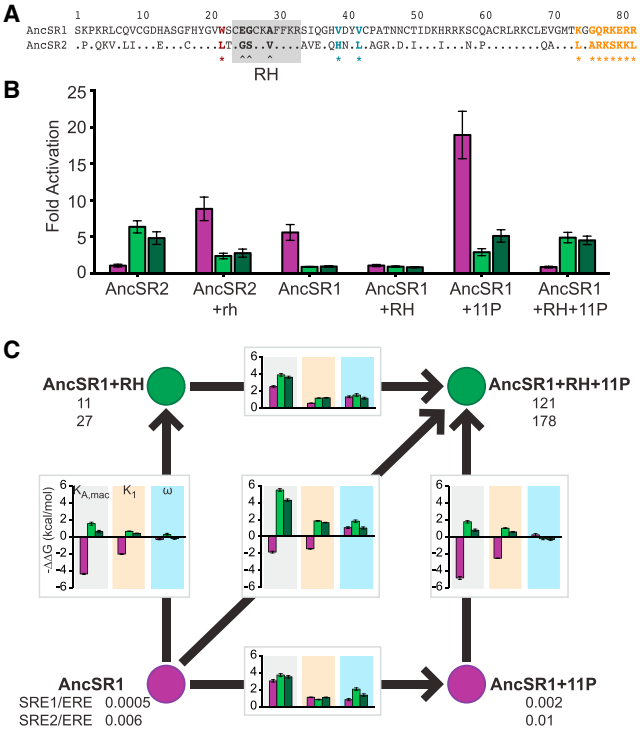


Figure 3. Genetic Basis for Evolution of New DNA Specificity

(A) AncSR1 and AncSR2 sequences. Substitutions between AncSR1 and AncSR2 are shown. Dots indicate conserved sites. The caret (^) indicates recognition helix (RH) and the asterisk (*) indicates permissive substitutions. Gray box, RH.

(B) Effect of RH and 11 permissive (11P) substitutions in luciferase reporter assays. Lower and uppercase letters denote ancestral and derived states, respectively. Fold activation over vector-only control is shown with SEM of three replicates.

(C) RH substitutions shift half-site affinity among REs, and permissive substitutions nonspecifically increase half-site affinity and cooperativity. The corners of the square represent genotypes of AncSR1 with or without RH and 11P substitutions. At each corner, circle color shows RE preference; numbers are the ratio of the K_{Amac} for binding to SRE1 (upper) or SRE2 (lower) versus ERE. Along each edge, vertical bar graphs show the effect of RH or permissive substitutions on the energy of association for the dimeric complex (gray background); contributions of effects on half-site binding (beige) and cooperativity (cyan) are shown. Bar color shows effects on binding to ERE (purple), SRE1, and SRE2 (light and dark green, respectively). Graphs in the square's center show the effect of 11P and RH combined. Mean ± SEM of three experimental replicates is shown.

See Figures S2, S3, and S4 and Tables S3 and S5.

Recognition Helix Substitutions Are Necessary, but Not Sufficient, for Evolution of the Derived Function

We next sought to identify the evolutionary genetic changes that caused specificity to change between AncSR1 and AncSR2. We focused first on the recognition helix because it makes the only direct contacts to bases in the DNA half-site. There are ten residues in the RH, but only three changed between AncSR1 and AncSR2—e25G, g26S, and a29V (Figure 3A, with lower and upper cases denoting ancestral and derived states, respectively). All three residues are strictly conserved in the AncSR1-like state in all ERs and the AncSR2-like state in all AR, PR,

GR, and MRs (Figure S3A). This region is known to play an important role in the specificity of extant SRs (Alroy and Freedman, 1992; Zilliacus et al., 1992).

To test the hypothesis that these three substitutions were the main determinants of the evolutionary change in DNA specificity, we first reversed them to their ancestral state in AncSR2 (generating AncSR2+rh). As predicted, these changes are sufficient to restore the ancestral preference for ERE over SREs in a luciferase assay (Figure 3B). They do so by restoring the DBD's capacity to activate transcription from ERE while dramatically decreasing SRE activation.

We also determined the crystal structure of AncSR2+rh on ERE at 2.2 Å and found that reversing these three substitutions largely restores the ancestral protein-DNA interface (Figures S2B and S2C). The interactions of AncSR2+rh with ERE-specific nucleotides are almost identical to those made by AncSR1. Only a few minor differences are apparent in nonspecific interactions to the DNA backbone and to nucleotides outside of the half-sites, presumably because of differences in crystallization conditions or protein sequence outside the RH. Taken together, these data indicate that the RH substitutions were the primary determinants of the evolutionary change in half-site specificity from ERE to SREs.

To determine whether the RH substitutions were also sufficient causes of the shift in specificity, we introduced the derived RH states into AncSR1. Surprisingly, activation was entirely abolished on all REs tested (Figure 3B). This result is robust to uncertainty about the ancestral sequence; introducing the RH substitutions—which are inferred unambiguously—into the reconstruction of AncSR1 containing all plausible alternative amino acids caused the same effect (Figure S2A). The lack of activity is not due to differences in protein expression between AncSR1 and AncSR1+RH (Figure S2D), implying that the RH substitutions strongly compromise DBD function when introduced into AncSR1, rather than depleting protein in the cell. The derived RH states, however, are conserved in AncSR2 and all of its descendants, all of which activate transcription. These data indicate that additional epistatic substitutions, which permitted the DBD to tolerate the RH substitutions, must have also occurred during the AncSR1/AncSR2 interval.

Permissive Substitutions Outside the DNA Interface Were Required for the Evolution of New Specificity

To identify these permissive substitutions, we divided the 35 other substitutions that occurred during the AncSR1/AncSR2 interval into eight groups based on contiguity in the linear sequence and tertiary structure (Figure S3A). We tested the hypotheses that each group contained permissive substitutions by reverting it to the ancestral state in AncSR2, because reversing a permissive substitution in the context of the derived RH should compromise function. We found that just three groups, containing a total of 16 amino acid replacements, significantly reduced activation when reversed, indicating that the derived states at these sites are necessary for full DBD function and therefore contribute to the permissive effect (Figure S3B and Table S5).

