Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Jennifer Katherine Colucci Date

Historical mechanisms driving the evolution of ligand specificity in steroid hormone nuclear receptors

By

Jennifer Katherine Colucci
Doctor of Philosophy

Graduate Division of Biological and Biomedical Science
Biochemistry, Cell, and Developmental Biology

_____
Eric A. Ortlund, Ph.D.
Advisor

_____
Christine M. Dunham, Ph.D.
Committee Member

_____
John R. Hepler, Ph.D.
Committee Member

_____
Ahsan Husain, Ph.D.
Committee Member

_____
Keith D. Wilkinson, Ph.D.
Committee Member

Accepted:

_____
Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

_____
Date

Historical mechanisms driving the evolution of ligand specificity in steroid hormone nuclear receptors

By

Jennifer Katherine Colucci
B. S., Emory University, 2008

Advisor: Eric A. Ortlund, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Graduate Division of Biological and Biomedical Science
Biochemistry, Cell, and Developmental Biology
2013

ABSTRACT

Mechanisms of the evolution of ligand specificity in steroid hormone nuclear receptors

By Jennifer Katherine Colucci

The genetic and biophysical mechanisms by which new protein functions evolve are central concerns in evolutionary biology and molecular evolution. Despite much speculation, we know little about how protein function evolves in natural proteins. Here, we use ancestral protein reconstruction (APR) to trace the evolutionary history of ligand recognition in steroid hormone nuclear receptors (SRs), an ancient family of ligand-regulated transcription factors that enable long-range cellular communication central to multicellular life. We found that the most ancestral SR, ancSR1, was regulated by estrogens (steroids with aromatic A rings and small substituents at their carbon 17 position). After a gene duplication event, the duplicate SR, ancSR2, evolved specificity towards progestagens and corticosteroids (nonaromatic 3-ketosteroids with bulky substituents at their carbon 17 position) while excluding estrogens from the binding pocket. We show that this switch from ancSR1- to ancSR2-specificity is mediated by the evolution of several large-effect substitutions within the ligand binding pocket (LBP) that confer a stable hydrogen-bond network for the A ring of nonaromatic 3-ketosteroids. We show that recognition of the hormone's carbon 17 substituent in ancSR2 was conferred via a series of epistatic interactions that served to reposition the ligand and exploit available hydrogen bond capabilities within the ligand binding pocket. Finally, we show that ancestral receptors can be modulated by modern pharmaceuticals and suggest that the SR antagonist mifepristone may act as receptor modulator at the proteins coactivator binding cleft.

Historical mechanisms driving the evolution of ligand specificity in steroid hormone nuclear receptors

By

Jennifer Katherine Colucci
B. S., Emory University, 2008

Advisor: Eric A. Ortlund, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Graduate Division of Biological and Biomedical Science
Biochemistry, Cell, and Developmental Biology
2013

Table of Contents

CHAPTER 1: INTRODUCTION 1

STEROID RECEPTORS REGULATE NORMAL AND DISEASE PHYSIOLOGIES 3

NUCLEAR RECEPTOR STRUCTURE 6

SR MECHANISM OF ACTIVATION 7

MOLECULAR EVOLUTION 9

ANCESTRAL GENE RESURRECTION 10

Figure 1.3: Nuclear Receptor Domain Architecture. 16

Figure 1.6: Model of the mechanism of NR activation. 19

A. PERCENT IDENTITY OF SR LBDS: 21

REFERENCES 22

CHAPTER 2: EVOLUTION OF MINIMAL SPECIFICITY AND PROMISCUITY IN STEROID HORMONE RECEPTORS 30

ABSTRACT 32

AUTHOR SUMMARY 33

INTRODUCTION 34

RESULTS AND DISCUSSION 37

Reconstruction and characterization of ancestral proteins. 37

Ancestral structure-activity criteria 39

Minimal specificity in SR evolution. 40

An evolutionary explanation for SR-mediated endocrine disruption 42

Structural causes of SR promiscuity. 43

Promiscuity, selection, and neutrality in the evolution of signaling. 44

METHODS 46

Phylogenetics and ancestral sequence reconstruction 46

Reporter activation assays 47

Alternative ancestral reconstructions 48

Protein expression 49

Crystallization and structural analysis 50

ACKNOWLEDGEMENTS 52

FIGURES 53

Figure 2.1: Evolutionary expansion of the steroid receptors and their ligands. *53*
Figure 2.2: Ligand-recognition rules of ancSR1 and ancSR2. ..... *55*
Figure 2.3: Evolution of minimal specificity. *57*
Figure 2.4: Structural causes of minimal specificity. *59*
Figure 2.5: Histogram of posterior probabilities for ancSR2 ..... *61*
Figure 2.6: Histogram of posterior probabilities for ancSR1 ..... *62*
Figure 2.7: Dose activation curves of ancSR1 ..... *63*
Figure 2.8: Dose activation curves of ancSR2 ..... *64*
Figure 2.9: The specificity of ancSR1 is robust to uncertainty in the reconstruction. *65*
Figure 2.10 The specificity of ancSR2 is robust to uncertainty in the reconstruction. *66*
Figure 2.11: Sensitivities of extant human receptors to an estrogen, androgen, progestagen, and corticosteroid. *67*
Figure 2.12: Activation of the estrogen receptor ligand binding domains of two annelids and human ERα. *68*
Figure 2.13: AncSR2 is not activated by the nonsteroidal ER agonists diethylstilbestrol and genistein and is not inhibited by ICI182870 and 4-hydroxytamoxifen *69*
Figure 2.14: ML steroid receptor phylogeny for ancSR2 ..... *70*
Figure 2.15: ML steroid receptor phylogeny for ancSR1 ..... *71*
Figure 2.16: Unreduced 184-taxon steroid receptor gene duplication phylogeny *73*
Figure 2.17: Omit maps of progesterone and 11-deoxycorticosterone *74*
Table 2.1: Reconstructed sequence of ancSR2 ..... *75*
Table 2.2: Reconstructed sequence of ancSR1 ..... *76*
Table 2.3: ancSR1 and ancSR2 percent similarities *77*
Table 2.4: CID numbers for synthetic and natural steroids used in this study *78*
Table 2.5: Fold preferences for hormone pairs *79*
Table 2.6: Data collection and refinement statistics *80*
Table 2.7: Receptors and organisms used for phylogenetic analyses *81*
Table 2.8: ancSR2 sequence comparison *82*
REFERENCES 83

CHAPTER 3: BIOPHYSICAL MECHANISMS FOR LARGE-EFFECT MUTATIONS IN THE EVOLUTION OF STEROID HORMONE RECEPTORS 88
INTRODUCTION 91
Protein biophysics and evolution. *91*
An evolutionary shift in hormone specificity. *92*

RESULTS AND DISCUSSION 94
Phylogenetic and structural analyses to identify causal mutations. *94*
Two large-effect replacements shifted hormone specificity. *94*
Structural mechanisms for the shift in specificity. *95*
Changes in the energetic landscape of ligand binding. *96*
Experimental analysis of changes in dynamics. *98*
Arg82 is necessary for ligand-specificity *99*
Evolution of proteins as complex physical systems. *100*
METHODS 102
Reporter activation assays. *102*
Sequence conservation analysis *103*
Molecular dynamics methods. *104*
Free energy landscapes. *105*
Characterization of water-penetrated states. *107*
HDX-MS. *107*
ACKNOWLEDGEMENTS 111
FIGURES 112
Figure 3.1: Evolution of ancSR1 and ancSR2 specificity. *112*
Figure 3.2: Large-effect historical mutations drove the evolution of new ligand specificity. *115*
Figure 3.3: Two historical mutations altered the energetic landscape of protein-ligand binding. *116*
Figure 3.4: Ligand-specific disruption of the A-ring hydrogen-bond network. *117*
Figure 3.5: Cognate steroids of the six human steroid receptors. *119*
Figure 3.6: A-ring ligand contacts are largely conserved between ancSR2 (magenta) and ancSR2 (blue) *120*
Figure 3.7: Representative dose activation curves of ancSR2/Q41e/M75l and ancSR2 wild-type *121*
Figure 3.8: Representative dose activation curves of ancSR1 and ancSR1/e41Q/l75M. *122*
Figure 3.9: A control MD simulation with the apo protein. *123*
Figure 3.10: Derived amino acids introduce a new direct contact with the norP 3-keto group *124*
Figure 3.11 Populated rotamers of Glu41 and Gln41 ..... *125*
Figure 3.12: Dependence of number of states on $\Delta G_{barrier}$. *126*

Figure 3.13: Non-aromatized steroid with 3-hydroxyl does not populate frustrated hydrogen bond networks. *127*
Figure 3.14: Historical mutations cause increased peptide solvent exchange in a ligand-dependent manner. *128*
Figure 3.15: Model fits to incorporation vs. time data for the five peptides which exhibited decreased NPT-specific protection factors in the derived state. *130*
Table 3.1: Pubmed compound identifier (CID) numbers for cholesterol and the synthetic and natural steroid hormones tested in this study. *131*
Table 3.2: Conservation analysis of extant naSRs and ERs. *137*
Table 3.3: Simulations display additivity: absolute free energies of barriers are the same for $i \rightarrow j$ versus $j \rightarrow i$ transitions. *138*
Table 3.4: Transition matrices for top 95% of observed states with 2 kcal/mol energy cutoff. *139*
Table 3.5: HDX-MS kinetics model selection. *140*
REFERENCES 141

CHAPTER 4: X-RAY CRYSTAL STRUCTURE OF THE ANCESTRAL 3-KETOSTEROID RECEPTOR – PROGESTERONE – MIFEPRISTONE COMPLEX SHOWS MIFEPRISTONE BOUND AT THE COACTIVATOR BINDING SURFACE.. 145
ABSTRACT 146
INTRODUCTION 147
MATERIALS AND METHODS 149
Reagents. *149*
Expression and Purification. *149*
Crystallization, data collection, structure determination and refinement. *150*
Reporter Gene Assays. *150*
RESULTS 152
Overall Structure. *152*
Mifepristone binds at two distinct surface sites. *153*
Improved resolution of the ancSR2-progesterone structure permits visualization of D-ring contacts. *154*
DISCUSSION 155
ACKNOWLEDGEMENTS 159
FIGURES 160

Figure 4.1: Crystals of the ancSR2–progesterone–mifepristone complex and in vitro activation data. *160*
Figure 4.2: Overall structure of the ancSR2–progesterone–mifepristone complex. *161*
Figure 4.3: Omit maps of bound ligands. *162*
Figure 4.4: Mifepristone binding site interactions. *163*
Figure 4.5: Mifepristone occupies the coactivator protein space. *164*
Figure 4.6: Global alignment of progesterone-bound steroid receptors *166*
Figure 4.7: Mifepristone and 4-hydroxytamoxifen show similar binding modes to the steroid receptor coactivator binding cleft. *167*
REFERENCES 168

CHAPTER 5: EXPRESSION, PURIFICATION, AND CRYSTALLIZATION OF THE ANCESTRAL ANDROGEN RECEPTOR-DHT COMPLEX 172
ABSTRACT 174
INTRODUCTION 175
MATERIALS AND METHODS 177
Reagents. *177*
Cloning. *177*
Expression and Purification. *178*
Crystallization and Data Collection. *178*
RESULTS AND DISCUSSION 180
ACKNOWLEDGEMENTS 181
FIGURES 182
Figure 5.1: Following a series of affinity columns, ancAR1–DHT was purified to homogeneity. *182*
Figure 5.2: Crystals of ancAR1–DHT. *183*
Figure 5.3: Diffraction image of an ancAR1–DHT crystal. *184*
Table 5.1: Data collection statistics for AncAR1-DHT-Tif2. ..... *185*
REFERENCES 186

CHAPTER 6: BEYOND MINIMAL SPECIFICITY: EVOLVING THE ABILITY TO DISCRIMINATE AMONG DIVERSE 3-KETOSTEROIDS 195
ABSTRACT 196
INTRODUCTION 197
MATERIALS AND METHODS 199

Reagents. ... *199*
Structural Analysis. ... *200*
Mutagenesis. ... *200*
Reporter activation assays. ... *200*
RESULTS ... 202
Comparison of estrogen versus progesterone recognition in the ligand binding pocket ... *202*
Which amino acid substitutions facilitate recognition of bulky carbon 17 substituents? ... *203*
DISCUSSION ... 205
Figure 6.1: Phylogeny of the Steroid Receptor lineage. ... *207*
Figure 6.2: Rotation of the 17-acetyl ligand in the binding pocket allows for exploitation of pre-existing hydrogen bond capacity. ... *208*
Figure 6.3: Forward Evolution of D-ring residues increased preference for 17-acetyl ligands. ... *209*
Figure 6.4: Epistatic interactions shaped ancSR2 evolution. ... *210*
Figure 6.5: Evolutionary pathway to the evolution of 17-acetyl recognition. ... *211*
Table 6.1: Hormone sensitivity of WT ancSR2 and mutants. ... *212*
REFERENCES ... 213
CHAPTER 7: DISCUSSION ... 215
HOW DID THE DIFFERENCES IN LIGAND SPECIFICITY BETWEEN THE ERS AND NASRS EVOLVE? ... 216
WHAT ARE THE MECHANISMS THAT DICTATE THE LIGAND PREFERENCES OF ERS AND NASRS? ... 217
HOW CAN ANCESTRAL PROTEINS BE USED TO UNDERSTAND MODERN PHARMACOLOGY? ... 219
HOW DO EPISTATIC INTERACTIONS INFLUENCE THE EVOLUTION OF LIGAND SPECIFICITY? ... 220
COMPOSITE DISCUSSION ... 222
FUTURE DIRECTIONS ... 224
Can we completely recapitulate the switch in hormone selectivity from ancSR1 to ancSR2? ... *224*
What factors contribute to the evolution of androgen specificity? ... *225*
REFERENCES ... 226

ABBREVIATIONS

AF-2: activation function surface 2

AGR: ancestral gene reconstruction

ancAR1: ancestral androgen receptor 1

ancCR: ancestral corticoid receptor

ancSR1: ancestral steroid receptor 1

ancSR2: ancestral steroid receptor 2

AGR: ancestral gene resurrection

AIS: androgen insensitivity syndrome

APR: ancestral protein reconstruction

AR: androgen receptor

CBI: coactivator binding inhibitor

CNS: central nervous system

CTD: C-terminal domain

DBD: DNA binding domain

ER: estrogen receptor

GR: glucocorticoid receptor

H3: helix 3

H4: helix 4

H12: helix 12

HDX/MS: hydrogen-deuterium exchange/ mass spectrometry

HNF4α: hepatocyte nuclear factor 4α

HSP70: heat shock protein 70

HSP90: heat shock protein 90

LBD: ligand binding domain

LBP: ligand binding pocket

MDD: major depressive disorder

MES: 2-(N-morpholino)ethanesulfonic acid

ML: maximum likelihood

MR: mineralocorticoid receptor

NorP: 19-norprogesterone

NorT: 19-nortestosterone

NPT: 19-nor-1, 3, 5(10)-pregnatriene-3-ol-20-one

NR: nuclear receptor

NTD: N-terminal domain

PEG: polyethylene glycol

PR: progesterone receptor

RMSD: root-mean-squared deviation

SR: steroid receptor

CHAPTER 1: INTRODUCTION

Steroid Receptors

Steroid Receptors (SRs) are key regulators of endocrine and metabolic processes that include cell growth, development, immune response, and reproduction [1]. They belong to a larger family of transcription factors known as nuclear receptors (NRs) that evolved at the onset of the metazoan lineage (Figure 1.1) [2]. There are 48 family members, which can either be ligand-regulated or "orphan" receptors, which have no known ligand [1, 3]. SRs are regulated by small cholesterol-derived hormones; these fat-soluble molecules are well suited for binding intracellular receptors due to their ability to easily pass through the lipid bilayer of cell membranes [1, 4]. The SR family consists of the androgen receptor (AR), progesterone receptor (PR), glucocorticoid receptor (GR), mineralocorticoid receptor (MR), and estrogen receptor (ER) (Figure 1.2).

Evolution of Steroid Receptors.

The most ancient NRs were discovered in the demosponge *Amphimedon queenslandica*, the most ancient branch of the metazoan lineage; this species contains two nuclear receptors most closely related to the hepatocyte nuclear factor 4 α (HNF4α) [2]. NRs have also been identified in several pre-Eumetazoan lineages [5]. SRs evolved via a series of two large-scale genome expansions, one prior to the evolution of eumetazoans and one after (Figure 1.1). The oldest SR was cloned from the mollusk Aplysia [6], a protostome, in 2003. Prior to this, it was thought that SRs were only found in deuterosomes and their absence from the ecdysozoas was the result of a genomic loss in the family. This finding implies that the lack of SRs in ecdysozoan lineage represents a loss in the genome [7]. The Aplysia SR has high homology to the ER, however it was surprisingly shown to be constitutively active and did not bind any estrogenic hormones [6]. Since later ERs are ligand sensitive, this raised a question about the ligand dependence of the seminal SR. To examine this, the seminal SR (ancSR1) was resurrected and shown to bind and be

activated by estrogenic compounds (see Figure 3.5) [6]. Following a gene duplication after the evolution of jawed vertebrates, the second SR, ancestral steroid receptor 2 (ancSR2) evolved specificity for an intermediate in the estrogen metabolic pathway, progesterone (Figure 1.2). Androgen and corticoid receptors evolved much later (see Chapters 2-6) (Figure 1.2) [8].

Steroid Receptors regulate normal and disease physiologies

Due to their diverse and important biological roles, SRs are major pharmaceutical targets [1]. For example, the estrogen receptor is a potent regulator across the reproduction continuum from sexual behavior to sperm production to parental behavior [9, 10]. ER is a major regulator in bone growth and maintenance; a deficiency of estrogen is linked with the development of increased bone reabsorption and osteoporosis [11]. Estrogens have the ability to offer neuroprotective effects in the case of stroke, Alzheimer's Disease (AD), and Parkinson's Disease (PD), where elevated estrogen levels lower the risk of these diseases in pre-menopausal women [12-15]. A similar phenomenon protects pre-menopausal women from cardiovascular disease, where estrogens regulate lipid and cholesterol levels, and directly aid in the recovery from vascular injury; men and post-menopausal women have fewer of these protective benefits due to decreased estrogen levels [11]. Reduced estrogen levels can also lead to increased deposition of white adipose tissue in the body's midsection, which can drive obesity [11].

Misregulation of ER can drive breast cancer in one of two manners [16, 17]. In the first, activation of ER leads to proliferation of mammary cells; rapid increase in these cell divisions can lead to genome replication errors which could drive a cancer phenotype. In the second mechanism, estrogen metabolism can lead to production of genotoxic by-products that can damage cellular DNA and promote cancer. The ER is also a factor in the development of ovarian cancers; overexpression of both ERα and ERβ are found in ovarian tumors with a variety of origins. The growth of ovarian cancer cells in culture can be slowed by the addition of ER

antagonists [18]. In contrast to breast and ovarian cancers, where ER activation drives cancer proliferation, the presence of ERβ is protective in colon cancers [19].

The androgen receptor (AR) is a critical component of the development of both male and female primary and secondary sexual characteristics [20]. Androgens are the primary hormones within the male prostate, Wolffian ducts, urogenital tubes, and hair follicles [20]; AR regulates female reproduction and ovulation [21]. The partial or complete inability for a cell to respond to androgens is known as Androgen Insensitivity Syndrome [22-24]. This syndrome results in the impairment or cessation of the development of secondary sexual characteristics [24]. AR is the primary driver of prostate cancer and is shown to have high expression levels throughout cancer progression [25]. Mutation of AR during prostate cancer progression can lead to constitutive activity of activation of the receptor with antiandrogens, making the cancer difficult to treat [25]. The AR has also been shown to contribute to the progression of ovarian cancer [26]. Administration of antiandrogens has been shown to inhibit ovarian cancer cell growth [27, 28]. In addition to affecting reproduction and cancer development, the loss of androgens has been linked to the progression of cardiovascular disease in men. AR has also been shown to demonstrate DNA-independent effects in an androgen-sensitive manner [29].

The progesterone receptor, AR's closest evolutionary relative, also exerts effects on the reproductive system and the progression of some cancer subtypes [30]. While AR can drive the progression of ovarian cancer, PR has been shown to exert protective effects on the ovaries; preventative administration of synthetic progesterone can decrease the likelihood of acquisition of this cancer [31]. The expression of PR has been shown to be protective in other cancer subtypes as well, including non small cell lung cancer (NSCLC) [32], colorectal cancers [33], endometrial cancer [34], and breast cancer [35].

In addition to its effects on cell growth, PR is a potent driver of various other metoazoan physiological processes. Progesterone is the primary hormone controlling reproductive processes in females, including function of the mammary glands, thymus, ovaries, and uterus [36]. PR is both necessary and sufficient to elicit the reproductive responses necessary for female fertility [37]. In addition to its reproductive effects, progesterone affects Central Nervous System (CNS) functions ranging from neurogenesis and myelination to cognition and inflammation [38].

The Glucocorticoid Receptor (GR) is the primary regulator of the stress pathway, guiding both short- and long-term stress responses [39]. GR responds selectively to the steroid hormone cortisol in a stress-induced and circadian manner to regulate a range of metabolic and homeostatic functions [40]. Glucocorticoids have been shown to mediate inflammatory processes in response to immune challenges [41] through the up-regulation of anti-inflammatory proteins and the down-regulation of pro-inflammatory proteins [42]. These processes are highly targeted for pharmaceutical intervention and immunosuppression through the use of agents such as hydrocortisone, cortisone, prednisone and other glucocorticoid agonists. These drugs are used to treat an array of autoimmune, inflammatory, and allergic disorders [43]. In addition to their role in inflammatory processes, glucocorticoids have been implicated in neurological function and have been shown to play a part in Major Depressive Disorder (MDD), bipolar mania, posttraumatic stress disorder, panic disorder, and schizophrenia [39, 44]. Finally, it has been suggested that GR can exert non-cannonical effects on a cell through interaction with membrane-located or mitochondrial receptors [45].

The closest evolutionary relative of GR is the mineralocorticoid receptor (MR). Unlike ER, PR, AR, and GR, MR has a broader ligand specificity. MR responds highly and equally to corticosterone, aldosterone, and cortisol [46]; it also shows high affinity for progesterone yet this hormone does normally not drive transactivation [47]. A prevalent serine-to-leucine mutation in the MR ligand binding domain (LBD), however, confers progesterone responsiveness

contributing to hypertension during pregnancy when circulating progesterone levels are high. In polarized epithelial tissue (such as the colon and kidneys), MR regulates sodium and water balance through the apical Epithelial Na Channel (ENaC) [48] and the basolateral Na^+, K^+-ATPase pump [49]. In addition to its epithelial functions, MR is a potent regulator of non-epithelial tissues, such as the heart and vasculature tissue. Dysregulation of the MR can result in cardiac failure and hypertension [50]. A greater understanding of SR biology and mechanisms of activation has widespread implication in the development of important therapies.

In addition to their canonical functions, steroid hormones have recently been found to possess secondary functional characteristics. The steroid estrogen is able to interact with the G-protein coupled receptor (GPCR) GPR30 at the cell membrane to transduce activation [51]. In addition, ER has the ability to mediate the inhibition of NF-κB through non-canonical pathways [52].

Nuclear Receptor structure

Nuclear Receptors have a modular architecture, consisting of a five domains: an N-terminal domain (NTD), DNA binding domain (DBD), flexible hinge region, a LBD, and a C-terminal domain (CTD) (Figure 1.3) [1]. The NR NTD is highly variable in size and sequence (Figure 1.3). The DBD is a small and highly conserved domain of NRs due to constraints needed to preserve DNA binding and overall structure (Figure 1.3, Table 1.1). It binds target DNA sequences (typically inverted repeats) [53, 54] through two zinc finger domains, which also serve to maintain domain architecture [55, 56]. The DBD also contains a nuclear export signal, allowing some receptors to cycle back into the cytoplasm after affecting transcription [57]. The NR hinge region is the least conserved domain and is a flexible connector that varies in size between the DBD and LBD. C-terminal to the LBD, a short CTD contains nuclear localization signal to bring the receptor into the nucleus upon ligand binding [57]. The size of NRs varies widely across the 48 members. The SR subfamily has a more highly conserved size; the total

molecular weight is approximately 100 kD for the naSRs and 66 kD for the estrogen receptors (Figure 1.3).

The NR LBD is structured like a three-layered alpha-helical sandwich (Figure 1.4) [7]. The top half of the receptor is very rigid (helices 1,4,5,8, and 10), as demonstrated by lower crystallographic temperature factors [1]. The bottom section of the receptor is missing the middle layer of the alpha helical sandwich, creating a cavity for the ligand. This region is less rigid, as seen through high crystallographic temperature factors and broadening of NMR spectra, suggesting that NRs have properties of molten globules [58, 59]. The sides of this cavity are enclosed by a two-stranded beta-sheet and the C-terminal helix, helix twelve (H12) (Figure 1.4). SRs are modular and each domain, including the LBD, has been shown to both fold and function independently of the remainder of the protein [60]. Thus, a large percent of research on LBD function utilizes this domain alone.

Promiscuous NRs, such as the pregnane X receptor, tend to have a large ligand cavity, up to 1000 $Å^3$, allowing a diverse array of ligands to bind in different orientations; more specific NRs, such as SRs, have much smaller ligand cavities that better compliment the shape of their ligands [61]. The interior of this cavity, also called the ligand binding pocket (LBP), is lined primarily with nonpolar amino acids, but contains several polar residues to correspond to the character of the ligand.

SR mechanism of activation

Prior to ligand binding, SRs are sequestered dormant in the cytoplasm by heat shock proteins (HSPs) (Figure 1.5). HSP90, HSP70, and p60 hold apoSRs in a ligand-ready state [62]. Upon ligand binding, the HSP complex is remodeled, exchanging HSP70 and p60 for HSP90, FKBP52, and p23, triggering nuclear import [63] (Figure 1.5). Once in the nucleus, SRs interact

with transcriptional coregulators (Figures 1.5, 1.6). These proteins can either be co-activator proteins or co-repressor proteins, which serve to promote or repress transcription, respectively (Figure 1.6).

Coactivator proteins interact with SRs via an alpha helix containing a short LxxLL motif (L- leucine, x- any amino acid) [64, 65], whereas corepressor proteins' interaction motif consists of LxxxIxxx[I/L] (Figure 1.6) [66-69]. These two types of coregulator proteins bind with mutual exclusivity to the same hydrophobic groove on the surface of the SR termed the activation function surface 2 (AF-2) (Figure 1.6) [66-71]. The AF-2 is comprised of helices 3, 4, and 12; it is held in place by a charge clamp formed by a lysine on H3 and a glutamate on H12 that interact with the helix dipole (Figure 1.6) [65, 72]. The choice of coregulator is dictated by the identity of the bound ligand as well as the levels of coregulators in a particular tissue. Apo receptors have a highly mobile H12, as shown by fluorescence anisotropy [73, 74]. This mobility is significantly diminished upon ligand binding [73, 74]. A ligand-induced shift allows H12 of the SR to undergo a conformational change that either places H12 across the middle of the coregulator binding groove or beside it [1]. When H12 is unbound from the coactivator cleft, a long corepressor interaction motif is free to bind (Figure 1.6) [72, 75]. When H12 is positioned across the groove in its "active" conformation, corepressor binding is excluded and coactivator binding occurs (Figure 1.6). In addition to ligand identity, coregulator choice is influenced by the relative amounts of corepressor and coactivators in a given cell type [76].

Ligand binding stabilizes the NR LBD through creating a less conformationally dynamic receptor (this will be further discussed in Chapter 3). Thermal denaturation studies show that ligand bound LBDs have higher melting points by promoting intramolecular interactions [77, 78]. Internal tryptophan fluorescence assays show that ligand bound SR LBDs are more resistant to chemical-induced denaturation than apoSRs [79]. Size exclusion chromatography demonstrates that ligand bound LBDs have a smaller hydrodynamic radius than apoLBDs [77]. ApoLBDs are

also more sensitive to proteolytic cleavage, demonstrating greater conformational dynamics [80, 81]. Despite this understanding, we still lack full comprehension of how SRs achieved their ligand specificity. Here, we address this challenge through the investigation of SRs' molecular evolution.

Molecular Evolution

Molecular evolution is the study of the process by which macromolecular families with diverse functions evolve. There are two primary types of molecular evolution within proteins: functional evolution and fold evolution. Functional evolution involves alteration of the ancient functionality and adaptation to the current molecular environment. There are cases when a single residue change is sufficient to alter protein function [82], and there are also cases where large structural and residue changes occur with little change in functionality [83]. For example, two amino acid changes in the human androgen receptor are sufficient to lower the receptor's sensitivity to androgens as well as increasing broad agonism by non-androgens [84]. Fold evolution involves structural changes to a molecule, with or without functional changes [85]. For example, bacterial luciferase and nonfluorescent flavoprotein of Photobacterium show 30 % sequence identity, belong to the same bacterial operon, and yet show highly different overall folds [86]. Protein structure is much more highly conserved than protein sequence [87] and evolves at a much slower rate than function [85, 88, 89]. Within any given protein family, only 5% of residue identities might be shared; however, protein families typically have highly similar folds, with up to 50% conservation in root mean square deviations (RMSD) within 3.0 Å [87, 90].

Gene families arise when a gene duplicates and one copy of the gene is allowed to drift and acquire new functionality. Once a duplication event occurs in an individual, this change has the potential to be fixed in the population (10). The second copy of this gene is considered a paralog

[91]. Since two functionally redundant genes provide no evolutionary advantage (except in the case where increased RNA or protein product is beneficial), the fixed paralogous gene has the opportunity to undergo drift [91]. These mutations can either generate a pseudogene, which is either unexpressed or functionless [92, 93] or attain new functionality. This new functionality can either stem from concerted evolution (where a new gene with similar, but not identical, function arises [94]), subfunctionalization (where the duplicated genes each take on a portion of the parental gene's function [95]) or neofunctionalization (where an entirely new function emerges) [91] (examples of these phenomena will be discussed in Chapters 2, 3, 4, and 6). The phenomenon of gene duplication was first published in 1936 by Bridges [96] but the extent of its prevalence in the genome wasn't fully understood until the late 1990s when genome sequencing became more widespread. It is estimated that approximately 70% and 90% of domains in prokaryotes and eukaryotes, respectively, have been duplicated [97, 98].

A fuller understanding of modern biology can be achieved through understanding both protein function and how that function arose. For example, simple analysis of our evolutionary history reveals conserved sites and functions and generates a broader understanding of extant function. Understanding the causes of protein evolution is key to universal integration of molecular evolution, protein biochemistry, and structural biology.

Ancestral Gene Resurrection

Protein families with diverse functions provide ideal systems for deciphering the evolution of novel functionality. The identification of the mechanisms by which these varied functionalities arose can be complex since we lack records or samples of ancestral proteins. A majority of sequence variation between proteins occurs via neutral drift, which has no apparent impact on function but tends to destabilize the protein over time [99]. These numerous neutral

substitutions can mask the function-shifting changes making them difficult to identify even when structural information is available [100]. In addition, large-effect changes that involve function-switching residues often destabilize the protein due in large part to the accumulation of neutral mutations [100]. Thus, it is easier to decipher functional changes between internal nodes on a phylogenetic tree than between paralogs that have accumulated neutral mutations. Horizontal amino acid mutation between protein family members (i.e. residue switching across extant proteins) often yields inactive proteins [94]. Through evolutionary history, proteins accumulate permissive mutations that buttress the structure allowing for new functional changes (an example of this will be discussed in Chapter 6). Simply switching large-effect function-dictating residues often doesn't yield a functional protein since the supporting permissive mutations are not in place. These challenges can be overcome through use of a vertical approach to studying protein family function through evolutionary history. Through resurrection of protein family ancestors, one can study the effects of historical mutations in an appropriate sequence background.

The idea of ancestral gene resurrection was first proposed by Drs. Linus Pauling and Emile Zuckerkandl in 1963 [99]. They proposed the possibility of determining the sequence of ancestral genes and the ability to "synthesize these presumed components of extinct organisms" as well as to "study the physico-chemical properties of these molecules" [99]. Decades later, when computing power increased, Pauling and Zuckerkandl's vision became feasible. At first, consensus methods were used for ancestral gene resurrection studies [101, 102]. This method stated that the most commonly occurring sequence was the ancestral state, irrespective of phylogenetic relationship. However, this method was flawed due to its high sensitivity to the genes selected for analysis [103], e.g. inclusion of multiple mammalian gene sequences would bias the analysis towards a mammalian sequence.

In 1970, the Maximum Parsimony method was proposed and was first applied in the 1980s to ancestral gene deduction. This method is based on seminal work done by Dr. Walter M.

Fitch in which the sequence at each internal node on a phylogenetic tree is deduced so as to result in the fewest number of sequence changes over time [104]. By taking into account the phylogenetic relationship between the selected orthologous sequences, this method is more accurate (98.6% experimentally) at inferring the correct ancestral sequence [105, 106]. By 1990, it became possible to synthesize these ancestral genes and functionally characterize their properties. At this time, Dr. Steven A. Benner showed that a resurrected five to ten million year old ribonuclease was able to degrade RNA with equivalent efficacy to extant ribonucleases [107, 108]. Shortly thereafter, Dr. Ahsan Husain showed a second example of a resurrected protein maintaining extant enzymatic capabilities [109], describing substrate specificity in the serine protease family.

While Maximum Parsimony was used in early resurrections, the method was intrinsically flawed. Maximum Parsimony treats every substitution as equally probable; however, evolutionary biases can cause substitutions to occur more or less frequently than others. Additionally, Maximum Parsimony is less efficacious when a sequence site mutates more than once over evolutionary time, and differentially reconstructed trees can be calculated as equivalently parsimonious. As branch lengths, and thus amino acid substitutions, increase, Maximum Parsimony is found to be increasingly less effective [91].

Most recently, Maximum Likelihood (ML) analyses have been developed as effective tools for phylogenetic analysis and ancestral gene reconstruction (AGR). This method uses robust Bayesian statistics to calculate the posterior probability of potential sequences; the posterior probability is the likelihood that a given hypothesis is true given and that the successive hypotheses are true. In this case, the posterior probability of an ancestral sequence being correct is directly dependent on identity of the extant sequences [110] . The ability to apply strong statistics to the resurrection confers confidence to the ML method, thus ML is more efficacious at resolving ambiguous phylogenies than the maximum parsimony method [105]. Unlike Maximum

Parsimony, ML takes into account evolutionary biases, such as conservative amino acid replacements, as well as phylogenetic branch lengths [103]. Finally, ML infers phylogenetic relationships rather than them being assigned. As such, the branch lengths of an evolutionary relationship are inferred to generate a phylogenetic tree that is most likely to be true given the sequence data [105]. This serves to restrict biases generated by incorrect phylogenetic models. Prior to the advent of ML, only genes up to 5-100 million years old could be resurrected. With ML analysis, genes that are 240 million to over 1 billion years old may be accurately reconstructed [103].

The aforementioned sequence analyses use amino acid sequences because they contain less background "noise" than DNA sequences [103]. Once the protein sequence is determined, the DNA sequence is later inferred using codon bias corresponding to the expression system (Figure 1.7) [103]. Since the advent of higher automation synthesis, longer oligonucleotides can now be synthesized to high purity. This advancement eases the generation of ancestral genes. Once generated, the DNA sequence can then be synthesized, and the protein can be expressed, purified, and functionally categorized.

Figures

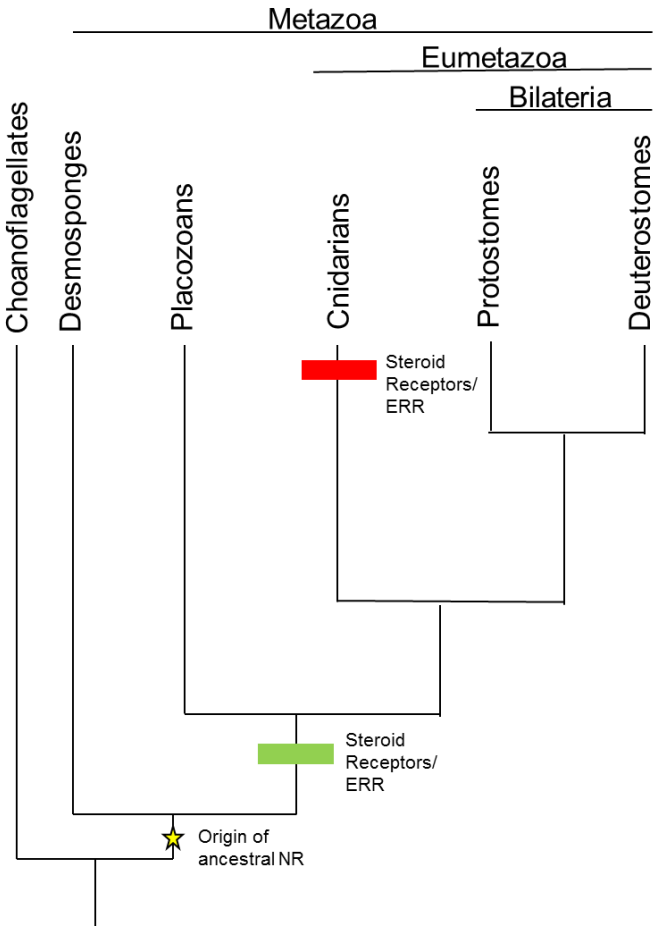


Figure 1.1: Origins of SR Evolution

Gene duplications represented by green boxes, gene loses represented by red boxes. Adapted from Bridgham, 2010 [2].

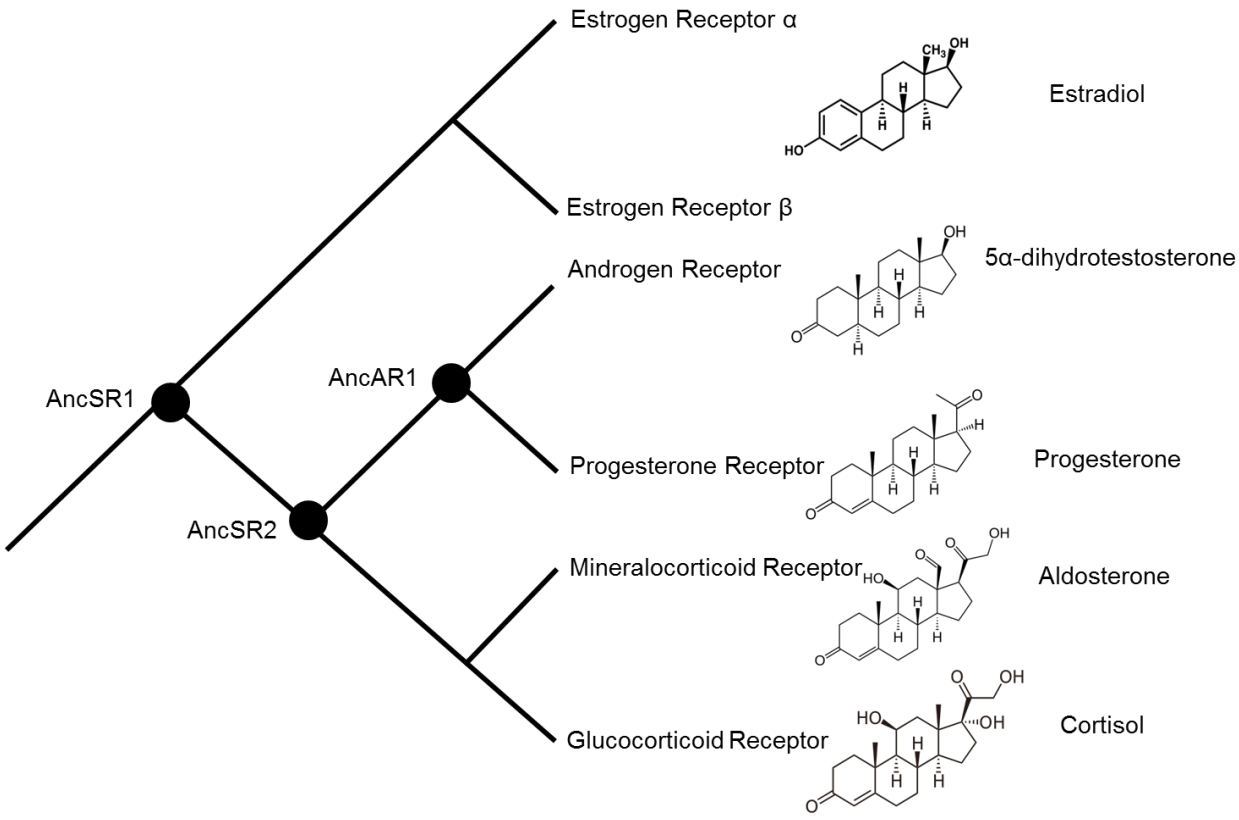


Figure 1.2: Cladogram showing the evolution of Steroid Receptors.

The evolutionary relationship between the SRs is shown. The cognate ligand for each SR is depicted. The three ancestral SRs that will be discussed in Chapters 2-7 are shown as black circles.

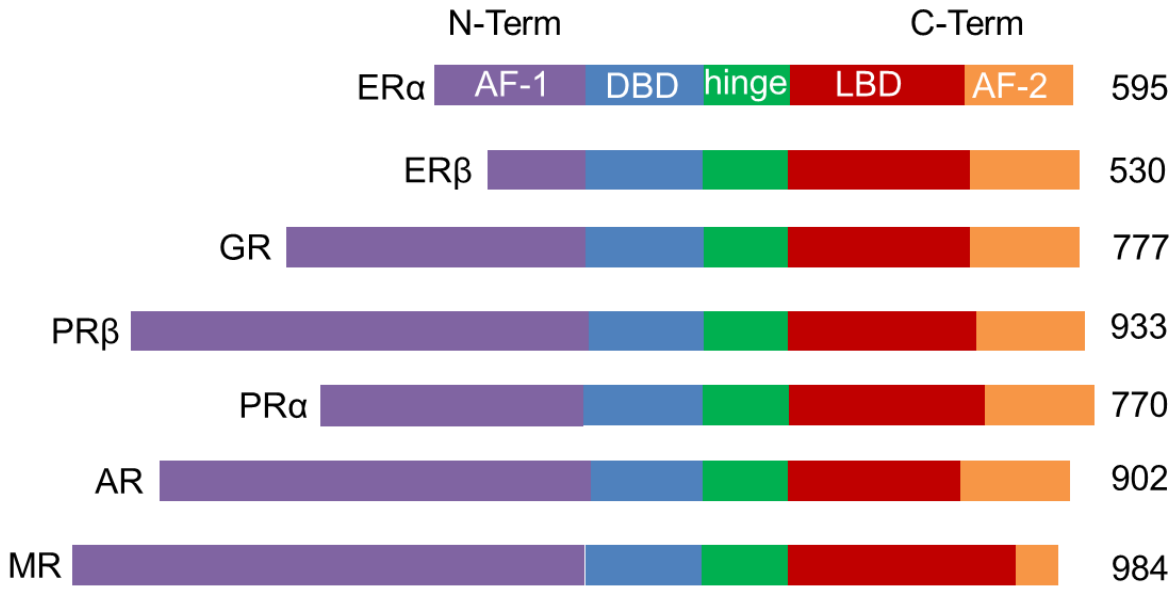


Figure 1.3: Nuclear Receptor Domain Architecture.

Nuclear Receptors have a modular domain architecture consisting of five parts. The AF-1 and DBD are separated from the LBD and AF-2 by a flexible hinge region. AF-1 – activation function helix 1, DBD – DNA-binding domain, LBD – ligand binding domain, AF-2 – activation function helix 2.