Using a series of forward and reverse genetic experiments testing the effects of the individual mutations within these groups, we ruled out a role for several substitutions and narrowed the set of permissive changes to 11 historical substitutions (11P) distributed among the three structural groups (Figures S4A–S4C and Table S5). When the derived residues at these sites are introduced into the nonfunctional AncSR1+RH, they rescue activation and recapitulate the evolution of the derived DNA specificity (Figures 3A and 3B). Their permissive effect is robust to uncertainty about the precise sequence of AncSR1 (Figure S2A). All three groups are necessary for the full permissive effect (Figure S4D and Table S5).

These substitutions are permissive in that they are required for the protein to tolerate the derived RH, but when introduced into AncSR1, they have no effect on specificity. Rather, they enhance activation nonspecifically on ERE and SREs alike (Figure 3B). Taken together, these data indicate that a large number of permissive mutations, which did not themselves affect specificity, were required for the specificity-switching substitutions to be tolerated.

The effect of these ancient permissive mutations persists to the present. We found that introducing the derived RH states from the human GR into human ERα results in a nonfunctional DBD, just as it did in AncSR1, which is consistent with the fact that the lineage leading to ERs branches from the rest of the SR phylogeny before AncSR2's permissive mutations occurred (Figure S2E). Adding the 11P into the nonfunctional ERα+RH protein, however, rescued activation and yielded a DBD with preference for SREs. Conversely, the ancestral RH states can be introduced into human GR, where they dramatically increase activation on ERE, just as they do in AncSR2 (Figure S2E; Alroy and Freedman, 1992; Zilliacus et al., 1991). Taken together, these results indicate that the ancient RH and permissive substitutions provide a sufficient genetic explanation for the evolution of the distinct DNA specificities of the two major classes of SRs in modern humans.

Evolution of Specificity by Negative Protein-DNA Interactions

Having identified the genetic changes that caused the evolution of AncSR2's new specificity, we sought to understand the biophysical mechanisms by which they did so. We first measured the effect of the RH substitutions on the energetics of sequence-specific DNA binding. We found that they improve the DBD's macroscopic binding preference for SREs by a factor of 30,000; this effect is caused by a 2,000-fold reduction in affinity for ERE and a 15-fold increase in SRE affinity (Figure 3C and Table S3). These effects are entirely attributable to changes in half-site binding affinity, as the RH substitutions do not affect cooperativity (Figure 3C).

To understand the atom-level mechanisms for the effects of the RH mutations, we compared crystal structures of the ancestral DBDs containing the ancestral or derived RH amino acids in complex with both ERE and SRE1; we also performed molecular dynamics (MD) simulations of AncSR1, AncSR1+RH, and AncSR2, each bound to ERE, SRE1, and SRE2. In principle, the evolutionary change in DNA specificity could have been caused by changes in positive interactions—hydrogen bonds or van der Waals attractions between protein and DNA atoms—or in negative interactions, such as electrostatic or steric clashes. If the change in specificity were solely due to changes in positive interactions, then the RH substitutions would reduce

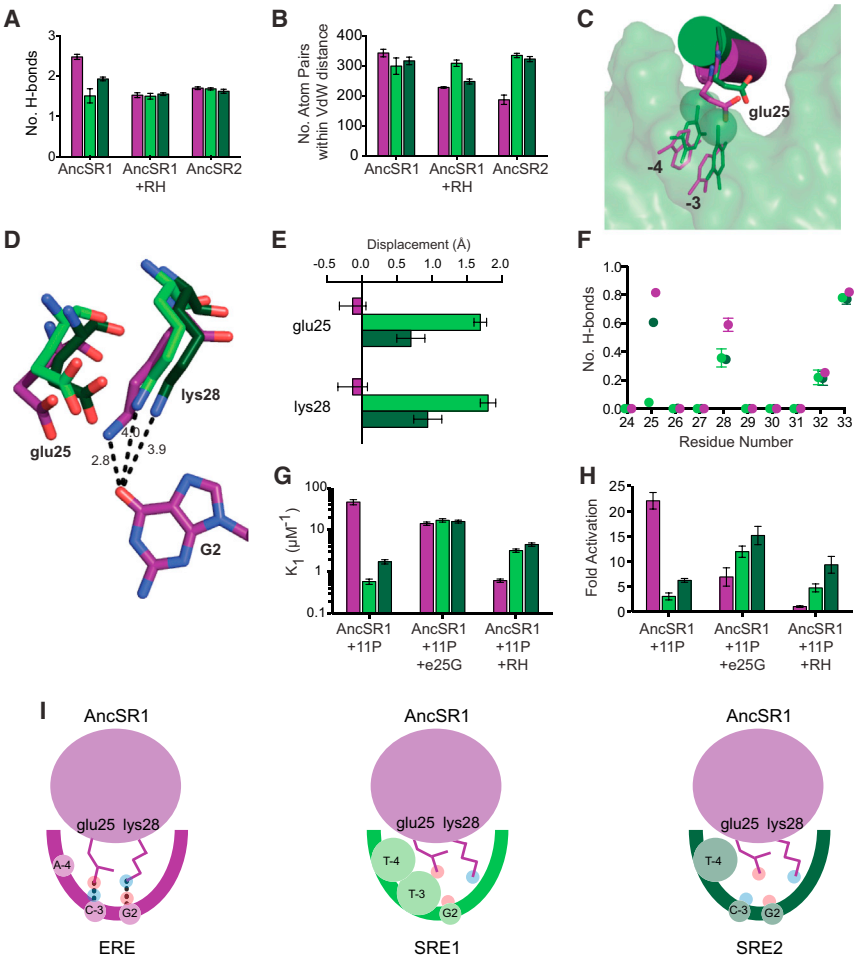


Figure 4. Recognition Helix Substitutions Change DNA Specificity by Altering Negative Interactions

(A) In MD simulations, RH substitutions reduce hydrogen bonds to ERE but do not increase hydrogen bonds to SREs. Bars show mean number of direct hydrogen bonds from all ten RH residues to DNA (purple, ERE; light green, SRE1; dark green, SRE2), each sampled across three MD trajectories with SEM.

(B) RH substitutions reduce packing efficiency at the protein-DNA interface on ERE but do not improve packing on SREs. Bars show the mean number of atoms in the ten RH residues within 4.5 Å of a DNA atom with SEM across three trajectories.

(C) Ancestral residue glu25 (sticks) shifts position due to steric clashes with T-4 and T-3 of SRE1. A representative sample frame from MD trajectories is shown for AncSR1 with ERE (purple) or SRE1 (green). DNA is shown as surface, with atoms in the variable bases −4 and −3 shown as lines; methyls of T−4 and T−3 are spheres.

(D–F) Repositioning of glu25 by SREs causes Lys28 to shift, reducing hydrogen bonds to DNA.

(D) The average position of these residues in MD trajectories of AncSR1 with various REs is shown when all atoms in the protein-DNA complex are aligned. Distance of lys28 from hydrogen bond acceptor G2 on ERE, measured in Å, is shown in black.

(E) Displacement of glu25 and lys28 of AncSR1 on SREs relative to their position on ERE. The mean positions of all atoms in each MD trajectory were calculated, and the DNA atoms in these "mean structures" were aligned in pairs. Bars show the average distances from the atoms in complexes with SRE1 (light green) or SRE2 (dark green) to the corresponding atom in ERE; error is SEM across three replicate trajectories.

(F) Lys28 forms fewer hydrogen bonds to DNA on SREs than on ERE. Points show the mean number of hydrogen bonds formed by each RH residue to different REs with SEM for three MD trajectories.

(G and H) Effect of introducing e25G and other RH substitutions on half-site binding affinity (G) and transcriptional activation (H). e25G enhances binding and activation to SRE without introducing new hydrogen bonds. See Figure S6 and Table S3.

(I) Summary of mechanisms by which ancestral RH excludes SREs. Ancestral glu25 and conserved residue Lys28 form hydrogen bonds (black dotted lines) with ERE bases. These side chains would sterically clash with methyl groups of bases T−3 and T−4 on SRE1 and SRE2, so they are repositioned and are unable to form hydrogen bonds to DNA, leaving unpaired donors (blue) and acceptors (red) at the DNA-RH interface. The RH substitutions resolve the steric clash and remove the unfulfilled donor on e25, increasing SRE affinity.

See Figures S5 and S6.

favorable interactions with ERE and increase favorable interactions with SREs.

Contrary to this prediction, we found that the RH substitutions primarily change negative interactions between the DBD and DNA binding sites, relieving clashes with SRE and establishing new ones with ERE. The ancestral RH does form more hydrogen bonds on ERE than on SREs, and the RH substitutions reduce the number of hydrogen bonds to ERE (Figure 4A); these observations are consistent with the view that positive interactions are the primary determinants of specificity. By removing hydrogen bond acceptors, however, these substitutions also establish unfavorable polar interactions, leaving polar groups on ERE-specific bases unpaired and leading to penetration of transient solvent molecules into the protein-DNA interface (Figures S5A–S5D). The effect of these negative interactions is expected to be much stronger than the loss of the positive interactions. Eliminating a protein-DNA hydrogen bond would reduce binding affinity only slightly because the same number of total hydrogen bonds would form whether or not the protein and DNA are bound to each other or free in solvent. In contrast, leaving an unpaired polar atom at the protein-DNA interface results in more hydrogen bonds in the unbound than the bound state, leading to a much larger difference in energy between the bound and unbound states and a more dramatic reduction in affinity (von Hippel and Berg, 1986).

The improvement in SRE binding also cannot be explained by an increase in SRE-specific positive interactions. The RH substitutions do not increase the total number of hydrogen bonds on

SRE1 and actually reduce the number of hydrogen bonds on SRE2 (Figure 4A). They do so by eliminating or weakening hydrogen bonds formed by the ancestral protein to SREs without forming enough new hydrogen bonds to compensate. Although the derived RH does establish one novel hydrogen bond from derived residue Ser26 to the DNA backbone, this interaction forms more frequently on ERE than on SREs (Figure S5E). Overall, AncSR1+RH (like AncSR2) forms equal numbers of hydrogen bonds with ERE and SREs, indicating that hydrogen bonding does not explain the evolution of preference for SREs. As for van der Waals interactions, the RH substitutions reduce the efficiency of packing on ERE, but they do not improve packing on SREs (Figure 4B). Taken together, these results indicate that changes in positive interactions—hydrogen bonds and van der Waals forces—do not explain AncSR2's increase in affinity or its preference for SREs.

If new SRE-specific positive interactions do not explain the increase in affinity for SREs caused by the RH substitutions, what mechanisms do mediate this effect? We found that the RH substitutions improve SRE affinity by relieving SRE-specific steric and electrostatic clashes with the ancestral RH. Crystal structures and MD simulations both show that the long side chain of glu25 sterically clashes with T-4 and T-3 of SREs; these bases contain large methyl groups that protrude into the DNA major groove of SREs but are absent from the corresponding bases in ERE (Figures 4C and S6A–S6E). As a result of this clash, glu25 is forced to move away from the major groove of SREs and, in turn, to displace the conserved residue Lys28, which in high-affinity complexes, forms hydrogen bonds to DNA bases that do not vary among REs (Figures 4D and 4E). As a result, Lys28 forms fewer hydrogen bonds on SREs compared to ERE (Figure 4F). Additionally, by pushing the negatively charged glu25 away from the bases in the center of the major groove, the SRE-protein interface is left with numerous unpaired hydrogen bond donors and acceptors, leading to water penetration into the interface with SREs (Figures S6F–S6H). The RH substitutions ameliorate this clash by replacing glu25 with the much smaller Gly, thus relieving the negative effect of the glu on SRE binding.

To test the hypothesis that removing glu25 improves SRE recognition by relieving negative interactions, we used site-directed mutagenesis to introduce e25G alone into AncSR1 containing the permissive mutations. We found, as predicted, that SRE affinity and activation were enhanced, despite the fact that Gly25 makes no apparent favorable interactions with SREs (Figures 4G and 4H).

The other two RH substitutions preferentially reduce recognition of ERE, apparently by establishing additional ERE-specific negative interactions. When g26S and a29V are added to e25G, yielding the derived RH genotype, they reduce affinity and activation on all REs but do so much more severely on ERE than SREs (Figures 4G and 4H). The mechanism for this effect is not obvious in the structures or simulations (Figures S6I and S6J), but it does not involve eliminating hydrogen bonds or van der Waals interactions with ERE; neither ancestral amino acid forms hydrogen bonds to ERE (Figure 4F), and they do not pack more efficiently against ERE than the derived amino acids do (Figure S6K).