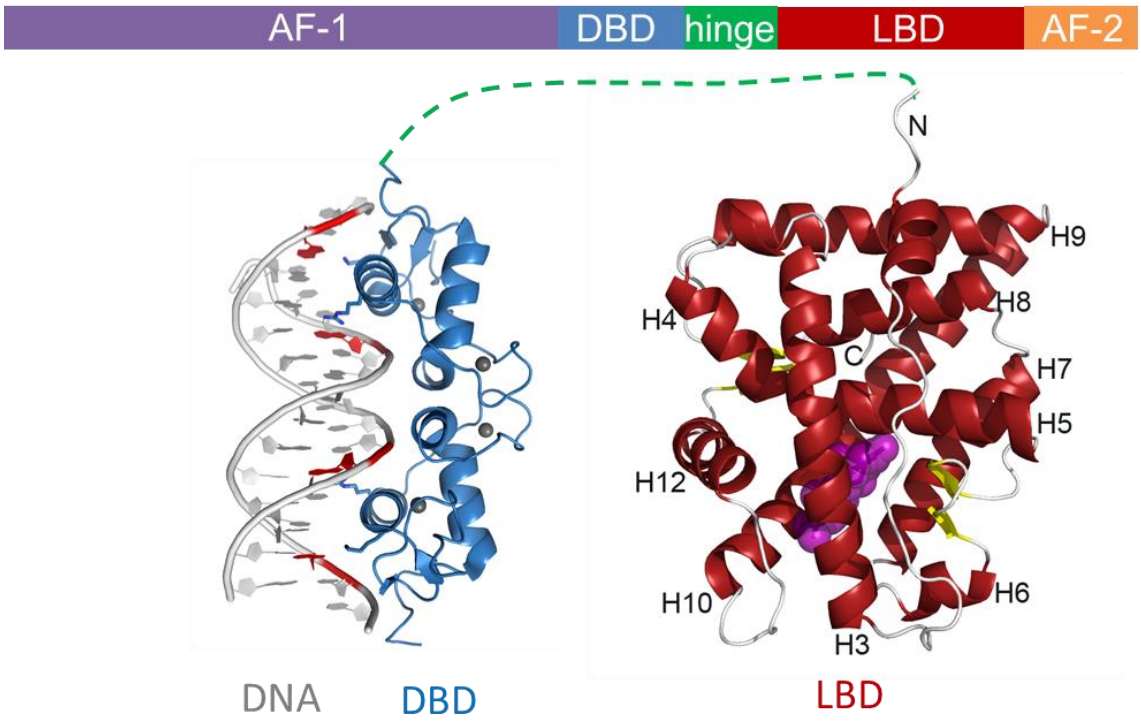


Figure 1.4: Representative Crystal Structures for the SR DBD and LBD

Crystal structure of an ancestral steroid receptor bound to ligand at 2.75 Å. Helices are labeled H1 through H12. Helices -- red; loops -- white; beta sheets -- yellow; ligand -- magenta.

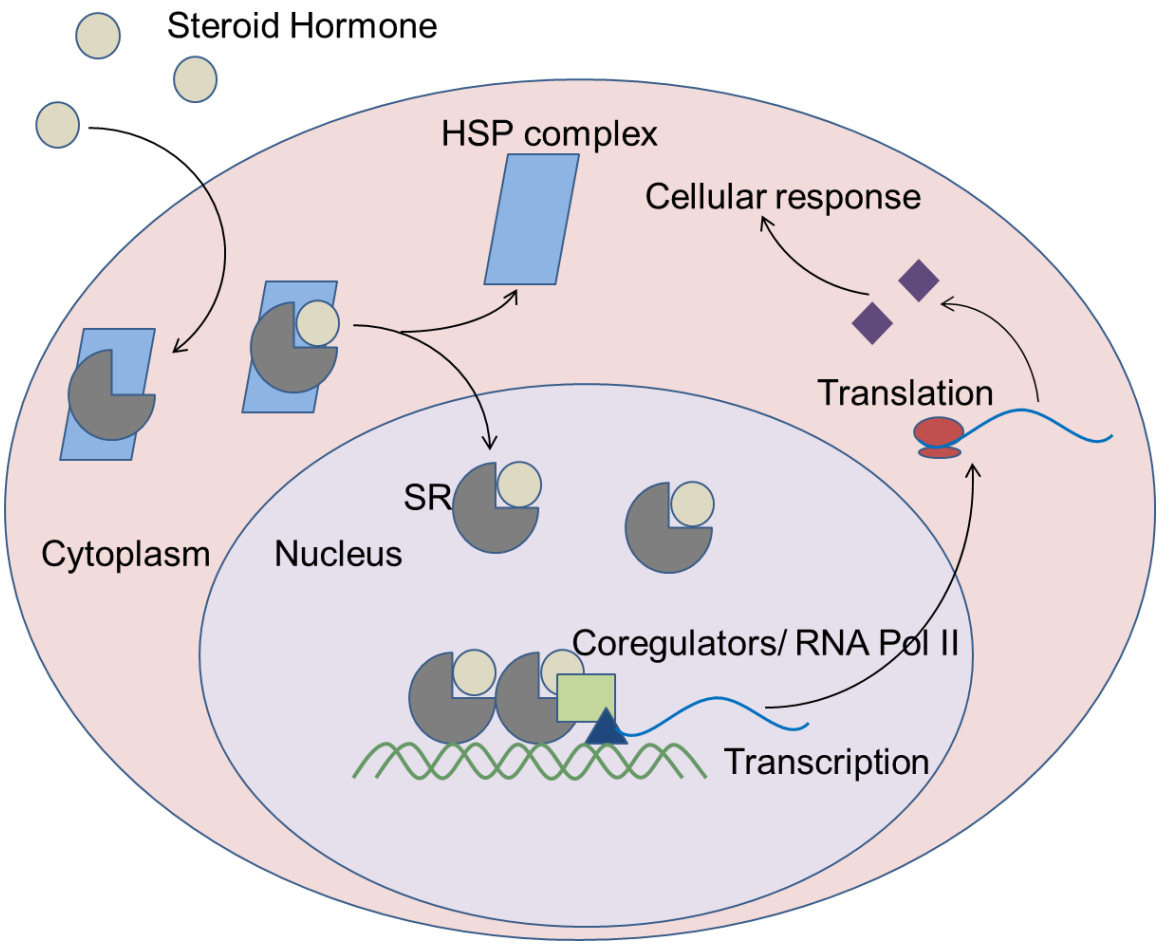


Figure 1.5: Cellular response to SR activation.

A steroid hormone passively diffuses into the cytoplasm of a cell, where SRs are held in an inactive state by a heat shock protein complex. Upon binding to hormone, this heat shock protein complex dissociates and the SR-hormone complex translocates to the nucleus. In the nucleus, the receptor-ligand complex binds DNA as a dimer, recruits coregulators and RNA Polymerase II to activate transcription of downstream genes. This mRNA travels out of the nucleus, where it is translated and those proteins elicit a cellular response to activation of the SR.

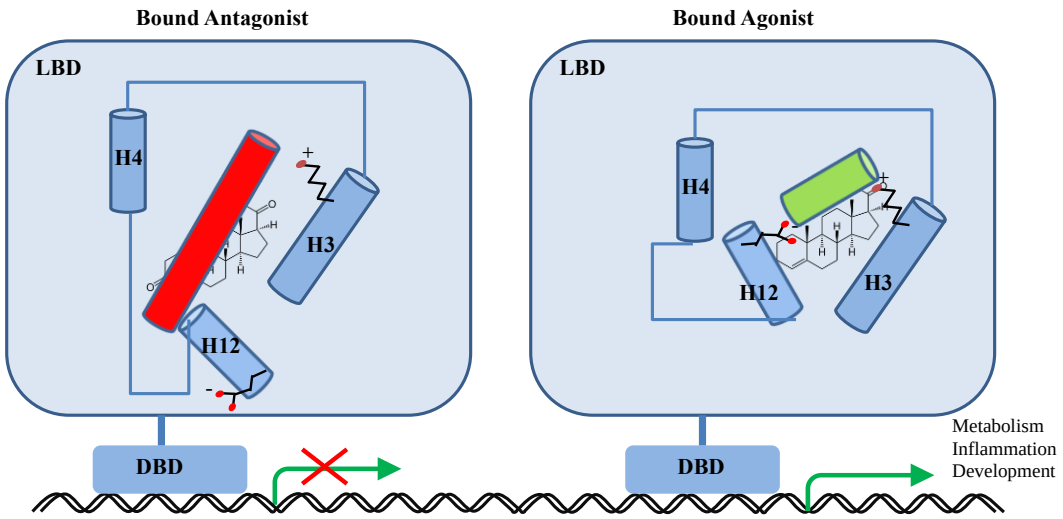


Figure 1.6: Model of the mechanism of NR activation.

The NR LBD undergoes conformational changes upon ligand binding, allowing for transcriptional regulation. Left figure shows a bound antagonist and co-repressor protein (red), right figure shows a bound agonist and coactivator protein (green).

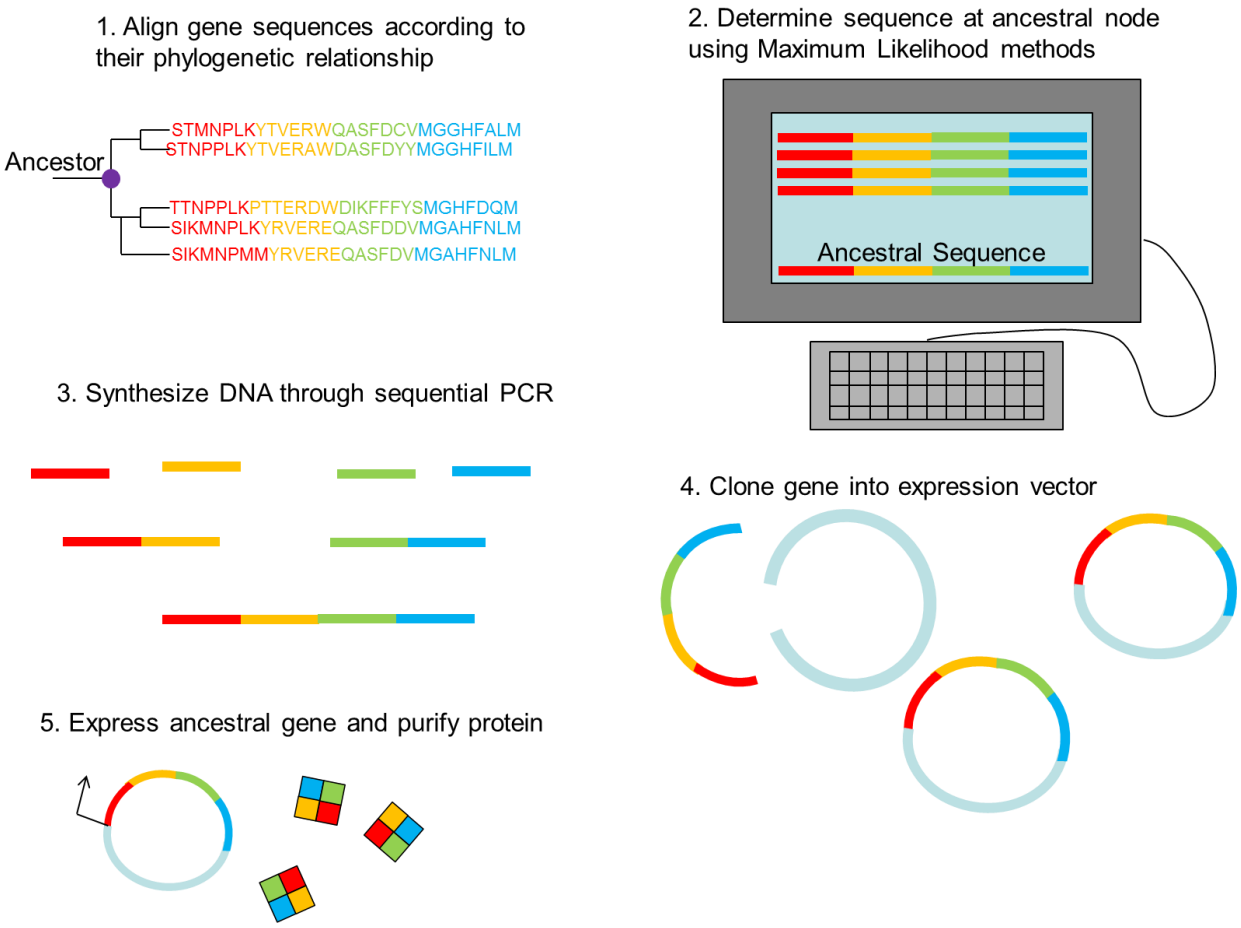


Figure 1.7: Ancestral Gene Resurrection Workflow Schematic

Ancestral genes can be resurrected through the process of aligning sequences based on their phylogenetic relationship, determining the sequence of the ancestral node through Maximum Likelihood analysis, synthesizing the gene through sequential PCR, cloning the gene into an expression vector, expressing the gene, and purifying the protein.

A. Percent identity of SR LBDs:

	ERα	**PR**	**GR**	**AR**	**MR**
ERα	100	22.8	25.7	19.6	23.2
PR	22.8	100	53.8	54.0	55.1
GR	25.7	53.8	100	49.0	56.1
AR	19.6	54.0	49.0	100	50.8
MR	23.2	55.1	56.1	50.8	100

B. Percent identity of SR DBDs:

	ERα	**PR**	**GR**	**AR**	**MR**
ERα	100	56.1	57.6	59.1	57.6
PR	56.1	100	90.9	79.5	90.9
GR	57.6	90.9	100	78.8	93.9
AR	59.1	79.5	78.8	100	78.8
MR	57.6	90.9	93.9	78.8	100

Table 1.1: Percent Identities of human SR domains

Shown are the percent amino acid identities across the human SR LBDs and DBDs. Table A adapted from Baker, 1997 [111].

References

1. Nagy, L. and J.W. Schwabe, *Mechanism of the nuclear receptor molecular switch.* Trends Biochem Sci, 2004. **29**(6): p. 317-24.
2. Bridgham, J.T., et al., *Protein evolution by molecular tinkering: diversification of the nuclear receptor superfamily from a ligand-dependent ancestor.* PLoS Biol, 2010. **8**(10).
3. Mangelsdorf, D.J. and R.M. Evans, *The RXR heterodimers and orphan receptors.* Cell, 1995. **83**(6): p. 841-50.
4. Kushiro, T., E. Nambara, and P. McCourt, *Hormone evolution: The key to signalling.* Nature, 2003. **422**(6928): p. 122.
5. Reitzel, A.M. and A.M. Tarrant, *Nuclear receptor complement of the cnidarian Nematostella vectensis: phylogenetic relationships and developmental expression patterns.* BMC Evol Biol, 2009. **9**: p. 230.
6. Thornton, J.W., E. Need, and D. Crews, *Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling.* Science, 2003. **301**(5640): p. 1714-7.
7. Schwabe, J.W. and S.A. Teichmann, *Nuclear receptors: the evolution of diversity.* Sci STKE, 2004. **2004**(217): p. pe4.
8. Thornton, J.W., *Evolution of vertebrate steroid receptors from an ancestral estrogen receptor by ligand exploitation and serial genome expansions.* Proc Natl Acad Sci U S A, 2001. **98**(10): p. 5671-6.
9. Carreau, S., S. Wolczynski, and I. Galeraud-Denis, *Aromatase, oestrogens and human male reproduction.* Philos Trans R Soc Lond B Biol Sci, 2010. **365**(1546): p. 1571-9.
10. Ogawa, S., et al., *Roles of estrogen receptor-alpha gene expression in reproduction-related behaviors in female mice.* Endocrinology, 1998. **139**(12): p. 5070-81.
11. Deroo, B.J. and K.S. Korach, *Estrogen receptors and human disease.* J Clin Invest, 2006. **116**(3): p. 561-70.
12. Garcia-Segura, L.M., I. Azcoitia, and L.L. DonCarlos, *Neuroprotection by estradiol.* Prog Neurobiol, 2001. **63**(1): p. 29-60.
13. Kajta, M. and C. Beyer, *Cellular strategies of estrogen-mediated neuroprotection during brain development.* Endocrine, 2003. **21**(1): p. 3-9.
14. Saunders-Pullman, R., *Estrogens and Parkinson disease: neuroprotective, symptomatic, neither, or both?* Endocrine, 2003. **21**(1): p. 81-7.
15. Currie, L.J., et al., *Postmenopausal estrogen use affects risk for Parkinson disease.* Arch Neurol, 2004. **61**(6): p. 886-8.

16. Shao, W. and M. Brown, *Advances in estrogen receptor biology: prospects for improvements in targeted breast cancer therapy.* Breast Cancer Res, 2004. **6**(1): p. 39-52.
17. Yue, W., et al., *Tamoxifen versus aromatase inhibitors for breast cancer prevention.* Clin Cancer Res, 2005. **11**(2 Pt 2): p. 925s-30s.
18. Syed, V., et al., *Expression of gonadotropin receptor and growth responses to key reproductive hormones in normal and malignant human ovarian surface epithelial cells.* Cancer Res, 2001. **61**(18): p. 6768-76.
19. Foley, E.F., et al., *Selective loss of estrogen receptor beta in malignant human colon.* Cancer Res, 2000. **60**(2): p. 245-8.
20. Sinisi, A.A., et al., *Sexual differentiation.* J Endocrinol Invest, 2003. **26**(3 Suppl): p. 23-8.
21. Walters, K.A., U. Simanainen, and D.J. Handelsman, *Molecular insights into androgen actions in male and female reproductive function from androgen receptor knockout models.* Hum Reprod Update, 2010. **16**(5): p. 543-58.
22. Oakes, M.B., et al., *Complete androgen insensitivity syndrome--a review.* J Pediatr Adolesc Gynecol, 2008. **21**(6): p. 305-10.
23. Hughes, I.A. and A. Deeb, *Androgen resistance.* Best Pract Res Clin Endocrinol Metab, 2006. **20**(4): p. 577-98.
24. Galani, A., et al., *Androgen insensitivity syndrome: clinical features and molecular defects.* Hormones (Athens), 2008. **7**(3): p. 217-29.
25. Heinlein, C.A. and C. Chang, *Androgen receptor in prostate cancer.* Endocr Rev, 2004. **25**(2): p. 276-308.
26. Risch, H.A., *Hormonal etiology of epithelial ovarian cancer, with a hypothesis concerning the role of androgens and progesterone.* J Natl Cancer Inst, 1998. **90**(23): p. 1774-86.
27. Judd, H.L., et al., *Endocrine function of the postmenopausal ovary: concentration of androgens and estrogens in ovarian and peripheral vein blood.* J Clin Endocrinol Metab, 1974. **39**(6): p. 1020-4.
28. Slotman, B.J. and B.R. Rao, *Response to inhibition of androgen action of human ovarian cancer cells in vitro.* Cancer Lett, 1989. **45**(3): p. 213-20.
29. Pang, T.P., et al., *A physiological role for androgen actions in the absence of androgen receptor DNA binding activity.* Mol Cell Endocrinol, 2012. **348**(1): p. 189-97.
30. Horwitz, K.B., et al., *Progestin action and progesterone receptor structure in human breast cancer: a review.* Recent Prog Horm Res, 1985. **41**: p. 249-316.

31. Modugno, F., *Ovarian cancer and polymorphisms in the androgen and progesterone receptor genes: a HuGE review.* Am J Epidemiol, 2004. **159**(4): p. 319-35.
32. Ishibashi, H., et al., *Progesterone receptor in non-small cell lung cancer--a potent prognostic factor and possible target for endocrine therapy.* Cancer Res, 2005. **65**(14): p. 6450-8.
33. Newcomb, P.A., et al., *Estrogen plus progestin use, microsatellite instability, and the risk of colorectal cancer in women.* Cancer Res, 2007. **67**(15): p. 7534-9.
34. Jongen, V., et al., *Expression of estrogen receptor-alpha and -beta and progesterone receptor-A and -B in a large cohort of patients with endometrioid endometrial cancer.* Gynecol Oncol, 2009. **112**(3): p. 537-42.
35. Pichon, M.F., et al., *Relationship of presence of progesterone receptors to prognosis in early breast cancer.* Cancer Res, 1980. **40**(9): p. 3357-60.
36. Conneely, O.M., et al., *Reproductive functions of the progesterone receptor isoforms: lessons from knock-out mice.* Mol Cell Endocrinol, 2001. **179**(1-2): p. 97-103.
37. Conneely, O.M., B. Mulac-Jericevic, and J.P. Lydon, *Progesterone-dependent regulation of female reproductive activity by two distinct progesterone receptor isoforms.* Steroids, 2003. **68**(10-13): p. 771-8.
38. Brinton, R.D., et al., *Progesterone receptors: form and function in brain.* Front Neuroendocrinol, 2008. **29**(2): p. 313-39.
39. Maletic, V., et al., *Neurobiology of depression: an integrated view of key findings.* Int J Clin Pract, 2007. **61**(12): p. 2030-40.
40. Kadmiel, M. and J.A. Cidlowski, *Glucocorticoid receptor signaling in health and disease.* Trends Pharmacol Sci, 2013. **34**(9): p. 518-30.
41. Silverman, M.N., et al., *Glucocorticoid receptor dimerization is required for proper recovery of LPS-induced inflammation, sickness behavior and metabolism in mice.* Mol Psychiatry, 2013. **18**(9): p. 1006-17.
42. Smoak, K.A. and J.A. Cidlowski, *Mechanisms of glucocorticoid receptor signaling during inflammation.* Mech Ageing Dev, 2004. **125**(10-11): p. 697-706.
43. Boumpas, D.T., et al., *Glucocorticoid therapy for immune-mediated diseases: basic and clinical correlates.* Ann Intern Med, 1993. **119**(12): p. 1198-208.
44. Yehuda, R., et al., *Glucocorticoid receptor number and cortisol excretion in mood, anxiety, and psychotic disorders.* Biol Psychiatry, 1993. **34**(1-2): p. 18-25.
45. Kino, T., E. Charmandari, and G.P. Chrousos, *Glucocorticoid receptor: implications for rheumatic diseases.* Clin Exp Rheumatol, 2011. **29**(5 Suppl 68): p. S32-41.

46. Krozowski, Z.S. and J.W. Funder, *Renal mineralocorticoid receptors and hippocampal corticosterone-binding species have identical intrinsic steroid specificity.* Proc Natl Acad Sci U S A, 1983. **80**(19): p. 6056-60.
47. Myles, K. and J.W. Funder, *Progesterone binding to mineralocorticoid receptors: in vitro and in vivo studies.* Am J Physiol, 1996. **270**(4 Pt 1): p. E601-7.
48. Rossier, B.C., et al., *Epithelial sodium channel and the control of sodium balance: interaction between genetic and environmental factors.* Annu Rev Physiol, 2002. **64**: p. 877-97.
49. Horisberger, J.D., et al., *Structure-function relationship of Na,K-ATPase.* Annu Rev Physiol, 1991. **53**: p. 565-84.
50. Bonvalet, J.P., *Regulation of sodium transport by steroid hormones.* Kidney Int Suppl, 1998. **65**: p. S49-56.
51. Maggiolini, M. and D. Picard, *The unfolding stories of GPR30, a new membrane-bound estrogen receptor.* J Endocrinol, 2010. **204**(2): p. 105-14.
52. Kalaitzidis, D. and T.D. Gilmore, *Transcription factor cross-talk: the estrogen receptor and NF-kappaB.* Trends Endocrinol Metab, 2005. **16**(2): p. 46-52.
53. Luisi, B.F., et al., *Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA.* Nature, 1991. **352**(6335): p. 497-505.
54. Tsai, S.Y., et al., *Molecular interactions of steroid hormone receptor with its enhancer element: evidence for receptor dimer formation.* Cell, 1988. **55**(2): p. 361-9.
55. Gronemeyer, H. and D. Moras, *Nuclear receptors. How to finger DNA.* Nature, 1995. **375**(6528): p. 190-1.
56. Schwabe, J.W. and D. Rhodes, *Beyond zinc fingers: steroid hormone receptors have a novel structural motif for DNA recognition.* Trends Biochem Sci, 1991. **16**(8): p. 291-6.
57. Black, B.E., et al., *DNA binding domains in diverse nuclear receptors function as nuclear export signals.* Curr Biol, 2001. **11**(22): p. 1749-58.
58. Nolte, R.T., et al., *Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor-gamma.* Nature, 1998. **395**(6698): p. 137-43.
59. Johnson, B.A., et al., *Ligand-induced stabilization of PPARgamma monitored by NMR spectroscopy: implications for nuclear receptor activation.* J Mol Biol, 2000. **298**(2): p. 187-94.
60. Moras, D. and H. Gronemeyer, *The nuclear receptor ligand-binding domain: structure and function.* Curr Opin Cell Biol, 1998. **10**(3): p. 384-91.

61. Watkins, R.E., et al., *The human nuclear xenobiotic receptor PXR: structural determinants of directed promiscuity.* Science, 2001. **292**(5525): p. 2329-33.
62. O'Malley, B.W. and M.J. Tsai, *Molecular pathways of steroid receptor action.* Biol Reprod, 1992. **46**(2): p. 163-7.
63. Pratt, W.B. and D.O. Toft, *Steroid receptor interactions with heat shock protein and immunophilin chaperones.* Endocr Rev, 1997. **18**(3): p. 306-60.
64. Heery, D.M., et al., *A signature motif in transcriptional co-activators mediates binding to nuclear receptors.* Nature, 1997. **387**(6634): p. 733-6.
65. Darimont, B.D., et al., *Structure and specificity of nuclear receptor-coactivator interactions.* Genes Dev, 1998. **12**(21): p. 3343-56.
66. Hu, X. and M.A. Lazar, *The CoRNR motif controls the recruitment of corepressors by nuclear hormone receptors.* Nature, 1999. **402**(6757): p. 93-6.
67. Nagy, L., et al., *Mechanism of corepressor binding and release from nuclear hormone receptors.* Genes Dev, 1999. **13**(24): p. 3209-16.
68. Perissi, V., et al., *Molecular determinants of nuclear receptor-corepressor interaction.* Genes Dev, 1999. **13**(24): p. 3198-208.
69. Webb, P., et al., *The nuclear receptor corepressor (N-CoR) contains three isoleucine motifs (I/LXXII) that serve as receptor interaction domains (IDs).* Mol Endocrinol, 2000. **14**(12): p. 1976-85.
70. Marimuthu, A., et al., *TR surfaces and conformations required to bind nuclear receptor corepressor.* Mol Endocrinol, 2002. **16**(2): p. 271-86.
71. Benko, S., et al., *Molecular determinants of the balance between co-repressor and co-activator recruitment to the retinoic acid receptor.* J Biol Chem, 2003. **278**(44): p. 43797-806.
72. Shiau, A.K., et al., *The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen.* Cell, 1998. **95**(7): p. 927-37.
73. Kallenberger, B.C., et al., *A dynamic mechanism of nuclear receptor activation and its perturbation in a human disease.* Nat Struct Biol, 2003. **10**(2): p. 136-40.
74. Tamrazi, A., K.E. Carlson, and J.A. Katzenellenbogen, *Molecular sensors of estrogen receptor conformations and dynamics.* Mol Endocrinol, 2003. **17**(12): p. 2593-602.
75. Xu, H.E., et al., *Structural basis for antagonist-mediated recruitment of nuclear co-repressors by PPARalpha.* Nature, 2002. **415**(6873): p. 813-7.
76. Germain, P., et al., *Co-regulator recruitment and the mechanism of retinoic acid receptor synergy.* Nature, 2002. **415**(6868): p. 187-92.

77. Watkins, R.E., et al., *Coactivator binding promotes the specific interaction between ligand and the pregnane X receptor.* J Mol Biol, 2003. **331**(4): p. 815-28.
78. Pavlik, E.J. and B.S. Katzenellenbogen, *The stability of the uterine estrogen receptor when complexed with estrogens or antiestrogens.* Mol Pharmacol, 1980. **18**(3): p. 406-12.
79. Gee, A.C. and J.A. Katzenellenbogen, *Probing conformational changes in the estrogen receptor: evidence for a partially unfolded intermediate facilitating ligand binding and release.* Mol Endocrinol, 2001. **15**(3): p. 421-8.
80. Leng, X., et al., *Ligand-dependent conformational changes in thyroid hormone and retinoic acid receptors are potentially enhanced by heterodimerization with retinoic X receptor.* J Steroid Biochem Mol Biol, 1993. **46**(6): p. 643-61.
81. Keidel, S., P. LeMotte, and C. Apfel, *Different agonist- and antagonist-induced conformational changes in retinoic acid receptors analyzed by protease mapping.* Mol Cell Biol, 1994. **14**(1): p. 287-98.
82. Todd, A.E., C.A. Orengo, and J.M. Thornton, *Evolution of function in protein superfamilies, from a structural perspective.* J Mol Biol, 2001. **307**(4): p. 1113-43.
83. Walker, J.M., *The proteomics protocols handbook*. 2005, Totowa, N.J.: Humana Press. xviii, 988 p.
84. Zhao, X.Y., et al., *Two mutations identified in the androgen receptor of the new human prostate cancer cell line MDA PCa 2a.* J Urol, 1999. **162**(6): p. 2192-9.
85. Grishin, N.V., *Fold change in evolution of protein structures.* J Struct Biol, 2001. **134**(2-3): p. 167-85.
86. Moore, S.A., et al., *Crystal structure of a flavoprotein related to the subunits of bacterial luciferase.* EMBO J, 1993. **12**(5): p. 1767-74.
87. Chothia, C. and A.M. Lesk, *The relation between the divergence of sequence and structure in proteins.* EMBO J, 1986. **5**(4): p. 823-6.
88. Tokuriki, N. and D.S. Tawfik, *Protein dynamism and evolvability.* Science, 2009. **324**(5924): p. 203-7.
89. Ingles-Prieto, A., et al., *Conservation of protein structure over four billion years.* Structure, 2013. **21**(9): p. 1690-7.
90. Orengo, C.A., et al., *Review: what can structural classifications reveal about protein evolution?* J Struct Biol, 2001. **134**(2-3): p. 145-65.
91. Zhang, J., *Evolution by gene duplication: An Update.* Trends in Ecology and Evolution, 2003. **18**(292).

92. Walsh, J.B., *How often do duplicated genes evolve new functions?* Genetics, 1995. **139**(1): p. 421-8.
93. Lynch, M., et al., *The probability of preservation of a newly arisen gene duplicate.* Genetics, 2001. **159**(4): p. 1789-804.
94. Li, W.-H., *Molecular evolution*. 1997, Sunderland, Mass.: Sinauer Associates. xv, 487 p.
95. M.A. Nowack, M.C.B., J. Cooke, J.M. Smith, *Evolution of genetic redundancy.* Nature, 1997. **388**(167).
96. Bridges, C.B., *The Bar "Gene" a Duplication.* Science, 1936. **83**(2148): p. 210-1.
97. Apic, G., J. Gough, and S.A. Teichmann, *Domain combinations in archaeal, eubacterial and eukaryotic proteomes.* J Mol Biol, 2001. **310**(2): p. 311-25.
98. Chothia, C., et al., *Evolution of the protein repertoire.* Science, 2003. **300**(5626): p. 1701-3.
99. L. Pauling, E.Z., *Chemical Paleogenetics Molecular Restoration Studies of Extinct Forms of Life.* Acta Chem Scand, 1963. **17**(9).
100. Zuckerkandl, E., *Neutral and Nonneutral Mutations: The Creative Mix-Evolution of Complexity in Gene Interaction Systems.* J Mol Evol, 1997. **44**(4): p. 470.
101. Ivics, Z., et al., *Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells.* Cell, 1997. **91**(4): p. 501-10.
102. Adey, N.B., et al., *Molecular resurrection of an extinct ancestral promoter for mouse L1.* Proc Natl Acad Sci U S A, 1994. **91**(4): p. 1569-73.
103. Thornton, J.W., *Resurrecting ancient genes: experimental analysis of extinct molecules.* Nat Rev Genet, 2004. **5**(5): p. 366-75.
104. Fitch, W.M., *Evolution of clupeine Z, a probable crossover product.* Nat New Biol, 1971. **229**(8): p. 245-7.
105. Zhang, J. and M. Nei, *Accuracies of ancestral amino acid sequences inferred by the parsimony, likelihood, and distance methods.* J Mol Evol, 1997. **44 Suppl 1**: p. S139-46.
106. Hillis, D.M., et al., *Experimental phylogenetics: generation of a known phylogeny.* Science, 1992. **255**(5044): p. 589-92.
107. Stackhouse, J., et al., *The ribonuclease from an extinct bovid ruminant.* FEBS Lett, 1990. **262**(1): p. 104-6.
108. Jermann, T.M., et al., *Reconstructing the evolutionary history of the artiodactyl ribonuclease superfamily.* Nature, 1995. **374**(6517): p. 57-9.
109. Wouters, M.A., et al., *A despecialization step underlying evolution of a family of serine proteases.* Mol Cell, 2003. **12**(2): p. 343-54.

110. Felsenstein, J., *Evolutionary trees from DNA sequences: a maximum likelihood approach.* J Mol Evol, 1981. **17**(6): p. 368-76.
111. Baker, M.E., *Steroid receptor phylogeny and vertebrate origins.* Mol Cell Endocrinol, 1997. **135**(2): p. 101-7.

CHAPTER 2: EVOLUTION OF MINIMAL SPECIFICITY AND PROMISCUITY IN STEROID HORMONE RECEPTORS

Geeta N. Eick[1,2], Jennifer K. Colucci[3], Michael J. Harms[1], Eric A. Ortlund[3], Joseph W. Thornton[1,2,4a]

1) Institute of Ecology and Evolution, University of Oregon, Eugene, OR 97403, USA 2) Howard Hughes Medical Institute, Eugene OR, 97403, USA

3) Biochemistry Department, Emory University School of Medicine, Atlanta GA 30322 USA

4) Department of Human Genetics and Department of Ecology & Evolution, University of Chicago, Chicago IL, 60637 USA

Eick GN, Colucci JK, Harms MJ, Ortlund EA, Thornton JW (2012) Evolution of minimal specificity and promiscuity in steroid hormone receptors. PLoS Genet 8(11).

In order to understand the evolution of ligand specificity from aromatized steroids to nonaromatized 3-keto steroids in steroid hormone nuclear receptors, it is crucial to visualize the ligand binding pockets of these receptors. In this work, I crystallized the ancestral steroid receptor 2 (ancSR2) in complex with both progesterone and 11-deoxycorticosterone (11-DOC). These structures highlight the properties of the ligand binding pocket as well as the molecular

[a] Conceived and designed the experiments: JWT GNE JKC MJH EAO. Performed the experiments: GNE JKC MJH. Analyzed the data: GNE JKC MJH JWT EAO. Wrote the paper: JWT GNE MJH JKC EAO.

interactions between receptor and ligand that are key for binding and activation. This work was previously published in PLoS Genetics.

Abstract

Most proteins are regulated by physical interactions with other molecules; some are highly specific, but others interact with many partners. Despite much speculation, we know little about how and why specificity/promiscuity evolves in natural proteins. It is widely assumed that specific proteins evolved from more promiscuous ancient forms and that most proteins' specificity has been tuned to an optimal state by selection. Here we use ancestral protein reconstruction to trace the evolutionary history of ligand recognition in the steroid hormone receptors (SRs), a family of hormone-regulated animal transcription factors. We resurrected the deepest ancestral proteins in the SR family and characterized the structure-activity relationships by which they distinguished among ligands. We found that that the most ancient split in SR evolution involved a discrete switch from an ancient receptor for aromatized estrogens -- including xenobiotics -- to a derived receptor that recognized non-aromatized progestagens and corticosteroids. The family's history, viewed in relation to the evolution of their ligands, suggests that SRs evolved according to a principle of minimal specificity: at each point in time, receptors evolved ligand recognition criteria that were just specific enough to parse the set of endogenous substances to which they were exposed. By studying the atomic structures of resurrected SR proteins, we found that their promiscuity evolved because the ancestral binding cavity was larger than the primary ligand and contained excess hydrogen bonding capacity, allowing adventitious recognition of larger molecules with additional functional groups. Our findings provide an historical explanation for the sensitivity of modern SRs to natural and synthetic ligands – including endocrine-disrupting drugs and pollutants -- and show that knowledge of history can contribute to ligand prediction. They suggest that SR promiscuity may reflect the limited power of selection within real biological systems to discriminate between perfect and "good enough."

Author Summary

The functions of most proteins are defined by their interactions with other biological substances, such as DNA, nutrients, hormones, or other proteins. Some proteins are highly specific, but others are more promiscuous and can interact with a variety of natural substances, as well as drugs and pollutants. Understanding molecular interactions is a key goal in pharmacology and toxicology, but there are few general principles to help explain or predict protein specificity. Because every biological entity is the result of evolution, understanding a protein's history might help explain why it interacts with the substances to which it is sensitive. In this paper, we used ancestral protein reconstruction to experimentally trace how specificity evolved in an ancient group of proteins, the steroid hormone receptors (SRs), a family of proteins that regulate reproduction and other biological processes in animals. We show that SRs evolved according to a principle of minimal specificity: at each point in time, these proteins evolved to be specific enough to distinguish among the substances to which they were naturally exposed, but not more so. Our findings provide an historical explanation for modern SRs' diverse sensitivities to natural and man-made substances; they show that knowledge of history can contribute to predicting the ligands to which a modern protein will respond and indicate that promiscuity reflects the limited power of natural selection to discriminate between perfect and "good enough."

Introduction

Cells, like biological entities at higher levels (*2*), can be viewed as information processing systems, because they change their state or activity in response to specific internal or external cues. This behavior is mediated by functional interactions among the proteins and other molecules that comprise the system (*3*). Some proteins are highly specific (*4, 5*), but others can be regulated by a broader array of molecular partners, including various endogenous ligands, drugs, and pollutants (*1*).

There has been much speculation about the evolutionary causes of specificity and promiscuity. It is widely believed that evolution usually proceeds from generalist ancestral proteins to more specific recent forms (*1, 6-9*). Both narrow and broad specificity are often assumed to be the result of optimization by natural selection; according to this view, the capacity of ancient molecules to interact with many partners allowed species with small protein repertoires to carry out a broad set of biological activities and promoted the future evolvability of new functions, while specialization in more recent proteins provides greater efficiency, finer regulation, or prevention of deleterious interactions (refs. (*6, 7, 10-13*), but see ref. (*9*)).

These hypotheses are largely untested, because there are few natural protein families for which the historical trajectory of changes in specificity has been carefully dissected, although the proximate mechanisms for promiscuous responses have been studied in some extant and engineered proteins (*8, 9, 14*). Further, although promiscuous interactions of proteins with exogenous substances are core issues in pharmacology and toxicology, the lack of strong historical case studies means that there are no general principles that explain why molecules have evolved their present-day ligand-recognition criteria. Without such principles, predicting the ligands to which proteins will be sensitive has proven difficult (*1, 15*).

Steroid hormone receptors (SRs) are an excellent model for the evolution of specificity. SRs are hormone-activated nuclear transcription factors with distinct specificities for endogenous steroid hormones and exogenous substances. In all SRs, the activating hormone binds in an internal cavity within a well-conserved ligand binding domain (LBD), causing the LBD to change conformation, attract coactivator proteins, and increase transcription of target genes (*16*). The SR family diversified through a series of gene duplications that took place during early chordate and vertebrate evolution (*17*). Humans have two phylogenetic classes of SRs, which correspond to the chemical classes of endogenous ligands that activate each receptor's LBD. In the first class – the estrogen receptors (ERs) -- the endogenous ligands are 18-carbon steroids with an aromatized A-ring and a hydroxyl attached to carbon 3 on the steroid skeleton (Figure 2.1A). The other class – the nonaromatized steroid receptors (naSRs) – includes receptors for androgens (AR), progestagens (PR), glucocorticoids (GR), and mineralocorticoids (MR); these ligands all contain a nonaromatized A-ring, an additional methyl at carbon 19, and, in most cases, a ketone at carbon 3. Each paralog within the naSR class has distinct specificity based on the size and polarity of the functional groups at carbon 17 and carbon 21 on the steroid's D-ring. Although functional groups at other positions may affect sensitivity, they do not distinguish the classes of ligands recognized by paralogous receptors. SRs also differ in their promiscuous sensitivity to exogenous substances: ERs can be activated by a large set of phenolic drugs and pollutants in diverse chemical classes with highly variant structures, whereas naSRs have far fewer synthetic agonists (*18, 19*).

Here we characterize in detail the evolutionary trajectory of changes in ligand specificity/promiscuity in the SR protein family, as well as the underlying structural mechanisms for promiscuous responses to non-target ligands. For this purpose, we use ancestral protein resurrection (APR), which uses computational phylogenetic techniques to infer ancestral protein sequences from an alignment of their present-day descendants, followed by gene synthesis, molecular functional assays, and experimental studies of protein structure to directly characterize

them. APR represents a powerful strategy for experimentally testing hypotheses about the structure and function of ancient proteins (*20, 21*). By dissecting the structure-activity criteria by which ancient receptors distinguished among ligands – and tracing how those criteria changed over time – we sought to gain insight into the evolution of specificity versus promiscuity in the SR family. We also sought to determine whether an understanding of a protein family's history can reveal explanatory principles for understanding and predicting the ligands to which its members will respond.

Results and Discussion

Reconstruction and characterization of ancestral proteins.

To understand how and why the differences in ligand specificity between the ERs and naSRs receptors evolved, we used ancestral protein resurrection (*20*) to experimentally characterize the LBDs of two key ancient members of the protein family. AncSR1 is the last common ancestral protein from which the entire SR family descends by a series of gene duplications; ancSR2 is the ancestral protein of all naSRs (Figure 2.1B). The family's phylogeny indicates that both proteins are hundreds of millions of years old: AncSR1 predates the divergence of vertebrates from other chordates, and ancSR2 predates the divergence of jawed vertebrates from jawless fishes (*17*).

From alignments of ~200 extant receptor proteins, we used likelihood-based phylogenetic methods to infer the best-fitting evolutionary model, phylogeny, and ancestral protein sequences. The sequence of ancSR2 was reconstructed with high confidence (mean posterior probability (PP) = 0.93 per site, Figure 2.5, Table 2.1), and even less ambiguity at ligand-contacting sites (mean PP = 0.96). AncSR1 was more ambiguously reconstructed (mean PP=0.70 overall, Figure 2.6, Table 2.2), but at ligand-contacting sites its reconstruction was considerably more robust (mean PP=0.90).

The AncSR1 sequence is most similar to those of the extant ERs, whereas that of ancSR2 is most similar to the naSRs, and this pattern is most pronounced at sites in the ligand-contacting pockets (Figure 2.1C, Table 2.3). These findings suggest that ancSR1 may have been activated by estrogens and ancSR2 by nonaromatized steroids, a scenario also supported by the phylogenetic distribution of ligand specificities among extant receptors – particularly the presence of estrogen-sensitive receptors in invertebrates such as annelids and cephalochordates (*17, 22*).

To experimentally test these hypotheses, we synthesized cDNAs for the ancSR1 and ancSR2 LBD protein sequences, expressed them as Gal4-DBD fusion constructs, and characterized their sensitivity to hormones using luciferase reporter gene assays. As predicted, we found that ancSR1 is a highly specific estrogen receptor, activating transcription in the presence of nanomolar concentrations of physiological estrogens. It was unresponsive to a broad array of androgens, progestagens, and corticosteroids, as well as cholesterol (Figure 2.2A, Figure 2.7). In contrast, ancSR2 was completely unresponsive to estrogens (and cholesterol) but strongly activated by low concentrations of diverse nonaromatized steroid hormones, including progestagens and corticosteroids and – to a lesser extent – androgens (Figure 2.2A, Figure 2.8). We also experimentally characterized numerous alternative reconstructions of ancSR1 and ancSR2 and found that these proteins' specificities for aromatized and nonaromatized steroids, respectively, are highly robust to uncertainty in the reconstruction (Figures 2.9, 2.10).

We conclude that a fundamental inversion of ligand specificity for endogenous steroid hormones – not a narrowing of specificity from a promiscuous ancestor – took place during the evolutionary interval between ancSR1 and ancSR2. This inversion must have occurred in the lineage leading to vertebrates after they diverged from cephalochordates, because cephalochordates possess a single naSR ortholog, which retains the ancestral specificity for estrogens (Figure 2.1B, see (*23*)). Subsequently, the promiscuous responses of ancSR2 to nonaromatized steroids were differentially partitioned among its descendant lineages to yield the more specific PR, GR, MR, and AR. In extant receptors, mutations that make these SRs sensitive to the ligands of other members of the family now cause deleterious phenotypes (*24-26*).