Taken together, these data indicate that differences in sequence-specific positive interactions do not explain the switch in specificity caused by the RH substitutions. Rather, negative interactions that interfered with SRE binding in the ancestral state were lost, and new negative interactions that impair binding to ERE were gained (Figure 4I). The result was to transform the DBD's ancestral ERE preference into AncSR2's derived SRE preference. A secondary effect was to reduce affinity for the preferred DNA sequence and thus to require permissive substitutions for activation to be maintained.

Permissive Substitutions Nonspecifically Improve Affinity for Both the Derived and Ancestral REs

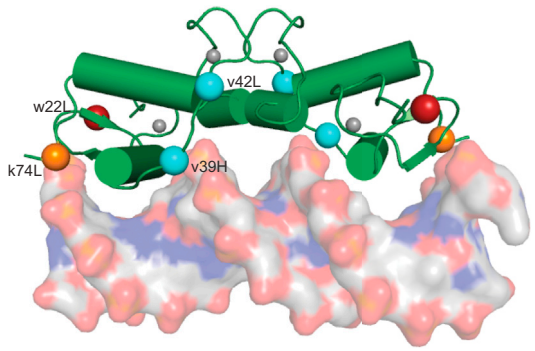
Permissive substitutions are often thought to act by increasing thermodynamic stability, allowing the protein to tolerate mutations that confer new functions but compromise stability (Bershtein et al., 2006; Gong et al., 2013). Using reversible chemical denaturation, however, we found that the 11P substitutions do not increase stability, and the RH substitutions do not decrease stability (Figures 5A and 5B).

Because the RH substitutions radically reduce affinity for ERE and only weakly increase affinity for SREs—yielding a low-affinity receptor for both kinds of element—we hypothesized that the permissive substitutions might offset these effects by increasing affinity in a non-sequence-specific manner. As predicted, introducing 11P into the ancestral background increases macroscopic binding affinity by increasing both cooperativity and half-site affinity on all REs (Figure 3C), indicating a tradeoff in the energetics of binding between the permissive and specificity-switching substitutions during evolution.

The crystal structures suggest that the permissive substitutions cause these effects by enhancing nonspecific protein-protein interactions at the dimerization interface and nonspecific interactions with the DNA backbone and minor groove. Two of the permissive substitutions (v39H and v42L) may facilitate dimer formation because they are located on the loop that links the RH to the dimerization surface (Figure 5A). In AncSR1, as in human ERα, the loop is unresolved, but it is fully resolved in complexes containing the derived state at these residues, including AncSR2, AncSR2+rh, and the human GR (Luisi et al., 1991). Using analytical ultracentrifugation, we found that the permissive substitutions do not measurably increase DBD dimerization in solution (Figures 5C and 5D). We therefore propose that v39H and v42L contribute to cooperativity by stabilizing the dimerization interface in a DNA-dependent manner. Consistent with this view, this loop has been shown in extant SRs to undergo functionally relevant conformational changes when DNA is bound (Wikström et al., 1999; Berglund et al., 1997; Watson et al., 2013; Meijsing et al., 2009). The remaining permissive substitutions may enhance nonspecific DNA binding because they are involved in contacts to the DNA backbone or other base-nonspecific interactions. Substitution w22L is adjacent to several backbone-contacting residues (Figure 5A), and the other permissive substitutions are in the C-terminal tail; although unresolved in our ancestral crystal structures, this region binds directly to the DNA backbone or minor groove just outside the core RE in other nuclear receptors (Helsen et al., 2012; Roemer et al., 2006; Meijsing et al., 2009).

Taken together, our findings indicate that numerous permissive substitutions, which increased nonspecific affinity, were

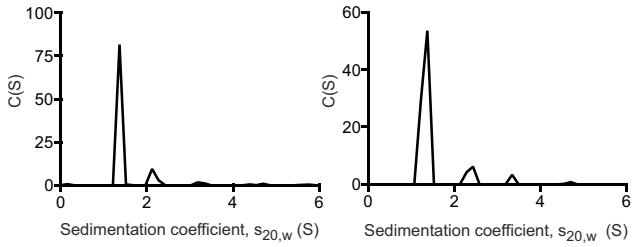
A



B

	ΔG_{H2O} (kcal/mol)	m (kcal/mol)	C_M (M)
AncSR1	-5.1 ± 0.1	1.53 ± 0.09	3.50 ± 0.2
AncSR1+11P	-5.1 ± 0.1	1.87 ± 0.01	2.73 ± 0.02
AncSR1+RH	-4.7 ± 0.1	1.65 ± 0.05	3.00 ± 0.04
AncSR1+RH+11P	-5.0 ± 0.1	1.99 ± 0.01	2.54 ± 0.00

C



D

	$s_{20,w}$	MW (kDa)	MW_{theo} (kDa)	% total	RMSD	f/f_0
AncSR1	1.364	10.5	9.62	83.7	0.013	1.3
AncSR1+11P	1.373	10.4	9.48	85	0.011	1.31

Figure 5. Permissive Substitutions Do Not Improve Protein Stability or Dimerization in the Absence of DNA

(A) Crystal structure of AncSR2 bound to SRE1. Sites of permissive substitutions are shown as Cα spheres; red, cyan, and orange spheres indicate clustered groups of permissive residues. Only one residue in the C-terminal group is shown.

(B) Permissive substitutions (11P) do not increase protein stability. Thermodynamic values for the reversible chemical denaturation of purified DBDs. ΔG_{H2O}, calculated Gibbs free energy of chemically induced unfolding; m, slope of the unfolding transition; C_M, denaturant concentration at which 50% of protein is folded. Error bars are SEM of three experimental replicates.

(C and D) Permissive substitutions do not increase protein dimerization in the absence of DNA, measured by analytical ultracentrifugation. Distribution (C) and best-fit values (D) of sedimentation velocity coefficients ($S_{20,w}$) for AncSR1 (left) or AncSR1+11P (right) at 0.5 mM. The fraction of the total signal under the dominant peak (% total), the estimated molecular weight of that peak (MW), and the expected molecular weight of the monomeric protein (MW_{theo}) show that AncSR1 and AncSR2 are both predominantly monomeric. RMSD indicates root mean square deviation of the data from the model; f/f_0 indicates total shape asymmetry. Signal at higher MW peaks may reflect aggregation due to high protein concentration.

necessary for the affinity-reducing effects of the RH mutations to be tolerated. The evolving DBD therefore traversed sequence space extensively without changing its specificity, reaching regions relatively distant from AncSR1, before the transition to a new function via the RH substitutions could be completed. Selection for the derived specificity could not have driven this exploration; either neutral chance processes (such as drift and linkage) or selection for functions unrelated to specificity must therefore have played crucial roles in the evolution of AncSR2's DNA recognition mechanism.

DISCUSSION

Evolution of a New Gene Regulatory Module

These results, together with our previous work on the evolution of the ancestral ligand binding domain, elucidate the mechanisms by which the distinct regulatory modules mediated by the two classes of extant SRs evolved from an ancestral module mediated by a single TF. We recently reported that AncSR1's LBD also had ER-like functions, responding specifically to estrogens; after duplication of AncSR1, AncSR2 lost estrogen sensitivity entirely and gained activation by nonaromatized steroids (Eick et al., 2012; Harms et al., 2013). Our present findings therefore establish that, during the interval after the duplication of AncSR1, AncSR2's LBD and DBD both evolved entirely new specificities for upstream stimuli and downstream DNA targets (Figure 6A). The other protein lineage produced by this duplication, which led to the present-day estrogen receptors, maintained the specificity of the ancestral signaling module essentially unchanged for hundreds of millions of years. During the period when AncSR2's new specificity evolved, androgens and progestagens were already produced as intermediates in the synthesis of estrogens (Eick and Thornton, 2011).

By evolving distinctly new specificities in both domains after gene duplication, a new regulatory module was established without interfering with the functional specificity of the ancestral module. If one domain of AncSR2 had retained the ancestral specificity while the other evolved new interactions, the information conveyed by the ancestral signaling system would have been compromised by noise—ancestral targets would have been activated by additional stimuli, or the ancestral stimuli would have activated additional targets (Figure 6B). A similar effect would have ensued if the DBD and/or LBD became promiscuous (Figures 6C and 6D). Because the new specificities for hormone and DNA evolved during the same phylogenetic interval, we cannot determine which appeared first. It is possible that a promiscuous DBD arose as an evolutionary intermediate during the transition between the distinct RE specificities of AncSR1 and AncSR2. If it did, however, it did so transiently, was abolished relatively rapidly, and left no promiscuous descendants that persist in present-day species. Thus, the distinct AncSR2-mediated signaling module arose by establishing new functional connections and, just as importantly, by actively erasing the ancestral connections.

Our findings indicate that negative determinants of specificity—mechanisms that actively prevent binding to "nontarget" partners—played key roles in the evolution of the new AncSR2-mediated regulatory module (Figure 6E). In both domains, just a few key mutations—three in the DBD and two in the LBD (Harms et al., 2013) —changed the protein's binding preferences by many orders of magnitude. These substitutions dramatically impaired interactions with the ancestral partner and, to a lesser extent, improved binding of the ancestral TF to the derived

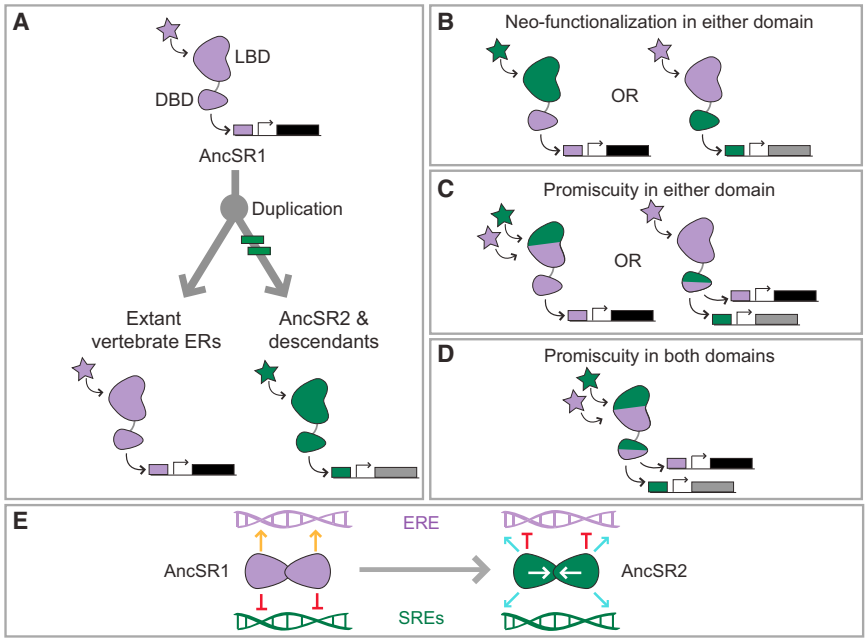


Figure 6. Evolution of a New Regulatory Module

(A) After duplication of AncSR1, the ancestral specificity for estrogens (purple stars) and ERE (purple box) was maintained to the present in the ER lineage. In the lineage leading to AncSR2, ancestral specificity for both DNA and hormone was lost, and novel sensitivity evolved for SREs (green box) and nonaromatized steroids (green star). A new set of target genes (light gray) was thus activated in response to different stimuli. Green hashes mark the branch on which these events occurred.

(B–D) Other potential evolutionary trajectories for evolving new functions would interfere with the ancestral signaling network. (B) Evolution of new specificity for DNA or ligand would cause activation of old targets by new stimuli or activation of new targets in response to ancestral stimuli.

(C and D) Evolution of promiscuity in one or both domains would cause similar effects.

(E) The shift in specificity from ERE (purple helices) to SREs (green helices) in AncSR2 involved losing favorable interactions (orange arrows) to ERE, losing unfavorable negative interactions (red bars) to SRE, and gaining unfavorable interactions to ERE. Offsetting the loss of positive interactions in the DNA major groove, AncSR2 evolved favorable nonspecific DNA contacts (blue arrows) and protein-protein interactions (white arrows in dimer interface) that increased cooperativity.

partner. In both domains, the biophysical mechanisms for this transition involved changes in negative determinants of specificity: the key mutations introduced unfavorable steric or electrostatic clashes with estrogens or ERE and removed clashes that in the ancestral state impaired binding to nonaromatized steroids and SREs (Harms et al., 2013).