Our findings, viewed in the context of the ancient pathway for steroid hormone synthesis, suggest that some hormone-receptor pairs were assembled during evolution by a process of molecular exploitation, whereby molecules with a different ancient function are recruited into new signaling partnerships after gene duplication and/or divergence (*22, 27*). That the ancient ancSR1 was specific for estrogens implies that progestagens and androgens, which are

intermediates in the synthesis of estrogens (Figure 2.1A), existed before steroid receptors evolved to transduce their signals. When ancSR2 and its descendants evolved the capacity to be activated by nonaromatized steroids, these biochemical steppingstones in estrogen synthesis were recruited into new, bona fide signaling partnerships.

Ancestral structure-activity criteria.

The specificity of a protein can be described by the biochemical criteria by which it distinguishes between functionally relevant binding partners and all other substances. To dissect more precisely how the ligand-recognition criteria of SRs evolved during the interval between ancSR1 and ancSR2, we applied a structure-activity approach. We characterized the specificity of these two ancestral proteins using a library of synthetic and natural steroids that differ from each other only by the aromatization of the A-ring or the functional groups at specific positions that vary among physiological steroids **(**Figure 2.2, Table 2.4**)**.

We found that ancSR1's specificity is determined primarily by a single major criterion: requirement for an aromatized A-ring. All aromatized steroids tested activated ancSR1, but no natural nonaromatized steroids were effective at nanomolar concentrations (Figure 2.2A, Figure 2.7). Comparisons using several matched pairs of aromatized/nonaromatized steroids confirm that ancSR1 distinguishes strongly among potential ligands based on its requirement for an aromatized A-ring, with EC50s that increase by orders of magnitude when only this aspect of the ligand is changed (Figure 2.2B, Table 2.5). Beyond this major criterion, ancSR1's specificity is rather loose. In particular, it tolerates different functional groups around the D-ring, as shown by its similar sensitivity to estradiol and estrone, which contain a 17-hydroxyl and ketone, respectively (Figures 2.2A, 2.2B). Even the "chimeric" steroid 19-nor-1, 3, 5(10)-pregnatriene-3-ol-20-one (NPT) – which has the much larger 17□-acetyl group found on progesterone and corticosteroids – is almost as potent an ancSR1 activator as endogenous estrogens (Figure 2.2B).

ancSR2's ligand-recognition criteria differ from ancSR1's in two major ways (Figure 2.2, Table 2.5). First, ancSR1's A-ring rule is inverted in ancSR2, which is more sensitive to nonaromatized steroids than to otherwise identical aromatized substances by two to three orders of magnitude (Figure 2.2C). Second, ancSR2 evolved an additional criterion: it prefers steroids with a 17□-acetyl group (such as progestagens and corticosteroids) to those with smaller hydroxyls or ketones (androgens and estrogens), as demonstrated by the 21- to 87-fold difference in EC50 values between pairs of hormones that differ only at this position (Figure 2.2D).

Beyond these two criteria, ancSR2's specificity is rather loose (Figures 2.2E-I). AncSR2 does not distinguish strongly between progestagens and corticosteroids because it has only a weak preference for steroids with a 21-hydroxyl (Figure 2.2F). The presence/absence of an 11-hydroxyl, present on many corticosteroids, does not strongly affect the receptor's sensitivity (Figure 2.2G). AncSR2 does not distinguish between 3-hydroxy and 3-ketosteroids, so long as the A-ring is not aromatized (Figure 2.2E), and it does not require the 19-methyl present on endogenous nonaromatized steroids (Figure 2.2H). Taken together, these data indicate that the evolution of ancSR2's ligand specificity entailed two major changes: inversion of ancSR1's fundamental ligand-recognition criterion for an aromatized A-ring and acquisition of an additional criterion at the D-ring.

Minimal specificity in SR evolution.

The evolving ligand recognition rules of ancSR1 and ancSR2 can be understood in light of existing knowledge concerning the biosynthesis and evolution of the ligands. Taken together, our findings suggest that the evolution of the SR family has been characterized by minimal specificity, a concept borrowed by analogy from information theory [30]: each receptor evolved to be specific enough to distinguish among the set of contemporaneous endogenous ligands to which it was exposed, but not more so.

The concept of minimal specificity provides an evolutionary explanation for the specificity and promiscuity possessed by each receptor. For example, ancSR1's single criterion – requiring an aromatized A-ring – provided minimally sufficient specificity for estrogens (Figure 2.3A). Estrogens are the only aromatized steroids produced in animals, because aromatization of the steroid A-ring is the final step in a conserved estrogen synthesis pathway beginning with cholesterol and proceeding via progestagens and androgens as intermediates (Figure 2.1A). ancSR1's simple criterion therefore allowed it to exclude all other endogenous steroids, including androgens, progestagens, and cholesterol and its metabolites. These hormones are all ancient: synthesis of estrogens via progestagens and androgens is as old as the ancestor of cephalochordates and vertebrates (*28*), and it may be even older, given the presence of all these hormones in mollusks (*29*).

Minimal specificity is also apparent in the evolution of ancSR2 and its descendants (Figure 2.3A, Figure 2.11). When ancSR2 became sensitive to nonaromatized steroids, it would have excluded estrogens but become sensitive to both progestagens and androgens; acquiring its second ligand-recognition rule restricted ancSR2's sensitivity to progestagens only. AncSR2 did not yet distinguish progestagens from corticosteroids, but endogenous synthesis of these steroids had not yet evolved; only later – during or after the same period of early vertebrate evolution when synthesis of corticosteroids first evolved due to the emergence of 21-hydroxylase activities in the CYP450 family (*28, 30*) – were ancSR2's promiscuous sensitivities partitioned among the PR, GR, and MR.

These data indicate that each receptor evolved ligand recognition criteria sufficiently complex to parse the repertoire of ligands present during its evolution, but those rules were not sufficient to prevent promiscuous responses to other substances that had not yet evolved. By evolving narrower specificity as the synthesis of new steroids emerged during vertebrate evolution, the various SRs presumably maintained the capacity to transduce specific signals despite the organism's increasing chemical repertoire (Figure 2.3A).

An evolutionary explanation for SR-mediated endocrine disruption.

Predicting ligands that interact with intended and secondary protein targets is an important goal in pharmacology and toxicology, but understanding from first principles which targets will respond more or less promiscuously has proven difficult (*1*, *15*). The concept of minimal specificity predicts that ER's capacity to be disrupted by exogenous phenolics is inherited from ancSR1. To test this possibility, we characterized the ability of several xenoestrogens to activate ancSR1. As predicted, we found that ancSR1 is activated by the strong nonsteroidal ER agonists diethylstilbestrol and genistein and is competitively inhibited by the ER antagonists 4-hydroxytamoxifen and ICI182780 (Figure 2.3B).

Our observations provide an historical explanation for the greater susceptibility of ERs than naSRs to activation by pollutants, pharmaceuticals, and dietary compounds. Extant ERs inherited ancSR1's simple ligand-recognition criterion requiring little more than an aromatized A-ring with a 3-hydroxyl (Figure 2.12). Although this rule provided sufficient specificity throughout virtually all of vertebrate evolution, ERs are now exposed to -- and fortuitously activated by -- a wide range of aromatized pharmaceutical, industrial, and agricultural substances of the appropriate size and shape that have come into large-scale production only in the last century (*18*).

In contrast, the more restrictive specificity of AR, PR, GR, and MR – which reflects the greater variety of endogenous potential activators to which they were exposed during evolution – makes them susceptible to activation by fewer synthetic substances than ERs, although they can still be disrupted by some novel substances, such as nonaromatized 19-norsteroids used as synthetic androgens. As predicted, we found that ancSR2, like its descendants, is insensitive to the aromatized xenoestrogens (Figure 2.13).

Taken together, our findings suggest that analysis of a protein's history and the chemical milieu in which it evolved can provide useful information for predicting the endogenous and exogenous ligands that can interact with it.

Structural causes of SR promiscuity.

Finally, we sought to understand the underlying features of protein structure that caused ancSR1's and ancSR2's promiscuous responses associated with minimal specificity. We first used X-ray crystallography to determine the structures of bacterially expressed ancSR2-LBD in complex with progesterone and with 11-deoxycorticosterone (DOC), at 2.75 and 2.82 Å resolution, respectively (Figure 2.4A, Table 2.6). The structures reveal why ancSR2 did not yet distinguish between progestagens and corticosteroids, which differ only in that the latter contain a 21-hydroxyl. The two protein backbones have nearly identical topologies (RMSD = 0.28 Å), and there are virtually no differences in the ways the ligands are bound (Figure 2.4A, Table 2.6). The ancSR2-progesterone complex contains ample room to accommodate the additional 21-hydroxyl of corticosteroids (Figure 2.4B). Further, Asn35 offers a perfectly positioned hydrogen bond partner, which is unpaired in the ancSR2-progesterone complex, for DOC's hydroxyl (Figure 2.4B). This additional favorable interaction explains why ancSR2 not only accommodates corticosteroids but is even more sensitive to them than progestagens.

To understand the structural causes of ancSR1's inability to distinguish between 17-hydroxyl and 17-acetyl steroids, we used homology modeling/energy minimization based on a human ERα template to predict the ancSR1-LBD structure in complex with estradiol and NPT. Despite differing by 172 amino acids, ancSR1 and ancSR2 have remarkably similar peptide backbone conformations (RMSD=0.87 Å). AncSR1's capacity to adventitiously accommodate larger 17-acetyl steroids appears to be due to excess volume and hydrogen bonding capacity in ancSR1's cavity near the ligand's D-ring. When NPT is docked in the ancSR1 cavity, virtually no adjustment is required in the position of nearby residues compared to those in the estradiol

complex: instead, the long axis of the ligand moves slightly towards H10, allowing NPT's larger acetyl group to slot into space that was unoccupied in the estradiol complex (Figure 2.4C). Further, the 20-keto of NPT accepts a hydrogen bond from His206, which can serve as a donor (as in the NPT complex), acceptor (as in the estradiol complex), or both, depending on its ionization state.

Taken together, these data indicate that the promiscuous responses of both ancSR1 and ancSR2 to non-target ligands are due in large part to unfilled volume in the internal cavity and untapped potential of polar side chains to form hydrogen bonds with polar atoms on the ligand (*1*, *8*).

Promiscuity, selection, and neutrality in the evolution of signaling.

The promiscuity we observed during SR evolution appears to reflect the fact that there is no functional difference between a receptor that excludes ligands to which the cell is never exposed and a more promiscuous receptor that does not possess such ligand recognition criteria. Although ancient and extant SRs are only minimally specific, their potential promiscuity would not have caused them to transduce noisy signals in their historical chemical environments, because such signals were not rampantly produced at the time; there would presumably have been be no fitness cost or benefit associated with the specific forms of promiscuity these receptors manifested. Rather than representing an optimum, then, the imperfect specificity of each SR appears to reflect the limited power of selection to distinguish between "perfect" and "good enough," given the chemical context in which these proteins evolved. Our findings are related to prior work suggesting that other protein properties, such as marginal stability, may not be uniquely adaptive states but may instead reflect the limited power of selection to optimize a property that affects fitness only when the property is near a threshold (*31*).

We predict that minimal specificity will be apparent in many other protein families. Protein engineering studies have shown that enzymes in the laboratory often neutrally evolve

promiscuous responses to substrates not yet present in the system (*8, 14*). Further, the limited specificity of natural proteins is what allows them to respond to novel drugs and xenobiotic pollutants. Direct study of historical evolution in other protein families and their ligands is necessary to determine the generality of the principle of minimal specificity and to characterize the dynamics that have shaped proteins' natural specificity and their responses to drugs and pollutants.

A phenomenon similar to minimal specificity is well known in biological information systems at higher levels, such as choice by individuals of conspecific mates (*32*) and mimics that lure prey or pollinators by exploiting a receiving species' signal recognition capacity (*33, 34*). In each case, the "receptor" distinguishes target from nontarget signals in the species' environment but fails to exclude novel signals to which it has not previously been exposed. Minimal specificity, reflecting evolution in the face of the limited set of stimuli present in real environments, may therefore be a general characteristic of signaling and information systems from molecular to community scales.

Methods

Phylogenetics and ancestral sequence reconstruction

Annotated protein sequences for nuclear receptors were downloaded from UniPROTKB/TrEMBL, GenBank, the JGI genome browser, and Ensemble (Table 2.7). For the reconstruction of ancSR2, 184 steroid and related receptor sequences containing both DNA binding and ligand binding domains were aligned using the Multiple Sequence Alignment by Log-Expectation (MUSCLE) program (*35*). The alignment was checked to ensure alignment of the nuclear receptor AF-2 domain and manually edited to remove lineage-specific indels. The N-terminal variable region and hinge region were removed from the alignment file, as these areas could not be aligned reliably among sequences. ancSR1 was reconstructed using an expanded alignment (213 sequences), reflecting the deposition of many new SR sequences in public databases since a much earlier study of ancSR1 (*22*).

Phylogenies (Figures 2.14, 2.15) were inferred from these alignments using PHYML v2.4.5 (*36*) and the Jones-Taylor-Thornton model with gamma-distributed among-site rate variation and empirical state frequencies, which was the best-fit evolutionary model selected using the Akaike Information Criterion implemented in PROTTEST software. Statistical support for each node was evaluated by obtaining the approximate likelihood ratio (the likelihood of the best tree with the node divided by the likelihood of the best tree without the node) and the chi-squared confidence statistic derived from that ratio (*37*).

AncSR1 and ancSR2 were initially reconstructed by the maximum likelihood method (*38*) on the ML phylogeny for each alignment using the Codeml module of PAML v3.14 (*39*) and Lazarus software (*40*), assuming a free eight-category gamma distribution of among-site rate variation and the Jones-Taylor-Thornton protein model. AncSR2 was also reconstructed on a single-branch rearrangement of the ML phylogeny that requires fewer gene duplications and

losses to explain the distribution of SRs in agnathans and jawed vertebrates (Figure 2.16, Table 2.8). Average probabilities were calculated across all LBD sites except those containing indels.

Reporter activation assays

cDNAs coding for the maximum likelihood ancSR2 LBD and ancSR1 LBD were synthesized (Genscript) and verified. The LBDs were then cloned into the Gal4-DBD-pSG5 vector; 31 amino acids of the GR hinge containing the nuclear localization signal-1 (*41*) were inserted between the DBD and LBD to ensure nuclear localization and conformational independence of the two domains. The hinge and ligand-binding domain (LBD) of the human progesterone receptor (hPR; aa 632-933; Swiss-Prot P06401), human estrogen receptor alpha (hERα, aa 435-595; Swiss-Prot P03372, (*42*)), human glucocorticoid receptor (hGR; aa 485-777; Swiss-Prot P04150, (*43*)), human mineralocorticoid receptor (hMR, aa 736-984, Swiss-Prot P08235; (*43*)) were cloned into the Gal4-DBD-pSG5 vector in frame with the Gal4 DBD. The human androgen receptor (hAR) LBD was cloned into the pFN26A (BIND) hRluc-neo Flexi Vector (Promega) without the hinge domain (aa 671-919; Swiss-Prot P10275), as the hinge domain of the hAR inhibits AF-2 dependent activation of the hAR (*44*).

The hormone-dependent transcriptional activity of resurrected ancestral receptors and their variants as well as the human receptor LBDs was assayed using a luciferase reporter system. CHO-K1 cells were grown in 96-well plates and transfected with 1 ng of receptor plasmid, 100 ng of a UAS-driven firefly luciferase reporter (pFRluc), and 0.1 ng of the constitutive pRLtk *Renilla* luciferase reporter plasmid, using Lipofectamine and Plus Reagent in OPTIMEM (Invitrogen). After 4 h, transfection medium was replaced with phenol-red-free αMEM supplemented with 10% dextran-charcoal stripped FBS (Hyclone). After overnight recovery, cells were incubated in triplicate with the hormone of interest from 10^-12 to 10^-5 M for 24 h, then assayed using Dual-Glo luciferase (Promega). Firefly luciferase activity was normalized by

Renilla luciferase activity. Dose-response relationships were estimated using nonlinear regression in Prism4 software (GraphPad Software, Inc.); fold increases in activation were calculated relative to the vehicle-only (ethanol) control.

Alternative ancestral reconstructions

To determine the robustness of functional inferences to statistical uncertainty in the reconstruction of ancSR1 and ancSR2, we used two approaches. AncSR1 had too many ambiguously reconstructed sites to examine each such residue individually, so we computationally sampled from the posterior probability distribution of reconstructed amino acid states to generate a cloud of possible ancestral sequences, each harboring a large number of alternate states. Specifically, we generated 1,000,000 possible ancestral sequences by sampling from the posterior probability distribution of states at each site. Of this sample, the five sequences with the highest total posterior probability differed from the ML reconstruction at 55 to 59 sites and from each other by 63 to 82 sites; these sequences had total posterior probabilities lower than ancSR1-ML by a factor of 10^{-23} to 10^{-24}. They differed from each other at several sites in the ligand pocket and included four unique combinations of ligand-contacting residues. We synthesized these five radically alternative ancestral reconstructions de novo and repeated the functional assays. Despite their extreme distance from ancSR1-ML, all five alternative reconstructions were sensitive to estrogens and did not respond to nonaromatized steroids (Figure 2.9).

For ancSR2, we identified all plausible alternate reconstructions (those with posterior probability >0.20 excluding biochemically similar K/R, D/E, S/T, and I/L differences) and introduced each alternate state individually into the ancSR2 background using the Quikchange Mutagenesis kit (Stratagene), verified clones by sequencing, and repeated the activation assays with each version of ancSR2 (Figure 2.10). The ML ancSR2 sequence reconstructed on the ML

tree had high baseline activation in the absence of ligand; this phenotype is almost certainly an artifact, because constitutive baseline activity is not present in any of ancSR2's extant descendants; it is well-established that some amino acid replacements can cause nuclear receptors to become constitutive by marginally stabilizing the active conformation in the absence of hormone (*45*). We therefore introduced all plausible alternate reconstructions into ancSR2-ML and found that one (L79M) eliminated this ligand-independent activity. The "constitutive" Leu79 state is weakly supported on the ML tree (PP=0.59), and has no support (PP=0.00) on the phylogeny that is most parsimonious in terms of gene duplications and losses; in contrast, the "non-constitutive" state Met79 has PP=0.41 on the ML tree and PP=1.00 on the rearranged gene duplication/loss tree (Figure 2.16, Table 2.8). The ancSR2 sequence used for all experiments reported in the text therefore contains state Met79. The other alternate reconstructions were then reintroduced into this ancSR2 sequence: none qualitatively changed the receptor's sensitivity to the various classes of steroid hormones, except for A171V, which conferred constitutive activity (Figure 2.10).

Protein expression

The ancSR2 ligand binding domain (LBD) cDNA (residues 1-252) was cloned into pLIC-MBP (provided by J. Sondek, Chapel Hill, NC), which contains a hexahistadine tag followed by the maltose binding protein (MBP) and a tobacco etch virus (TEV) protease site N-terminal to the protein. AncSR2 was expressed as a fusion protein in BL21(DE3) pLys cells in the presence of 50 μM ligand using standard methods, and initially purified using affinity chromatography (HisTrap columns, GE Healthcare). Following TEV cleavage, the tagged MBP was removed by an additional nickel affinity column. AncSR2 was purified to homogeneity via gel filtration. Pure ancSR2 LBD was dialyzed against 150 mM sodium chloride, 20 mM Tris HCl (pH 7.4), 5% glycerol, and 50 □M CHAPS and concentrated to 2-5 mg/mL.

Crystallization and structural analysis

Crystals of ancSR2-LBD with ligand were grown by hanging drop vapor diffusion at 22°C from solutions containing 1.0 μL of protein at 2-5 mg/mL protein and 1.0 μL of the following crystallant: 0.8-1.2 M $MgSO_4$, 6-12% glycerol, and 100 mM MES, pH 5.4-6.4. Orthorhombic crystals of the ancSR2 – progesterone and 11-DOC complex grew in $P2_12_12_1$ and $C222_1$ spacegroups with either two monomers or one monomer in the asymmetric unit, respectively.

Crystals were cryoprotected in crystallant containing 20% glycerol and were flash-cooled in liquid N_2. Data to 2.75 Å and 2.82 Å resolution were collected for the ancSR2-progesterone and ancSR2-deoxycorticosterone complexes, respectively (Table 2.6). All data were collected at South East Regional Collaborative Access Team (SER-CAT) 22-ID at the Advanced Photon Source at Argonne National Laboratory in Chicago, IL, and were processed and scaled with HKL2000 (HKL Inc.). Initial phases for the ancSR2- progesterone complex were determined using a homology model to the progesterone receptor (1A28) as the initial search model in Phenix (Phenix) (*46*). Subsequent structures were solved using the best available ancSR2 structure for initial phases. All structures were refined using standard methods in the CCP4 suite of programs and COOT v0.9 was used for model building (*47*). Omit maps were generated by removing coordinates corresponding to the ligand and running 10 rounds of restrained refinement in CCP4. Maps are contoured to 1 σ (Figure 2.17). Figures were generated using PyMol (Schrödinger, LLC). AncSR2 structures with progesterone and DOC have PDB accessions 4FN9 and 4FNE, respectively. Structures were rendered for display using Pymol software.

The structure of ancSR1-LBD was predicted by homology modeling, based on a human ERα:estradiol structure (1ERE), the most similar human receptor in sequence and function. We used Modeller software (*48*) to infer the ancSR1-LBD structure 100 times, chose the lowest-energy iteration from these structures, and verified it using RAMPAGE software (*49*), which

showed only 4/237 Ramachandran outliers, all of which were in surface loops. Cavity volumes were inferred using VOIDOO software (*50*) by calculating the volume accessible to a probe 1.4 Å in diameter.

Acknowledgements

We thank Jamie Bridgham and other members of the Thornton lab for comments and support. We thank E. Wilson for generously providing a clone of human AR.

Figures

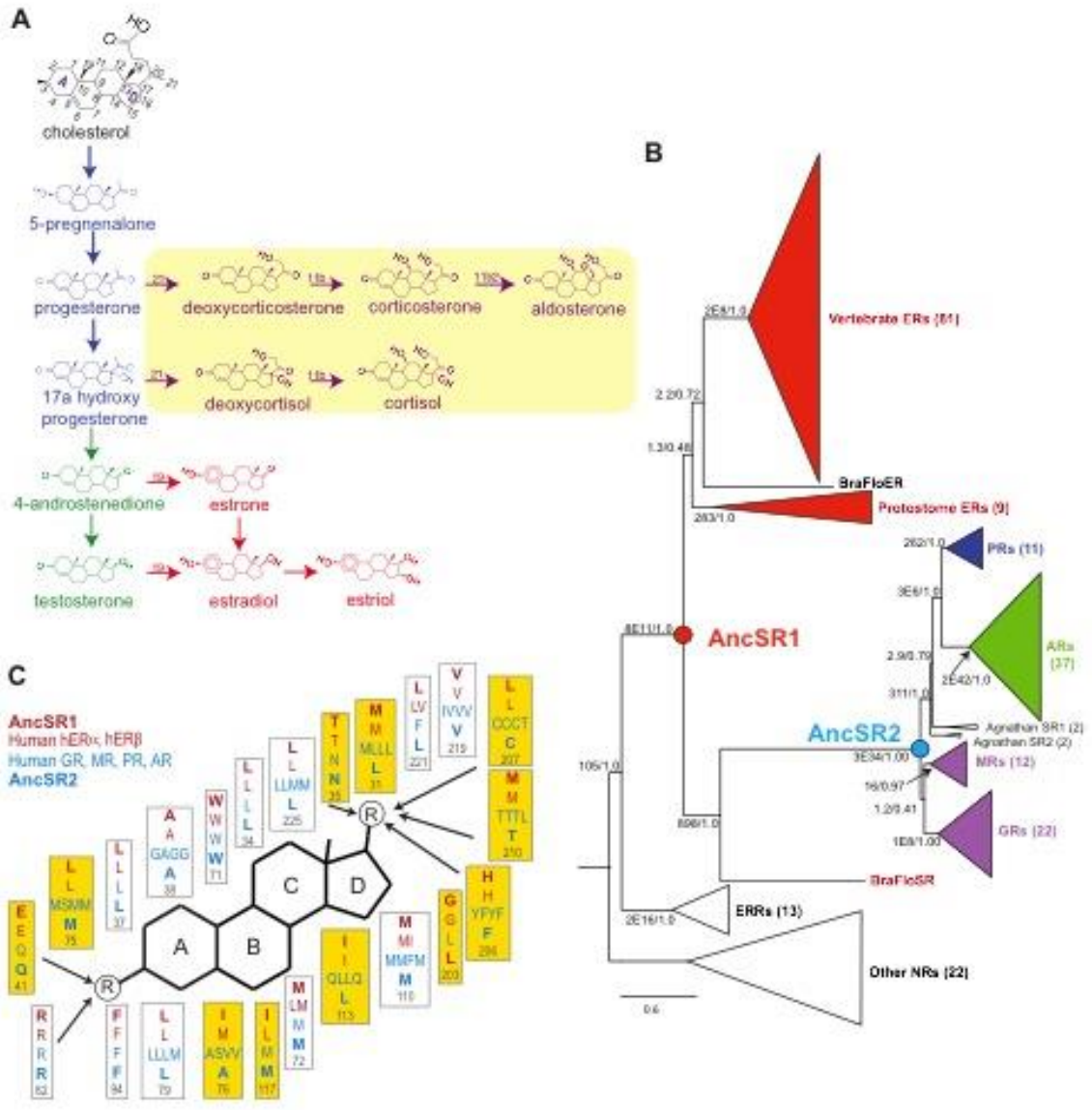


Figure 2.1: Evolutionary expansion of the steroid receptors and their ligands.

A, Pathway for synthesis of vertebrate steroid hormones. The main pathway – synthesis of estrogens (red) via progestagens (blue) and androgens (green) – is at least as ancient as the chordate ancestor. Yellow box, synthesis pathway to corticosteroids (purple), is a later evolutionary novelty found only in vertebrates. The numbering system on the steroid backbone is

shown in black. **B**, Phylogeny of the SR gene family. Receptors are color-coded by the classes of ligands to which they are most sensitive. Ancestral steroid receptors (ancSR1 and ancSR2) resurrected in this study are marked as circles. The number of sequences in each clade is shown in parentheses. Branch supports show approximate likelihood ratios and chi-square confidence metrics for each clade compared to the best phylogeny without that clade. Estrogen-responsive receptors are shown in red. For unreduced phylogenies and a list of sequences, see Figures 2.14, 2.15 and Table 2.7**.** **C**, Maximum likelihood reconstruction of ligand-contacting amino acids in ancSR1 and ancSR2, along with residues at homologous sites in extant human SRs. The steroid rings are labeled; circled R indicates polar functional groups at which the major steroid classes differ from each other; arrows indicate residues within hydrogen bonding distance. Residues that differ between ancSR1 and ancSR2 are highlighted in yellow.

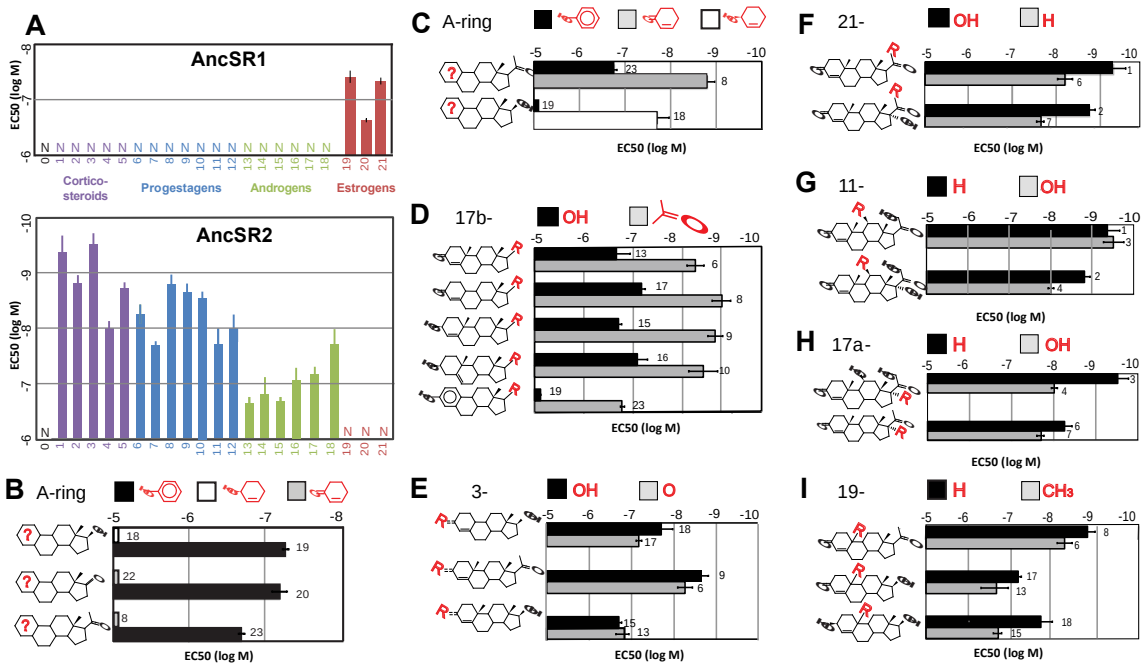


Figure 2.2: Ligand-recognition rules of ancSR1 and ancSR2.

A, The sensitivity of ancSR1-LBD (top panel) and ancSR2-LBD (bottom panel) to various hormones (Table 2.4) was characterized in a triplicate luciferase reporter assay and is displayed as EC50, the concentration at which half-maximal reporter activation is achieved. Error bars, 95% confidence interval. Sets of hormones are grouped by color and are numerically labeled according to the list below. **B,** AncSR1's ligand recognition criteria. Each pair of bars shows the EC50 of ancSR1 to a pair of hormones that differ only by aromatization of the A-ring (shown in red on the ligand structure and in the key). Unlike aromatization, substitution of a 17-keto or acetyl for estradiol's hydroxyl has only a weak effect on sensitivity, as shown by the small differences among pairs. **C-I,** AncSR2's ligand recognition criteria. Each pair of bars shows the sensitivity of the receptor to hormones that differ only in the functional group at specified positions or aromatization of the A-ring. Bar labels indicate the substance tested: **0**, cholesterol, **1**, 11-deoxycorticosterone, **2**, 11-deoxycortisol; **3**, corticosterone; **4**, cortisol; **5**, aldosterone; **6**,

progesterone; **7**, 17α-hydroxyprogesterone; **8**, 19-norprogesterone; **9**, 4-pregnenolone; **10**, 5-pregnenolone; **11**, 20α hydroxyprogesterone; **12**, 20β hydroxyprogesterone; **13**, testosterone; **14,** dihydrotestosterone; **15**, 4-androstenediol; **16**, 5-androstenediol; **17**, 19-nortestosterone; **18**, bolandiol; **19**, estradiol; **20**, estrone; **21**, estriol; **22**, 4-androstenedione; **23**, 19-nor-1, 3, 5(10)-pregnatriene-3-ol-20-one (NPT).

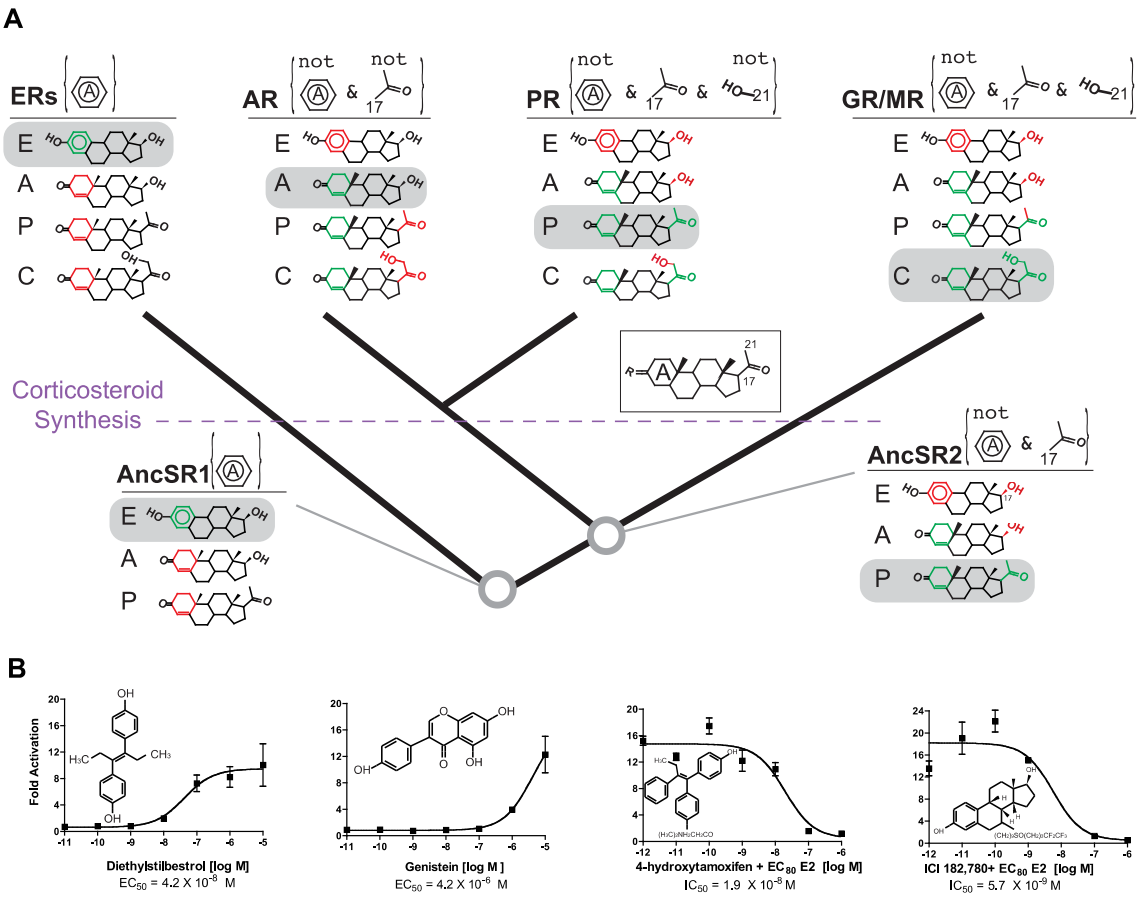


Figure 2.3: Evolution of minimal specificity.

A, Evolution of ligand-recognition criteria on the SR phylogeny. For each ancient and extant receptor, the criteria that distinguish activating ligands from other endogenous steroids are shown in brackets. Rules labeled "not" indicate significantly strongly reduced sensitivity when the specified moiety is present; other rules indicate strongly increased sensitivity when the moiety is present. The structures of representative endogenous hormones – estrogens (E), androgens (A), progestagens (P) and corticosteroids (C) – that were synthesized at each point in time are shown. Green portions of each hormone show moieties that satisfy the receptor's rules; red portions violate rules. Each receptor's rules are sufficient to allow activation by only a single class of hormones (gray boxes). The evolution of corticosteroid synthesis is indicated; ancSR2's criteria

would not have been sufficient to distinguish corticosteroids from progestagens. Inset: common steroid structure with A-ring and key carbons labeled. Dose-response curves for extant receptors are shown in Figure 2.11. **B**, AncSR1 is activated/antagonized by xenoestrogens in a luciferase reporter assay. IC50, concentration at which half-maximal inhibition was achieved in the presence of estradiol (EC_{80} = 200 nM). Each point shows the mean and SEM of three replicates.

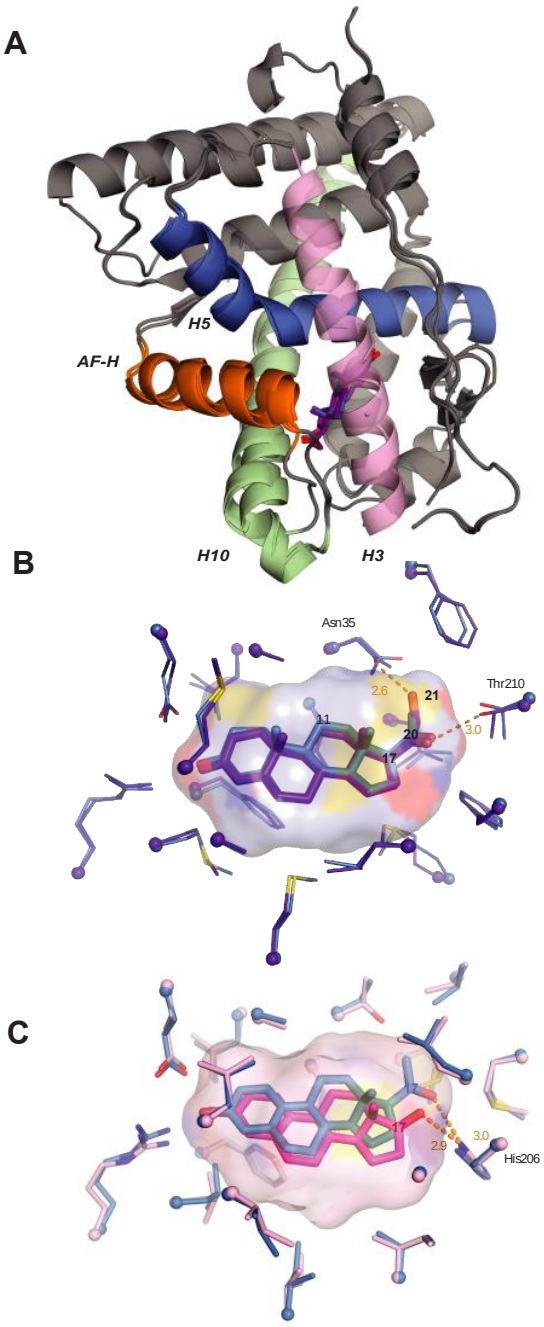
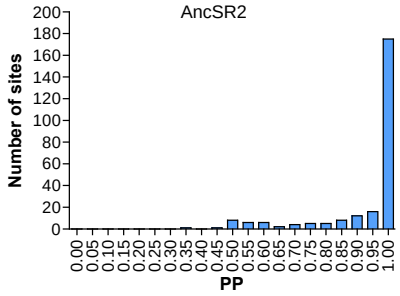


Figure 2.4: Structural causes of minimal specificity.

A, X-ray crystal structures of ancSR2 with progesterone (blue) and DOC (purple) are superimposed. Ligands are shown as sticks. Helices making major ligand contacts and the activation-function helix (AF-H) are shown in contrasting colors. **B,** Structural causes of promiscuity in ancSR2. The ligand cavity of the ancSR2-progesterone structure, shown as a surface, has adequate volume to accommodate the 21-hydroxyl of DOC. Ligand contacts in the crystal structures of ancSR2 with progesterone (blue) and DOC (purple) are shown. Thick sticks, ligand; thin sticks, side chains that contact ligand; balls, α-carbons. Steroid carbons 11, 17, 20, and 21 are numbered. Hydrogen bonds are shown as orange dotted lines. **C,** Structural basis for promiscuity in ancSR1. Ligand contacts in the ancSR1 model with estradiol (magenta) and NPT (blue) are shown. The cavity of the ancSR1-estradiol complex, which has adequate room to accommodate the 17-acetyl of NPT, is shown. Two side chains between the viewer and the ligand are hidden for clarity.

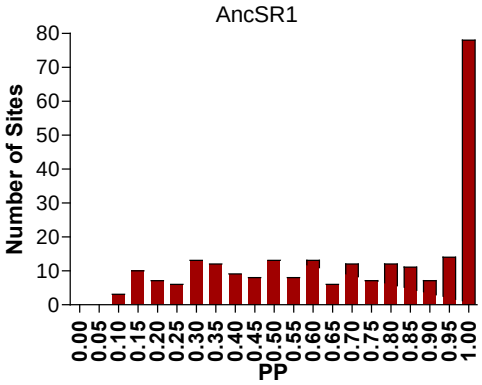


AncSR2 Binding Pocket

Position	Reconstructed Amino Acid	Probability	Alt State #1	Probability
41	**Q**	1.00		
75	**M**	1.00		
37	**L**	1.00		
38	**A**	0.87	C	0.06
71	**W**	1.00		
34	**L**	1.00		
225	**L**	0.77	M	0.23
35	**N**	1.00		
31	**L**	1.00		
221	**F**	1.00		
219	**V**	1.00		
207	**C**	1.00		
210	**T**	1.00		
206	**F**	1.00		
203	**L**	1.00		
110	**M**	1.00		
113	**L**	1.00		
72	**M**	1.00		
117	**M**	1.00		
76	**A**	0.88	S	0.05
79	**L**	0.59	M	0.40
94	**F**	1.00		
82	**R**	1.00		
	Mean Posterior Probability	**0.96**		

Figure 2.5: Histogram of posterior probabilities for ancSR2

Histogram of distribution of posterior probabilities for ancSR2 and posterior probabilities of amino acid residues lining the binding pocket.



AncSR1 binding pocket

Position	Reconstructed Amino Acid	Probability	Alt. State #1	Probability
41	**E**	1.00		
75	**L**	1.00		
37	**L**	1.00		
38	**A**	1.00		
71	**W**	1.00		
34	**L**	0.99	I	0.00
225	**L**	1.00	M	0.00
35	**T**	0.97	S	0.02
31	**M**	0.62	L	0.27
221	**L**	0.84	M	0.10
219	**V**	1.00		
207	**L**	0.97	F	0.03
210	**M**	0.47	I	0.44
206	**H**	1.00	Y	0.00
203	**G**	0.88	A	0.11
110	**M**	0.98	L	0.02
113	**I**	0.99	L	0.00
72	**M**	0.77	L	0.23
117	**I**	0.79	V	0.16
76	**I**	0.50	M	0.46
79	**L**	0.99	M	0.01
94	**F**	1.00		
82	**R**	1.00		
Mean Posterior Probability		0.90		

Figure 2.6: Histogram of posterior probabilities for ancSR1

Histogram of distribution of posterior probabilities for ancSR1 and posterior probabilities of amino acid residues lining the binding pocket.

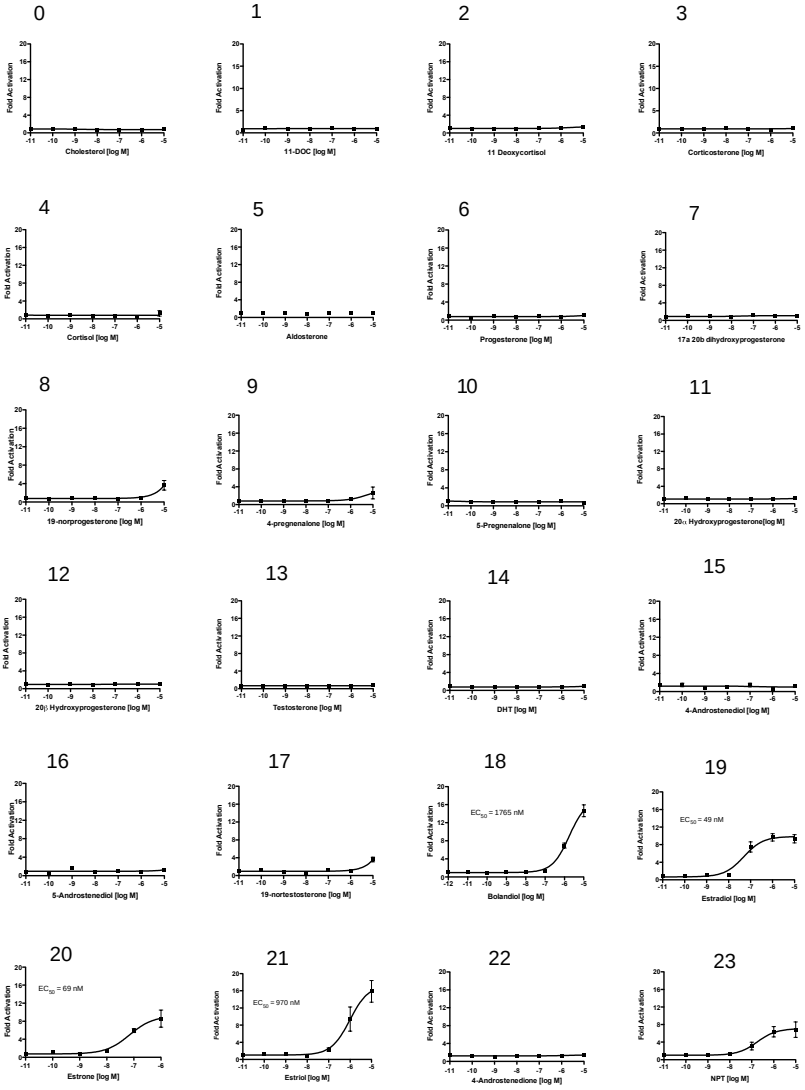


Figure 2.7: Dose activation curves of ancSR1

Representative dose activation curves of ancSR1 in response to cholesterol and a library of hormones (#0-23).