Negative Determinants of Specificity: Mutational Constraints on TF Evolution

AncSR2's new DNA specificity was conferred by a complex set of changes: three RH-mediated mutations that changed exclusionary interactions and a large number of permissive mutations that offset the affinity-reducing effects of the specificity-switching mutations. Why did evolution not utilize a simpler mechanism to cause the shift in specificity, such as gains and losses of positive interactions? We propose that differences in the abundance of mutational opportunities to establish negative versus positive mechanisms of specificity determined the evolutionary trajectory by which AncSR2's new mode of DNA recognition evolved.

As a protein evolves, it drifts through a "neutral network" of neighboring genotypes with similar functional outputs; it may cross into a network that encodes different functions, if one is accessible by mutation and compatible with selective constraints (Smith, 1970; Wagner, 2008). Biophysical considerations suggest that there may be few mutational opportunities to increase affinity in a sequence-specific fashion. Establishing a new sequence-specific positive interaction in the complex, heterogeneous interface with DNA would require introducing a side chain of fairly precise length, angle, volume, polarity, and charge to interact favorably with a feature of DNA that is unique to the target sequence, all without disrupting other aspects of the protein-DNA complex. In contrast, the requirements to establish a negative interaction via a steric or electrostatic clash are likely to be considerably less precise, as are those to abolish a hydrogen bond and thereby leave unpaired polar atoms in an interface. Thus, just as the integrated architecture of protein folds makes mutations that stabilize proteins more rare than those that destabilize them (Bloom et al., 2006), the biophysical architecture of protein-DNA interactions should make mutations that shift specificity by establishing new sequence-specific positive interactions much more rare than those that do so by reducing affinity for nontarget sequences.

Evolutionary trajectories that utilize predominantly negative mechanisms to achieve specificity—like those during the evolution of AncSR2's DBD and LBD—should therefore be more likely to be realized than those that change specificity by establishing new, sequence-specific positive interactions. Consistent with this view, directed evolution experiments that select for specific binding to a new DNA target typically reduce affinity (Rockah-Shmuel and Tawfik, 2012). Further, studies that select for binding without selecting for specificity usually increase affinity in a nonspecific fashion (Cohen et al., 2004), indicating that increased affinity often evolves because of nonspecific positive interactions, but specificity is realized largely through sequence-specific negative interactions.

Although they are more numerous, mutations that shift specificity by negative, exclusionary interactions would be eliminated by natural selection if they were to reduce affinity to a level below that required for target gene activation, as the RH substitutions do if introduced directly into AncSR1. The historical permissive mutations, by increasing cooperativity and nonspecific affinity, moved the evolving AncSR2 into a region of its neutral network

in which the historical specificity-inducing mutations could be tolerated. This evolutionary dynamic is similar to that observed for permissive mutations that increase protein stability and therefore allow destabilizing mutations that confer new functions to be tolerated (Bloom et al., 2006). In the present case, however, the critical parameter is the binding affinity of a protein-DNA complex, rather than the stability of the protein fold. Because macroscopic binding affinity is determined by both half-site affinity and cooperativity, permissive mutations that enhance either parameter—or both, as is the case for the evolution of the SR DBD—could facilitate the evolution of new TF specificity and the rewiring of transcriptional circuits (Tuch et al., 2008; Li and Johnson, 2010).

Because of the limitations imposed by mutational opportunities and purifying selection, AncSR2 evolved distinct, high-affinity DNA binding using a mechanism that is not the simplest or most elegant form imaginable for a TF-DNA complex. But it was the mechanism that happened to be available, given AncSR2's chance wanderings through sequence space and the constraints imposed by the physical architecture of SR proteins, DNA, and the interaction between them. That ancient, awkward mechanism persists to the present.

EXPERIMENTAL PROCEDURES

Ancestral sequences and posterior probability distributions for AncSR1 and AncSR2 DBDs were inferred using maximum-likelihood phylogenetics from an alignment of 213 peptide sequences of extant steroid and related receptors, using the maximum likelihood gene family phylogeny and the best-fit evolutionary model (JTT+G) (see Eick et al., 2012). Complementary DNAs coding for these peptides were synthesized, subcloned, and expressed as fusion constructs with the NFκB-activation domain in CV-1 cells. Activation was measured using a dual luciferase assay in which firefly luciferase expression was driven by four copies of ERE or SRE. Variant proteins were generated using Quikchange mutagenesis and verified by sequencing. Tagged DBDs were expressed in *E. coli* and purified by affinity chromatography; we measured the change in fluorescence polarization of 6-FAM-labeled double-stranded DNA oligos as protein concentration increased. Oligos containing a single half-site or a full palindromic element were assayed, and the data were globally fit to a two-site model with a cooperativity parameter to determine the half-site affinity and the cooperativity coefficient (the fold-increase in the K_A of dimeric binding compared to the expected value if the monomers bind independently [Härd et al., 1990]). To measure protein stability, we used circular dichroism to measure the reversible loss of secondary structure in increasing guanidinium chloride. Protein dimerization was assayed by sedimentation velocity analytical centrifugation. For crystallography, purified DBDs were crystallized in complex with palindromic DNA oligos and diffracted at the Advanced Photon Source; structures were determined using molecular replacement. Atomic coordinates were deposited as AncSR1:ERE (PDB 4OLN, 1.5 Å), AncSR2:SRE1 (4OOR, 2.7 Å), AncSR2+rh:ERE (4OND, 2.2 Å), and AncSR2+rh:SRE1 (4OV7, 2.4 Å). Molecular interactions were characterized with molecular dynamics simulations using Gromacs, TIP3P waters and AMBER FF03 parameters for protein and DNA, as well as custom terms for atoms involved in the Cys coordination of Zn atoms (Table S6). For each condition, three replicate 50 ns simulations were run, starting from crystal structures of ancestral proteins; historical mutations were introduced and energy minimized before MD simulation. For details, see the Extended Experimental Procedures.

ACCESSION NUMBERS

The GenBank accession numbers for the AncSR1 DBD and AncSR2 DBD reported in this paper are KM516100 and KM516101, respectively. Atomic coordinates for protein crystal structures reported in this paper were deposited in the RCSB Protein Data Bank with the following accession numbers: AncSR1:ERE, 4OLN; AncSR2:SRE1, 4OOR; AncSR2+rh:ERE, 4OND; and AncSR2+rh:SRE1, 4OV7.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, six figures, and six tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2014.09.003.

AUTHOR CONTRIBUTIONS

A.N.M., J.T.B., and J.W.T. conceived the project. All authors designed the experiments and analyzed data. J.T.B. performed the functional characterizations of ancestral proteins and their variants and identified key historical substitutions; A.N.M. performed the biochemical and biophysical characterizations of ancestral proteins and their variants; D.W.A. performed the molecular dynamics simulations; M.N.M. and E.A.O. performed X-ray crystallography and preliminary biophysical characterizations. A.N.M. and J.W.T. wrote the paper with contributions from all authors.

ACKNOWLEDGMENTS

We thank Geeta Eick for phylogenetic analysis and Mike Harms for extensive advice. We thank Vincent Lynch, Pete von Hippel, members of the Thornton lab, and Will Hudson for comments on the manuscript. The University of Oregon ACISS cluster provided computing resources. This work was supported by NIH R01-GM104397 (J.W.T. and J.T.B.), a Howard Hughes Medical Institute Early Career Scientist Award (J.W.T.), NIH training grant 5-T32-GM-7759-33 (A.N.M.), and AHA award 11PRE7510085 (D.W.A.).

Received: April 28, 2014
Revised: July 9, 2014
Accepted: September 3, 2014
Published: September 25, 2014

REFERENCES

Alroy, I., and Freedman, L.P. (1992). DNA binding analysis of glucocorticoid receptor specificity mutants. Nucleic Acids Res. *20*, 1045–1052.

Bain, D.L., Heneghan, A.F., Connaghan-Jones, K.D., and Miura, M.T. (2007). Nuclear receptor structure: implications for function. Annu. Rev. Physiol. *69*, 201–220.

Bain, D.L., Yang, Q., Connaghan, K.D., Robblee, J.P., Miura, M.T., Degala, G.D., Lambert, J.R., and Maluf, N.K. (2012). Glucocorticoid receptor-DNA interactions: binding energetics are the primary determinant of sequence-specific transcriptional activity. J. Mol. Biol. *422*, 18–32.

Baker, C.R., Tuch, B.B., and Johnson, A.D. (2011). Extensive DNA-binding specificity divergence of a conserved transcription regulator. Proc. Natl. Acad. Sci. USA *108*, 7493–7498.

Baker, C.R., Booth, L.N., Sorrells, T.R., and Johnson, A.D. (2012). Protein modularity, cooperative binding, and hybrid regulatory states underlie transcriptional network diversification. Cell *151*, 80–95.

Beato, M., and Sánchez-Pacheco, A. (1996). Interaction of steroid hormone receptors with the transcription initiation complex. Endocr. Rev. *17*, 587–609.

Beato, M., Chalepakis, G., Schauer, M., and Slater, E.P. (1989). DNA regulatory elements for steroid hormones. J. Steroid Biochem. *32*, 737–747.

Bentley, P.J. (1998). Comparative Vertebrate Endocrinology (Cambridge: Cambridge University Press).

Berglund, H., Wolf-Watz, M., Lundbäck, T., van den Berg, S., and Härd, T. (1997). Structure and dynamics of the glucocorticoid receptor DNA-binding domain: comparison of wild type and a mutant with altered specificity. Biochemistry *36*, 11188–11197.

Bershtein, S., Segal, M., Bekerman, R., Tokuriki, N., and Tawfik, D.S. (2006). Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. Nature *444*, 929–932.

Bloom, J.D., Labthavikul, S.T., Otey, C.R., and Arnold, F.H. (2006). Protein stability promotes evolvability. Proc. Natl. Acad. Sci. USA *103*, 5869–5874.

Brayer, K.J., Lynch, V.J., and Wagner, G.P. (2011). Evolution of a derived protein-protein interaction between HoxA11 and Foxo1a in mammals caused by changes in intramolecular regulation. Proc. Natl. Acad. Sci. USA *108*, E414–E420.

Carroll, S.B. (2008). Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell *134*, 25–36.

Chusacultanachai, S., Glenn, K.A., Rodriguez, A.O., Read, E.K., Gardner, J.F., Katzenellenbogen, B.S., and Shapiro, D.J. (1999). Analysis of estrogen response element binding by genetically selected steroid receptor DNA binding domain mutants exhibiting altered specificity and enhanced affinity. J. Biol. Chem. *274*, 23591–23598.

Cohen, H.M., Tawfik, D.S., and Griffiths, A.D. (2004). Altering the sequence specificity of HaeIII methyltransferase by directed evolution using in vitro compartmentalization. Protein Eng. Des. Sel. *17*, 3–11.

Eick, G.N., and Thornton, J.W. (2011). Evolution of steroid receptors from an estrogen-sensitive ancestral receptor. Mol. Cell. Endocrinol. *334*, 31–38.

Eick, G.N., Colucci, J.K., Harms, M.J., Ortlund, E.A., and Thornton, J.W. (2012). Evolution of minimal specificity and promiscuity in steroid hormone receptors. PLoS Genet. *8*, e1003072.

Garvie, C.W., and Wolberger, C. (2001). Recognition of specific DNA sequences. Mol. Cell *8*, 937–946.

Gong, L.I., Suchard, M.A., and Bloom, J.D. (2013). Stability-mediated epistasis constrains the evolution of an influenza protein. Elife *2*, e00631.

Härd, T., Dahlman, K., Carlstedt-Duke, J., Gustafsson, J.A., and Rigler, R. (1990). Cooperativity and specificity in the interactions between DNA and the glucocorticoid receptor DNA-binding domain. Biochemistry *29*, 5358–5364.

Harms, M.J., and Thornton, J.W. (2010). Analyzing protein structure and function using ancestral gene reconstruction. Curr. Opin. Struct. Biol. *20*, 360–366.