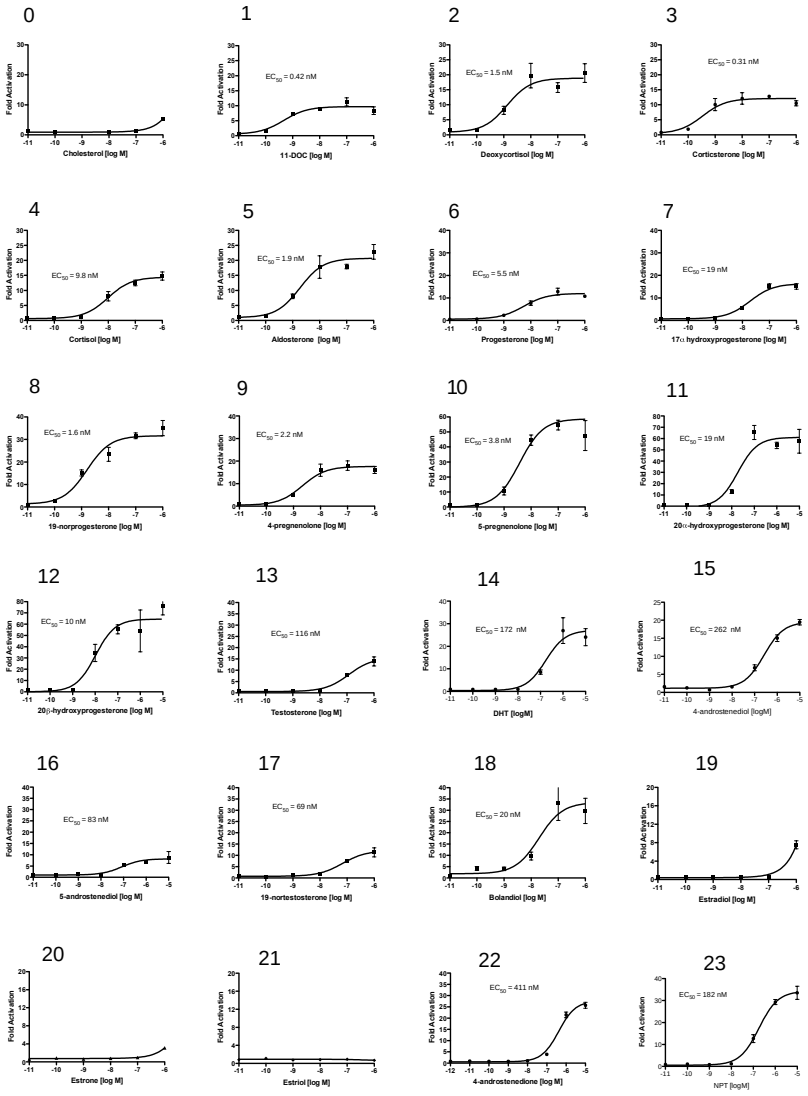
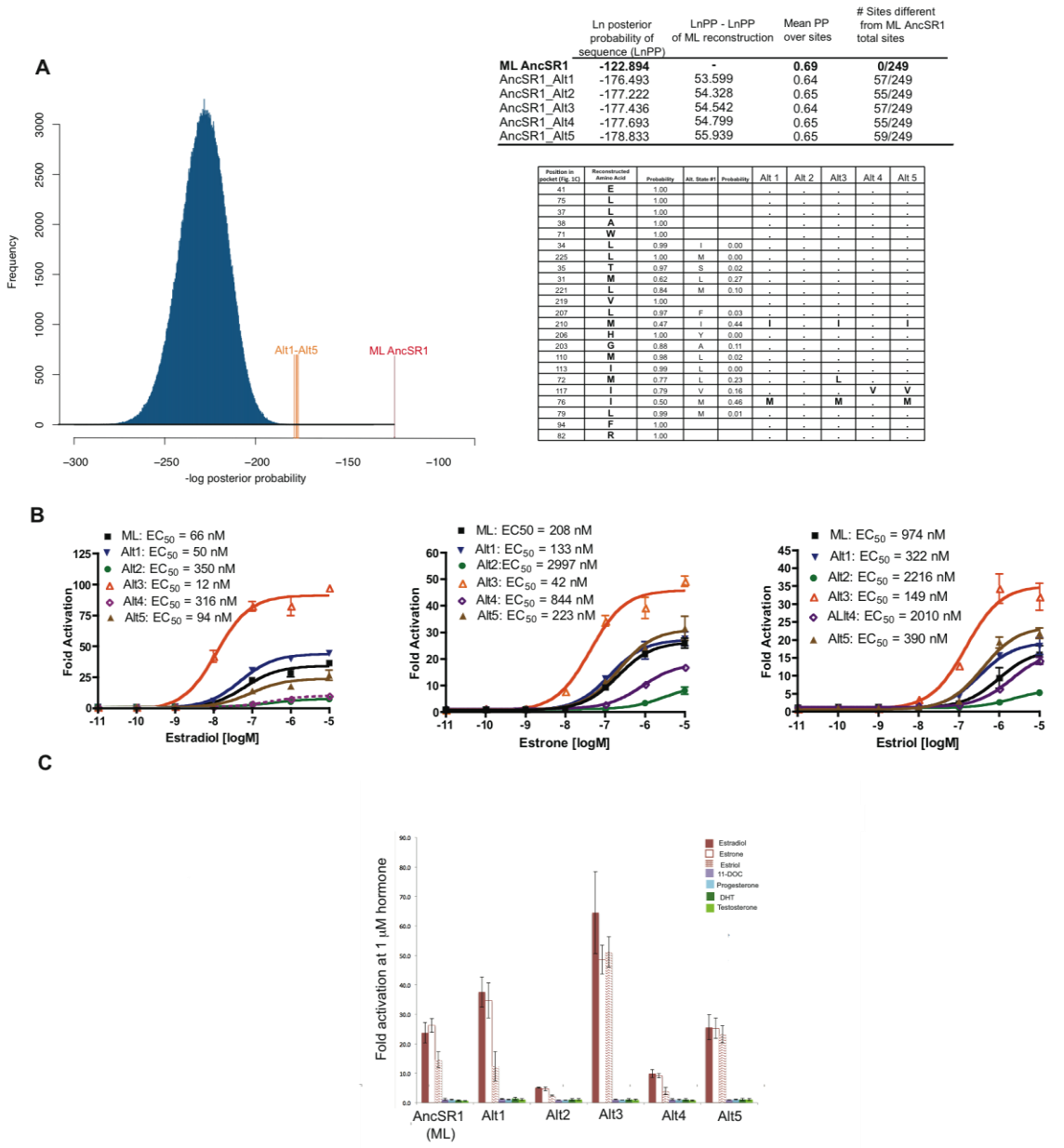


Figure 2.8: Dose activation curves of ancSR2

Representative dose activation curves of ancSR2 in response to cholesterol and a library of hormones (#0-23).



	Ln posterior probability of sequence (LnPP)	LnPP - LnPP of ML reconstruction	Mean PP over sites	# Sites different from ML AncSR1 total sites
ML AncSR1	**-122.894**	**-**	**0.69**	**0/249**
AncSR1_Alt1	-176.493	53.599	0.64	57/249
AncSR1_Alt2	-177.222	54.328	0.65	55/249
AncSR1_Alt3	-177.436	54.542	0.64	57/249
AncSR1_Alt4	-177.693	54.799	0.65	55/249
AncSR1_Alt5	-178.833	55.939	0.65	59/249

Position in pocket (Fig. 1C)	Reconstructed Amino Acid	Probability	Alt. State #1	Probability	Alt 1	Alt 2	Alt3	Alt 4	Alt 5
41	**E**	1.00			.	.	.	.	.
75	**L**	1.00			.	.	.	.	.
37	**L**	1.00			.	.	.	.	.
38	**A**	1.00			.	.	.	.	.
71	**W**	1.00			.	.	.	.	.
34	**L**	0.99	I	0.00	.	.	.	.	.
225	**L**	1.00	M	0.00	.	.	.	.	.
35	**T**	0.97	S	0.02	.	.	.	.	.
31	**M**	0.62	L	0.27	.	.	.	.	.
221	**L**	0.84	M	0.10	.	.	.	.	.
219	**V**	1.00			.	.	.	.	.
207	**L**	0.97	F	0.03	.	.	.	.	.
210	**M**	0.47	I	0.44	**I**	.	**I**	.	**I**
206	**H**	1.00	Y	0.00	.	.	.	.	.
203	**G**	0.88	A	0.11	.	.	.	.	.
110	**M**	0.98	L	0.02	.	.	.	.	.
113	**I**	0.99	L	0.00	.	.	.	.	.
72	**M**	0.77	L	0.23	.	.	**L**	.	.
117	**I**	0.79	V	0.16	.	.	.	**V**	**V**
76	**I**	0.50	M	0.46	**M**	.	**M**	.	**M**
79	**L**	0.99	M	0.01	.	.	.	.	.
94	**F**	1.00			.	.	.	.	.
82	**R**	1.00			.	.	.	.	.

Figure 2.9: The specificity of ancSR1 is robust to uncertainty in the reconstruction.

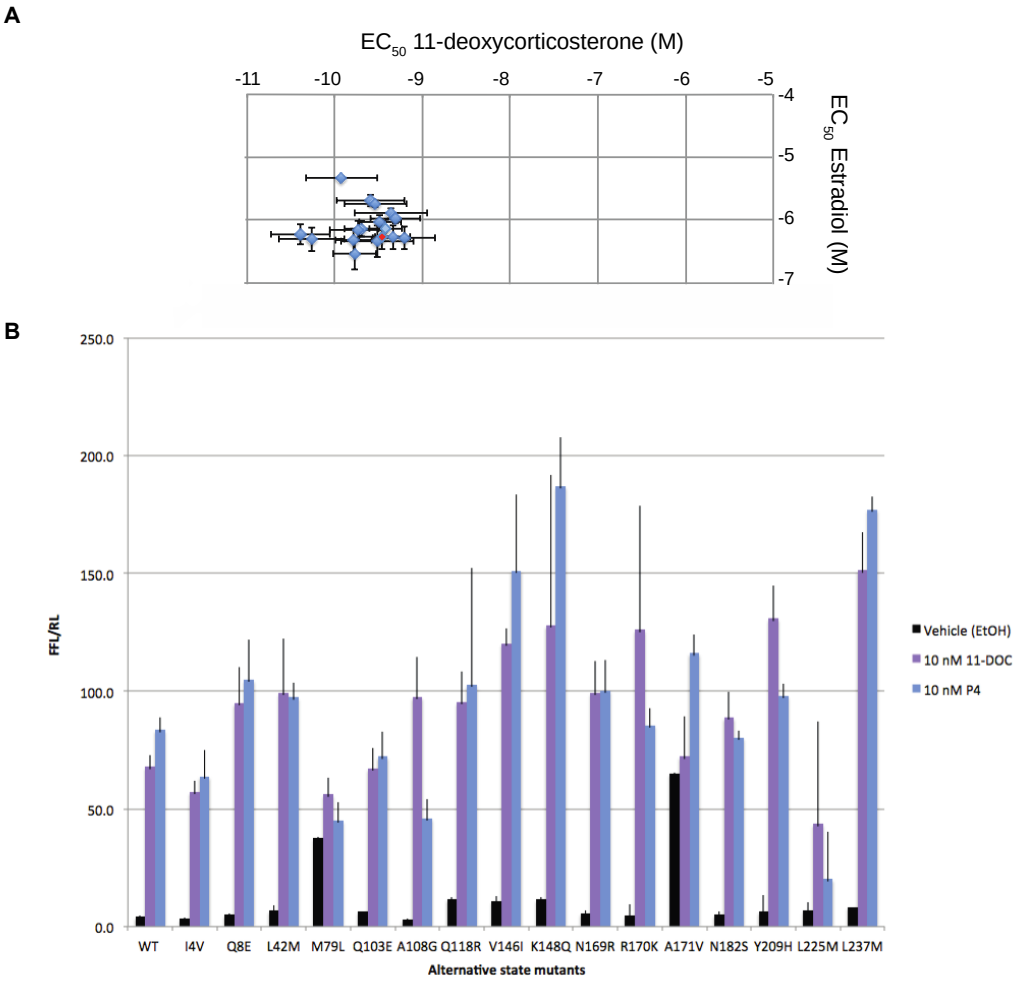


Figure 2.10 The specificity of ancSR2 is robust to uncertainty in the reconstruction.

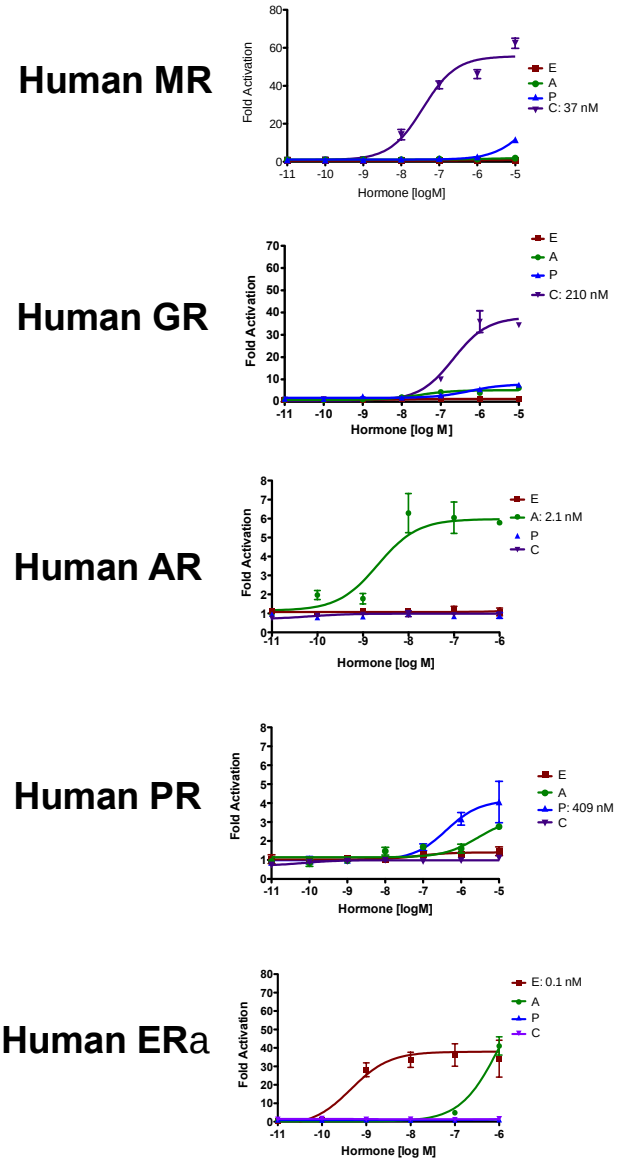


Figure 2.11: Sensitivities of extant human receptors to an estrogen, androgen, progestagen, and corticosteroid.

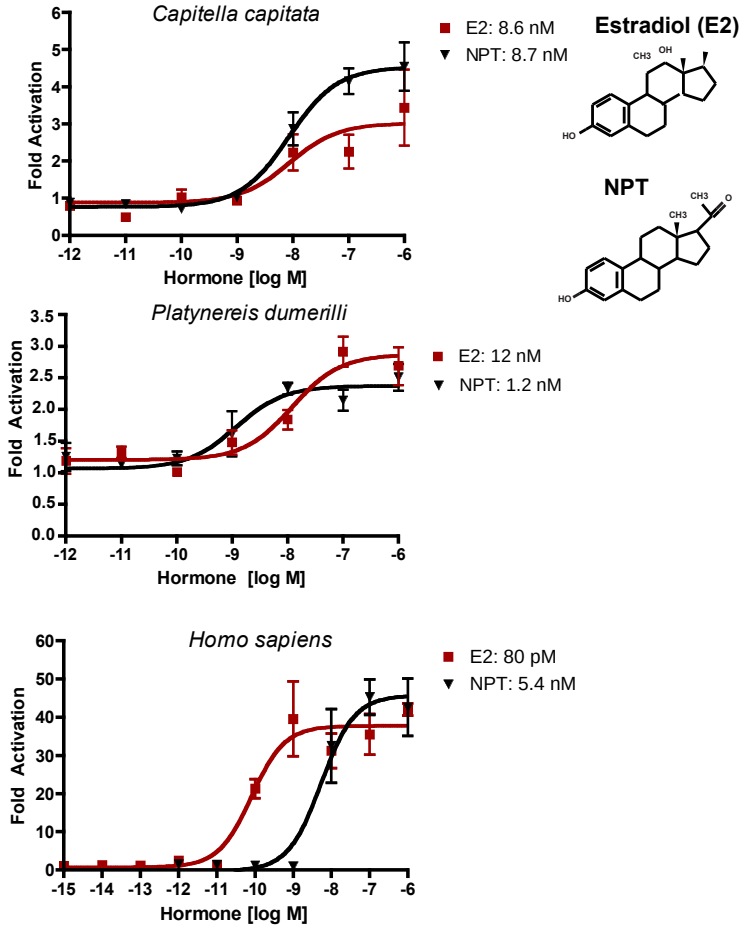


Figure 2.12: Activation of the estrogen receptor ligand binding domains of two annelids and human ERα.

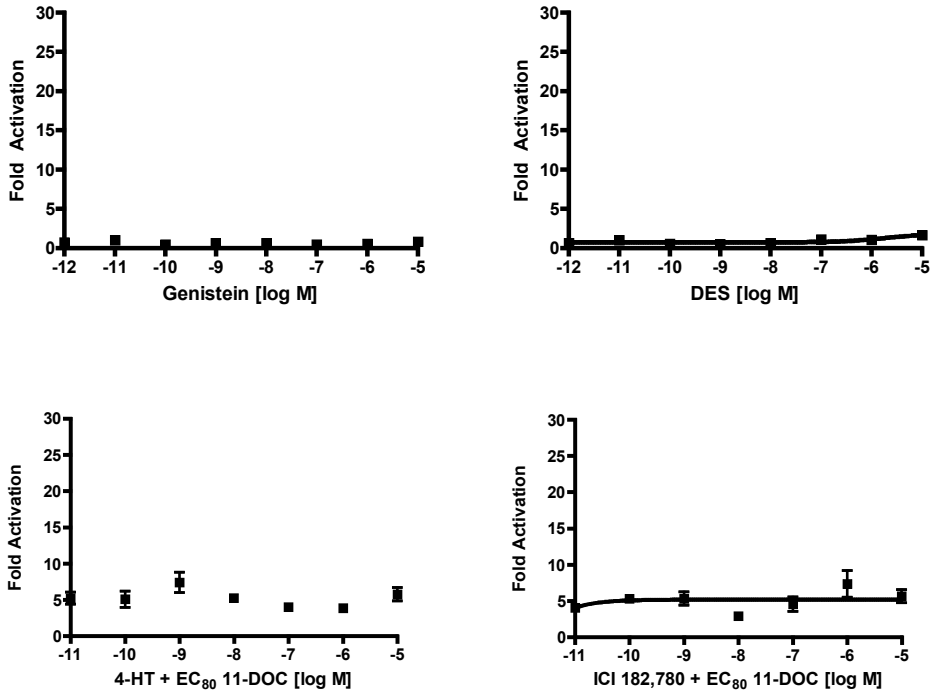


Figure 2.13: AncSR2 is not activated by the nonsteroidal ER agonists diethylstilbestrol and genistein and is not inhibited by ICI182870 and 4-hydroxytamoxifen.

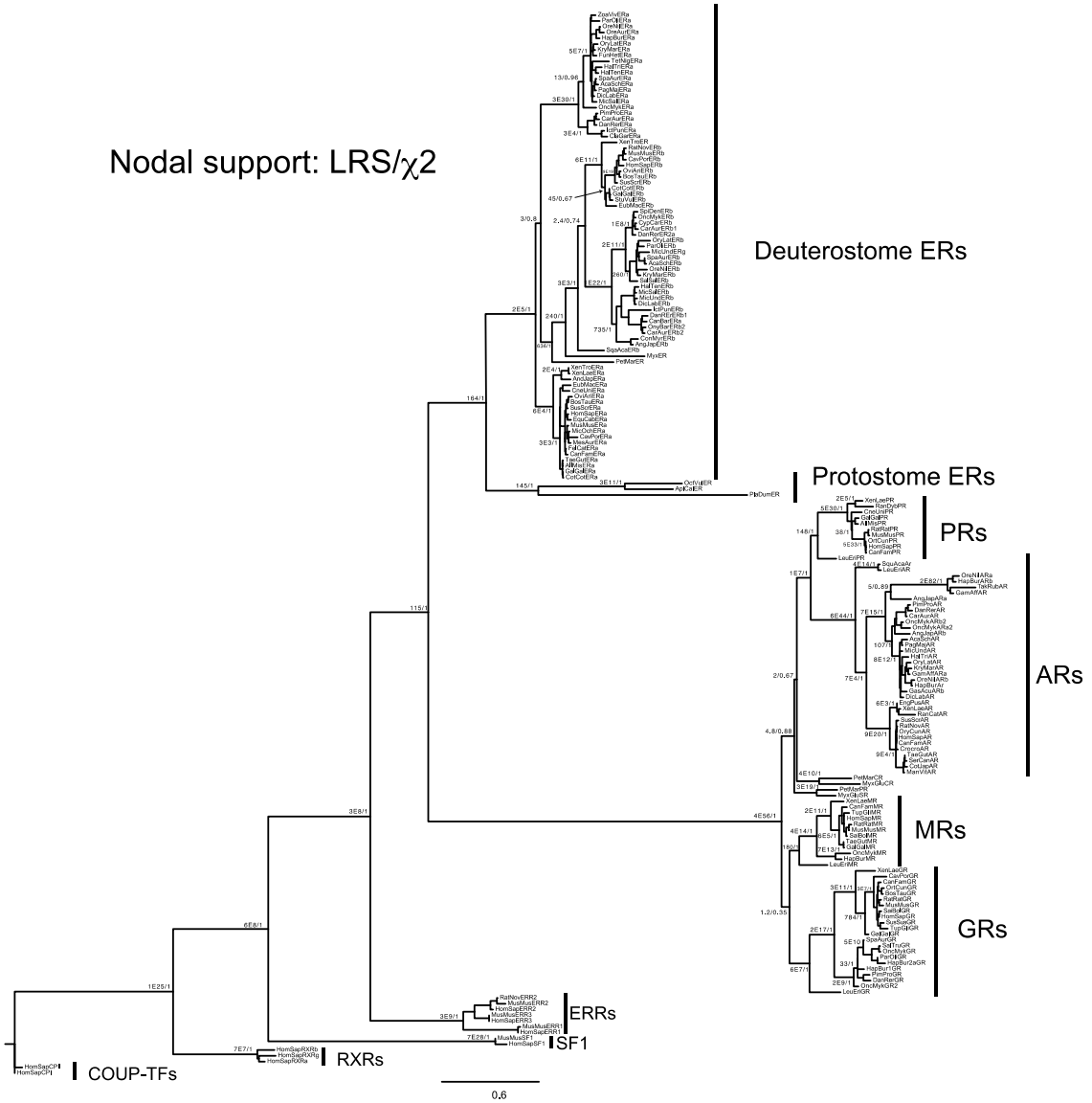


Figure 2.14: ML steroid receptor phylogeny for ancSR2

Unreduced ML steroid receptor phylogeny based on alignment of 184 steroid receptors and related sequences used to reconstruct ancSR2.

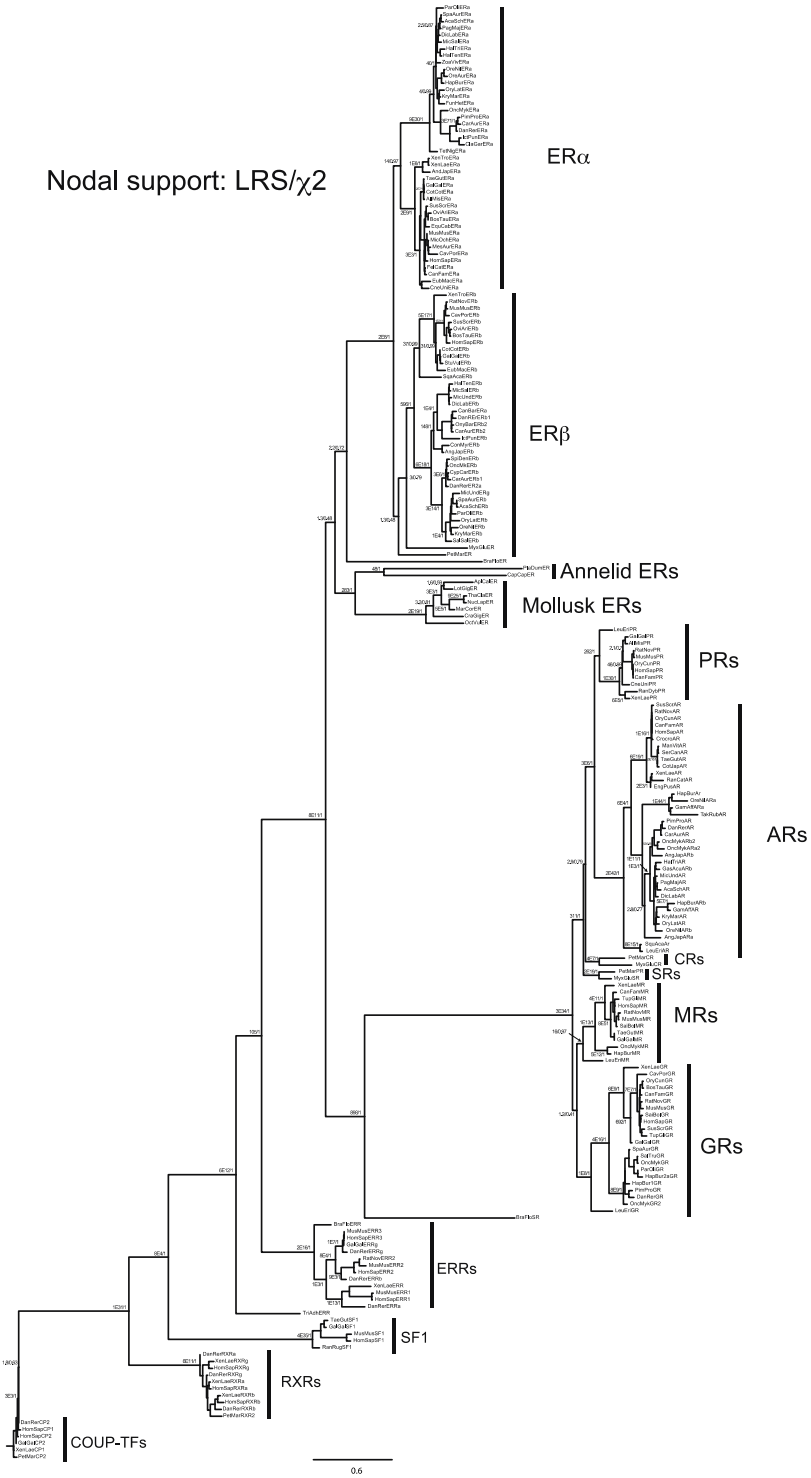


Figure 2.15: ML steroid receptor phylogeny for ancSR1

Unreduced ML steroid receptor phylogeny based on alignment of 213 steroid receptors and related sequences used to reconstruct ancSR1.

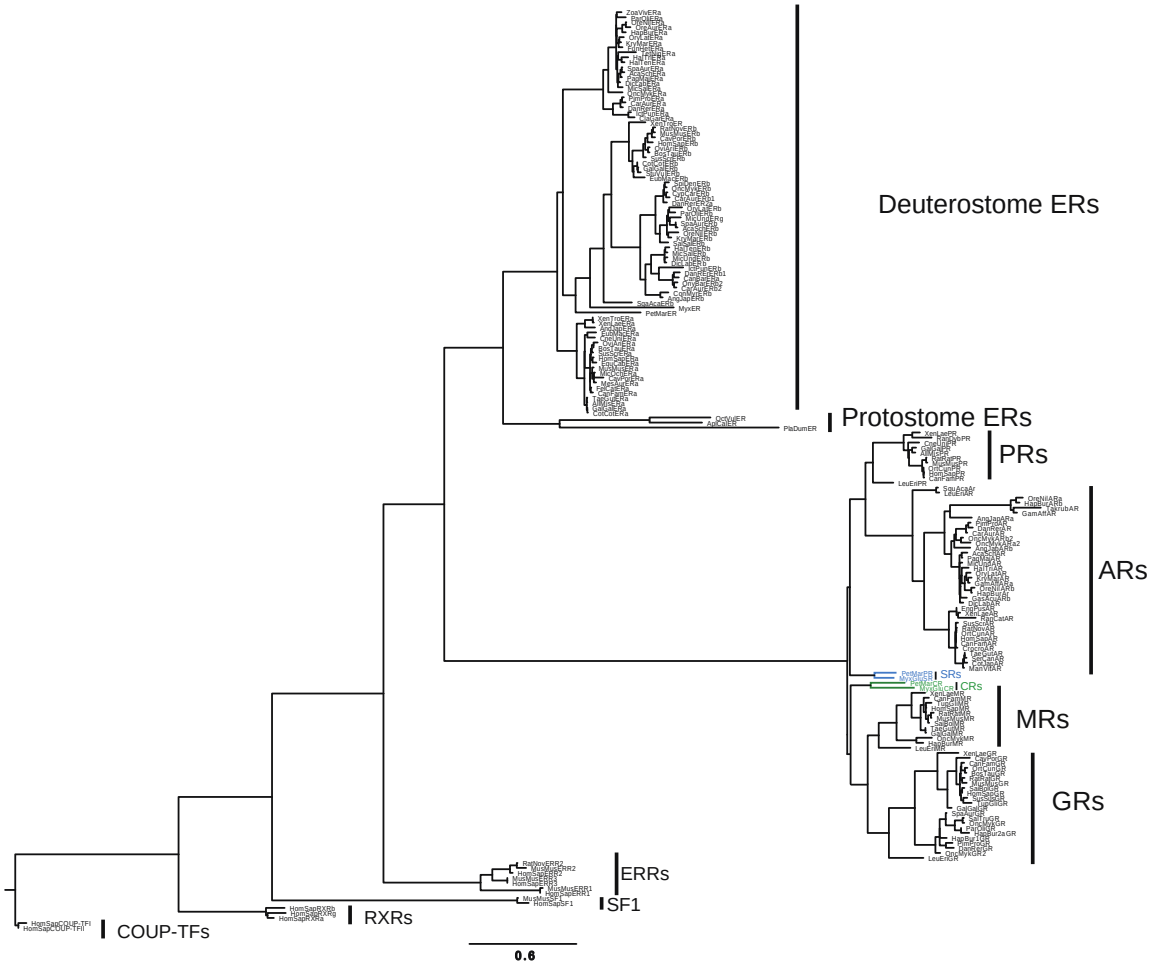


Figure 2.16: Unreduced 184-taxon steroid receptor gene duplication phylogeny.

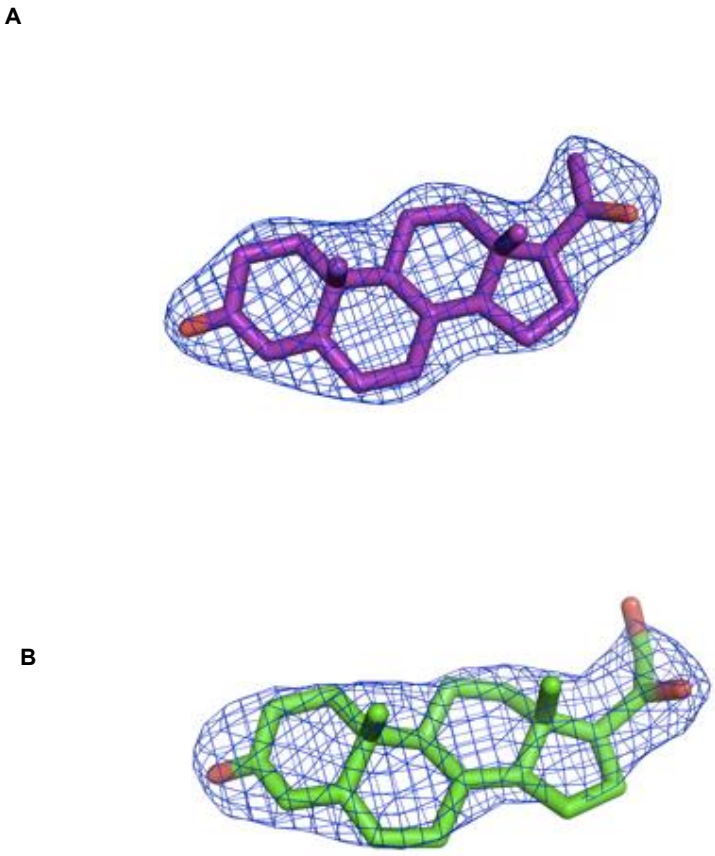


Figure 2.17: Omit maps of progesterone and 11-deoxycorticosterone

F_o-F_c electron density (green) contoured to 2 σ showing evidence for bound ligand (progesterone – blue, DOC – green). Omit maps were generated by removal of the ligand from the structure and running 3 cycles of gradient energy minimization and B-factor optimization in PHENIX (version dev-1423) to minimize model bias

AncSR2 reconstructed on 184-taxa ML tree

Residue No.	Reconstructed Amino Acid	PP	Alt state 1	PP	Alt state 2	PP
1	P	1.00				
2	S	0.81	T	0.19	P	0.01
3	L	0.91	I	0.04	M	0.03
4	I	0.78	V	0.21	M	0.00
5	S	0.94	T	0.05	A	0.01
6	I	0.94	V	0.03	L	0.02
7	L	1.00				
8	Q	0.50	E	0.49	K	0.01
9	A	0.56	V	0.19	S	0.13
10	I	1.00				
11	E	1.00				
12	P	1.00				
13	E	0.99	D	0.01		
14	V	0.96	I	0.04	M	0.00
15	V	0.98	I	0.02	L	0.00
16	Y	1.00	F	0.00		
17	A	1.00				
18	G	1.00				
19	Y	1.00	F	0.00		
20	D	1.00				
21	N	0.81	S	0.19	T	0.00
22	T	0.61	S	0.38	A	0.01
23	Q	0.97	R	0.02	H	0.00
24	P	1.00				
25	D	0.95	E	0.04	M	0.00
26	T	1.00				
27	T	0.99	S	0.01	A	0.00
28	N	0.97	S	0.02	D	0.01
29	Y	0.80	H	0.19	R	0.01
30	L	1.00	M	0.00		
31	L	1.00				
32	S	0.90	T	0.10	A	0.00
33	S	1.00	T	0.00		
34	L	1.00				
35	N	1.00				
36	R	0.99	K	0.01	Q	0.00
37	L	1.00				
38	A	0.87	C	0.06	G	0.05
39	E	0.85	G	0.13	D	0.02
40	K	0.61	R	0.40		
41	Q	1.00				
42	L	0.77	M	0.23		
43	V	1.00	I	0.00		
44	S	0.92	R	0.05	P	0.02
45	V	1.00	I	0.00		
46	V	1.00				
47	K	1.00	R	0.00		
48	W	1.00				
49	A	1.00				
50	K	1.00				
51	A	0.88	V	0.08	S	0.04
52	L	1.00	M	0.00		
53	P	1.00				
54	G	1.00				
55	F	1.00				
56	R	1.00	K	0.00		
57	N	0.99	S	0.01	D	0.00
58	L	1.00				
59	H	0.98	P	0.01	Q	0.01
60	L	0.79	I	0.21	M	0.00
61	D	0.99	E	0.01	N	0.00
62	D	1.00				
63	Q	1.00				
64	M	1.00				
65	T	1.00				
66	L	1.00				
67	I	0.52	L	0.48	M	0.00
68	Q	1.00				
69	Y	1.00				
70	S	1.00				
71	W	1.00				
72	M	1.00				
73	G	0.75	S	0.14	C	0.10
74	L	1.00				
75	M	1.00				
76	A	0.88	S	0.05	T	0.03
77	F	1.00				
78	A	0.84	S	0.15	G	0.00
79	L	0.59	M	0.41	I	0.00
80	G	0.92	S	0.07	A	0.02
81	W	1.00				
82	R	1.00				
83	S	1.00				
84	Y	1.00				
85	K	0.99	R	0.01	Q	0.00
86	H	1.00				
87	T	0.82	A	0.07	V	0.06
88	N	1.00				
89	G	1.00	S	0.00		
90	Q	0.76	K	0.13	N	0.04
91	M	1.00	L	0.00		
92	L	1.00				
93	Y	1.00	F	0.00		
94	F	1.00				
95	A	1.00				
96	P	1.00				
97	D	1.00				
98	L	1.00				
99	I	0.98	V	0.02		
100	F	1.00				
101	N	1.00				
102	E	1.00				
103	Q	0.63	E	0.36	D	0.01
104	R	1.00				
105	M	1.00				
106	Q	0.99	H	0.00	K	0.00
107	Q	1.00				
108	S	1.00				
109	A	1.00	T	0.00	S	0.00
110	M	1.00				
111	Y	1.00	F	0.00	H	0.00
112	D	0.73	E	0.27	N	0.00
113	L	1.00				
114	C	1.00				
115	Q	0.48	L	0.18	M	0.09
116	G	1.00				
117	M	1.00				
118	Q	0.48	R	0.48	H	0.02
119	Q	0.99	K	0.01	N	0.00
120	I	1.00	V	0.00		
121	S	0.99	A	0.01	T	0.00
122	Q	0.47	L	0.11	K	0.09
123	E	1.00	D	0.00		
124	F	1.00				
125	V	0.87	I	0.07	M	0.03
126	R	0.98	K	0.02	H	0.00
127	L	1.00				
128	Q	1.00				
129	V	0.97	L	0.03	I	0.01
130	T	0.98	S	0.02	A	0.00
131	Q	0.53	H	0.28	Y	0.17
132	E	1.00	D	0.01		
133	E	1.00				
134	F	0.99	Y	0.01		
135	L	1.00				
136	C	1.00				
137	M	1.00				
138	K	1.00				
139	A	0.87	V	0.13		
140	L	0.98	I	0.02	M	0.00
141	L	1.00				
142	L	1.00				
143	L	1.00				
144	S	0.99	N	0.01	G	0.00
145	T	1.00				
146	V	0.67	I	0.33		
147	P	1.00				
148	K	0.61	Q	0.34	R	0.04
149	E	0.54	D	0.46	N	0.00
150	G	1.00				
151	L	1.00				
152	K	1.00	R	0.00		
153	S	1.00	T	0.00	N	0.00
154	Q	1.00	H	0.00		
155	A	0.83	T	0.13	S	0.04
156	S	0.34	Y	0.17	A	0.15
157	F	1.00				
158	D	0.95	E	0.05	N	0.00
159	E	1.00	D	0.00	Q	0.00
160	M	0.99	I	0.01	L	0.00
161	R	1.00				
162	M	1.00	I	0.00	L	0.00
163	N	0.98	S	0.02	D	0.00
164	Y	1.00				
165	I	1.00				
166	K	0.68	R	0.32	Q	0.00
167	E	1.00				
168	L	1.00				
169	N	0.43	R	0.25	G	0.18
170	R	0.66	K	0.34	Q	0.00
171	A	0.68	V	0.31	T	0.01
172	I	0.94	V	0.06		
173	A	0.68	V	0.20	G	0.05
174	K	0.55	R	0.34	Q	0.06
175	K	0.55	R	0.24	Q	0.14
176	E	0.98	D	0.02	Q	0.00
177	N	0.85	K	0.08	G	0.04
178	N	1.00	S	0.00	D	0.00
179	S	0.53	T	0.24	A	0.16
180	A	0.47	G	0.24	S	0.19
181	Q	0.99	E	0.01		
182	N	0.56	S	0.41	G	0.01
183	W	1.00				
184	Q	1.00	H	0.00		
185	R	1.00				
186	F	1.00				
187	Y	1.00				
188	Q	1.00				
189	L	1.00				
190	T	1.00				
191	K	1.00				
192	L	0.99	M	0.01		
193	L	1.00				
194	D	1.00				
195	S	1.00				
196	M	1.00				
197	H	1.00				
198	D	0.97	E	0.03		
199	L	1.00	M	0.00		
200	V	1.00				
201	G	1.00	E	0.00	S	0.00
202	G	0.93	K	0.05	N	0.01
203	L	1.00				
204	L	1.00				
205	Q	0.99	E	0.01		
206	F	1.00				
207	C	1.00				
208	F	1.00				
209	Y	0.60	H	0.30	N	0.07
210	T	1.00				
211	F	1.00				
212	V	0.99	I	0.01	M	0.01
213	Q	0.90	E	0.10		
214	S	1.00	N	0.00		
215	Q	0.82	K	0.17	R	0.01
216	A	0.95	T	0.03	S	0.01
217	L	1.00	M	0.00		
218	S	0.96	K	0.01	N	0.01
219	V	1.00				
220	E	1.00				
221	F	1.00				
222	P	1.00				
223	E	1.00	D	0.00		
224	M	1.00				
225	L	0.77	M	0.23	I	0.00
226	V	0.63	S	0.18	A	0.14
227	E	1.00				
228	I	1.00	M	0.00		
229	I	1.00				
230	S	1.00	N	0.00	T	0.00
231	A	0.90	D	0.07	N	0.03
232	Q	1.00				
233	L	1.00	M	0.00	F	0.00
234	P	1.00				
235	K	0.95	R	0.05	Q	0.00
236	V	0.86	I	0.14	M	0.00
237	L	0.49	M	0.22	T	0.14
238	A	0.92	S	0.05	T	0.03
239	G	1.00				
240	M	0.97	L	0.01	I	0.01
241	A	0.96	T	0.03	V	0.00
242	K	1.00	R	0.00		
243	P	1.00	S	0.00		
244	L	1.00				
245	L	0.99	F	0.00	M	0.00
246	F	1.00				
247	H	1.00				
248	K	0.91	R	0.07	Q	0.02
249	K	1.00	R	0.00		
	Mean Posterior Probability	0.93				

Table 2.1: Reconstructed sequence of ancSR2

The reconstructed sequence of ancSR2. Ambiguously reconstructed sites with a posterior probability (PP)>0.2 (grey shading) were introduced into the ML ancestor and their responses to 11-deoxycorticosterone and estradiol evaluated using a luciferase reporter assay.

AncSR1 reconstructed on 213-taxon gene duplication tree

Residue No.	Reconstructed Amino Acid	PP	Alt state 1	PP	Alt state 2	PP
1	E	0.14	K	0.13	R	0.09
2	K	0.17	R	0.14	S	0.10
3	P	0.21	S	0.12	A	0.10
4	L	0.12	P	0.11	A	0.10
5	S	0.12	K	0.09	A	0.09
6	S	0.13	A	0.09	K	0.08
7	L	0.16	S	0.09	A	0.08
8	P	0.31	S	0.26	T	0.15
9	A	0.38	S	0.15	T	0.13
10	N	0.22	E	0.16	K	0.12
11	Q	0.28	S	0.15	P	0.11
12	L	0.45	I	0.30	V	0.16
13	I	0.62	V	0.32	L	0.03
14	S	0.40	N	0.14	A	0.13
15	A	0.57	S	0.09	T	0.07
16	L	1.00				
17	L	0.67	M	0.13	Q	0.05
18	E	0.47	K	0.16	Q	0.07
19	A	0.99	V	0.01	T	0.00
20	E	1.00				
21	P	1.00				
22	P	0.94	S	0.02	Q	0.01
23	V	0.29	I	0.28	T	0.10
24	L	0.51	V	0.21	I	0.14
25	Y	0.79	H	0.15	F	0.01
26	S	0.49	A	0.44	T	0.06
27	R	0.16	Q	0.14	H	0.11
28	H	0.97	Y	0.03	Q	0.00
29	D	0.99	N	0.01	E	0.00
30	P	0.99	S	0.00	H	0.00
31	S	0.13	A	0.09	D	0.09
32	K	0.12	L	0.12	R	0.11
33	P	1.00				
34	D	0.16	P	0.16	S	0.15
35	T	0.98	S	0.02	A	0.00
36	E	0.75	D	0.17	K	0.03
37	A	0.48	V	0.14	S	0.10
38	H	0.36	N	0.22	S	0.13
39	L	0.50	M	0.37	I	0.07
40	M	0.62	L	0.27	I	0.10
41	T	0.88	A	0.06	S	0.03
42	S	0.81	T	0.18	A	0.01
43	L	0.99	I	0.00	V	0.00
44	T	0.97	S	0.02	I	0.00
45	N	0.37	D	0.29	E	0.14
46	L	1.00				
47	A	1.00				
48	D	1.00	E	0.00	N	0.00
49	R	0.98	K	0.02		
50	E	1.00				
51	L	1.00				
52	V	0.99	I	0.01		
53	H	0.33	G	0.12	D	0.12
54	I	0.52	V	0.36	M	0.09
55	I	0.99	V	0.01		
56	N	0.68	D	0.20	S	0.05
57	W	1.00				
58	A	1.00				
59	K	1.00				
60	K	0.39	R	0.36	H	0.15
61	I	0.96	V	0.03	L	0.01
62	P	1.00				
63	G	1.00				
64	Y	0.64	F	0.36		
65	S	0.82	T	0.12	A	0.04
66	D	0.48	E	0.28	N	0.12
67	L	1.00				
68	S	0.86	P	0.04	A	0.03
69	L	0.99	M	0.01	I	0.00
70	N	0.64	H	0.09	D	0.08
71	D	1.00				
72	Q	1.00				
73	V	1.00	M	0.00		
74	H	0.63	N	0.14	S	0.08
75	L	1.00				
76	L	0.95	I	0.05	M	0.00
77	Q	0.87	E	0.13	R	0.00
78	S	0.61	C	0.35	A	0.02
79	C	0.60	S	0.31	A	0.06
80	W	1.00				
81	M	0.77	L	0.23	I	0.00
82	E	0.98	D	0.02		
83	L	0.69	I	0.16	V	0.14
84	L	1.00				
85	I	0.50	M	0.46	V	0.04
86	L	0.70	I	0.22	V	0.05
87	G	1.00	D	0.00		
88	L	0.99	M	0.01	V	0.00
89	A	0.68	V	0.29	I	0.02
90	W	1.00				
91	R	1.00				
92	S	1.00				
93	M	0.83	I	0.12	L	0.04
94	D	0.42	E	0.37	Q	0.06
95	H	0.98	Y	0.02		
96	E	0.36	K	0.21	Q	0.14
97	G	0.97	D	0.02	N	0.01
98	K	0.80	E	0.08	R	0.07
99	L	1.00	M	0.00		
100	V	0.50	I	0.39	L	0.05
101	F	1.00				
102	A	1.00				
103	P	1.00	S	0.00	A	0.00
104	D	1.00	N	0.00		
105	L	0.99	M	0.01	F	0.00
106	I	0.61	V	0.25	L	0.08
107	L	0.73	M	0.20	F	0.04
108	D	1.00	N	0.00	E	0.00
109	R	0.91	K	0.06	Q	0.02
110	E	0.29	D	0.21	N	0.16
111	Q	0.69	R	0.23	H	0.04
112	S	0.82	G	0.07	A	0.05
113	K	0.59	R	0.37	Q	0.02
114	C	0.68	S	0.09	A	0.05
115	V	0.59	I	0.17	A	0.12
116	A	0.91	S	0.07	T	0.02
117	G	1.00				
118	M	0.98	L	0.02	I	0.00
119	E	0.22	D	0.15	A	0.13
120	E	0.82	D	0.17	Q	0.01
121	I	0.99	L	0.00	V	0.00
122	C	0.74	S	0.17	F	0.06
123	D	0.42	E	0.30	N	0.06
124	Q	0.79	H	0.13	P	0.03
125	I	0.79	V	0.16	M	0.04
126	L	0.91	M	0.03	F	0.02
127	E	0.35	A	0.30	Q	0.14
128	I	0.46	L	0.24	V	0.19
129	A	0.87	S	0.09	T	0.02
130	S	0.17	Q	0.17	R	0.15
131	Q	0.56	R	0.22	K	0.22
132	F	1.00	Y	0.00		
133	R	0.61	K	0.19	Q	0.04
134	E	0.59	Q	0.15	D	0.12
135	L	0.98	F	0.02	I	0.00
136	K	0.37	Q	0.20	R	0.16
137	V	0.43	L	0.25	I	0.24
138	Q	0.37	N	0.15	E	0.12
139	K	0.53	R	0.44	Q	0.03
140	E	0.96	D	0.03	Q	0.01
141	E	1.00				
142	F	0.93	Y	0.07		
143	V	0.95	I	0.02	L	0.02
144	C	1.00				
145	L	1.00				
146	K	1.00	R	0.00		
147	A	1.00				
148	I	0.94	M	0.03	V	0.02
149	T	0.61	A	0.18	V	0.11
150	L	1.00				
151	L	0.45	V	0.43	I	0.10
152	N	1.00	S	0.00		
153	S	0.55	A	0.38	G	0.04
154	G	0.57	S	0.08	A	0.06
155	V	0.33	I	0.15	L	0.15
156	F	0.34	Y	0.20	C	0.12
157	T	0.31	S	0.14	A	0.11
158	F	0.27	S	0.18	L	0.12
159	L	0.30	S	0.13	M	0.06
160	S	0.42	N	0.11	A	0.09
161	S	0.43	A	0.24	T	0.15
162	D	0.25	A	0.25	E	0.19
163	A	0.23	V	0.22	S	0.16
164	K	0.32	E	0.32	R	0.16
165	R	0.32	G	0.16	K	0.14
166	L	1.00	M	0.00	I	0.00
167	E	0.94	D	0.04	Q	0.01
168	D	0.69	N	0.17	E	0.07
169	H	0.16	Y	0.12	F	0.10
170	E	0.21	D	0.19	G	0.08
171	Q	0.38	K	0.34	R	0.19
172	V	0.63	I	0.36	M	0.00
173	Q	0.70	E	0.15	H	0.09
174	K	0.31	Q	0.23	E	0.16
175	L	0.54	I	0.27	M	0.17
176	Q	0.99	R	0.01	H	0.00
177	D	0.61	E	0.26	Q	0.09
178	K	0.19	N	0.18	T	0.15
179	I	0.83	V	0.11	L	0.06
180	T	0.42	H	0.33	N	0.07
181	D	0.78	E	0.22	N	0.00
182	A	1.00				
183	L	1.00				
184	V	0.30	I	0.12	Q	0.10
185	D	0.68	E	0.16	H	0.08
186	T	0.43	A	0.18	S	0.15
187	V	0.30	C	0.25	T	0.16
188	A	0.54	S	0.18	T	0.08
189	K	0.62	R	0.25	Q	0.04
190	S	0.29	R	0.11	N	0.10
191	H	0.84	Y	0.06	Q	0.05
192	P	0.26	L	0.13	S	0.09
193	D	0.26	E	0.25	N	0.14
194	S	0.35	N	0.20	P	0.17
195	P	0.98	S	0.01	A	0.00
196	Q	0.51	E	0.16	R	0.09
197	Q	0.50	R	0.08	K	0.07
198	S	0.48	P	0.12	A	0.08
199	R	0.87	K	0.08	Q	0.02
200	R	1.00				
201	L	0.67	F	0.13	I	0.11
202	A	0.98	S	0.01	G	0.00
203	Q	0.76	K	0.20	R	0.03
204	L	0.98	M	0.01	I	0.01
205	L	1.00	M	0.00		
206	M	0.51	L	0.48	I	0.00
207	L	0.86	I	0.07	M	0.05
208	L	1.00	M	0.00		
209	S	0.77	P	0.23	T	0.00
210	H	0.87	Q	0.11	E	0.00
211	I	0.68	V	0.23	L	0.07
212	R	1.00				
213	Q	0.98	H	0.01	E	0.01
214	V	0.59	I	0.27	M	0.07
215	S	0.98	A	0.02	T	0.00
216	S	0.27	N	0.26	T	0.21
217	K	0.57	R	0.42	Q	0.01
218	G	0.88	A	0.11	S	0.01
219	I	0.76	V	0.17	M	0.06
220	E	0.78	D	0.13	Q	0.06
221	H	1.00	Y	0.00		
222	L	0.97	F	0.03	I	0.00
223	Y	0.91	F	0.08	H	0.01
224	S	0.49	N	0.18	K	0.07
225	M	0.47	I	0.44	V	0.05
226	K	0.96	R	0.04	Q	0.00
227	S	0.37	C	0.15	N	0.10
228	E	0.78	A	0.07	Q	0.07
229	G	0.69	N	0.15	S	0.11
230	R	0.20	K	0.18	A	0.13
231	V	1.00				
232	P	1.00	S	0.00		
233	L	0.84	M	0.10	F	0.03
234	Y	0.70	H	0.26	C	0.01
235	D	0.96	N	0.02	E	0.02
236	L	1.00				
237	L	1.00	M	0.00		
238	L	0.84	S	0.04	M	0.04
239	E	1.00				
240	M	1.00	I	0.00		
241	L	1.00	I	0.00	V	0.00
242	D	0.82	E	0.18	N	0.00
243	A	1.00				
244	Q	0.89	H	0.09	K	0.01
245	T	0.35	S	0.11	N	0.10
246	S	0.41	P	0.37	A	0.06
247	Q	0.20	H	0.14	A	0.13
248	S	0.30	A	0.27	T	0.17
249	P	0.13	S	0.10	A	0.09

Mean Posterior Probability 0.69

Table 2.2: Reconstructed sequence of ancSR1

The reconstructed sequence of ancSR1 and associated posterior probability (PP) values.

	AncSR1	AncSR2
AncSR1	100%	
hERa	61.80%	24.80%
hERb	56.70%	25.70%
hAR	25.60%	62.60%
hPR	29.80%	66.50%
hGR	29.80%	64.80%
hMR	29%	71.70%
AncSR2	30.70%	100%

Table 2.3: ancSR1 and ancSR2 percent similarities

Percent similarity of the ligand-binding domains of ancSR1 and ancSR2 to those of extant steroid receptors in humans.

Number in text	Steroid Name in Text	Abbreviation	Pubmed Compound Identifier Number (CID)	Steraloids Catalogue #	Steraloids name
1	11-Deoxycorticosterone	11-DOC	6166		
2	11-Deoxycortisol		440707		
3	Corticosterone		5753		
4	Cortisol		5754		
5	Aldosterone		5839		
6	Progesterone		5994		
7	17-Hydroxyprogesterone		6238		
8	19-Norprogesterone		228864		
9	4-Pregnenolone		-	Q3540-000	4-PREGNEN-3β-OL-20-ONE
10	5-Pregnenolone		8955		
11	20α Hydroxyprogesterone		440204		
12	20β Hydroxyprogesterone		92747		
13	Testosterone		6013		
14	Dihydrotestosterone	DHT	10635		
15	4-Androstenediol		136297		
16	5-Androstenediol		10634		
17	19-Nortestosterone		9904		
18	Bolandiol		16141		
19	Estradiol		5757		
20	Estrone		5870		
21	Estriol		5756		
22	4-Androstenedione		6128		
23	19-nor-1, 3, 5(10)-pregnatriene-3-ol-20-one	NPT	-	N0600-000	19-NOR-1, 3, 5(10)-PREGNATRIEN-3-OL-20-ONE
0	Cholesterol		5997		
	Diethylstilbestrol	DES	448537		
	Genistein		5280961		
	4-Hydroxytamoxifen		449459		
	ICI 182,780	fulvestrant	104741		

Table 2.4: CID numbers for synthetic and natural steroids used in this study

Pubmed compound identifier (CID) numbers for cholesterol and the synthetic and natural steroid hormones tested in this study. For the two substances that a CID number was not available, the Steraloids catalogue number and name are provided.

		Hormone 1	EC_{50} (nM)	Hormone 2	EC_{50} (nM)	Ratio H1/H2	Ratio H2/H1
b	preference for aromatized A-ring (H1/H2)	18	1765	19	49	36.02	0.03
		22	>10 000	20	69	144.93	0.01
		8	>10 000	23	179	55.87	0.02
c	preference for non-aromatized A-ring (H1/H2)	23	182	8	1.6	113.75	0.01
		19	>10 000	18	20	500.00	0.00
d	preference for 17-acetyl vs. 17-hydroxy group (H1/H2)	13	116	6	5.5	21.09	0.05
		17	69	8	1.6	43.13	0.02
		15	262	9	2.2	119.09	0.01
		16	83	10	3.8	21.84	0.05
		19	>10 000	23	182	54.95	0.02
e	preference for 3-hydroxy vs. 3-keto group (H2/H1)	18	20	17	69	0.29	3.45
		9	2.2	6	5.5	0.40	2.50
		15	262	13	116	2.26	0.44
f	preference for 21-OH group vs. H group (H2/H1)	1	0.42	6	5.5	0.08	13.10
		2	1.5	7	19	0.08	12.67
g	preference for 11-OH vs. H group (H1/H2)	1	0.42	3	0.31	1.35	0.74
		2	1.5	4	9.8	0.15	6.53
h	preference for 17a-H group vs. -OH group (H2/H1)	3	0.31	4	9.8	0.03	31.61
		6	5.5	7	19	0.29	3.45
i	preference for 19-H vs. 19-methyl group (H2/H1)	8	1.6	6	5.5	0.29	3.44
		17	69	13	116	0.59	1.68
		18	20	15	262	0.08	13.10

1	11-deoxycorticosterone
2	11-deoxycortisol
3	corticosterone
4	cortisol
5	aldosterone
6	progesterone
7	17a-hydroxyprogesterone
8	19-norprogesterone
9	4-pregnenolone
10	5-pregnenolone
11	20a hydroxyprogesterone
12	20b hydroxyprogesterone
13	testosterone
14	dihydrotestosterone
15	4-androstenediol
16	5-androstenediol
17	19-nortestosterone
18	bolandiol
19	estradiol
20	estrone
21	estriol
22	4-androstenedione
23	NPT

Table 2.5: Fold preferences for hormone pairs

Fold preferences of ancSR1 and ancSR2 for the hormone pairs indicated in Figure 2.

Data Collection and Refinement Statistics		
	AncSR2-Progesterone	AncSR2-11-DOC
Resolution (Å)	2.75(2.85-2.75)	2.82 (2.92-2.82)
Space Group	$P2_12_12_1$	$C222_1$
Unit Cell Dimensions	53.47, 112.11, 132.85	52.80, 111.62, 130.77
a, b, c (Å)	90, 90, 90	90, 90, 90
α, β, γ (°)		
No. of Reflections	20584	9198
R^a_{sym}	8.9% (44.3%)	7.1% (34.5%)
Completeness	99.4% (96.1%)	92.6% (70.0%)
Ave. Redundancy	6.8 (5.2)	3.9 (3.1)
I/σ	25.1 (3.5)	19.4 (3.1)
Monomers per asymmetric unit (AU)	2	1
No. of protein atoms/AU	4228	2069
No. of ligand atoms/AU	2	2
No. of waters/AU	65	31
$R^b_{working}$ (R^c_{free})	23.3 (29.1)	23.1 (30.6)
Ave. B-factors (Å^2)		
Protein	63.06	72.12
Ligand	55.63	70.04
Water	63.21	68.44
r.m.s. deviations		
Bond lengths, Å	0.014	0.009
Bond angles, °	1.792	1.298

[a] $R_{sym} = \Sigma|I- \langle I\rangle|/ \Sigma|I|$, where I is the observed intensity and <I> is the average intensity of several symmetry-related observations.
[b] $R_{working} = \Sigma||Fo|-|Fc||/ \Sigma|Fo$, where Fo and Fc are the observed and calculated structure factors, respectively.
[c] $R_{free} = \Sigma||Fo|-|Fc||/ \Sigma|Fo$ for 7% of the data not used at any stage of the structural refinement.
*Highest resolution shell is shown in parentheses.

Table 2.6: Data collection and refinement statistics

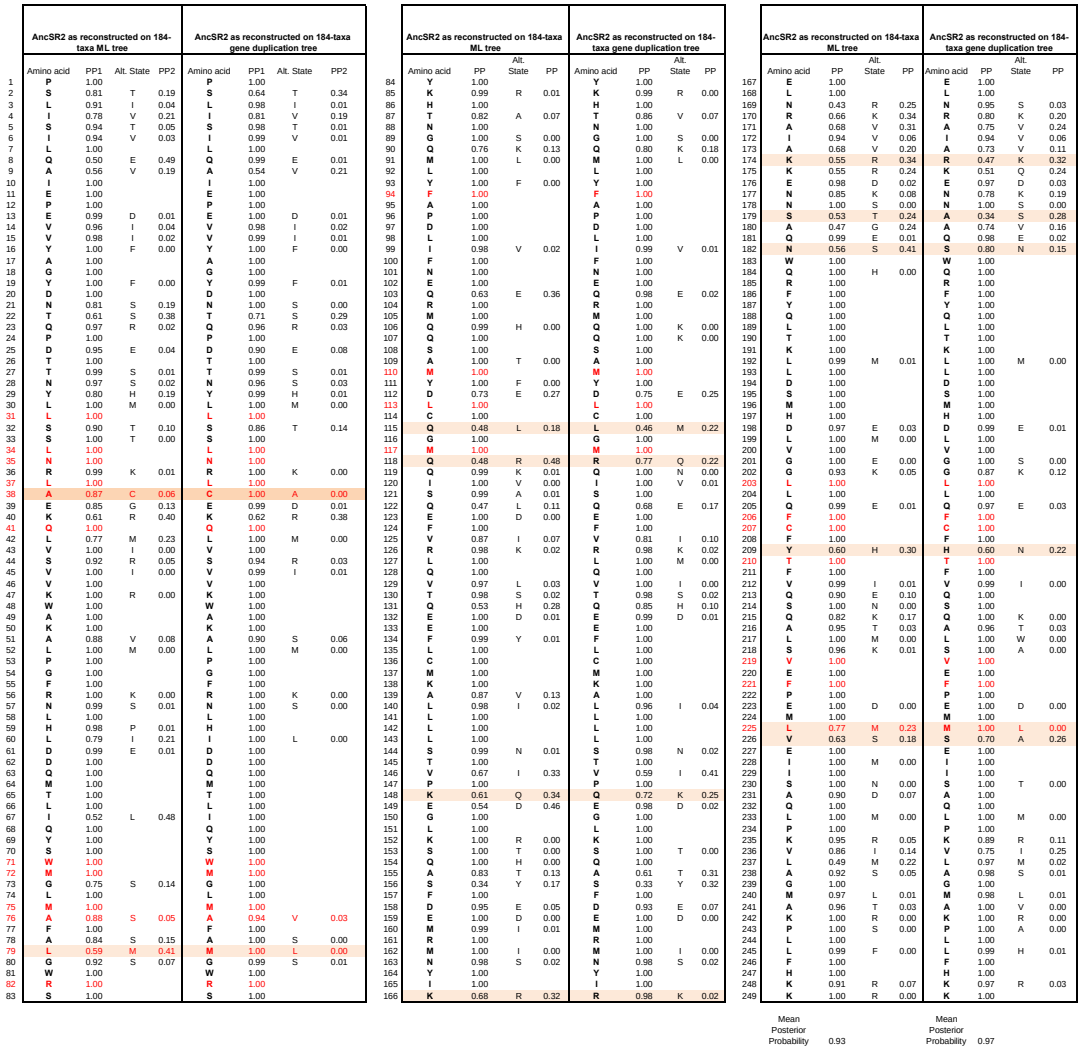
Data collection and refinement statistics for the ancSR2 crystal structure in complex with 11-deoxycorticosterone (11-DOC) and progesterone.

Species Name	Common Name	Phylum	Receptor Type	Abbreviation	Accession #	Database
Acanthopagrus schlegeli	black porgy	Vertebrata, Teleostei	AR	AcaSchAR	Q800S7	UniPROTKB/TrEMBL
			ERα	AcaSchERa	Q804Q6	UniPROTKB/TrEMBL
			ERβ	AcaSchERb	Q804Q7	UniPROTKB/TrEMBL
Alligator mississippiensis	American alligator	Vertebrata, Crocodylidae	ERα	AlMisERa	Q76SN7	UniPROTKB/TrEMBL
			PR	AlMisPR	Q76SN5	UniPROTKB/TrEMBL
Anguilla japonica	Japanese eel	Vertebrata, Teleostei	ARa	AngJapARa	Q2YGV9	UniPROTKB/TrEMBL
			ARb	AngJapARb	Q9PWG5	UniPROTKB/TrEMBL
			ERα	AngJapERa	Q13012	UniPROTKB/TrEMBL
			ERβ	AngJapERb	O13012	UniPROTKB/TrEMBL
Aplysia californica	sea slug	Mollusca, Gastropoda	ER	ApCalER	Q6VU64	UniPROTKB/TrEMBL
Bos taurus	cow	Vertebrata, Certartiodactyla	ERα	BosTauERa	P49884	UniPROTKB/TrEMBL
			ERβ	BosTauERb	Q9XSB5	UniPROTKB/TrEMBL
			GR	BosTauGR	Q865Y6	UniPROTKB/TrEMBL
Branchiostoma floridae	amphioxus	Vertebrata, Cephalochordata	ER	BraFlaER	B5L586	UniPROTKB/TrEMBL
			ERR	BraFlaERR	Q5XTR0	UniPROTKB/TrEMBL
			SR	BraFlaSR	B5SVV7	UniPROTKB/TrEMBL
Capitella capitata	bristle worm	Annelida, Polychaeta	ER	CapCapER	C3ZRV2	UniPROTKB/TrEMBL
Candidia barbatus	Taiwanese cyprinid fish	Vertebrata, Teleostei	ERα	CanBarERa	Q7T3U5	UniPROTKB/TrEMBL
Canis familaris	dog	Vertebrata, Carnivora	AR	CanFamAR	Q9TT90	UniPROTKB/TrEMBL
			ERα	CanFamERa	AAS20260	Genbank
			GR	CanFamGR	Q3HTT7	UniPROTKB/TrEMBL
			MR	CanFamMR	Q3HTT9	UniPROTKB/TrEMBL
			PR	CanFamPR	Q9GLW0	UniPROTKB/TrEMBL
Carassius auratus	goldfish	Vertebrata, Teleostei	AR	CarAurAR	Q9GFV2	UniPROTKB/TrEMBL
			ERα	CarAurERa	Q90WV1	UniPROTKB/TrEMBL
			ERβ1	CarAurERb1	Q9W669	UniPROTKB/TrEMBL
			ERβ2	CarAurERb2	Q9PLL9	UniPROTKB/TrEMBL
Cavia porcellus	guinea pig	Vertebrata, Euarchontoglires	ERα	CavPorERa	Q8CGK8	UniPROTKB/TrEMBL
			ERβ	CavPorERb	Q8CGK9	UniPROTKB/TrEMBL
			GR	CavPorGR	P49115	UniPROTKB/TrEMBL
Clarias gariepinus	African catfish	Vertebrata, Teleostei	ERα	ClaGarERa	Q90WH6	UniPROTKB/TrEMBL
Cnemidophorus uniparens	whip-tailed lizard	Vertebrata, Squamata	ERα	CneUniERa	Q8UWA9	UniPROTKB/TrEMBL
			PR	CneUniPR	Q91425	UniPROTKB/TrEMBL
Conger myriaster	conger eel	Vertebrata, Teleostei	ERβ	ConMyrERb	Q76ZD8	UniPROTKB/TrEMBL
Coturnix coturnix	common quail	Vertebrata, Aves	ERα	CotCotERa	Q8AYH0	UniPROTKB/TrEMBL
			ERβ	CotCotERb	Q63511	UniPROTKB/TrEMBL
Coturnix japonica	Asian quail	Vertebrata, Aves	AR	CotJapAR	Q689Y8	UniPROTKB/TrEMBL
Crassostrea gigas	Pacific oyster	Mollusca, Bivalvia	ER	CraGigER	A2A113	UniPROTKB/TrEMBL
Crocuta crocuta	hyena	Vertebrata, Carnivora	AR	CroCroAR	Q8MIK0	UniPROTKB/TrEMBL
Cyprinus carpio	common carp	Vertebrata, Teleostei	ERβ	CypCarERb	Q8JJB9	UniPROTKB/TrEMBL
Danio rerio	zebrafish	Vertebrata, Teleostei	AR	DanRerAR	Q6EAL5	UniPROTKB/TrEMBL
			COUP-TFII	DanRerCP2	O42532	UniPROTKB/TrEMBL
			ERα	DanRerERa	P57717	UniPROTKB/TrEMBL
			ERα2	DanRerER2a	Q7ZU32	UniPROTKB/TrEMBL
			ERβ1	DanRerERb1	Q90WS9	UniPROTKB/TrEMBL
			ERRα	DanRerERRa	Q6Q6F6	UniPROTKB/TrEMBL
			ERRβ	DanRerERRb	Q6Q6F5	UniPROTKB/TrEMBL
			ERRγ	DanRerERRg	Q6Q6F4	UniPROTKB/TrEMBL
			GR	DanRerGR	Q1XHK0	UniPROTKB/TrEMBL
			RXRα	DanRerRXRa	Q80416	UniPROTKB/TrEMBL
			RXRβ	DanRerRXRb	Q80417	UniPROTKB/TrEMBL
			RXRγ	DanRerRXRg	Q6DHP9	UniPROTKB/TrEMBL
Dicentrarchus labrax	European sea bass	Vertebrata, Teleostei	AR	DicLabAR	Q4G497	UniPROTKB/TrEMBL
			ERα	DicLabERa	Q6A4C2	UniPROTKB/TrEMBL
			ERβ	DicLabERb	Q6KF57	UniPROTKB/TrEMBL
Engystomops pustulosus	Lepodactylid frog	Vertebrata, Anura	AR	EngPusAR	Q2PKD4	UniPROTKB/TrEMBL
Equus caballus	horse	Vertebrata, Perissodactyla	ERα	EquCabERa	Q2TV98	UniPROTKB/TrEMBL
Eublepharis macularius	leopard gecko	Vertebrata, Squamata	ERα	EubMacERa	Q2HWK5	UniPROTKB/TrEMBL
			ERβ	EubMacERb	Q2HWK4	UniPROTKB/TrEMBL
Felis catus	cat	Vertebrata, Carnivora	ERα	FelCatERa	Q53AD2	UniPROTKB/TrEMBL
Fundulus heteroclitus	killfish	Vertebrata, Teleostei	ERα	FunHetERa	Q7SZ8I	UniPROTKB/TrEMBL
Gallus gallus	chicken	Vertebrata, Aves	COUP-TFII	GalGalCP2	Q60733	UniPROTKB/TrEMBL
			ERa	GalGalERa	P06212	UniPROTKB/TrEMBL
			ERβ	GalGalERb	Q9PTU5	UniPROTKB/TrEMBL
			ERRg	GalGalERRg	Q5UKY7	UniPROTKB/TrEMBL
			GR	GalGalGR	Q38HX7	UniPROTKB/TrEMBL
			MR	GalGalMR	C1KBB4	UniPROTKB/TrEMBL
			PR	GalGalPR	P07812	UniPROTKB/TrEMBL
			SF1	GalGalSF1	Q5FWY7	UniPROTKB/TrEMBL
Gambusia affinis	western mosquitofish	Vertebrata, Teleostei	AR	GamAffAR	Q60I32	UniPROTKB/TrEMBL
			ARa	GamAffARa	Q60GT9	UniPROTKB/TrEMBL
Gasterosteus aculeatus	three-spined stickelback	Vertebrata, Teleostei	ARb	GasAcuARb	Q801Z2	UniPROTKB/TrEMBL
Halichoeres tenuispinis	wrasse	Vertebrata, Teleostei	ERα	HalTenERa	Q7T2K8	UniPROTKB/TrEMBL
			ERβ	HalTenERb	Q7T2K7	UniPROTKB/TrEMBL
Halichoeres trimaculatus	protogynous wrasse	Vertebrata, Teleostei	AR	HalTriAR	Q9DDJ4	UniPROTKB/TrEMBL
			ERα	HalTriERa	Q9DDJ3	UniPROTKB/TrEMBL
Haplochromis burtoni	African cichlid	Vertebrata, Teleostei	AR	HapBurAR	Q9W6F4	UniPROTKB/TrEMBL
			ARb	HapBurARb	Q8QFV7	UniPROTKB/TrEMBL
			ERα	HapBurERa	Q6TGB3	UniPROTKB/TrEMBL
			GR1	HapBurGR1	Q8JJ92	UniPROTKB/TrEMBL
			GR2	HapBur2aGR	Q8JJ91	UniPROTKB/TrEMBL
			MR	HapBurMR	Q8JJ89	UniPROTKB/TrEMBL
Homo sapiens	human	Vertebrata, Primates	AR	HomSapAR	P10275	UniPROTKB/TrEMBL
			COUP-TF1	HomSapCP1	P10589	UniPROTKB/TrEMBL
			COUP-TFII	HomSapCP2	P24468	UniPROTKB/TrEMBL
			ERα	HomSapERa	P03372	UniPROTKB/TrEMBL
			ERβ	HomSapERb	Q9UHD3	UniPROTKB/TrEMBL
			ERRα	HomSapERR1	P11474	UniPROTKB/TrEMBL
			ERRβ	HomSapERR2	O95718	UniPROTKB/TrEMBL
			ERRγ	HomSapERR3	P62508	UniPROTKB/TrEMBL
			GR	HomSapGR	P04150	UniPROTKB/TrEMBL
			MR	HomSapMR	P08235	UniPROTKB/TrEMBL
			PR	HomSapPR	P06401	UniPROTKB/TrEMBL
			RXRα	HomSapRXRa	P19793	UniPROTKB/TrEMBL
			RXRβ	HomSapRXRb	P28702	UniPROTKB/TrEMBL
			RXRγ	HomSapRXRg	P48443	UniPROTKB/TrEMBL
			SF1	HomSapSF1	Q13285	UniPROTKB/TrEMBL
Ictalurus punctatus	channel catfish	Vertebrata, Teleostei	ERa	IctPunERa	Q97YU7	UniPROTKB/TrEMBL
			ERβ	IctPunERb	Q6IAK1	UniPROTKB/TrEMBL
Kryptolebias marmoratus	mangrove rivulus	Vertebrata, Teleostei	AR	KryMarAR	Q1I9HT9	UniPROTKB/TrEMBL
			ERα	KryMarERa	Q0KKB8	UniPROTKB/TrEMBL
			ERβ	KryMarERb	Q0KKB9	UniPROTKB/TrEMBL
Leucoraja erinacea	little skate	Vertebrata, Elasmobranchii	AR	LeuErIAR	Q1KXY2	UniPROTKB/TrEMBL
			GR	LeuErIGR	Q1KXY4	UniPROTKB/TrEMBL
			MR	LeuErIMR	Q1KXY3	UniPROTKB/TrEMBL
			PR	LeuErIPR	Q1KXY1	UniPROTKB/TrEMBL

Species Name	Common Name	Phylum	Receptor Type	Abbreviation	Accession #	Database
Lottia gigantea	limpet	Mollusca, Gastropoda	ER	LotGigER	Lotgi1:132166	JGI
Menacus vitellinus	gold-collared manakin	Vertebrata, Aves	AR	ManVitAR	Q4U102	UniPROTKB/TrEMBL
Marisa cornuarietis	apple snail	Mollusca, Gastropoda	ER	MarCorER	Q56AL2	UniPROTKB/TrEMBL
Mesocricetus auratus	golden hamster	Vertebrata, Rodentia	ERα	MesAurERa	Q9QZJ5	UniPROTKB/TrEMBL
Microtus ochrogaster	prairie vole	Vertebrata, Rodentia	ERα	MicOchERa	Q1ED86	UniPROTKB/TrEMBL
Micropterus salmoides	largemouth bass	Vertebrata, Teleostei	ERα	MicSalERa	Q6D2D24	UniPROTKB/TrEMBL
			ERβ	MicSalERb	Q6XSH2	UniPROTKB/TrEMBL
Micropogonias undulatus	Atlantic croaker	Vertebrata, Teleostei	AR	MicUndAR	Q66VR6	UniPROTKB/TrEMBL
			ERβ	MicUndERb	P57781	UniPROTKB/TrEMBL
			ERγ	MicUndERg	P57783	UniPROTKB/TrEMBL
Mus musculus	mouse	Vertebrata, Rodentia	ERα	MusMusERa	P19785	UniPROTKB/TrEMBL
			ERβ	MusMusERb	O08537	UniPROTKB/TrEMBL
			ERRα	MusMusERR1	O08580	UniPROTKB/TrEMBL
			ERRβ	MusMusERR2	Q61539	UniPROTKB/TrEMBL
			ERRg	MusMusERR3	P62509	UniPROTKB/TrEMBL
			GR	MusMusGR	P06537	UniPROTKB/TrEMBL
			MR	MusMusMR	Q61IB	UniPROTKB/TrEMBL
			PR	MusMusPR	Q00175	UniPROTKB/TrEMBL
			SF1	MusMusSF1	P33242	UniPROTKB/TrEMBL
Myxine glutinosa	Atlantic hagfish	Vertebrata, Teleostei	CR	MyxGluCR	Q1KXY6	UniPROTKB/TrEMBL
			ER	MyxGluER	EU439936	Genbank
			SR	MyxGluSR	Q1KXY5	UniPROTKB/TrEMBL
Nucella lapillus	Atlantic dogwinkle	Mollusca, Gastropoda	ER	NucLapER	A5YYJ5	UniPROTKB/TrEMBL
Octopus vulgaris	octopus	Mollusca, Cephalopoda	ER	OctVulER	Q19AB0	UniPROTKB/TrEMBL
Oncorhynchus mykiss	rainbow trout	Vertebrata, Teleostei	AR	OncMykAR2	O93244	UniPROTKB/TrEMBL
			AR	OncMykARb2	O93245	UniPROTKB/TrEMBL
			ERα	OncMykERa	P16058	UniPROTKB/TrEMBL
			ERβ	OncMykERb	P57782	UniPROTKB/TrEMBL
			GR	OncMykGR	P49843	UniPROTKB/TrEMBL
			GR	OncMykGR2	Q6RK03	UniPROTKB/TrEMBL
			MR	OncMykMR	Q5WP02	UniPROTKB/TrEMBL
Onychostoma barbatula	cyprinid fish	Vertebrata, Teleostei	ERβ	OnyBarERb2	Q2PDJ6	UniPROTKB/TrEMBL
Oreochromis aureus	blued tilapia	Vertebrata, Teleostei	ERα	OreAurERa	P50240	UniPROTKB/TrEMBL
Orechromis niloticus	Nile tilapia	Vertebrata, Teleostei	AR	OreNilARa	Q8UWB8	UniPROTKB/TrEMBL
			AR	OreNilARb	Q8UWB7	UniPROTKB/TrEMBL
			ERα	OreNilERa	Q6YH33	UniPROTKB/TrEMBL
			ERβ	OreNilERb	Q7T3U4	UniPROTKB/TrEMBL
Oryctolagus cuniculus	rabbit	Vertebrata, Glires	AR	OrtCunAR	P49699	UniPROTKB/TrEMBL
			GR	OrtCunGR	P59667	UniPROTKB/TrEMBL
			PR	OrtCunPR	P06186	UniPROTKB/TrEMBL
Oryzias latipes	Japanese ricefish	Vertebrata, Teleostei	AR	OryLatAR	Q76LM5	UniPROTKB/TrEMBL
			ERα	OryLatERa	P50241	UniPROTKB/TrEMBL
			ERβ	OryLatERb	Q8UW75	UniPROTKB/TrEMBL
Ovis aries	sheep	Vertebrata, Ruminantia	ERα	OviAriERa	Q95L13	UniPROTKB/TrEMBL
			ERβ	OviAriERb	Q9TU15	UniPROTKB/TrEMBL
Pagrus major	red sea bream	Vertebrata, Teleostei	AR	PagMajAR	O93497	UniPROTKB/TrEMBL
			ERα	PagMajERa	O42132	UniPROTKB/TrEMBL
Paralichthys olivaceus	Japanses flounder	Vertebrata, Teleostei	ERα	ParOliERa	Q8QHL0	UniPROTKB/TrEMBL
			ERβ	ParOliERb	Q8QHK9	UniPROTKB/TrEMBL
			GR	ParOliGR	O73673	UniPROTKB/TrEMBL
Petromyzon marinus	sea lamprey	Vertebrata, Hyperoartia	COUP-TFII	PetMarCP2	Q42509	UniPROTKB/TrEMBL
			CR	PetMarCR	Q90ZM7	UniPROTKB/TrEMBL
			ER	PetMarER	Q90ZM8	UniPROTKB/TrEMBL
			PR	PetMarPR	Q90ZM6	UniPROTKB/TrEMBL
			RXR2	PetMarRXR2	Q80Y01	UniPROTKB/TrEMBL
Pimephales promelas	fathead minnow	Vertebrata, Teleostei	AR	PimProAR	Q8I8F5	UniPROTKB/TrEMBL
			ERα	PimProERa	Q5XXP1	UniPROTKB/TrEMBL
			GR	PimProGR	Q2TU59	UniPROTKB/TrEMBL
Platynereis dumerili	polychaete worm	Annelida, Polychaeta	ER	PlaDumER	EU482033	Genbank
Rana catesbeiana	bullfrog	Vertebrata, Anura	AR	RanCatAR	Q7T1K4	UniPROTKB/TrEMBL
Rana dybowskii	Korean brown frog	Vertebrata, Anura	PR	RanDybPR	Q8AYI2	UniPROTKB/TrEMBL
Rana rugosa	wrinkled frog	Vertebrata, Anura	SF1	RanRugSF1	Q9YI95	UniPROTKB/TrEMBL
Rattus norvegicus	rat	Vertebrata, Rodentia	AR	RatNovAR	P15207	UniPROTKB/TrEMBL
			ERβ	RatNovERb	Q62986	UniPROTKB/TrEMBL
			ERRβ	RatNovERR2	P11475	UniPROTKB/TrEMBL
			GR	RatNovGR	P06536	UniPROTKB/TrEMBL
			MR	RatNovMR	P22199	UniPROTKB/TrEMBL
			PR	RatNovPR	Q63449	UniPROTKB/TrEMBL
Saimiri boliviensis	squirrel monkey	Vertebrata, Primates	GR	SaiBolGR	O13186	UniPROTKB/TrEMBL
			MR	SaiBolMR	Q4JM28	UniPROTKB/TrEMBL
Salmo salar	Atlantic salmon	Vertebrata, Teleostei	ERβ	SalSalERb	Q6R7S4	UniPROTKB/TrEMBL
Salmo trutta	brown salmon	Vertebrata, Teleostei	GR	SalTruGR	Q5I1Z4	UniPROTKB/TrEMBL
Serinus canarius	canary	Vertebrata, Aves	AR	SerCanAR	Q91445	UniPROTKB/TrEMBL
Sparus aurata	gilthead seabream	Vertebrata, Teleostei	ERα	SpaAurERa	Q9PVZ9	UniPROTKB/TrEMBL
			ERβ	SpaAurERb	Q9W6M2	UniPROTKB/TrEMBL
			GR	SpaAurGR	Q1HGK6	UniPROTKB/TrEMBL
Spinibarbus denticulatus	cyprinid fish	Vertebrata, Teleostei	ERβ	SpiDenERb	Q1HCL3	UniPROTKB/TrEMBL
Squalus acanthias	spiny dogfish	Vertebrata, Elasmobranch	AR	SquAcaAR	Q56VU2	UniPROTKB/TrEMBL
			ERβ	SquAcaERb	Q90ZE6	UniPROTKB/TrEMBL
Sturnus vulgaris	European starling	Vertebrata, Aves	ERβ	StuVulERb	Q9PVE2	UniPROTKB/TrEMBL
Sus scrofa	pig	Vertebrata, Certartiodactyl	AR	SusScrAR	Q9GKL7	UniPROTKB/TrEMBL
			ERα	SusScrERa	Q29040	UniPROTKB/TrEMBL
			ERβ	SusScrERb	Q9XSW2	UniPROTKB/TrEMBL
			GR	SusScrGR	Q9N1U3	UniPROTKB/TrEMBL
Taeniopygia guttata	zebrafinch	Vertebrata, Aves	AR	TaeGutAR	Q2VP86	UniPROTKB/TrEMBL
			ERα	TaeGutERa	Q91250	UniPROTKB/TrEMBL
			MR	TaeGutMR	Q157N1	UniPROTKB/TrEMBL
			SF1	TaeGutSF1	Q90XD4	UniPROTKB/TrEMBL
Takifugu rubripes	pufferfish	Vertebrata, Teleostei	AR	TakRubAR	NEWSINFRUP000001563 18	Ensembl
Tetraodon nigrovirids	green-spotted pufferfish	Vertebrata, Teleostei	ERα	TetNigERa	Q4S6Y6	UniPROTKB/TrEMBL
Thais clavigera	whelk	Mollusca, Gastropoda	ER	ThaClaER	Q86827	UniPROTKB/TrEMBL
Trichoplax adhaerens		Placozoa	ER	TriAdhERR	Triad1:16718	JGI
Tupaia glis	common tree shrew	Vertebrata, Scandentia	GR	TupGliGR	Q95267	UniPROTKB/TrEMBL
			MR	TupGliMR	Q29131	UniPROTKB/TrEMBL
Xenopus laevis	African clawed toad	Vertebrata, Anura	AR	XenLaeAR	P70048	UniPROTKB/TrEMBL
			CP1	XenLaeCP1	Q9W745	UniPROTKB/TrEMBL
			ERα	XenLaeERa	P81559	UniPROTKB/TrEMBL
			ERR	XenLaeERR	Q6AX97	UniPROTKB/TrEMBL
			GR	XenLaeGR	P49844	UniPROTKB/TrEMBL
			MR	XenLaeMR	Q66J29	UniPROTKB/TrEMBL
			PR	XenLaePR	Q9DDU9	UniPROTKB/TrEMBL
			RXRα	XenLaeRXRa	P51128	UniPROTKB/TrEMBL
			RXRβ	XenLaeRXRb	Q91840	UniPROTKB/TrEMBL
			RXRγ	XenLaeRXRg	P51129	UniPROTKB/TrEMBL
Xenopus tropicalis	clawed toad	Vertebrata, Anura	ERα	XenTroERa	Q6WSG9	UniPROTKB/TrEMBL
			ERβ	XenTroERb	Q6WSG8	UniPROTKB/TrEMBL
Zoarces viviparus	Eelpout	Vertebrata, Teleostei	ERα	ZoaViviERa	Q8DIQ2	UniPROTKB/TrEMBL

Table 2.7: Receptors and organisms used for phylogenetic analyses

List of receptors and the organisms they were isolated from used in the phylogenetic analyses. Receptors shaded in grey are those that were additionally incorporated into the phylogeny used to reconstruct ancSR2. ERs, estrogen receptors; PRs, progesterone receptors; ARs, androgen receptors; MRs; mineralocorticoid receptors; GRs, glucocorticoid receptors; ERRs, estrogen-related receptors; SF1, steroidogenic factor 1 receptor; RXR, retinoid X receptor; COUP-TFs, chicken ovalbumin upstream promoter transcription factors.