Harms, M.J., and Thornton, J.W. (2013). Evolutionary biochemistry: revealing the historical and physical causes of protein properties. Nat. Rev. Genet. *14*, 559–571.

Harms, M.J., Eick, G.N., Goswami, D., Colucci, J.K., Griffin, P.R., Ortlund, E.A., and Thornton, J.W. (2013). Biophysical mechanisms for large-effect mutations in the evolution of steroid hormone receptors. Proc. Natl. Acad. Sci. USA *110*, 11475–11480.

Helsen, C., Kerkhofs, S., Clinckemalie, L., Spans, L., Laurent, M., Boonen, S., Vanderschueren, D., and Claessens, F. (2012). Structural basis for nuclear hormone receptor DNA binding. Mol. Cell. Endocrinol. *348*, 411–417.

Kumar, V., and Chambon, P. (1988). The estrogen receptor binds tightly to its responsive element as a ligand-induced homodimer. Cell *55*, 145–156.

Li, H., and Johnson, A.D. (2010). Evolution of transcription networks—lessons from yeasts. Curr. Biol. *20*, R746–R753.

Luisi, B.F., Xu, W.X., Otwinowski, Z., Freedman, L.P., Yamamoto, K.R., and Sigler, P.B. (1991). Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. Nature *352*, 497–505.

Lundbäck, T., Cairns, C., Gustafsson, J.A., Carlstedt-Duke, J., and Härd, T. (1993). Thermodynamics of the glucocorticoid receptor-DNA interaction: binding of wild-type GR DBD to different response elements. Biochemistry *32*, 5074–5082.

Lynch, V.J., May, G., and Wagner, G.P. (2011). Regulatory evolution through divergence of a phosphoswitch in the transcription factor CEBPB. Nature *480*, 383–386.

Meijsing, S.H., Pufall, M.A., So, A.Y., Bates, D.L., Chen, L., and Yamamoto, K.R. (2009). DNA binding site sequence directs glucocorticoid receptor structure and activity. Science *324*, 407–410.

Rockah-Shmuel, L., and Tawfik, D.S. (2012). Evolutionary transitions to new DNA methyltransferases through target site expansion and shrinkage. Nucleic Acids Res. *40*, 11627–11637.

Roemer, S.C., Donham, D.C., Sherman, L., Pon, V.H., Edwards, D.P., and Churchill, M.E. (2006). Structure of the progesterone receptor-deoxyribonucleic acid complex: novel interactions required for binding to half-site response elements. Mol. Endocrinol. *20*, 3042–3052.

Rohs, R., Jin, X., West, S.M., Joshi, R., Honig, B., and Mann, R.S. (2010). Origins of specificity in protein-DNA recognition. Annu. Rev. Biochem. *79*, 233–269.

Sayou, C., Monniaux, M., Nanao, M.H., Moyroud, E., Brockington, S.F., Thévenon, E., Chahtane, H., Warthmann, N., Melkonian, M., Zhang, Y., et al. (2014). A promiscuous intermediate underlies the evolution of LEAFY DNA binding specificity. Science *343*, 645–648.

Schwabe, J.W., and Rhodes, D. (1991). Beyond zinc fingers: steroid hormone receptors have a novel structural motif for DNA recognition. Trends Biochem. Sci. *16*, 291–296.

Schwabe, J.W., Chapman, L., Finch, J.T., and Rhodes, D. (1993). The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements. Cell *75*, 567–578.

Smith, J.M. (1970). Natural selection and the concept of a protein space. Nature *225*, 563–564.

So, A.Y., Chaivorapol, C., Bolton, E.C., Li, H., and Yamamoto, K.R. (2007). Determinants of cell- and gene-specific transcriptional regulation by the glucocorticoid receptor. PLoS Genet. *3*, e94.

Stern, D.L., and Orgogozo, V. (2009). Is genetic evolution predictable? Science *323*, 746–751.

Teichmann, S.A., and Babu, M.M. (2004). Gene regulatory network growth by duplication. Nat. Genet. *36*, 492–496.

Thornton, J.W. (2004). Resurrecting ancient genes: experimental analysis of extinct molecules. Nat. Rev. Genet. *5*, 366–375.

Tuch, B.B., Li, H., and Johnson, A.D. (2008). Evolution of eukaryotic transcription circuits. Science *319*, 1797–1799.

Umesono, K., and Evans, R.M. (1989). Determinants of target gene specificity for steroid/thyroid hormone receptors. Cell *57*, 1139–1146.

von Hippel, P.H., and Berg, O.G. (1986). On the specificity of DNA-protein interactions. Proc. Natl. Acad. Sci. USA *83*, 1608–1612.

Wagner, A. (2008). Neutralism and selectionism: a network-based reconciliation. Nat. Rev. Genet. *9*, 965–974.

Watson, L.C., Kuchenbecker, K.M., Schiller, B.J., Gross, J.D., Pufall, M.A., and Yamamoto, K.R. (2013). The glucocorticoid receptor dimer interface allosterically transmits sequence-specific DNA signals. Nat. Struct. Mol. Biol. *20*, 876–883.

Welboren, W.J., Sweep, F.C., Span, P.N., and Stunnenberg, H.G. (2009). Genomic actions of estrogen receptor alpha: what are the targets and how are they regulated? Endocr. Relat. Cancer *16*, 1073–1089.

Wikström, A., Berglund, H., Hambraeus, C., van den Berg, S., and Härd, T. (1999). Conformational dynamics and molecular recognition: backbone dynamics of the estrogen receptor DNA-binding domain. J. Mol. Biol. *289*, 963–979.

Wray, G.A. (2007). The evolutionary significance of cis-regulatory mutations. Nat. Rev. Genet. *8*, 206–216.

Zilliacus, J., Dahlman-Wright, K., Wright, A., Gustafsson, J.A., and Carlstedt-Duke, J. (1991). DNA binding specificity of mutant glucocorticoid receptor DNA-binding domains. J. Biol. Chem. *266*, 3101–3106.

Zilliacus, J., Wright, A.P., Norinder, U., Gustafsson, J.A., and Carlstedt-Duke, J. (1992). Determinants for DNA-binding site recognition by the glucocorticoid receptor. J. Biol. Chem. *267*, 24941–24947.