	AncSR2 as reconstructed on 184-taxa ML tree				AncSR2 as reconstructed on 184-taxa gene duplication tree			
	Amino acid	PP1	Alt. State	PP2	Amino acid	PP1	Alt. State	PP2
1	P	1.00			P	1.00		
2	S	0.81	T	0.19	S	0.64	T	0.34
3	L	0.91	I	0.04	L	0.98	I	0.01
4	I	0.78	V	0.21	I	0.81	V	0.19
5	S	0.94	T	0.05	S	0.98	T	0.01
6	I	0.94	V	0.03	I	0.99	V	0.01
7	L	1.00			L	1.00		
8	Q	0.50	E	0.49	Q	0.99	E	0.01
9	A	0.56	V	0.19	A	0.54	V	0.21
10	I	1.00			I	1.00		
11	E	1.00			E	1.00		
12	P	1.00			P	1.00		
13	E	0.99	D	0.01	E	1.00	D	0.01
14	V	0.96	I	0.04	V	0.98	I	0.02
15	V	0.98	I	0.02	V	0.99	I	0.01
16	Y	1.00	F	0.00	Y	1.00	F	0.00
17	A	1.00			A	1.00		
18	G	1.00			G	1.00		
19	Y	1.00	F	0.00	Y	0.99	F	0.01
20	D	1.00			D	1.00		
21	N	0.81	S	0.19	N	1.00	S	0.00
22	T	0.61	S	0.38	T	0.71	S	0.29
23	Q	0.97	R	0.02	Q	0.96	R	0.03
24	P	1.00			P	1.00		
25	D	0.95	E	0.04	D	0.90	E	0.08
26	T	1.00			T	1.00		
27	T	0.99	S	0.01	T	0.99	S	0.01
28	N	0.97	S	0.02	N	0.96	S	0.03
29	Y	0.80	H	0.19	Y	0.99	H	0.01
30	L	1.00	M	0.00	L	1.00	M	0.00
31	L	1.00			L	1.00		
32	S	0.90	T	0.10	S	0.86	T	0.14
33	S	1.00	T	0.00	S	1.00		
34	L	1.00			L	1.00		
35	N	1.00			N	1.00		
36	R	0.99	K	0.01	R	1.00	K	0.00
37	L	1.00			L	1.00		
38	A	0.87	C	0.06	C	1.00	A	0.00
39	E	0.85	G	0.13	E	0.99	D	0.01
40	K	0.61	R	0.40	K	0.62	R	0.38
41	Q	1.00			Q	1.00		
42	L	0.77	M	0.23	L	1.00	M	0.00
43	V	1.00	I	0.00	V	1.00		
44	S	0.92	R	0.05	S	0.94	R	0.03
45	V	1.00	I	0.00	V	0.99	I	0.01
46	V	1.00			V	1.00		
47	K	1.00	R	0.00	K	1.00		
48	W	1.00			W	1.00		
49	A	1.00			A	1.00		
50	K	1.00			K	1.00		
51	A	0.88	V	0.08	A	0.90	S	0.06
52	L	1.00	M	0.00	L	1.00	M	0.00
53	P	1.00			P	1.00		
54	G	1.00			G	1.00		
55	F	1.00			F	1.00		
56	R	1.00	K	0.00	R	1.00	K	0.00
57	N	0.99	S	0.01	N	1.00	S	0.00
58	L	1.00			L	1.00		
59	H	0.98	P	0.01	H	1.00		
60	L	0.79	I	0.21	I	1.00	L	0.00
61	D	0.99	E	0.01	D	1.00		
62	D	1.00			D	1.00		
63	Q	1.00			Q	1.00		
64	M	1.00			M	1.00		
65	T	1.00			T	1.00		
66	L	1.00			L	1.00		
67	I	0.52	L	0.48	I	1.00		
68	Q	1.00			Q	1.00		
69	Y	1.00			Y	1.00		
70	S	1.00			S	1.00		
71	W	1.00			W	1.00		
72	M	1.00			M	1.00		
73	G	0.75	S	0.14	G	1.00		
74	L	1.00			L	1.00		
75	M	1.00			M	1.00		
76	A	0.88	S	0.05	A	0.94	V	0.03
77	F	1.00			F	1.00		
78	A	0.84	S	0.15	A	1.00	S	0.00
79	L	0.59	M	0.41	M	1.00	L	0.00
80	G	0.92	S	0.07	G	0.99	S	0.01
81	W	1.00			W	1.00		
82	R	1.00			R	1.00		
83	S	1.00			S	1.00		

	AncSR2 as reconstructed on 184-taxa ML tree				AncSR2 as reconstructed on 184-taxa gene duplication tree			
	Amino acid	PP	Alt. State	PP	Amino acid	PP	Alt. State	PP
84	Y	1.00			Y	1.00		
85	K	0.99	R	0.01	K	0.99	R	0.00
86	H	1.00			H	1.00		
87	T	0.82	A	0.07	T	0.86	V	0.07
88	N	1.00			N	1.00		
89	G	1.00	S	0.00	G	1.00	S	0.00
90	Q	0.76	K	0.13	Q	0.80	K	0.18
91	M	1.00	L	0.00	M	1.00	L	0.00
92	L	1.00			L	1.00		
93	Y	1.00	F	0.00	Y	1.00		
94	F	1.00			F	1.00		
95	A	1.00			A	1.00		
96	P	1.00			P	1.00		
97	D	1.00			D	1.00		
98	L	1.00			L	1.00		
99	I	0.98	V	0.02	I	0.99	V	0.01
100	F	1.00			F	1.00		
101	N	1.00			N	1.00		
102	E	1.00			E	1.00		
103	Q	0.63	E	0.36	Q	0.98	E	0.02
104	R	1.00			R	1.00		
105	M	1.00			M	1.00		
106	Q	0.99	H	0.00	Q	1.00	K	0.00
107	Q	1.00			Q	1.00	K	0.00
108	S	1.00			S	1.00		
109	A	1.00	T	0.00	A	1.00		
110	M	1.00			M	1.00		
111	Y	1.00	F	0.00	Y	1.00		
112	D	0.73	E	0.27	D	0.75	E	0.25
113	L	1.00			L	1.00		
114	C	1.00			C	1.00		
115	Q	0.48	L	0.18	L	0.46	M	0.22
116	G	1.00			G	1.00		
117	M	1.00			M	1.00		
118	Q	0.48	R	0.48	R	0.77	Q	0.22
119	Q	0.99	K	0.01	Q	1.00	N	0.00
120	I	1.00	V	0.00	I	1.00	V	0.01
121	S	0.99	A	0.01	S	1.00		
122	Q	0.47	L	0.11	Q	0.68	E	0.17
123	E	1.00	D	0.00	E	1.00		
124	F	1.00			F	1.00		
125	V	0.87	I	0.07	V	0.81	I	0.10
126	R	0.98	K	0.02	R	0.98	K	0.02
127	L	1.00			L	1.00	M	0.00
128	Q	1.00			Q	1.00		
129	V	0.97	L	0.03	V	1.00	I	0.00
130	T	0.98	S	0.02	T	0.98	S	0.02
131	Q	0.53	H	0.28	Q	0.85	H	0.10
132	E	1.00	D	0.01	E	0.99	D	0.01
133	E	1.00			E	1.00		
134	F	0.99	Y	0.01	F	1.00		
135	L	1.00			L	1.00		
136	C	1.00			C	1.00		
137	M	1.00			M	1.00		
138	K	1.00			K	1.00		
139	A	0.87	V	0.13	A	1.00		
140	L	0.98	I	0.02	L	0.96	I	0.04
141	L	1.00			L	1.00		
142	L	1.00			L	1.00		
143	L	1.00			L	1.00		
144	S	0.99	N	0.01	S	0.98	N	0.02
145	T	1.00			T	1.00		
146	V	0.67	I	0.33	V	0.59	I	0.41
147	P	1.00			P	1.00		
148	K	0.61	Q	0.34	Q	0.72	K	0.25
149	E	0.54	D	0.46	E	0.98	D	0.02
150	G	1.00			G	1.00		
151	L	1.00			L	1.00		
152	K	1.00	R	0.00	K	1.00		
153	S	1.00	T	0.00	S	1.00	T	0.00
154	Q	1.00	H	0.00	Q	1.00		
155	A	0.83	T	0.13	A	0.61	T	0.31
156	S	0.34	Y	0.17	S	0.33	Y	0.32
157	F	1.00			F	1.00		
158	D	0.95	E	0.05	D	0.93	E	0.07
159	E	1.00	D	0.00	E	1.00	D	0.00
160	M	0.99	I	0.01	M	1.00		
161	R	1.00			R	1.00		
162	M	1.00	I	0.00	M	1.00	I	0.00
163	N	0.98	S	0.02	N	0.98	S	0.02
164	Y	1.00			Y	1.00		
165	I	1.00			I	1.00		
166	K	0.68	R	0.32	R	0.98	K	0.02

	AncSR2 as reconstructed on 184-taxa ML tree				AncSR2 as reconstructed on 184-taxa gene duplication tree			
	Amino acid	PP	Alt. State	PP	Amino acid	PP	Alt. State	PP
167	E	1.00			E	1.00		
168	L	1.00			L	1.00		
169	N	0.43	R	0.25	N	0.95	S	0.03
170	R	0.66	K	0.34	R	0.80	K	0.20
171	A	0.68	V	0.31	A	0.75	V	0.24
172	I	0.94	V	0.06	I	0.94	V	0.06
173	A	0.68	V	0.20	A	0.73	V	0.11
174	K	0.55	R	0.34	R	0.47	K	0.32
175	K	0.55	R	0.24	K	0.51	Q	0.24
176	E	0.98	D	0.02	E	0.97	D	0.03
177	N	0.85	K	0.08	N	0.78	K	0.19
178	N	1.00	S	0.00	N	1.00	S	0.00
179	S	0.53	T	0.24	A	0.34	S	0.28
180	A	0.47	G	0.24	A	0.74	V	0.16
181	Q	0.99	E	0.01	Q	0.98	E	0.02
182	N	0.56	S	0.41	S	0.80	N	0.15
183	W	1.00			W	1.00		
184	Q	1.00	H	0.00	Q	1.00		
185	R	1.00			R	1.00		
186	F	1.00			F	1.00		
187	Y	1.00			Y	1.00		
188	Q	1.00			Q	1.00		
189	L	1.00			L	1.00		
190	T	1.00			T	1.00		
191	K	1.00			K	1.00		
192	L	0.99	M	0.01	L	1.00	M	0.00
193	L	1.00			L	1.00		
194	D	1.00			D	1.00		
195	S	1.00			S	1.00		
196	M	1.00			M	1.00		
197	H	1.00			H	1.00		
198	D	0.97	E	0.03	D	0.99	E	0.01
199	L	1.00	M	0.00	L	1.00		
200	V	1.00			V	1.00		
201	G	1.00	E	0.00	G	1.00	S	0.00
202	G	0.93	K	0.05	G	0.87	K	0.12
203	L	1.00			L	1.00		
204	L	1.00			L	1.00		
205	Q	0.99	E	0.01	Q	0.97	E	0.03
206	F	1.00			F	1.00		
207	C	1.00			C	1.00		
208	F	1.00			F	1.00		
209	Y	0.60	H	0.30	H	0.60	N	0.22
210	T	1.00			T	1.00		
211	F	1.00			F	1.00		
212	V	0.99	I	0.01	V	0.99	I	0.00
213	Q	0.90	E	0.10	Q	1.00		
214	S	1.00	N	0.00	S	1.00		
215	Q	0.82	K	0.17	Q	1.00	K	0.00
216	A	0.95	T	0.03	A	0.96	T	0.03
217	L	1.00	M	0.00	L	1.00	W	0.00
218	S	0.96	K	0.01	S	1.00	A	0.00
219	V	1.00			V	1.00		
220	E	1.00			E	1.00		
221	F	1.00			F	1.00		
222	P	1.00			P	1.00		
223	E	1.00	D	0.00	E	1.00	D	0.00
224	M	1.00			M	1.00		
225	L	0.77	M	0.23	M	1.00	L	0.00
226	V	0.63	S	0.18	S	0.70	A	0.26
227	E	1.00			E	1.00		
228	I	1.00	M	0.00	I	1.00		
229	I	1.00			I	1.00		
230	S	1.00	N	0.00	S	1.00	T	0.00
231	A	0.90	D	0.07	A	1.00		
232	Q	1.00			Q	1.00		
233	L	1.00	M	0.00	L	1.00	M	0.00
234	P	1.00			P	1.00		
235	K	0.95	R	0.05	K	0.89	R	0.11
236	V	0.86	I	0.14	V	0.75	I	0.25
237	L	0.49	M	0.22	L	0.97	M	0.02
238	A	0.92	S	0.05	A	0.98	S	0.01
239	G	1.00			G	1.00		
240	M	0.97	L	0.01	M	0.98	L	0.01
241	A	0.96	T	0.03	A	1.00	V	0.00
242	K	1.00	R	0.00	K	1.00	R	0.00
243	P	1.00	S	0.00	P	1.00	A	0.00
244	L	1.00			L	1.00		
245	L	0.99	F	0.00	L	0.99	H	0.01
246	F	1.00			F	1.00		
247	H	1.00			H	1.00		
248	K	0.91	R	0.07	K	0.97	R	0.03
249	K	1.00	R	0.00	K	1.00		
	Mean Posterior Probability	0.93			Mean Posterior Probability	0.97		

Table 2.8: ancSR2 sequence comparison

Comparison of the sequence of ancSR2 as reconstructed on the ML phylogeny and gene duplication phylogeny (see Figs 2.14, 2.16). Residues in the ligand binding pocket are indicated in red text. Differences between the two ancestral reconstructions are highlighted in orange.

References

1. B. Alberts, *Molecular biology of the cell*. (Garland Science, New York, ed. 4th, 2002), pp. xxxiv, 1548 p.
2. R. P. Bhattacharyya, A. Remenyi, B. J. Yeh, W. A. Lim, Domains, motifs, and scaffolds: the role of modular interactions in the evolution and wiring of cell signaling circuits. *Annual review of biochemistry* **75**, 655 (2006).
3. R. Rohs *et al.*, Origins of specificity in protein-DNA recognition. *Annual review of biochemistry* **79**, 233 (2010).
4. R. J. Lefkowitz, The superfamily of heptahelical receptors. *Nature cell biology* **2**, E133 (Jul, 2000).
5. S. D. Copley, Enzymes with extra talents: moonlighting functions and catalytic promiscuity. *Current opinion in chemical biology* **7**, 265 (Apr, 2003).
6. R. A. Jensen, Enzyme recruitment in evolution of new function. *Annual review of microbiology* **30**, 409 (1976).
7. P. J. O'Brien, D. Herschlag, Catalytic promiscuity and the evolution of new enzymatic activities. *Chemistry & biology* **6**, R91 (Apr, 1999).
8. O. Khersonsky, D. S. Tawfik, Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annual review of biochemistry* **79**, 471 (2010).
9. D. S. Tawfik, Messy biology and the origins of evolutionary innovations. *Nature chemical biology* **6**, 692 (Oct, 2010).
10. Y. Yoshikuni, T. E. Ferrin, J. D. Keasling, Designed divergent evolution of enzyme function. *Nature* **440**, 1078 (Apr 20, 2006).
11. D. A. Liberles, M. D. Tisdell, J. A. Grahnen, Binding constraints on the evolution of enzymes and signalling proteins: the important role of negative pleiotropy. *Proceedings. Biological sciences / The Royal Society* **278**, 1930 (Jul 7, 2011).
12. S. Bershtein, K. Goldin, D. S. Tawfik, Intense neutral drifts yield robust and evolvable consensus proteins. *Journal of molecular biology* **379**, 1029 (Jun 20, 2008).
13. J. D. Bloom, P. A. Romero, Z. Lu, F. H. Arnold, Neutral genetic drift can alter promiscuous protein functions, potentially aiding functional evolution. *Biology direct* **2**, 17 (2007).

14. J. D. Bloom, F. H. Arnold, In the light of directed evolution: pathways of adaptive protein evolution. *Proceedings of the National Academy of Sciences of the United States of America* **106 Suppl 1**, 9995 (Jun 16, 2009).
15. M. J. Keiser *et al.*, Predicting new molecular targets for known drugs. *Nature* **462**, 175 (Nov 12, 2009).
16. H. Gronemeyer, J. A. Gustafsson, V. Laudet, Principles for modulation of the nuclear receptor superfamily. *Nature reviews. Drug discovery* **3**, 950 (Nov, 2004).
17. G. N. Eick, J. W. Thornton, Evolution of steroid receptors from an estrogen-sensitive ancestral receptor. *Molecular and cellular endocrinology* **334**, 31 (Mar 1, 2011).
18. J. A. Katzenellenbogen, The structural pervasiveness of estrogenic activity. *Environmental health perspectives* **103 Suppl 7**, 99 (Oct, 1995).
19. D. O. Norris, J. A. Carr, *Endocrine disruption : biological basis for health effects in wildlife and humans*. (Oxford University Press, New York, 2006), pp. xiv, 477 p.
20. J. W. Thornton, Resurrecting ancient genes: experimental analysis of extinct molecules. *Nature reviews. Genetics* **5**, 366 (May, 2004).
21. D. A. Liberles, *Ancestral sequence reconstruction*. (Oxford University Press, Oxford ; New York, 2007), pp. xiii, 252 p.
22. J. W. Thornton, E. Need, D. Crews, Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling. *Science* **301**, 1714 (Sep 19, 2003).
23. J. T. Bridgham, J. E. Brown, A. Rodriguez-Mari, J. M. Catchen, J. W. Thornton, Evolution of a new function by degenerative mutation in cephalochordate steroid receptors. *PLoS genetics* **4**, e1000191 (2008).
24. D. S. Geller *et al.*, Activating mineralocorticoid receptor mutation in hypertension exacerbated by pregnancy. *Science* **289**, 119 (Jul 7, 2000).
25. J. Veldscholte *et al.*, The androgen receptor in LNCaP cells contains a mutation in the ligand binding domain which affects steroid binding characteristics and response to antiandrogens. *The Journal of steroid biochemistry and molecular biology* **41**, 665 (Mar, 1992).
26. X. Y. Zhao *et al.*, Glucocorticoids can promote androgen-independent growth of prostate cancer cells through a mutated androgen receptor. *Nature medicine* **6**, 703 (Jun, 2000).
27. J. W. Thornton, Evolution of vertebrate steroid receptors from an ancestral estrogen receptor by ligand exploitation and serial genome expansions. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 5671 (May 8, 2001).

28. T. Mizuta, K. Asahina, M. Suzuki, K. Kubokawa, In vitro conversion of sex steroids and expression of sex steroidogenic enzyme genes in amphioxus ovary. *Journal of experimental zoology. Part A, Ecological genetics and physiology* **309**, 83 (Mar 1, 2008).
29. A. D'Aniello *et al.*, Occurrence of sex steroid hormones and their binding proteins in Octopus vulgaris lam. *Biochemical and biophysical research communications* **227**, 782 (Oct 23, 1996).
30. D. A. Close, S. S. Yun, S. D. McCormick, A. J. Wildbill, W. Li, 11-deoxycortisol is a corticosteroid hormone in the lamprey. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 13942 (Aug 3, 2010).
31. D. M. Taverna, R. A. Goldstein, Why are proteins so robust to site mutations? *Journal of molecular biology* **315**, 479 (Jan 18, 2002).
32. M. J. Ryan, J. H. Fox, W. Wilczynski, A. S. Rand, Sexual selection for sensory exploitation in the frog Physalaemus pustulosus. *Nature* **343**, 66 (Jan 4, 1990).
33. W. Wickler, *Mimicry in plants and animals*. World university library (McGraw-Hill, New York,, 1968), pp. 253 p.
34. D. P. Edwards, The roles of tolerance in the evolution, maintenance and breakdown of mutualism. *Die Naturwissenschaften* **96**, 1137 (Oct, 2009).
35. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**, 1792 (2004).
36. S. Guindon, O. Gascuel, A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic biology* **52**, 696 (Oct, 2003).
37. M. Anisimova, O. Gascuel, Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Systematic biology* **55**, 539 (Aug, 2006).
38. Z. Yang, S. Kumar, M. Nei, A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics* **141**, 1641 (Dec, 1995).
39. Z. Yang, PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer applications in the biosciences : CABIOS* **13**, 555 (Oct, 1997).
40. V. Hanson-Smith, B. Kolaczkowski, J. W. Thornton, Robustness of ancestral sequence reconstruction to phylogenetic uncertainty. *Molecular biology and evolution* **27**, 1988 (Sep, 2010).
41. D. Picard, K. R. Yamamoto, Two signals mediate hormone-dependent nuclear localization of the glucocorticoid receptor. *The EMBO journal* **6**, 3333 (Nov, 1987).

42. J. Keay, J. T. Bridgham, J. W. Thornton, The Octopus vulgaris estrogen receptor is a constitutive transcriptional activator: evolutionary and functional implications. *Endocrinology* **147**, 3861 (Aug, 2006).
43. J. T. Bridgham, S. M. Carroll, J. W. Thornton, Evolution of hormone-receptor complexity by molecular exploitation. *Science* **312**, 97 (Apr 7, 2006).
44. L. Clinckemalie, D. Vanderschueren, S. Boonen, F. Claessens, The hinge region in androgen receptor control. *Molecular and cellular endocrinology* **358**, 1 (Jul 6, 2012).
45. J. T. Bridgham *et al.*, Protein evolution by molecular tinkering: diversification of the nuclear receptor superfamily from a ligand-dependent ancestor. *PLoS biology* **8**, (2010).
46. P. D. Adams *et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography* **66**, 213 (Feb, 2010).
47. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta crystallographica. Section D, Biological crystallography* **60**, 2126 (Dec, 2004).
48. N. Eswar, D. Eramian, B. Webb, M. Y. Shen, A. Sali, Protein structure modeling with MODELLER. *Methods in molecular biology* **426**, 145 (2008).
49. S. C. Lovell *et al.*, Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins* **50**, 437 (Feb 15, 2003).
50. G. J. Kleywegt, T. A. Jones, Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta crystallographica. Section D, Biological crystallography* **50**, 178 (Mar 1, 1994).

CHAPTER 3: BIOPHYSICAL MECHANISMS FOR LARGE-EFFECT MUTATIONS IN THE EVOLUTION OF STEROID HORMONE RECEPTORS

Michael J. Harms[1]**, Geeta N. Eick[1]**, Devrishi Goswami[2], Jennifer K. Colucci[3], Patrick R. Griffin[2], Eric A. Ortlund[3], and Joseph W. Thornton[1,4b]

[1]*Institute of Ecology and Evolution, University of Oregon, Eugene, OR 97403.*

[2]*Department of Molecular Therapeutics, The Scripps Research Institute, Jupiter, FL, 33458*

[3]*Biochemistry Department, Emory University School of Medicine, Atlanta, GA 30322.*

[4]*Departments of Human Genetics and Ecology & Evolution, University of Chicago, Chicago, IL 60637.*

M. J. Harms et al., *Biophysical mechanisms for large-effect mutations in the evolution of steroid hormone receptors.* Proceedings of the National Academy of Sciences of the United States of America ***110***, *11475 (Jul 9, 2013).*

Steroid hormone nuclear receptors evolved specificity from aromatic steroids to non-aromatic 3-keto steroids. To understand the biophysical mechanisms underlying this change, this work uses molecular dynamics paired with hydrogen-deuterium exchange mass spectrometry (HDX-MS) to analyze the change in bond networks between receptor and ligand. For this work, I aided in the conception of HDX-MS assays. I expressed and purified four novel complexes of the

---

[b] Author contributions: M.J.H., G.N.E., and J.W.T. conceived the project; M.J.H., G.N.E., and J.W.T. designed all aspects of the research; D.G., J.K.C., P.R.G., and E.A.O. designed and performed the HDX research; G.N.E. performed the phylogenetic analysis and functional/genetic experiments; M.J.H. performed the molecular dynamics simulations and led analysis of the HDX data; M.J.H., G.N.E., and J.W.T. analyzed the data and interpreted the findings; and M.J.H., G.N.E., and J.W.T. wrote the paper.

** These authors made equal contributions

ancestral 3-keto steroid receptor (ancSR2) and a mutant of ancSR2 with two steraloids each and performed the initial HDX-MS analysis. This work has been previously published in the Proceedings of the National Academy of Sciences of the United States of America.

Abstract

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 shifts in protein function can be triggered by a few mutations of large effect and, if so, the mechanisms by which they do so. 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 (SRs), a family of biologically essential hormone-controlled transcription factors. We previously found that the ancestor of the entire SR family was highly specific for estrogens, but its immediate phylogenetic descendant was sensitive only to androgens, progestagens, 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 biochemically subtle mutations into major shifts in the energetics and function of the protein.

Introduction

Protein biophysics and evolution.

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, like the role and mechanisms of large-effect mutations in phenotypic evolution (4-8), and in biophysics, like 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.

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 carbon 3 on the steroid backbone (Figure 3.1A). The other clade—the nonaromatized steroid receptors (naSRs)—includes receptors for androgens, progestagens, mineralocorticoids, and glucocorticoids, all of which have a nonaromatized A-ring and a keto or hydroxyl at carbon 3 (Figure 3.5). We used maximum likelihood phylogenetics and >200 extant sequences to infer the LBD sequences of the progenitor of the entire family (ancSR1) and the progenitor 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 (Figure 3.1B). 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 3.2), suggesting these sites are functionally constrained. To further narrow the set of candidate sites, we examined these diagnostic replacements 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 (Figure 3.6), but two substituted residues contact the A-ring or its C3 functional group—glu41GLN and leu75MET (Figure 3.2B, using upper and lower case 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 norprogesterone (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 (Figure 3.2C).

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 (Figure 3.2D,E). Similar effects were observed with other matched pairs of aromatized and nonaromatized

steroids, irrespective of whether they contain a 3-hydroxyl or keto (Figure 3.7). 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, while 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 non-aromatized A-rings (Figure 3.8).

These data indicate that replacements glu41GLN and leu75MET 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 (Figure 3.8), and ancSR2-glu41/leu75 retains some sensitivity to nonaromatized steroids (Figure 3.2D,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 to the ancSR1/estradiol structural model (19). Although the basic architecture of the ligand cavity is conserved (Figure 3.2A), 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 (Figure 3.2F). Replacement with GLN41, which substitutes a donor functional group for an acceptor, allows this side chain to hydrogen bond directly to 3-keto ligands (Figure 3.2F), explaining why glu41GLN causes a ~100-fold increase in sensitivity to 3-keto ligands (Figure 3.2D,E).

The two mutations' most significant effect, however, is to radically reduce activation by aromatized steroids (Figure 3.2D,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 (Figure 3.2F). MET75 may exacerbate this effect by adding a new weak hydrogen bond acceptor above the A-ring, further complicating an already over-constrained network (Figure 3.2F). 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 simulations to study the effects of the two key mutations on protein-ligand interactions. We conducted triplicate 50 nanosecond 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 (Figure 3.9), the purpose of this analysis is to elucidate how mutations change the ligands' interaction with the protein's active conformation.

We first tested the hypothesis that the glu41GLN mutation introduced a new, favorable interaction with the nonaromatized steroid norP (Figure 3.2F). As predicted, the ancestral glu41 cannot interact directly with norP's 3-keto. Instead, water is brought into the mouth to satisfy the hydrogen-bonding potential of both gln41 and the ligand (Figure 3.10). Introducing the derived amino acids causes the amine donor of GLN41 to form a new hydrogen bond to the ligand's

carbonyl (Figure 3.10), supporting the view that this interaction contributes to ancSR2's increased sensitivity to non-aromatized ligands.

We then tested the hypothesis that replacements 41 and 75 disrupted the A-ring hydrogen bond network (Figure 3.2F). To quantitatively analyze the network's behavior, we clustered the conformations populated during each simulation based on the state of the A-ring subnetwork, 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 (Figure 3.11-3.12, Table 3.3-3.4). 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 41, 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 (Figure 3.3A, ref. (20)). Just two subtle variants of this network were observed: in one, all 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 (see Movie S1 in (*51*)). This stable network allows the 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 changes this network dramatically. The A-ring system now transitions among seven distinct and suboptimal states, reflecting a rugged conformational free energy landscape with multiple basins separated by free energy barriers (Figure 3.3B, 3.4A). None of these configurations allow 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 (Figure 3.4B). 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 (Figures 3.3, 3.4C, see Movie S2 in (*51*)). Internal water penetration occurs because the unsatisfied polar atoms on GLN41 cannot interact with water from the pocket's mouth, because this residue's interactions with the ligand and MET75 cause its polar groups to face inward, away from bulk solvent (Figure 3.3). 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 (Figure 3.4C). Helices H3, H5, and H12 separate, disrupting the geometry of the coactivator interface (Figure 3.4D). 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 (Figure 3.13).

Taken together, these observations suggest a mechanistic model that explains the effects of the two key replacements on ligand specificity. The derived residues established a new favorable interaction with 3-ketosteroids. They also established a frustrated 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 these predictions, we employed hydrogen-deuterium exchange mass spectrometry (HDX-MS), which quantifies local solvent accessibility and dynamics across a protein by characterizing the rate of deuterium exchange exhibited for peptides across the protein (21). We performed HDX-MS on the complexes of ancSR2 and ancSR2-glu41/leu75, each with NPT or norP, and identified regions of each protein in which the two ligands resulted in different

rates of proton/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-glu41/leu75 (Figures 3.14-3.15, Table 3.5). 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 deuterium exchange in four regions: the C-terminus of H3, H6/H7, the loop connecting H10 to AF-H, and the AF-H helix (Figures 3.4C, 3.14). 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 (Figures 3.4C). These experimental observations corroborate the MD predictions and indicate that 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 (Figures 3.4C); the causes of this effect are unclear.

Arg82 is necessary for ligand-specificity.

The biophysical rationale derived from the structure (Figure 3.2F) and the MD simulations (Figure 3.3) includes a key role for the conserved, donor-rich Arg82 in mediating the effects of the two key historical replacements on aromatized steroid activation. 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 SR2 and characterized the effect of the glu41 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, glu41GLN/leu75MET triggers a mere 4-fold shift in preference (Figure 3.4E). 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 new 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 hydrogen-bond network when the ancestral ligand was present.

Our observations underscore the fact 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 very few mutations may cause states with different conformations and couplings to become energetically favorable (23). This can magnify subtle changes in the biochemistry of a few amino acids into large perturbations in the protein's biophysical behavior and, in turn, cause major evolutionary shifts in function. A protein's biophysical architecture 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 pSG5-DBD, mutated with Stratagene Quikchange, cotransfected into 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 SPC waters (25, 26). For each ligand/receptor complex, three independent trajectories were initiated, beginning with the ancSR2-progesterone crystal structure coordinates (PDB 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 frequency 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 mass spectrometer. Spectra were collected at six time points between 10 and 3,600 s. Deuterium incorporation kinetics were extracted using non-linear 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.

Reporter activation assays.

cDNAs coding for the maximum likelihood ancSR2 LBD and ancSR1 LBD were synthesized (Genscript) and verified. The LBDs were then cloned into the Gal4-DBD-pSG5 vector; 31 amino acids of the GR hinge containing the nuclear localization signal-1 (1) were inserted between the GAL4 DBD and LBD to ensure nuclear localization and conformational independence of the two domains. The hormone-dependent transcriptional activity of resurrected ancestral receptors and their variants was assayed using a luciferase reporter system. CHO-K1 cells were grown in 96-well plates and transfected with 1 ng of receptor plasmid, 100 ng of a UAS-driven firefly luciferase reporter (pFRluc), and 0.1 ng of the constitutive phRLtk Renilla luciferase reporter plasmid, using Lipofectamine and Plus Reagent in OPTIMEM (Invitrogen). After 4 h, transfection medium was replaced with phenol-red-free MEM supplemented with 10% dextran-charcoal stripped FBS (Hyclone). After overnight recovery, cells were incubated in triplicate with the hormone of interest from 10^-12 to 10^-5 M for 24 h, then assayed using Dual-Glo luciferase (Promega). Firefly luciferase activity was normalized by *Renilla* luciferase activity. Dose-response relationships were estimated using nonlinear regression in Prism4 software (GraphPad Software, Inc.); fold increase in activation was calculated relative to vehicle-only (ethanol) control.

Sequence conservation analysis.

Diagnostic residues were identified through analysis of the sequence alignment used to generate the phylogeny in the accompanying paper (2). Two subsets of the alignment, one containing all modern ERs and the other containing all modern naSRs, were generated. The Shannon information at each position in these sub-alignments was calculated using the Weblogo 3.1 python library (http://code.google.com/p/weblogo/) (3). Columns with information > 3.5 bits were considered "conserved." Diagnostic sites were defined as those where the ER and naSR clades possessed different, conserved states at that site and the transition between those states

occurred over the interval from ancSR1 to ancSR2 (Table 3.2). Of the 171 changes between ancSR1 and ancSR2, 52 were conserved in the ERs, 54 in the naSRs, and 22 proved diagnostic.

Molecular dynamics methods.

Models of each ligand/receptor complex were generated starting with the ancSR2-progesterone crystal structure (4FN9). Amino acid replacements were made using PyMOL, and the rotamer that minimized steric clashes was chosen visually. Ligands were placed in the pocket by aligning the ligand B, C, and D rings to progesterone in the crystal structure. Initial ligand models were generated using the PRODRG server (4). Owing to the lack of experimental transfer free energy parameters for these ligands, a sensitivity approach was taken to ligand parameterization. Ligand parameters were generated both by chemical analogy to existing moieties in the GROMACS parameter sets, as well as parameterized using explicit quantum mechanical calculations in GAMESS Oct12010R1 (5). These two parameterizations gave similar charge sets and qualitatively similar results in test simulations. For the simulations reported in this paper, the quantum mechanical charge sets were used. Molecular dynamics simulations were performed using GROMACS 4.5.1 (6) and the GROMOS96 53a6 force field (7) with the SPC water model. Calculations were done at 300 K and 1 bar. Three independent simulations were performed for each protein/ligand pair. For each simulation, the protein/ligand model was placed in the center of a cubic water box 20 Å larger than the maximum protein dimension, followed by energy minimization. Velocities were assigned from a Maxwell distribution and the system was equilibrated for 1 ns with heavy protein atoms (except for mutated residues) fixed. This was followed by 50 ns of unrestrained MD. The initial 10 ns were excluded as burn-in. The trajectory time step was 2 fs, with frames recorded every 5 ps. Final analyses were performed on frames taken every 12.5 ps. All bonds were treated as constraints and fixed using LINCS (8). Electrostatics were treated with the Particle Mesh Ewald model (9), using an FFT spacing of 12

Å, interpolation order of 4, tolerance of 1e-5, and a Coulomb cutoff of 9 Å. van der Waals forces were treated with a simple cutoff at 9 Å. We used velocity rescaled temperature coupling with a τ of 0.1 ps and Berendsen pressure coupling with a τ of 0.5 ps and a compressibility of 4.5e-5 bar^{-1}. Analyses were performed using VMD 1.9.0 (10)—with its built-in TCL scripting utility—as well as a set of in-house Python and R scripts.

Free energy landscapes.

To generate free energy landscapes, we assigned each frame from the trajectories to a state and then calculated the population of each state and the transition frequencies between states. We first defined residue 41 conformations based on rotamer and hydrogen bonds. Glu41 and GLN41 populated distinct rotamers, which could be identified using simple angle cutoffs in χ_1, χ_2 and χ_3 (Figure 3.11). Glu41 populated five rotamers, while GLN41 populated ten rotamers. We then counted the number and identity of the hydrogen bonds formed by the residue 41 side chain using a geometric criterion of $\theta < 45°$ and heavy-atom separation of < 3.5 Å. Over all simulations with ancestral residues, the Glu41 carboxylic acid accepted from the ligand O3 hydroxyl, up to two hydrogens from Arg82, and up to four hydrogens from water. Over all simulations with derived residues, the GLN41 carbonyl accepted from the ligand O3 hydroxyl, from Arg82, and up to two hydrogens from water, while the GLN41 amine donated to the ligand O3, the MET75 sulfur, and up to two waters. Given the rotamer and hydrogen bonds formed in each frame, each frame was then placed in a conformational bin.

We could then calculate the frequencies of each conformational bin, as well as the frequencies of transitions between them. We defined "states" as sets of bins whose members could reach any other bin within the state without crossing an energy barrier of defined height. To identify these states, we represented the transition matrix as a graph—where each bin was a

node and each transition probability greater than exp(-$\Delta G_{barrier}$/RT) was an edge—then found the strong connected components of this graph using Tarjan's algorithm. Each strongly connected component corresponds to a state separated from all other states by an energy barriers > $\Delta G_{barrier}$.

The number of states is strongly dependent on the choice of energy barrier (Figure 3.12). At extreme values of $\Delta G_{barrier}$, all protein/ligand pairs exhibit similar behavior. For high energy barriers (~2 to 3 kcal/mol), most conformations collapse into a single state. This indicates that almost all conformations are connected by some path—however tortuous—that does not require passing an energy barrier greater than 3 kcal/mol.

For low barriers (0 to ~1 kcal/mol), each conformational bin is assigned to its own state. Put another way, members of a bin are, on average, 5 times more likely to remain within that bin rather than transition to another. This implies that these bins are physically relevant: if the bins were arbitrary, transitions between conformational bins would be just as likely as remaining within a given conformational bin. Different protein/ligand pairs populate different numbers of conformations, ranging from 10 (glu41/leu75/norP) to 76 (GLN41/MET75/NPT). This is because there are fewer possible conformations for the glu41/le75 proteins than the GLN41/MET75 proteins because of the equivalent glutamic oxygen atoms, as well as fewer conformations with norP than NPT/NPR because it can only act as a donor, whereas NPT and NPR can both act as donors and acceptors.

Once frames are assigned to states, the free energy landscape follows directly from the populations of each state—giving the depth of the wells—and the transition probabilities between states—giving the barrier heights. The free energy of a state *i* is the log probability of observing that state:

$$\Delta G_i = -RT\, ln\left(\frac{n_i}{\sum n}\right)$$

The free energy barrier between states *i* and *j* is the log probability of observing a transition between states *i* and *j* given the population of state *i*:

$$\Delta G_{i\to j} = -RT\, ln\left(\frac{n_{i\to j}}{n_i}\right)$$

To test whether the system is at equilibrium with regard to these transitions, we checked for additivity:

$$\Delta G_i + \Delta G_{i\to j} = \Delta G_j + \Delta G_{j\to i}$$

This was observed (Table 3.3). We also verified that the number of transitions from state *i* to state *j* was equal to the number of transitions from state *j* to state *i*. As is evident from Table 3.4, good sampling was observed for all but the highest energy barriers.

Characterization of water-penetrated states.

The extent of water penetration was visualized using VMD's built in volmap command. The protein Cα atoms from each frame were aligned to the starting structure. Water oxygen atoms within 5 Å of any ligand atoms were used to generate a density map with a resolution of 0.5 Å. Plots were contoured to show density > 0.01 (Figure 3.4). To characterize the changes in the structures that occurred upon water penetration, we measured distances between helices at the coactivator interface. For H3/ H5, we used the distance between the Cαs of residues 41 and 75; for H5/H12, the distance is between the Cαs of residues 75 and 234 (Figure 3.4D).

HDX-MS.

We performed hydrogen-deuterium exchange mass-spectrometry (HDX-MS) on SR2 Q41e/M75l and WT SR2 in the presence of NPT and norP. Solution-phase amide HDX was

carried out with a fully automated system as described previously (11) with slight modifications. The automation system (CTC HTS PAL, LEAP Technologies, Carrboro, NC) was housed inside a chromatography cabinet held at 4 °C. Briefly, 5 µl of 10 µM protein was diluted to 25 µl with D_2O-containing HDX buffer and incubated at 4°C for 10s, 30s, 60s, 900s or 3,600s. Following on exchange, unwanted forward or back exchange was minimized and the protein was denatured by dilution to 50 µl with 0.1% (v/v) TFA in 3 M urea. Samples were then passed across an immobilized pepsin column (prepared in house) at 50 µl min^{-1} (0.1% v/v TFA, 15 °C); the resulting peptides were trapped on a C_8 trap cartridge (Hypersil Gold, Thermo Fisher). Peptides were then gradient-eluted (4% (w/v) CH_3CN to 40% (w/v) CH_3CN, 0.3% (w/v) formic acid over 5 min, at 4 °C) across a 1mm × 50mm C_{18} HPLC column (Hypersil Gold, Thermo Fisher) and subjected to electrospray ionization directly coupled to a high resolution Orbitrap mass spectrometer (LTQ Orbitrap XL with ETD, Thermo Fisher). MS/MS data were acquired in separate experiments with a 60 minute gradient. Data-dependent MS/MS was performed in the absence of exposure to deuterium and the amino acid sequence of each peptide used in the HDX peptide set were confirmed if they had a MASCOT score of 20 or greater and had no ambiguous hits using a decoy (reverse) database. For on-exchange experiments, the intensity weighted average m/z value (centroid) of each peptide's isotopic envelope was calculated and corrected for back-exchange using software developed in-house (12).

We extracted rate constants for deuterium exchange for each peptide from its deuterium incorporation time series. We excluded 36 peptides (of 242) that exhibited <15% deuterium incorporation over 3,600 s. Fitting was done using the nls function within the R 2.15.0 statistics package (13). Fit standard errors were used as reported by nls and then propagated for downstream analysis. Appropriate left- or right-tailed 95% confidence intervals were used to assess the significance of any changes in exchange rate/protection factors.

To identify the simplest model that adequately described the data, we tested four kinetics models against the data (Table 3.5). The simplest model (model #1) was a single exponential with two free parameters describing the exchange rate (k) and total incorporation of deuterium at infinite time (M). The most complex model (model #4) treated each peptide as consisting of two classes of sites, each with its own exchange rate (k_1 and k_2) and total incorporation (M_1 and M_2), giving four free parameters. We also constructed two intermediate, three parameter models. Model #2 treated the entire peptide as a single block that could exchange by two different processes, and therefore had a single total incorporation value (M) with two exchange rates (k_1 and k_2). Model #3 is a reduction of Model #4 in which k_1 is asserted to be much greater than the timescale of the experiment. In this scenario, an initial fast-exchange process occurs before the first time point is collected, followed by a slower exchange process captured by the rate constant k_2. This model therefore has a single exchange rate (k_2) with two incorporation values (M_1 and M_2).

We fit the parameters of these models to the experimental data and calculated the global root mean squared difference between the models and the data. We then employed the Akaike Information Criterion (AIC), which penalizes models that have excess parameters, to identify the model that best described the data with the fewest parameters. This analysis strongly favors model #3: using AIC-corrected likelihoods, the observed data are 3.7×10^{39} more likely given model #3 than the next-best model. On the basis of this analysis, we used model #3 for all further analyses.

Using extracted exchange rates, we calculated ligand-specific peptide protection factors by taking the ratio of the exchange rates for matched peptides from the same protein in the presence of either NPT or norP (Figure 3.14A). This measures whether, relative to norP, NPT increases or decreases protection of labile protons on that peptide. In the ancestral background (SR2/Q41e/M75l), NPT decreased protection ($p < 0.05$) for 11 peptides (19%) and increased

protection for 7 peptides (12%) (Figure 3.14C). In the derived background (SR2), NPT decreased protection for 15 peptides (23%) and increased protection for 10 peptides (16%) (Figure 3.14D). The final step of the analysis was to ask how the introduction of the derived amino acids altered these ligand-specific protection factors. The experiments using SR2 and SR2/Q41e/M75l shared 29 peptides, covering 78.6% of the protein sequence. We subtracted the derived ligand-specific protection factors from the ancestral factors for each peptide, finding that 5 (17%) exhibited decreased protection and 4 (14%) exhibited increased protection (Figure 3.14E). The raw fits used to calculate the protection factors for the peptides that correlate with the positions of water penetration in the MD simulations are show in Fig S11.

Acknowledgements

We thank J. Bridgham and J.W.T. laboratory members for comments. This work was supported by National Institutes of Health Grants R01-GM081592 and F32-GM090650, National Science Foundation Grants IOB-0546906 and DEB-0516530, and the Howard Hughes Medical Institute.

Figures

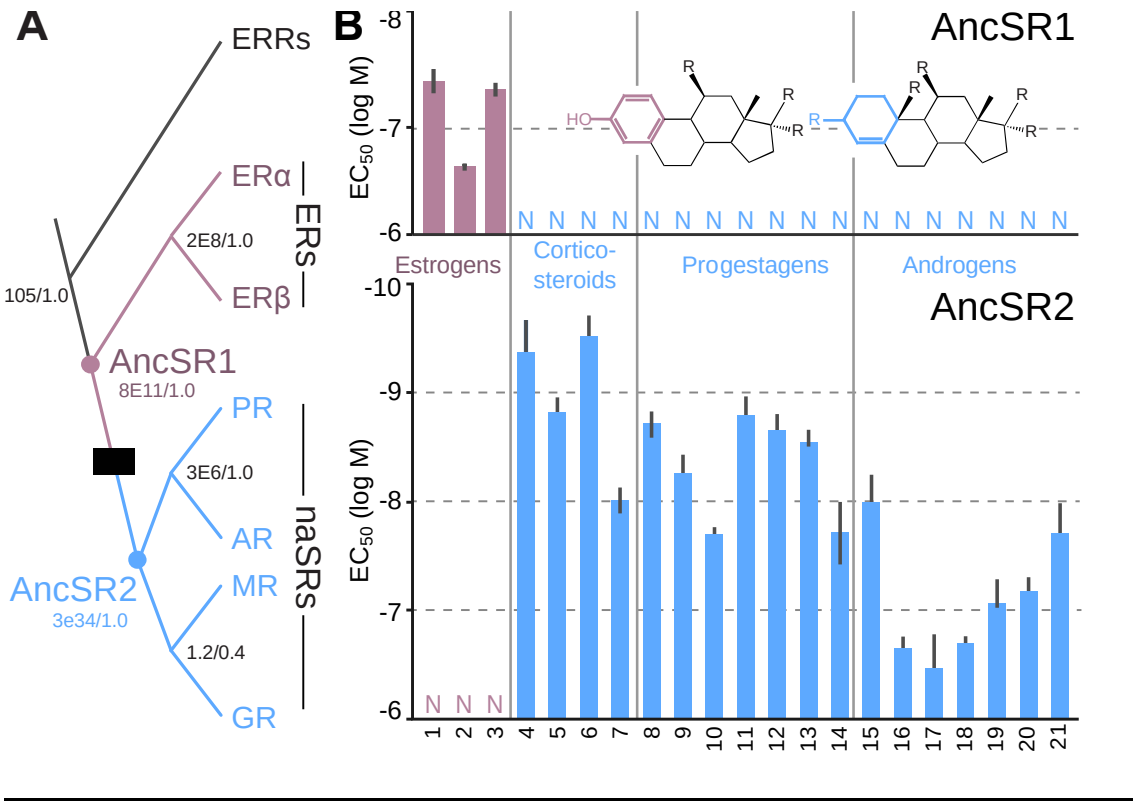


Figure 3.1: Evolution of ancSR1 and ancSR2 specificity.

A) Schematic phylogeny showing ancSR1, ancSR2, and the preference of each for steroids with aromatized (pink) vs. nonaromatized (blue) A-rings. Node labels show statistical confidence as approximate likelihood ratios (in scientific notation) and associated □-square support values (27). Black box, shift in specificity. For full tree, see (19). **B)** Hormone sensitivity of ancSR1-LBD (top panel) and ancSR2-LBD (bottom panel) as characterized in a triplicate luciferase reporter assay is displayed as EC50, the concentration at which half-maximal reporter activation is achieved (Table 3.1). Error bars, 95% c.i. Hormones in pink or blue have aromatized or nonaromatized A-rings, respectively: **1**, estradiol; **2**, estrone; **3**, estriol; **4**, 11-deoxycorticosterone; **5**, 11-deoxycortisol; **6**, corticosterone; **7**, cortisol; **8**, aldosterone; **9**, progesterone; **10**, 17α-

hydroxyprogesterone; **11**, 19-norprogesterone; **12**, 4-pregnenolone; **13**, 5-pregnenolone; **14**, 20α-hydroxyprogesterone; **15**, 20β-hydroxyprogesterone; **16**, testosterone; **17,** dihydrotestosterone; **18**, 4-androstenediol; **19**, 5-androstenediol; **20**, 19-nortestosterone; **21**, bolandiol. Figure 3.5 shows the major physiological ligands for extant receptors.

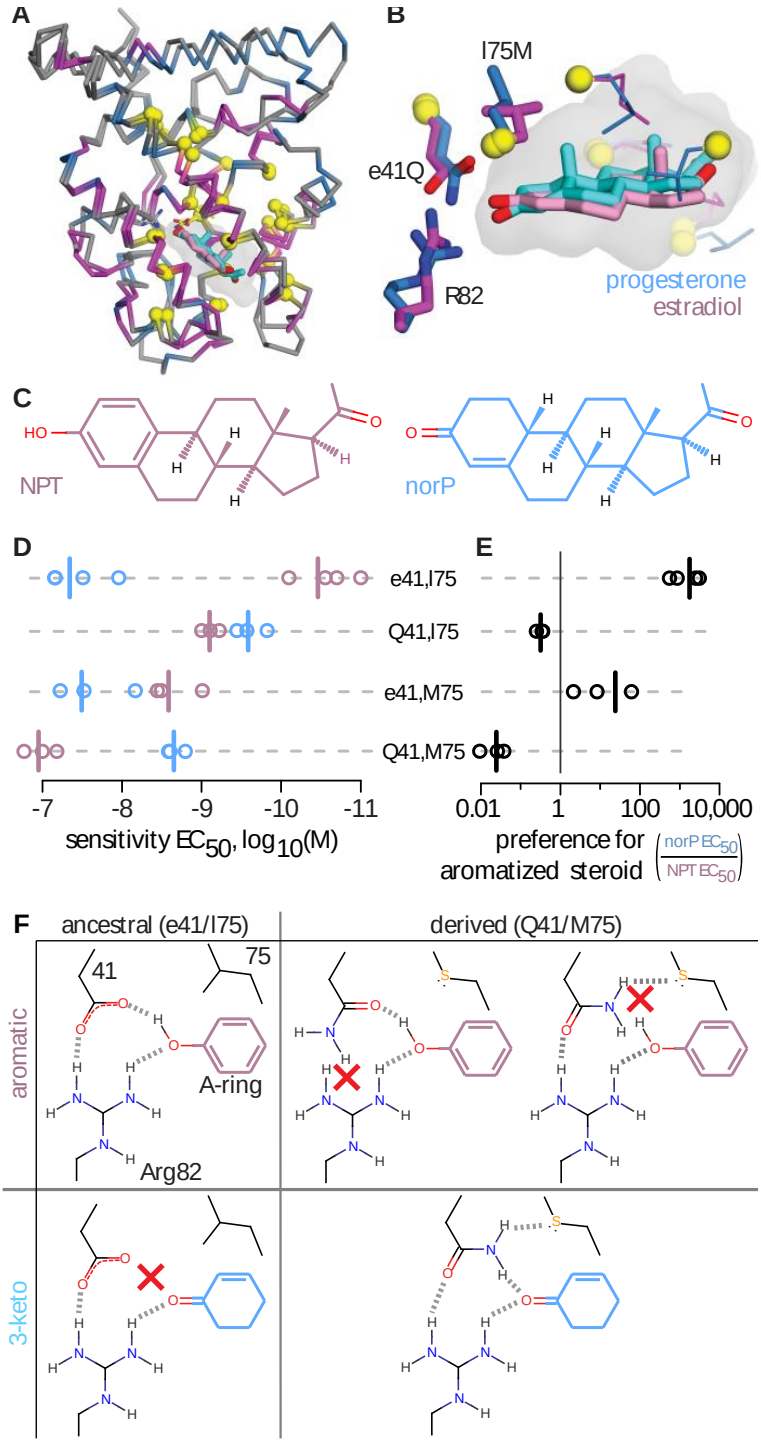

A
B
I75M
e41Q
R82
progesterone
estradiol
C
HO
NPT
norP
D
E
e41,I75
Q41,I75
e41,M75
Q41,M75
-7
-8
-9
-10
-11
sensitivity EC50, log10(M)
0.01
1
100
10,000
preference for aromatized steroid (norP EC50 / NPT EC50)
F
ancestral (e41/I75)
derived (Q41/M75)
aromatic
3-keto
41
75
A-ring
Arg82

Figure 3.2: Large-effect historical mutations drove the evolution of new ligand specificity.

A) Crystal structure of ancSR2 (4FN9) bound to progesterone is superimposed on the homology model of ancSR1 bound to estradiol (19). Progesterone (cyan) and estradiol (pink) are shown as sticks. Colors indicate site-specific conservation pattern: yellow, diagnostic sites conserved in different states in ERs and naSRs; blue, conserved among naSRs; magenta, conserved among ERs; gray, uninformative sites (constant in both or not conserved within either group). Spheres, Cα atoms of diagnostic residues. **B)** Diagnostic residues within 5 Å of the ligand in the ancSR1 homology model (magenta) and ancSR2 crystal structure (blue). Diagnostic residues near the A-ring and conserved Arg82 are shown as sticks and labeled, other residues as lines. Surface of the ligand cavity is also shown. **C)** Chemical structures of NPT (pink) and norP (blue). **D)** Effect of ancestral or derived amino acids (upper and lower case, respectively) at two key sites on sensitivity of ancSR2 to the aromatized steroid NPT (pink) and the nonaromatized steroid norP (blue). Each point shows the EC50 in a dose-response luciferase reporter assay; vertical lines show means across repeated experiments. **E)** Effect of key mutations on ancSR2's preference for aromatized steroids, expressed as the ratio of the EC50 for norP to the EC50 for NPT. **F)** Predicted hydrogen bond networks for ancestral (left) and derived (right) amino acids with aromatized (top row) or 3-keto (bottom row) steroids. The steroid A-ring and amino acids at position 41, 75, and 82 are shown as lines; predicted hydrogen bonds are shown in blue; red x, absent/unfavorable interactions.

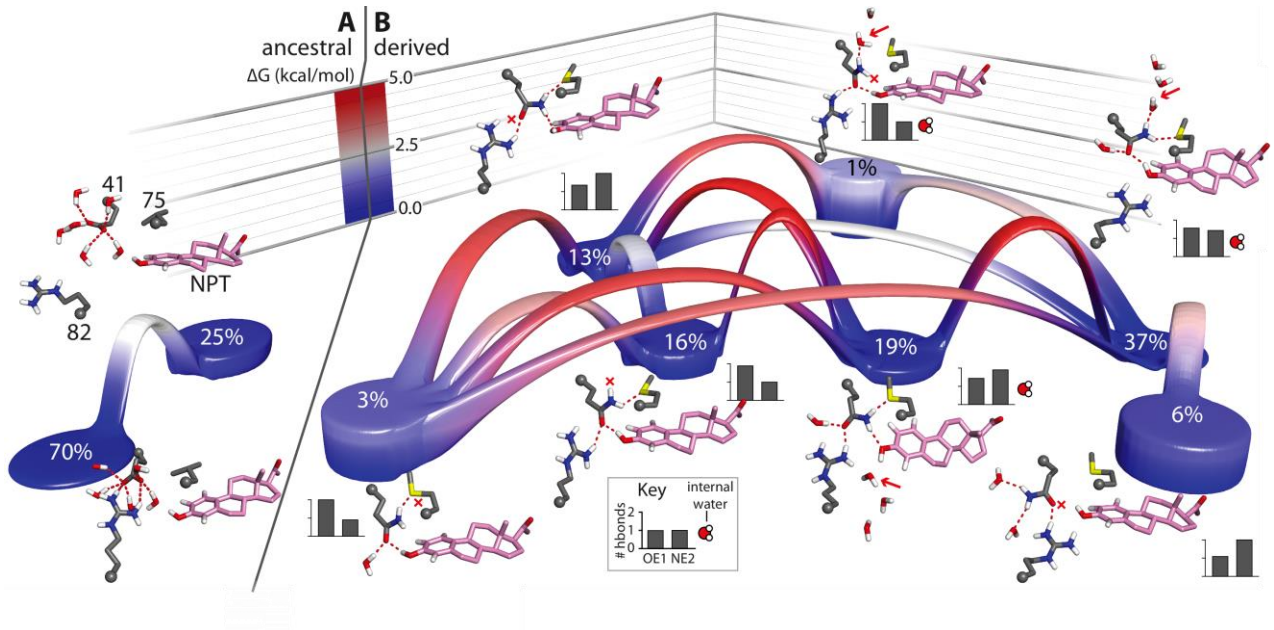


Figure 3.3: Two historical mutations altered the energetic landscape of protein-ligand binding.

Thermodynamic landscapes of the ensemble of interaction networks of the synthetic estrogen NPT and residues 41, 75, and 82 are shown for ancSR2 containing **A)** the ancestral states glu41 and leu75, or **B)** the derived states GLN41 and MET75. Each platform represents a conformational state, and each bridge a transition between states; height and coloring encode free energy from 0 (blue) to 5 kcal/mol (red) relative to the lowest free energy conformation on each landscape. Free energies were calculated from the frequency of conformations and transitions observed across replicate MD trajectories. Labels show the frequency of each conformation. Bar graphs indicate the number of hydrogen bonds formed by OE1 and NE2 of residue 41 in each state. Water icon indicates waters penetrating through the core of the protein. Molecular structures show representative samples of each conformation, including residues 41, 75, and 82 (gray carbons), NPT (pink), and waters. Red dashes, hydrogen bonds formed by residue 41. Red

crosses, unfulfilled polar atoms on residue 41. Red arrows point to water molecules that penetrate into the protein interior.

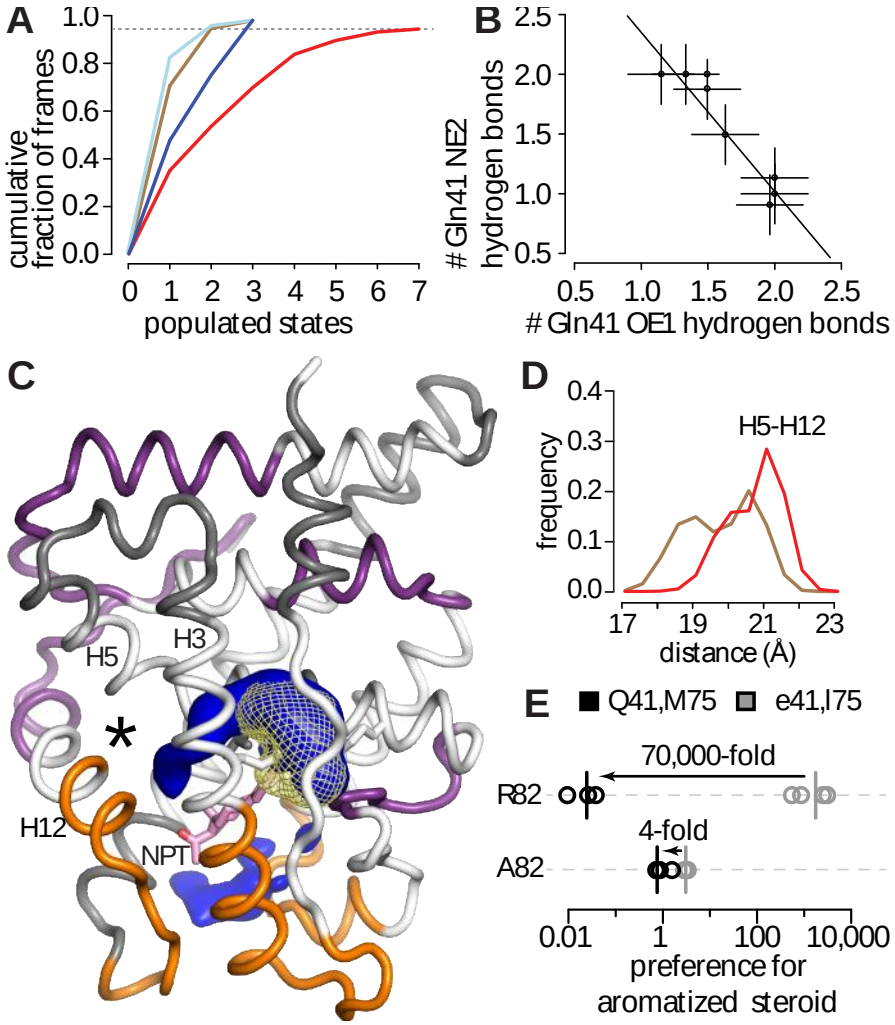


Figure 3.4: Ligand-specific disruption of the A-ring hydrogen-bond network.

A) The A-ring network of the aromatized steroid NPT with ancSR2 populates more conformational states than other protein-ligand complexes. Each line shows the cumulative frequency of conformations in the top 95% of states sampled by one protein-ligand complex:

ancSR2 with NPT (red) or norP (dark blue); ancSR2 with ancestral residues glu41/leu75 and NPT (brown) or norP (light blue). **B)** The OE1 and NE2 atoms of GLN41 compete for hydrogen bond partners in the ancSR2 complex with NPT. Each point represents one conformational state, plotted by the mean number of hydrogen bonds formed by OE1 and NE2 in that state; bars are standard deviations. **C)** Derived residues GLN41 and MET75 cause penetration of water into the protein interior in the complex with NPT. Average solvent occupancy in the protein interior is shown for MD trajectories when residues 41 and 75 are in the ancestral (yellow mesh) or derived (blue surface) states. Helices H3, H5, and H12 are indicated. NPT and the side chains of residues 41, 75, and 82 are shown as sticks. The backbone is colored by the effect of the historical mutations on ligand-specific hydrogen-deuterium protection factors, measured by HDX-MS: increased exchange with NPT (orange) or norP (pink), no effect (white), or no data (gray). **D)** Separation of helices that compose the coactivator interface in the ancSR2:NPT complex. The distribution of distances between H5 and H12 observed across trajectories is plotted for the complexes containing the ancestral (brown) vs. derived (red) residues at sites 41 and 75. **E)** Preference for aromatized steroids in luciferase reporter assays, defined as norP EC_{50}/NPT EC_{50}. Preference is shown when sites 41 and 75 are in the ancestral (gray) or derived (black) states, in the presence of Arg82 or Ala82.

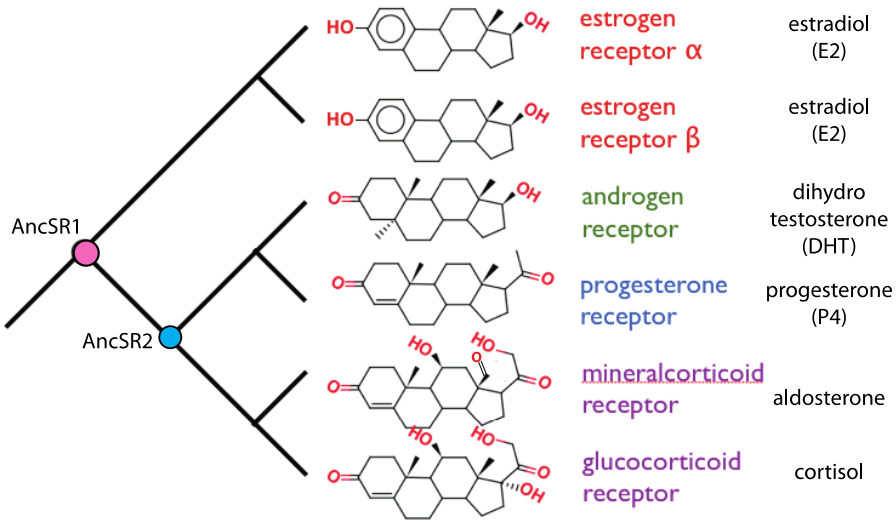


Figure 3.5: Cognate steroids of the six human steroid receptors.

Location of ancSR1 and ancSR2 are shown as circles on the schematic phylogenetic tree. Receptor name and cognate steroids are shown to the right.

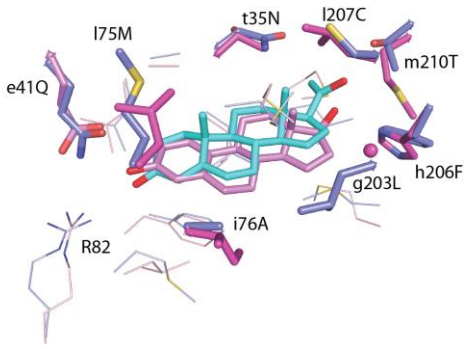


Figure 3.6: A-ring ligand contacts are largely conserved between ancSR2 (magenta) and ancSR2 (blue).

Overlay of ancSR1 homology model on the crystal structure of ancSR2 bound to progesterone (4FN9). Estradiol (pink) and progesterone (cyan) are shown as sticks. Residues conserved in one state in ERs and another in SRs are shown as sticks. Conserved residues are shown as lines. Lower and uppercase amino acid names represent ancestral and derived states, respectively.

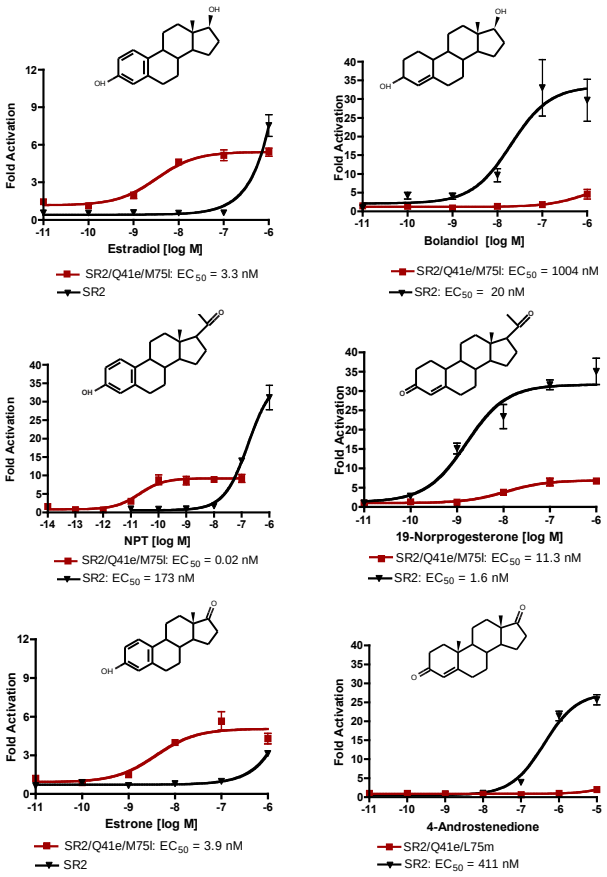


Figure 3.7: Representative dose activation curves of ancSR2/Q41e/M75l and ancSR2 wild-type.

Representative dose activation curves of ancSR2/Q41e/M75l (red curves) and ancSR2 wild-type (black curves) in response to three pairs of matched aromatized versus non-aromatized steroids (1, 3, 4-norprogestatrienelone vs. 19-Norprogesterone, Estradiol vs. Bolandiol, and Estrone vs. 4-Androstenedione). Activation is shown as the fold activation of a luceriferase reporter above vehicle-only treatment. EC50 values are shown for each curve. The steroid structures are shown above the graphs. Error bars indicate SEM.

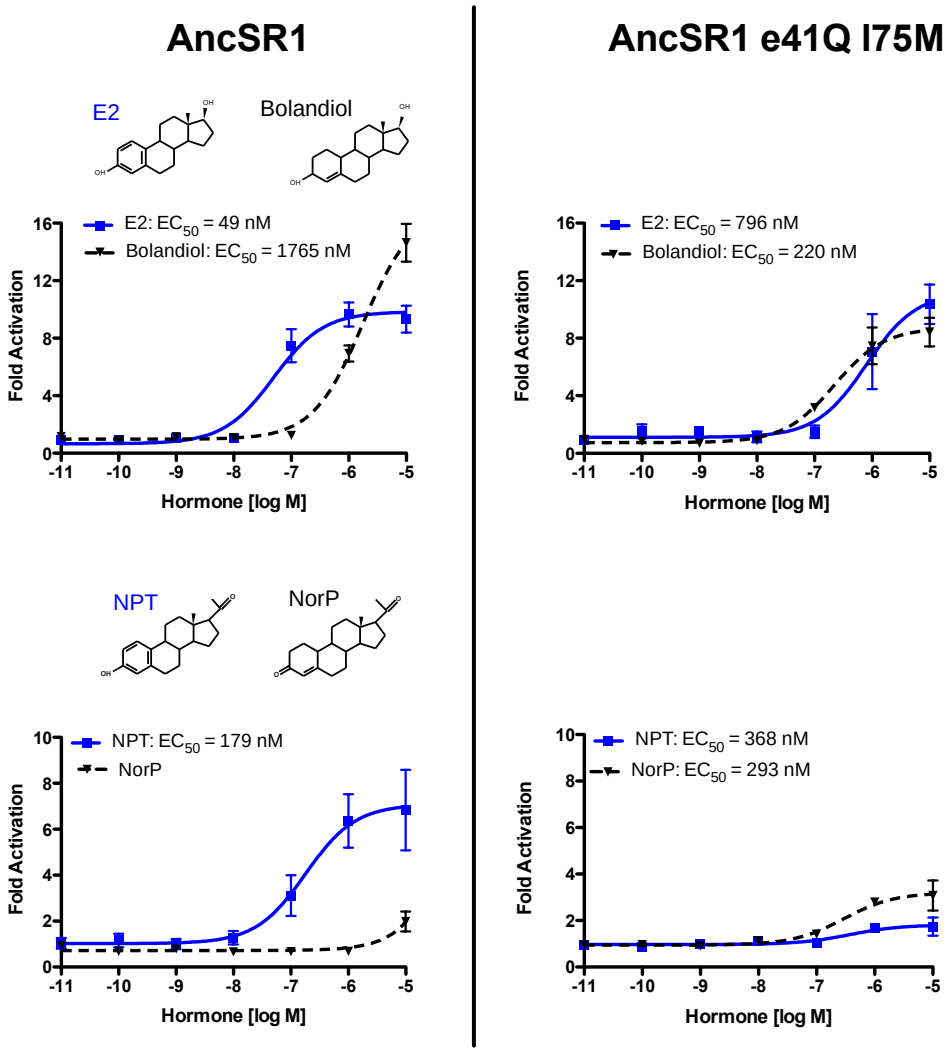


Figure 3.8: Representative dose activation curves of ancSR1 and ancSR1/e41Q/l75M.

Representative dose activation curves of ancSR1 and ancSR1/e41Q/l75M in response to two pairs of matched aromatized (blue) versus non-aromatized steroids (Estradiol vs. Bolandiol and 1, 3, 5-norprogestatrienelone vs. 19-Norprogesterone). Activation is shown as the fold activation of a luciferase construct above vehicle-only treatment. EC50 values are shown for each curve and the steroid structures are shown above the graphs on the left. Error bars indicate SEM.

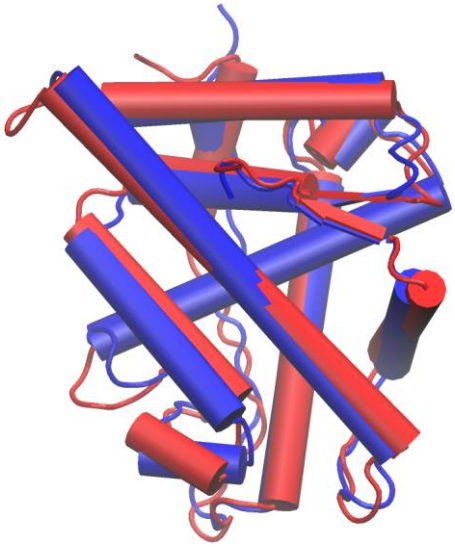


Figure 3.9: A control MD simulation with the apo protein.

A control MD simulation with the apo protein revealed that the protein does not relax to the inactive conformation – characterized by a large-scale conformational change – over the course of the calculation. Initial structure (red) overlaid on simulation after 50 ns (blue) in ancSR2 without ligand.

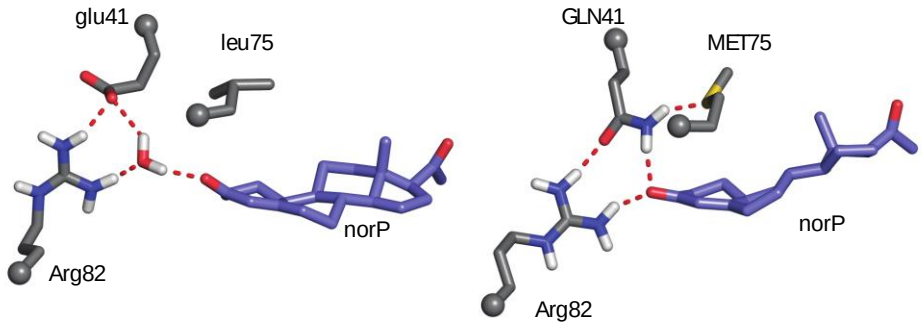


Figure 3.10: Derived amino acids introduce a new direct contact with the norP 3-keto group.

A) A representative snapshot of simulations with e41/leu75 and norP, highlighting hydrogen bonds formed in the A-ring network. B) A representative snapshot of simulation with GLN41/MET75 and norP. The direct GLN41 amine/norP 3-keto hydrogen bond is formed 45% of the time in the derived simulations.

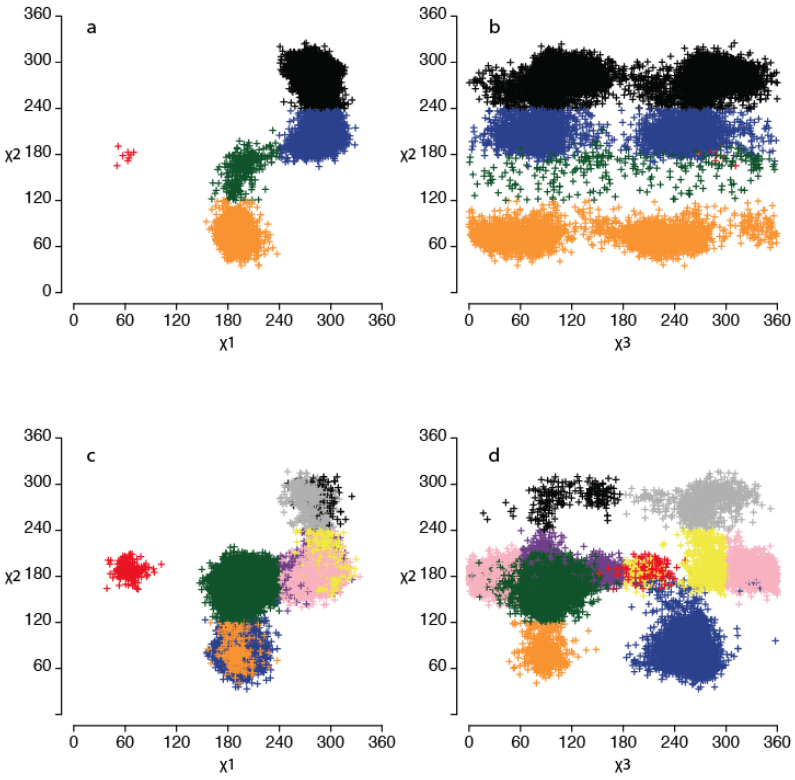


Figure 3.11 Populated rotamers of Glu41 and Gln41.

Rotamers populated by glu41 (panels A and B) and GLN41 (panels C and D) over the course of the MD simulations. Rotamers were defined by using simple cutoffs in χ1, χ2, and χ3. Colors represent the bin in which each rotamer was placed.

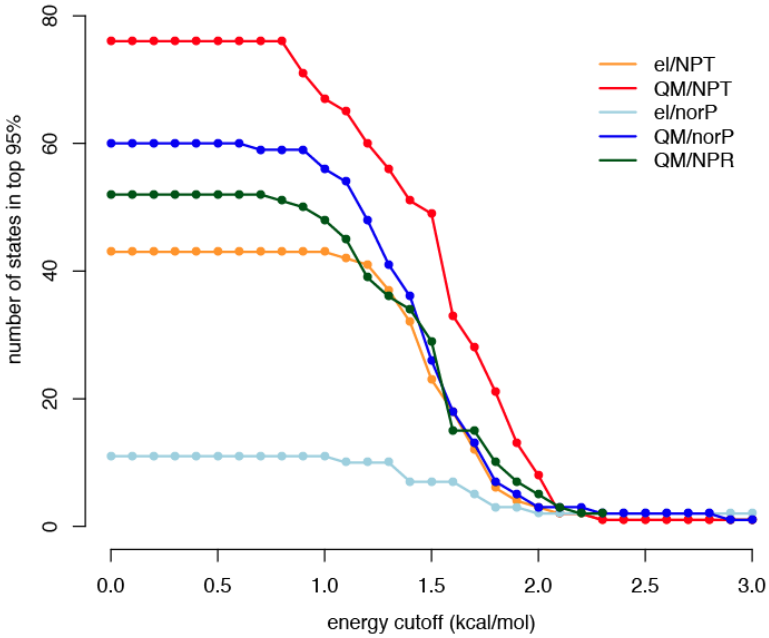


Figure 3.12: Dependence of number of states on $\Delta G_{barrier}$.

Number of states that make up the top 95% of the populated states as a function of energy barrier. States at sites 41 and 75 are shown as el (ancestral) or QM (derived). Steroids are: NPT: 1,3,5-norprogestatrienelone, norP: nor-progesterone, and NPR: 3-hydroxysteroid.

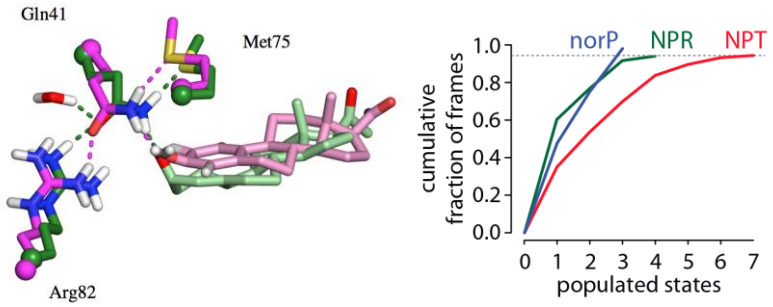


Figure 3.13: Non-aromatized steroid with 3-hydroxyl does not populate frustrated hydrogen bond networks.

A) Snapshots of the A-ring pocket residues from simulations with NPT (pink ligand, magenta sidechains) and 19-nor-4-pregnenalone (pale green ligand, green side chains). In simulations with 19-nor-4-pregnenalone (NPR), the GLN41 amine stability donates to the NPR O3 and MET75 sulfur, while the GLN41 carbonyl accepts from Arg82 and a water molecule. With NPT, GLN41 can donate to the ligand O3 and MET75; however, it accepts from Arg82 alone. This difference arises because of the geometry of the NPT and NPR hydroxyls, which cause GLN41 to populate different rotamers. When NPR is bound, the GLN41 OE1 can interact with a local water molecule, thus fulfilling all four possible hydrogen bonds. The different GLN41 rotamer with NPT—the OE1 is in plane with Arg82 – makes interaction with the water impossible, leading to an unfulfilled hydrogen bond despite populating a similar configuration. B) Number of populated states of the A-ring network in simulations of ancSR2 for norP (blue), NPR (green), and NPT (red).

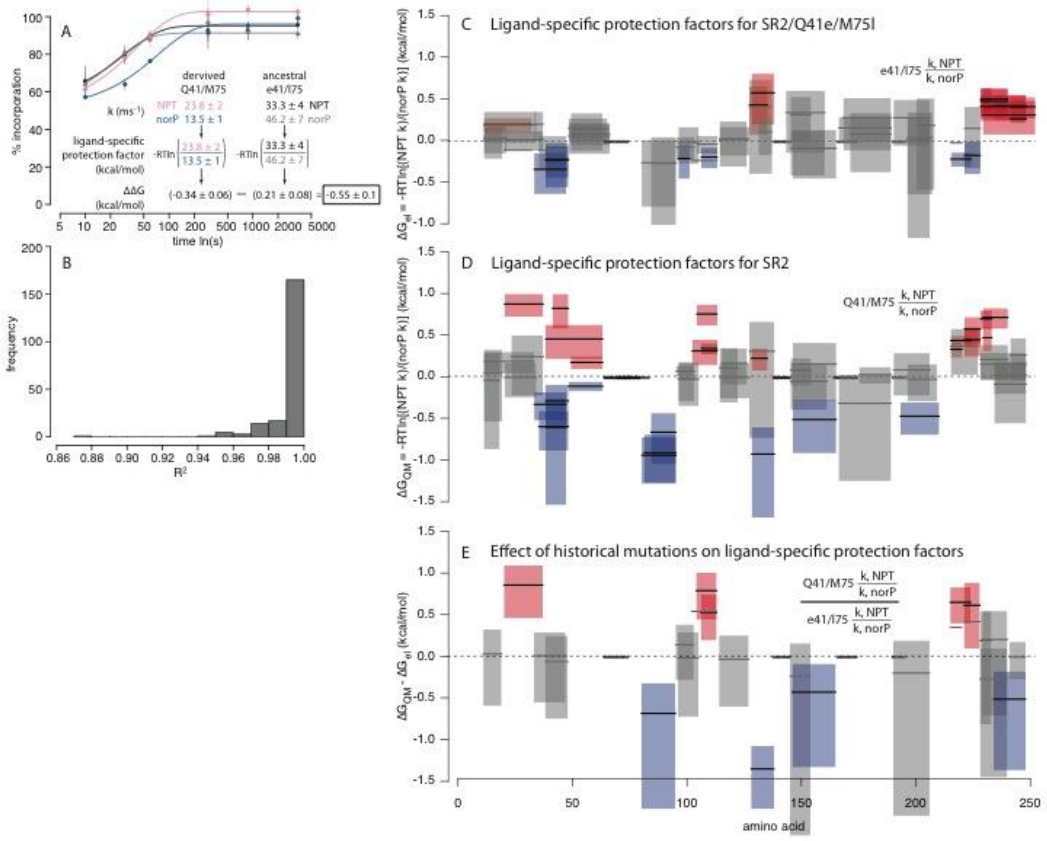


Figure 3.14: Historical mutations cause increased peptide solvent exchange in a ligand-dependent manner.

A) Analysis of deuterium incorporation kinetics for a representative peptide seen in all four protein/ligand pairs. Points shown mean incorporation for replicate experiments; vertical lines show standard errors; solid lines are fits of a first-order kinetic model. Equations show how extracted exchange rates are converted to ligand-specific protection factor (panels C and D) and finally to mutational effects on those values (panel E). B) Distribution of R^2 values for fits for all peptide time-series. C) Ligand-specific protection factors for ancestral state (SR2/Q41e/M75l). Lines denote peptide coverage; vertical position denotes ligand-specific protection factor – RTln[(rate with NPT)/(rate with norP)]; shaded boxes show propagated fit 95% confidence interval. Color denotes relative rate: less protection with NPT than norP (red); more protection

with NPT than norP (blue); or no significant difference (grey). Asterisks denote peptides that exhibited no incorporation. D) Ligand-specific protection factors for derived state (SR2). Colors/elements as in panel C. E) Effect of historical mutations on ligand-specific protection factors. Elements as in panel C. Colors denote peptides for which historical mutations: decrease protection with NPT more than norP (red); increase protection with NPT more than norP (blue); have no effect (grey).

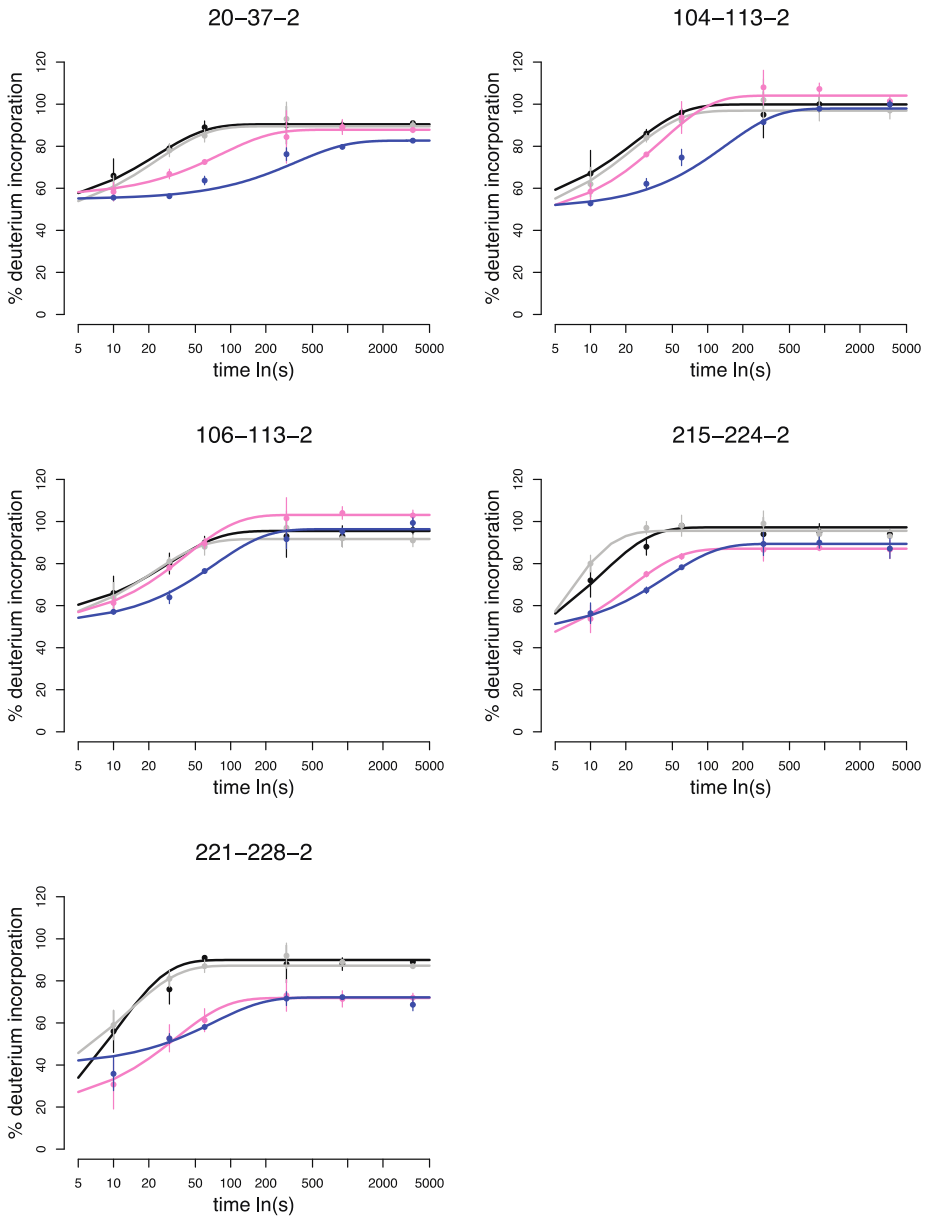


Figure 3.15: Model fits to incorporation vs. time data for the five peptides which exhibited decreased NPT-specific protection factors in the derived state.

Series show incorporation curves for e41/l75 NPT (black); e41/l75 norP (gray); Q41/M75 NPT (pink); and Q41/M75 norP (blue). Numbers above each panel denote [starting residue]-[ending residue]-[peptide charge].

Steroid Name in Text	Abbreviation	Pubmed Compound Identifier Number (CID)	Steraloids Catalogue #	Steraloids name
11-Deoxycorticosterone	11-DOC	6166		
11-Deoxycortisol		440707		
Corticosterone		5753		
Cortisol		5754		
Aldosterone		5839		
Progesterone		5994		
17-Hydroxyprogesterone		6238		
19-Norprogesterone		228864		
4-Pregnenolone		-	Q3540-000	4-PREGNEN-3β-OL-20-ONE
5-Pregnenolone		8955		
20α Hydroxyprogesterone		440204		
20β Hydroxyprogesterone		92747		
Testosterone		6013		
Dihydrotestosterone	DHT	10635		
4-Androstenediol		136297		
5-Androstenediol		10634		
19-Nortestosterone		9904		
Bolandiol		16141		
Estradiol		5757		
Estrone		5870		
Estriol		5756		
4-Androstenedione		6128		
19-nor-1, 3, 5(10)-pregnatriene-3-ol-20-one	NPT	-	N0600-000	19-NOR-1, 3, 5(10)-PREGNATRIEN-3-OL-20-ONE
Cholesterol		5997		
Diethylstilbestrol	DES	448537		
Genistein		5280961		
4-Hydroxytamoxifen		449459		
ICI 182,780	fulvestrant	104741		

Table 3.1: Pubmed compound identifier (CID) numbers for cholesterol and the synthetic and natural steroid hormones tested in this study.

For the two substances that a CID number was not available, the Steraloids catalog number and name are provided.

	ancestors		bits		conserved		
residue	ancSR1	ancSR2	ER	naSR	ER	naSR	diagnostic
-1	P	N	2.28	1.95			
0	A	A	2.83	1.76			
1	N	P	2.98	3.00			
2	Q	S	3.47	1.76			
3	L	L	2.26	2.23			
4	I	I	2.67	2.52			
5	S	S	1.66	2.39			
6	A	I	1.77	2.68			
7	L	L	3.30	4.32		L	
8	L	Q	2.34	2.81			
9	E	A	2.37	1.97			
10	**A**	**I**	**4.05**	**4.32**	**A**	**I**	**y**
11	E	E	4.16	4.32	E	E	
12	P	P	3.95	4.32	P	P	
13	P	E	4.14	2.77	P		
14	V	V	1.15	2.98			
15	L	V	2.48	2.98			
16	Y	Y	2.39	2.59			
17	S	A	2.56	3.76		A	
18	R	G	1.74	4.23		G	
19	H	Y	1.89	3.14			
20	D	D	2.25	4.32		D	
21	P	N	1.67	2.17			
22	S	T	2.01	2.17			
23	K	Q	2.12	2.24			
24	P	P	4.23	3.97	P	P	
25	D	D	2.57	3.65		D	
26	T	T	3.19	3.25			
27	E	T	3.95	2.12	E		
28	A	N	2.66	1.55			
29	H	Y	2.52	1.38			
30	L	L	2.88	3.64		L	
31	**M**	**L**	**3.98**	**3.51**	**M**	**L**	**y**

32	T	S	2.20	3.28			
33	S	S	3.08	3.19			
34	L	L	4.23	4.32	L	L	
35	**T**	**N**	**3.98**	**4.32**	**T**	**N**	**y**
36	N	R	2.65	2.19			
37	L	L	3.50	4.32	L	L	
38	A	A	4.14	3.23	A		
39	D	E	4.23	3.19	D		
40	R	K	3.24	3.27			
41	**E**	**Q**	**4.32**	**4.32**	**E**	**Q**	**y**
42	L	L	4.32	2.96	L		
43	V	V	4.14	2.93	V		
44	H	S	3.37	1.31			
45	I	V	3.98	3.18	M		
46	**I**	**V**	**4.23**	**4.32**	**I**	**V**	**y**
47	N	K	2.02	4.05		K	
48	W	W	4.32	4.32	W	W	
49	A	A	4.23	3.80	A	A	
50	K	K	4.32	4.32	K	K	
51	K	A	3.29	2.18			
52	I	L	2.82	3.32			
53	P	P	4.32	4.23	P	P	
54	G	G	4.32	4.32	G	G	
55	Y	F	4.10	4.23	F	F	
56	S	R	2.32	3.87		R	
57	D	N	2.52	3.48		N	
58	L	L	4.10	4.10	L	L	
59	S	H	2.38	3.58		H	
60	L	L	4.23	2.72	L		
61	N	D	1.91	3.51		D	
62	D	D	4.14	4.32	D	D	
63	Q	Q	4.07	4.14	Q	Q	
64	**V**	**M**	**3.95**	**3.51**	**V**	**M**	**y**
65	H	T	2.41	3.17			
66	L	L	4.23	3.08	L		
67	L	I	4.14	3.44	L		
68	Q	Q	4.05	4.23	E	Q	
69	S	Y	3.25	2.73			
70	C	S	2.69	4.01		S	
71	W	W	4.32	4.23	W	W	
72	M	M	3.83	3.83	L	M	
73	E	G	4.00	2.25	E		
74	L	L	3.13	3.34			
75	**L**	**M**	**4.32**	**3.76**	**L**	**M**	**y**
76	I	A	3.87	2.59	M		
77	L	F	2.68	4.23		F	
78	G	A	4.14	3.12	G		
79	L	M	4.05	3.49	L	L	

80	A	G	2.56	3.36			
81	W	W	4.14	4.32	W	W	
82	R	R	4.23	4.23	R	R	
83	S	S	4.32	4.32	S	S	
84	M	Y	2.59	3.56		Y	
85	D	K	1.98	2.63			
86	H	H	2.77	2.46			
87	E	T	3.89	2.24	P		
88	G	N	4.14	3.25	G		
89	-	G	0.00	2.74			
90	-	Q	0.00	2.09			
91	K	M	3.88	3.12	K		
92	L	L	4.05	4.23	L	L	
93	V	Y	2.69	3.44			
94	F	F	4.23	4.32	F	F	
95	A	A	3.22	4.32		A	
96	P	P	3.12	4.32		P	
97	D	D	3.39	4.32		D	
98	L	L	3.86	4.32	L	L	
99	I	I	1.81	3.42			
100	L	F	3.98	3.04	L		
101	D	N	3.04	4.01		N	
102	**R**	**E**	**3.98**	**3.51**	**R**	**E**	**y**
103	E	Q	2.14	1.89			
104	Q	R	3.06	3.91		R	
105	S	M	3.73	4.32	G	M	
106	K	Q	1.56	2.07			
107	C	-	4.23	0.00	C		
108	V	Q	4.05	1.55	V		
109	A	S	3.21	3.36			
110	G	A	4.14	1.60	G		
111	M	M	2.55	3.80		M	
112	E	Y	2.02	3.29			
113	E	D	4.00	2.53	E		
114	I	L	4.16	2.69	I		
115	C	C	3.95	4.23	F	C	
116	D	Q	3.98	1.51	D		
117	Q	G	3.89	1.74	M		
118	I	M	3.94	4.23	L	M	
119	L	Q	3.56	2.04	L		
120	E	Q	3.64	1.85	A		
121	I	I	2.79	2.68			
122	A	S	2.57	3.14			
123	S	Q	2.98	2.17			
124	Q	E	3.79	3.42	R		
125	F	F	3.61	3.57	F	F	
126	R	V	4.05	1.47	R		
127	E	R	2.37	2.06			

128	L	L	3.36	4.16		L	
129	K	Q	2.82	3.91		Q	
130	V	V	4.01	2.74	L		
131	Q	T	2.87	3.26			
132	K	Q	1.82	1.80			
133	E	E	3.43	3.20			
134	E	E	4.32	4.23	E	E	
135	F	F	3.16	3.37			
136	V	L	3.10	3.57		L	
137	C	C	4.10	3.57	C	C	
138	**L**	**M**	**3.77**	**4.23**	**L**	**M**	**y**
139	K	K	4.16	4.16	K	K	
140	A	A	3.24	2.82			
141	I	L	2.99	3.97		L	
142	T	L	3.85	3.82	I	L	
143	L	L	4.23	3.95	L	L	
144	L	L	3.91	3.27	L		
145	**N**	**S**	**4.23**	**3.48**	**N**	**S**	**y**
146	S	-	3.85	0.00	S		
147	G	-	2.79	0.00			
148	V	-	2.05	0.00			
149	F	-	2.66	0.00			
150	T	-	2.70	0.00			
151	F	-	2.46	0.00			
152	L	-	2.39	0.00			
153	S	T	2.50	2.78			
154	S	V	1.71	2.94			
155	D	P	2.06	4.32		P	
156	A	K	1.26	2.55			
157	K	E	2.72	3.26			
158	R	G	2.02	4.32		G	
159	L	L	2.77	4.16		L	
160	E	K	2.20	3.69		K	
161	D	S	2.41	3.10			
162	H	Q	1.75	4.23		Q	
163	E	A	1.26	1.82			
164	Q	S	2.00	1.40			
165	V	F	2.61	4.32		F	
166	Q	D	2.03	3.35			
167	K	E	1.64	4.00		E	
168	L	M	2.43	2.78			
169	Q	R	3.75	4.32	L	R	
170	D	M	3.28	1.96			
171	K	N	1.63	2.62			
172	I	Y	2.95	4.32		Y	
173	**T**	**I**	**3.48**	**4.32**	**T**	**I**	**y**
174	D	K	4.10	2.88	D		
175	**A**	**E**	**3.48**	**4.23**	**A**	**E**	**y**

176	L	L	4.16	4.32	L	L	
177	V	N	2.81	1.86			
178	D	R	2.39	3.32			
179	T	A	1.37	2.07			
180	V	I	3.21	3.30			
181	A	A	2.69	1.71			
182	K	K	2.84	1.82			
183	S	K	2.19	1.87			
184	H	E	3.70	1.45	G		
185	P	N	2.06	1.80			
186	D	N	2.57	2.95			
187	S	S	1.56	2.21			
188	P	A	3.10	1.78			
189	Q	Q	3.62	2.00	Q		
190	Q	N	3.80	2.34	Q		
191	S	W	2.07	2.69			
192	R	Q	2.18	3.04			
193	R	R	4.32	4.23	R	R	
194	L	F	3.24	3.80		F	
195	A	Y	3.71	3.44	A		
196	Q	Q	2.74	4.06		Q	
197	L	L	4.05	4.32	L	L	
198	**L**	**T**	**4.05**	**4.23**	**L**	**T**	**y**
199	M	K	3.09	3.33			
200	L	L	3.45	3.67		L	
201	L	L	4.32	3.69	L	L	
202	**S**	**D**	**3.91**	**4.23**	**S**	**D**	**y**
203	**H**	**S**	**4.14**	**3.51**	**H**	**S**	**y**
204	I	M	3.24	2.92			
205	R	H	4.32	3.24	R		
206	Q	D	3.96	2.11	H		
207	V	L	2.39	2.13			
208	S	V	4.14	3.38	S		
209	S	G	3.79	1.76	N		
210	K	G	4.07	1.76	K		
211	**G**	**L**	**4.07**	**4.13**	**G**	**L**	**y**
212	I	L	3.51	3.10	M		
213	E	Q	3.18	2.05			
214	H	F	4.14	3.19	H		
215	L	C	4.05	3.22	L		
216	Y	F	2.32	3.41			
217	S	Y	2.13	1.93			
218	M	T	3.91	3.18	M		
219	**K**	**F**	**4.05**	**3.73**	**K**	**F**	**y**
220	S	V	2.83	2.17			
221	E	Q	3.98	2.23	K		
222	G	S	3.66	3.13	N		
223	-	Q	0.00	2.40			

224	-	A	0.00	2.20			
225	-	L	0.00	2.92			
226	R	S	2.38	2.20			
227	V	V	4.14	3.66	V	V	
228	P	E	3.74	2.64	P		
229	L	F	3.34	4.04		F	
230	**Y**	**P**	**4.05**	**4.32**	**Y**	**P**	**y**
231	**D**	**E**	**4.23**	**3.51**	**D**	**E**	**y**
232	**L**	**M**	**4.23**	**4.32**	**L**	**M**	**y**
233	L	L	4.23	2.76	L		
234	L	V	4.05	2.38	L		
235	E	E	4.23	4.32	E	E	
236	**M**	**I**	**4.14**	**3.66**	**M**	**I**	**y**
237	**L**	**I**	**4.23**	**4.09**	**L**	**I**	**y**
238	D	S	3.40	3.00			
239	A	A	4.07	2.29	A		
240	Q	Q	3.28	3.48		Q	
241	T	L	2.23	2.81			
242	S	P	1.98	4.22		P	
243	Q	K	2.34	3.80		K	
244	S	V	1.64	2.42			
245	P	L	1.57	2.32			

Table 3.2: Conservation analysis of extant naSRs and ERs.

The sequences of the ancestors were compared to the conserved sequences of their dependents. The "bits" column shows the information content of the columns in alignments of modern ERs and naSRs. If the information is > 3.5 bits, the position is considered conserved. Diagnostic sites were defined as those where the ER and naSR clades possessed different, conserved states at that site and the transition between those states occurred over the interval from ancSR1 to ancSR2. The diagnostic sites are highlighted in bold. The residues selected for analysis—e41Q and l75M—are highlighted in yellow.

el/N06	1	2
1	---	2.78
2	2.78	---

el/nor P	1	2
1	---	5.37
2	5.37	---

QM/N06	1	2	3	4	5	6	7
1	---	4.95	3.06	5.36	3.46	4.21	3.33
2	4.95	---	4.95	>5.37	>5.37	4.54	>5.37
3	3.12	4.21	---	2.96	>5.37	4.41	4.30
4	5.36	5.36	2.98	---	>5.37	3.65	>5.37
5	3.43	>5.37	>5.37	>5.37	---	>5.37	>5.37
6	4.00	4.71	4.21	3.72	>5.37	---	>5.37
7	3.26	>5.37	5.36	>5.37	>5.37	>5.37	---

QM/nor P	1	2	3
1	---	2.95	>5.37
2	2.96	---	3.09
3	>5.37	3.02	---

QM/npr	1	2	3	4
1	---	4.55	2.60	>5.37
2	4.72	---	3.15	>5.37
3	2.59	3.09	---	4.96
4	>5.37	>5.37	5.37	---

Entries are $\Delta G_{row} + \Delta G_{row' column}$ in kcal/mol. For a perfectly additive system, i ! j = j ! i. Additivity holds to within 5% for all but two transitions (QM/N06 3 ! 7 (4.30) versus 7 ! 3 (5.36) and QM/N06 2 ! 3 (4.95) versus 3 ! 2 (4.21)). The maximum barrier height accessible would be a transition only observed once (-RT*ln(1/9200 frames) = 5.37 kcal/mol). Any transition that was not observed thus has a barrier height >5.37 kcal/mol.

Table 3.3: Simulations display additivity: absolute free energies of barriers are the same for i→j versus j→i transitions.

el/N06	1	2
1	6517	80
2	80	2205

el/norP	1	2
1	7784	1
2	1	1246

QM/N06	1	2	3	4	5	6	7
1	3168	2	49	1	25	7	31
2	2	1632	2	0	0	4	0
3	44	7	1409	58	0	5	6
4	1	1	56	1242	0	18	0
5	26	0	0	0	521	0	0
6	10	3	7	16	0	307	0
7	35	0	1	0	0	0	76

QM/norP	1	2	3
1	4459	62	0
2	61	2481	49
3	0	55	2102

QM/npr	1	2	3	4
1	5645	4	109	0
2	3	1479	43	0
3	111	47	1241	2
4	0	0	1	224

To generate this matrix, we walked through the frames of the trajectories for each protein/ligand pair, then recorded M[state(frame *i*),state(frame (*i*+*1*))]. This is a subset of the total matrix representing the transitions between the states that, together, account for ! 95% of the frames for each protein/ligand pair.

Table 3.4: Transition matrices for top 95% of observed states with 2 kcal/mol energy cutoff.

description	Model[1]	# free param.	global RMSD	AIC corrected Δln(L)	AIC weight
1	$y(t) = M(1 - e^{-kt})$	2	56354.8	1	0.00
2	$y(t) = M(2 - e^{-k_1 t} - e^{-k_2 t})$	3	6544.7	49,087	0.00
3	$\boldsymbol{y(t) = M_1 + M_2(1 - e^{-k_2 t})}$	**3**	**908.2**	**55,653**	**1.00**
4	$y(t) = M_1(1 - e^{-k_1 t}) + M_2(1 - e^{-k_2 t})$	4	469.4	55,471	0.00

1: M: deuterium incorporation at infinite time (%); k: exchange rate (s^{-1}); t: time (s)

2: AIC-corrected Δln(L) = num_peptides*(num_par_i – num_par_1) – ($RMSD_i$ – $RMSD_1$)

3: AIC weight = exp $(AIC_i)/\sum_j \exp(AIC_j)$

Model #3 was used for the analyses.

Table 3.5: HDX-MS kinetics model selection.

References

1. Golding GB, Dean AM (1998) The structural basis of molecular adaptation. *Mol Biol Evol* 15:355–369.

2. Dean AM, Thornton JW (2007) Mechanistic approaches to the study of evolution: the functional synthesis. *Nat Rev Genet* 8:675–688.

3. Harms MJ, Thornton JW (2010) Analyzing protein structure and function using ancestral gene reconstruction. *Curr Opin Struct Biol* 20:360–366.

4. Fisher RA (1990) *The Genetical Theory of Natural Selection: A Complete Varorium Edition* (Oxford University Press, Oxford).

5. Charlesworth D, Charlesworth B (1975) Theoretical genetics of Batesian mimicry I. single-locus models. *J Theor Biol* 55:283–303.

6. Gould SJ (1977) The return of hopeful monsters. *Natural History* 86:24–30.

7. Orr HA (2005) The genetic theory of adaptation: a brief history. *Nat Rev Genet* 6:119–127.

8. Eyre-Walker A, Keightley PD (2007) The distribution of fitness effects of new mutations. *Nat Rev Genet* 8:610–618.

9. Wang J, Verkhivker GM (2003) Energy landscape theory, funnels, specificity, and optimal criterion of biomolecular binding. *Phys Rev Lett* 90:188101.

10. McCammon JA (1998) Theory of biomolecular recognition. *Curr Opin Struct Biol* 8:245–249.

11. Mobley DL, Dill KA (2009) Binding of small-molecule ligands to proteins: "what you see" is not always "what you get". *Structure* 17:489–498.

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

13. Bridgham JT, Carroll SM, Thornton JW (2006) Evolution of hormone-receptor complexity by molecular exploitation. *Science* 312:97–101.

14. Yokoyama S, Tada T, Zhang H, Britt L (2008) Elucidation of phenotypic adaptations: Molecular analyses of dim-light vision proteins in vertebrates. *Proc Natl Acad Sci U S A* 105:13480–13485.

15. Field SF, Matz MV (2010) Retracing evolution of red fluorescence in GFP-like proteins from Faviina corals. *Mol Biol Evol* 27:225–233.

16. Lynch VJ, May G, Wagner GP (2011) Regulatory evolution through divergence of a phosphoswitch in the transcription factor CEBPB. *Nature* 480:383–386.

17. Bentley PJ (1998) *Comparative vertebrate endocrinology* (Cambridge University Press, Cambridge, UK ; New York).

18. Gronemeyer H, Gustafsson JA, Laudet V (2004) Principles for modulation of the nuclear receptor superfamily. *Nat Rev Drug Discov* 3:950–964.

19. Eick GN, Colucci JK, Harms MJ, Ortlund EA, Thornton JW (2012) Evolution of minimal specificity and promiscuity in steroid hormone receptors. *PLoS Genet* 8:e1003072.

20. Tanenbaum DM, Wang Y, Williams SP, Sigler PB (1998) Crystallographic comparison of the estrogen and progesterone receptor's ligand binding domains. *Proc Natl Acad Sci U S A* 95:5998–6003.

21. Zhang Z, Smith DL (1993) Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation. *Protein Sci* 2:522–531.

22. Alexander PA, He Y, Chen Y, Orban J, Bryan PN (2009) A minimal sequence code for switching protein structure and function. *Proc Natl Acad Sci U S A* 106:21149–21154.

23. Weber G (1975) Energetics of ligand binding to proteins. *Adv Protein Chem* 29:1–83.

24. Worth CL, Gong S, Blundell TL (2009) Structural and functional constraints in the evolution of protein families. *Nat Rev Mol Cell Biol* 10:709–720.

25. Oostenbrink C, Villa A, Mark AE, van Gunsteren WF (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J Comput Chem* 25:1656–1676.

26. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ (2005) GROMACS: fast, flexible, and free. *J Comput Chem* 26:1701–1718.

27. Anisimova M, Gascuel O (2006) Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol* 55:539–552.

CHAPTER 4: X-RAY CRYSTAL STRUCTURE OF THE ANCESTRAL 3-KETOSTEROID RECEPTOR – PROGESTERONE – MIFEPRISTONE COMPLEX SHOWS MIFEPRISTONE BOUND AT THE COACTIVATOR BINDING SURFACE

Jennifer K. Colucci[1] and Eric A. Ortlund[1c]

[1]*Department of Biochemistry and Winship Cancer Institute, Emory University School of Medicine, Atlanta, GA 30033, USA*

Colucci JK, Ortlund EA (2013) X-Ray Crystal Structure of the Ancestral 3-Ketosteroid Receptor–Progesterone–Mifepristone Complex Shows Mifepristone Bound at the Coactivator Binding Interface. PLoS ONE 8(11).

Steroid receptors are key regulators of metazoan physiology and are heavily targeted for pharmaceutical intervention. However, most currently approved pharmaceuticals for SRs have off-target pharmacology due to the high conservation of three-dimensional structures within the SR family. A better understanding of the protein-ligand relationship may enable the design of better pharmaceutical ligands. To investigate this, I attempted to crystalize ancSR2 with the SR antagonist mifepristone. This effort resulted in a higher resolution ancSR2 – progesterone structure due to lack of ligand exchange. Fortuitously, in the crystal structure we found mifepristone bound at the activation function 2 surface (which is responsible for mediating the coregulator interaction) and compared this interaction with that of other SR coactivator cleft binding small molecules. This work was accepted for publication in PLoS One in October 2013.

---

[c] Conceived and designed the experiments: JKC EAO. Performed the experiments: JKC. Analyzed the data: JKC EAO. Wrote the paper: JKC EAO.

Abstract

Steroid receptors are a subfamily of nuclear receptors found throughout all metazoans. They are highly important in the regulation of development, inflammation, and reproduction and their misregulation has been implicated in hormone insensitivity syndromes and cancer. Steroid binding to SRs drives a conformational change in the ligand binding domain that promotes nuclear localization and subsequent interaction with coregulator proteins to affect gene regulation. SRs are important pharmaceutical targets, yet most SR-targeting drugs have off-target pharmacology leading to unwanted side effects. A better understanding of the structural mechanisms dictating ligand specificity and the evolution of the forces that created the SR-hormone pairs will enable the design of better pharmaceutical ligands. In order to investigate this relationship, we attempted to crystallize the ancestral 3-ketosteroid receptor (ancSR2) with mifepristone, a SR antagonist. Here, we present the x-ray crystal structure of the ancestral 3-keto steroid receptor (ancSR2)–progesterone complex at a resolution of 2.05 Å. This improves upon our previously reported structure of the ancSR2-progesterone complex, permitting unambiguous assignment of the ligand conformation within the binding pocket. Surprisingly, we find mifepristone, fortuitously docked at the protein surface, poised to interfere with coregulator binding. Recent attention has been given to generating pharmaceuticals that block the coregulator binding site in order to obstruct coregulator binding and achieve tissue-specific SR regulation independent of hormone binding. Mifepristone's interaction with the coactivator cleft of this SR suggests that it may be a useful molecular scaffold for further coactivator binding inhibitor development.

Introduction

Steroid hormones play a crucial role in all but the most basic metazoans, orchestrating the cell-cell communication required to coordinate development, growth, metabolism, immunity and more (*45*). These hormones are small lipophilic molecules that act directly on a class of transcription factors termed steroid hormone receptors to mediate their down stream effects. Misregulation of steroid signaling leads to metabolic, immune, and neoplastic diseases. Thus, the steroid receptors (SRs), consisting of the estrogen receptor (ER), progesterone receptor (PR), androgen receptor (AR), mineralocorticoid receptor (MR), and glucocorticoid receptor (GR), are highly targeted for therapeutic intervention.

SRs have a modular domain architecture consisting of a highly variable N-terminal domain, a DNA-binding domain (DBD), a short hinge region, and a ligand binding domain (LBD) (*61*). ApoSRs are sequestered in the cytoplasm by heat shock proteins (HSPs), which hold them in a ligand-ready state; they are activated when a steroid hormone binds the ligand binding pocket, remodeling the HSP complex and triggering nuclear import (*61, 68*). Agonist binding drives a conformational change, whereby helices 3, 4, and 12 (H3, H4, H12, respectively) create a docking surface for coregulatory proteins termed the activation function surface (AF-H) (*69, 70*). Antagonist binding on the other hand prevents proper packing of H12 against H3 and H4 favoring corepressor interaction. Mutations within the AF-H can disrupt coregulator interaction causing ligand insensitivity (*71-73*). Coactivators, interact with SRs via helical LXXLL (L—leucine, X—any amino acid) motifs, and act as intermediaries to RNA Polymerase II and other transcriptional machinery (*69, 70*). The recruitment of any given coregulatory protein exhibits both ligand- and tissue-specificity dictated by the available coregulator pools and state of the cell (*74*). Coregulatory proteins act as conduits to all further transcriptional activation or repression, thus their regulation remains a highly desirable pharmaceutical target.

Recent efforts to achieve tissue-specific SR-mediated regulation has been focused on developing compounds to block the SR-coactivator interface to modulate certain SR-mediated gene activity. These small molecules, dubbed Coactivator Binding Inhibitors (CBIs), are effective at competing for coactivator binding space and altering downstream transcription (*75*). These compounds typically contain heterocyclic cores and possess substituents that mimic the three trussing leucine residues of coactivator proteins (*75*).

In the absence of HSPs, SRs are inherently unstable, complicating efforts to identify the mechanisms driving ligand specificity and our ability to build robust structure-function relationships. Recent studies have utilized ancestral steroid receptors (ancSRs) to identify the molecular mechanisms that dictated the evolution of ligands specificity among SRs (*64, 76, 77*). AncSRs display a greater tolerance to mutation while preserving faithful ligand specificity and activation in cells (*20, 63*). These qualities make ancSRs useful tools to study the selectivity and mechanisms of action of SR-targeting pharmaceuticals (*78*).

Here we report the structure of the ancestral 3-ketosteroid receptor, ancestral steroid receptor 2 (ancSR2), the ancestor of the AR, PR, MR, and GR. This structure shows the ancSR2—progesterone complex with the SR antagonist mifepristone bound at two surface sites (*67*). This structure (2.05 Å) improves the resolution of a previously published ancSR2—progesterone complex (2.75 Å) (*67*). Surprisingly, one of the bound mifepristone molecules occupies the coactivator binding space, suggesting a potential use of this drug as a molecular framework for further CBI development. A second bound mifepristone molecule sits at the base of the receptor and interacts with crystallographic symmetry mates; this molecule alters the crystal packing conditions from a previously published structure of the ancSR2—progesterone complex (PDB accession code: 4FN9) (*67*).

Materials and Methods

Reagents.

Chemicals were purchased from Sigma (St. Louis, MO) or Fisher (Hampton, NH). Mifepristone was purchased from Tocris Biosciences (Bristol, UK). Progesterone was purchased from MP Biomedicals (Santa Ana, CA). The vector for His tagged TEV was a gift from David Waugh (National Cancer Institute). The pLIC_MBP vector was a gift from John Sondek (UNC, Chapel Hill). The ancSR2 LBD was resurrected using well-established protocols and was kindly provided by Dr. Joseph Thornton (University of Oregon, OR) (*67*).

Expression and Purification.

AncSR2 LBD was expressed as a 6xHis-MBP fusion protein in BL21(DE3) *E. coli*. Cultures (1.0 L in TB) were grown to an OD_{600} of 0.8 and induced with a final concentration of 400 μM IPTG and 50 μM progesterone at 30 °C for 4 hours. Ancestral SRs, like the extant receptors are inherently unstable in the absence of ligand and adding ligand at induction is required to for soluble overexpression of the recombinant protein. Cell mass was collected by centrifugation at 4 krpm for 20 minutes, resuspended in 150 mM NaCl, 20 mM Tris HCl (pH 7.4), 5 % glycerol, 25 mM imidazole, 0.1 % PMSF and lysed using sonication on ice. The 6xHis-MBP-ancSR2-MBP was initially purified using Ni^{2+} affinity chromatography (HisTrap column, GE Healthcare). Fractions containing ancSR2-MBP were identified by denaturing polyacrylamide gel electrophoresis (SDS-PAGE), pooled and dialyzed against 150 mM NaCl, 20 mM Tris HCl (pH 7.4), 5 % glycerol and 1 mg TEV. Following TEV cleavage, the 6xHis-tagged MBP was removed by an additional Ni^{2+} affinity column. The flow though containing untagged ancSR2 LBD was concentrated using an Amicon Ultra 10K centrifugal filter device (Millipore), concentrated to 3-5 mg ml^{-1}, and dialyzed against 150 mM sodium chloride, 20 mM Tris HCl (pH 7.4), and 5 % glycerol. The final purity of the ancSR2 LBD was assessed using SDS-PAGE. In an attempt to exchange progesterone for mifepristone in the ligand binding pocket (LBP),

mifepristone (50 μM; approximately 500-fold molar excess) was added to the ancSR2-progesterone complex for 30 minutes at 4 °C and then centrifuged at 14 K rpm for 1 minute to clarify the solution prior to crystallization trails.

Crystallization, data collection, structure determination and refinement.

Orthorhombic crystals of the ternary ancSR2 LBD–progesterone–mifepristone complex were grown by hanging drop vapor diffusion at 22 °C from solutions containing 1.0 μL of protein at 3-5 mg mL^{-1} protein and 1.0 μL of the following crystallant 0.8 M $MgSO_4$, 10 % glycerol, 0.1 M MES (pH 6.0) (Figure 4.1A). Crystals of the ancSR2 complex grew in $P2_12_12_1$ space group with one monomer in the asymmetric unit. Crystals were cryoprotected by transient soaking in crystallant containing 20 % glycerol and were flash-cooled in liquid N_2 at 100 K. Data to 2.05 Å resolution were collected at the South East Regional Collaborative Access Team (SER-CAT) 22-BM at the Advanced Photon Source (APS) at Argonne National Laboratory in Chicago, IL using a wavelength of 0.97 Å. Data were processed and scaled with HKL2000 (Table 1) (*65*). Initial phases were determined using the previously published ancSR2–progesterone complex (PBD accession code: 4FN9) as the initial search model in Phenix-MR v1.7.1 (*46, 79*). Residues 2 through 248 were modeled and $R_{factors}$ for the final model are 17.9% and 21.2% for R_{work} and R_{free} respectively. MolProbity was used for model validation, indicating that 98.8% of the residues fall in the most favored regions of the Ramachandran plot with none in disallowed regions (*80, 81*). The overall MolProbity score was 1.64, placing the structure in the 100^{th} percentile for overall geometric quality among protein crystal structures of comparable resolution (*80, 81*).

Reporter Gene Assays.

AncSR2 LBD, was cloned into the Gal4-DBD-pSG5 vector; 31 amino acids of the glucocorticoid receptor hinge containing the nuclear localization signal-1 were inserted between the DBD and the LBD to ensure nuclear localization and conformational independence of the two domains (*67*). The hormone-dependent transcriptional activity of resurrected ancestral receptors

and their variants was assayed using a luciferase reporter system. CHO-K1 cells were grown in 96-well plates and transfected with 1.0 ng of receptor plasmid, 100 ng of a UAS-driven firefly luciferase reporter (pFRluc), and 0.1 ng of the constitutive phRLtk Renilla luciferase reporter plasmid, using Lipofectamine and Plus Reagent in OPTIMEM (Invitrogen). After 4 hours, transfection medium was replaced with phenol-red-free αMEM supplemented with 10% dextran-charcoal stripped FBS (Hyclone). After overnight recovery, cells were incubated in triplicate with the hormone of interest from 10^{-12} to 10^{-6} M for 24 hours, then assayed using Dual-Glo luciferase (Promega). Firefly luciferase activity was normalized by *Renilla* luciferase activity. Luminescence was read using a Synergy 4 microplate reader (BioTek). Dose-response relationships were estimated using nonlinear regression in Prism4 software (GraphPad Software, Inc.); fold increase in activation was calculated relative to vehicle-only (DMSO) control.

Results

Overall Structure.

Mifepristone is a highly potent SR antagonist, with strong antiprogestagen, antiglucocorticoid, and antiandrogen properties (*82, 83*). Clinically, it is used as an abortifacient, emergency contraceptive, and as treatment for Cushing's Syndrome (*82, 84*). Mifepristone has been shown to bind SRs within their LBP, leading to the recruitment of corepressor proteins (*85*) or the stabilization of the SR-heat shock protein 90 (HSP90) complex (*86*). To gain insight into mechanism by which mifepristone represses 3-keto SRs, we tested mifepristone's effect on ancSR2-driven gene expression via a luciferase reporter assay. Surprisingly, mifepristone activates ancSR2 with an EC_{50} of 56 nM (Figure 4.1B). Although this is several orders of magnitude less potent than progesterone activation, the EC_{50} is on par with the EC_{50} value for other SR-targeting pharmaceuticals. This weak or partial agonism has been observed for GR in certain cell types and at high enough receptor concentration (*87, 88*). Thus, mifepristone agonism may be a relic ancSR2. To gain insight into how mifepristone specifically activates ancSR2 while repressing all modern 3-keto SRs, we attempted generate a structure of the ancSR2-mifepristone complex via ligand exchange.

The crystal structure of the ancSR2–progesterone–mifepristone complex (PDB accession number 4LTW) shows that the receptor maintains the canonical steroid receptor fold consisting of a three-layered alpha-helical bundle with four beta strands (Figure 4.2). Our previously published lower resolution structure of the ancSR2—progesterone complex contained two ancSR2 monomers in the asymmetric unit within the $P2_12_12_1$ space group (*67*). Despite highly similar crystallization conditions, addition of mifepristone altered crystal packing and reduced the number of monomers within the asymmetric unit from two to one.

Mifepristone binds at two distinct surface sites.

Surprisingly, F_o-F_c omit electron density shows clear evidence for the presence of progesterone within the LBP (Figure 4.3) suggesting that mifepristone failed to exchange with this steroid hormone *in vitro* despite being in nearly 500-fold molar excess. However, initial F_o-F_c electron density clearly showed two well-ordered mifepristone molecules located at distinct surface sites on the receptor which we refer to as "site-one" and "site-two" (Figure 4.3). Site-one mifepristone makes extensive hydrophobic contacts along helices 3, 7, and 10 of the monomer within the asymmetric unit (AU) and with helices 9 and 10 and the C-terminus of a crystallographic symmetry mate (Figure 4.4). Site-one mifepristone buries a total surface area of 413.5 $Å^2$ between both the monomer located in the AU and the crystallographic symmetry mate (*89*). Analyses using the Proteins, Interfaces, Structures, and Assemblies (PISA) server shows a complex significance score of 0.00, suggesting that this interaction plays a role in crystal packing but is not biologically significant (*89*).

Surprisingly, a second surface mifepristone was bound to the interface of helices 3, 4, and 12, which is used to recruit coregulator proteins to drive transcriptional activation. Superposition with the ancestral corticoid receptor (ancCR)–deoxycorticosterone–small heterodimer partner (SHP; NR0B2) NRBox1 peptide complex (PDB accession code 2Q3Y), the most closely-related SR, reveals that mifepristone occupies the same position as a coregulator peptide (Figure 4.5A) (*64*). Thus, site-two mifepristone would compete with a coactivator for binding to the coactivator cleft. Site-two mifepristone is coordinated by extensive hydrophobic interactions at the coactivator cleft with an interface surface area of 402.6 $Å^2$ (*89*). Twelve of the thirteen residues making these hydrophobic contacts are conserved across 3-ketoseteroid receptors (Figure 4.5B). The carbon 17-hydroxyl group of mifepristone makes a hydrogen bond with the amine group of the conserved Gln68 (Figures 4.4B, 4.5B).

Unlike site-one mifepristone, site-two mifepristone does not make any crystal contacts, suggesting its binding may be biologically significant. However, there is a very low predicted free

energy of binding (1.4 kcal/mol) between the ligand and receptor, indicating low affinity binding (*89*). Further, site-two mifepristone was found to have a refined occupancy of 0.84, indicating that the receptor is not fully saturated with mifepristone at the coactivator cleft, despite a final concentration of approximately 25 μM in the crystallization drop.

Improved resolution of the ancSR2-progesterone structure permits visualization of D-ring contacts.

The structure overlays very closely with the structures of both the ancSR2–progesterone complex (PDB accession code: 4FN9) and the PR–progesterone complex (PDB accession code: 1A28); the root mean squared deviation values for all atoms between these structures are 0.3 Å and 0.6 Å, respectively (Figure 4.6A). Progesterone sits within the LBP, adopting an identical position and orientation as the ligand within the PR-progesterone complex structure (PDB accession code: 1A28) (Figure 4.6A). Progesterone makes extensive hydrophobic interactions with LBP in addition to two key hydrogen bonds. The first is between glutamate 41 and the 3-keto group of progesterone. It is this interaction that allows ancSR2 to differentiate between estrogen-like compounds and progestagens, androgens, and corticosteroids. Finally, the increased resolution structure allowed for unambiguous modeling of the progesterone's carbon 20 carbonyl, which is a critical moiety dictating ligand binding and receptor activation.

Discussion

This structure is higher resolution than the previously published 2.75 Å ancSR2–progesterone structure (*67*). Proper assignment of the hydrogen bond network guiding ligand binding to the LBP is absolutely critical to understand the conserved mechanism of activation across all 3-ketosteroid receptors. In the previous ancSR2–progesterone complex, the orientation of progesterone's C20 carbonyl was ambiguous and two equally probable H-bonding interactions were possible with asparagine 35 and threonine 210. A central question in the evolution of steroid hormone specificity is whether the allosteric networks that drive modern SR activation were present in the ancestral state or derived in modern proteins. Since the previous structure was not at sufficient resolution to orient the C20 group, we relied on molecular dynamics simulations to guide the final modeling of C20 (*76*). Our high-resolution structure reveals that indeed C20 is indeed oriented to accept a H-bond from threonine 210. This is consistent with molecular dynamics simulations of the ancSR2-progesterone complex and crystal structures of the progesterone receptor–progesterone, ancCR–deoxycorticosterone, ancestral glucocorticoid receptor 1–deoxycorticosterone, ancestral glucocorticoid receptor 2–dexamethasone, and mineralocorticoid receptor–aldosterone complexes (PDB accession codes 1A28, 2Q3Y, 3RY9, 3GN8, and 2AA2 respectively) (*64, 76, 77, 90-92*). Resolving the orientation of the ligand C20 confirms that the mechanism of activation among all 3-keto SRs originated in ancSR2 over 500 million years ago.

Attempts to determine the crystal structure of the ancSR2–mifepristone complex via ligand exchanged resulted instead in a crystal structure of ancSR2–progesterone with mifepristone bound at two surface sites. While the inability of mifepristone to exchange for progesterone was unexpected, it prompted us to examine the ability of ancSR2 to bind ligands and coregulators *in vitro*. Despite ancSR2's ability to respond to a wide array of ligands in mammalian cells (*67, 76*), attempts to measure ancSR2 ligand *in vitro* were not successful, even upon the addition of recombinant HSP90 (data not shown). It is possible that the full HSP90-

HSP70-p23-FKBP52-p60 complex is required for ligand binding or exchange *in vitro* (*93*). This inability to exchange ligands *in vitro* appears unique to ancSR2 since other ancestral SRs, including ancCR, ancGR1, and ancGR2, as well as the modern SRs are ligand exchangeable (*78*). AncSR2 represents the oldest 3-keto SR resurrected thus far; therefore, it is possible that an issue inherent to the reconstruction of the receptor has led to an extremely slow K_{off} preventing *in vitro* ligand exchange. It is well known that SRs display a narrow thermal window of activity and that their active *versus* inactive states are dictated by subtle thermodynamic changes. This is a consequence of both natural selection and neutral drift permitting the fine balance needed to allow the relatively small energetics of hormone binding to drive allosteric changes within the protein to propagate a signal. While too little stability prevents protein folding, too much stability may drive constitutive activation or prevent a dynamic response to ligand. It is possible that the ancSR2 is overstabilized (i.e. samples the active conformation too frequently when complexed to a ligand). This would explain its activation in the presence of mifepristone and its inability to exchange ligand *in vitro*. In line with these observations, ancSR2 is also unable to bind to coregulator peptides *in vitro*. It is well known that there is allosteric communication between the ligand binding sites and coregulator biding site (*61, 94*). Despite multiple attempts, we were unable to detect coregulator peptide binding to the ancSR2-progesterone complex by fluorescence polarization, while peptide binding to younger SR ancestors (i.e. ancGR1, ancGR2) is robust (*78*). Thus, ancSR2 is not able adjust its conformation/ dynamics to accommodate interaction with isolated peptides despite the fact that the structure is superimposable with closely related ancestral SR-ligand-coregulator peptide complexes (*77*).

Given the inability of ancSR2 to bind coactivator or ligand *in vitro*, it was surprising to see a weak secondary mifepristone interaction site at the AF-H surface. It is unclear whether this site is physiologically relevant *in vivo* however, the binding of mifepristone to the coactivator cleft in the crystal structure suggests there may be potential to use mifepristone as a scaffold for designing coactivator binding inhibitors. These novel pharmaceuticals would be instrumental in

the treatment of a range of diseases. Currently, there is a struggle to design effective peptidomimetic CBIs due to their inability to permeate the cell membranes (*95*). Mifepristone is already well established as having effective extracellular to intracellular transport and thus shows strong scaffold potential. Further work is needed to determine whether mifepristone or mifepristone derivatives are able to compete for the coactivator cleft in extant steroid receptors.

This structure is only the second to show a small molecule bound to the coactivator cleft of a steroid receptor. The first showed the anti-cancer drug 4-hydroxytamoxifen (HT) bound at the coactivator cleft of estrogen receptor beta (ERβ) (PDB accession code 2FSZ) (*96*). Overlaying these two structures reveals that the ligands adopt nearly identical positions at the coactivator cleft (Figure 4.7). Both insert phenolic substituents into the H3/H4 gap and are held in place primarily by hydrophobic interactions. Unlike the ERβ-HT interaction, in which HT doesn't contact H12 residues, mifepristone makes van der Waals contacts with H12 residues Met224, Glu227 and Ile228 (Figure 4.4). Together, these structures suggest that small hydrophobic molecules with perpendicular phenolic substituents are prime candidates for CBI development.

Our present structure suggests that this may extend to small molecules targeting the coactivator cleft binders. The AF-H surface presented in ancSR2 may represent a less dynamic, low energy target for such molecules. Future work should be devoted to investigating whether mifepristone shows potential for acting as a scaffold for further CBI development. Adaptations of previously approved drugs may be a successful avenue for development of pharmaceuticals that are otherwise difficult to craft.

Given their increased stability, ancestral steroid receptors are ideal tools for the extensive mutagenesis required to build robust structure-function relationships for both endogenous and synthetic ligands (*64, 78*). This same property makes them ideal tools to obtain crystal structures low affinity or weak SR modulators (i.e. lead compounds) that have been recalcitrant to crystallization. However, for ancSR2, given the very high affinity for progesterone and decreased

ability for exchange, it may be necessary to express the receptor in the presence of the desired target ligand prior to protein purification and crystallization.

Acknowledgements

We'd like to thank Dr. Geeta Eick and Dr. Joseph W. Thornton's (University of Oregon, OR) for providing the ancSR2 gene. Data were collected at Southeast Regional Collaborative Access Team (SER-CAT) 22-BM beamline at the Advanced Photon Source, Argonne National Laboratory. Supporting institutions may be found at www.ser-cat.org/members.html.

Figures

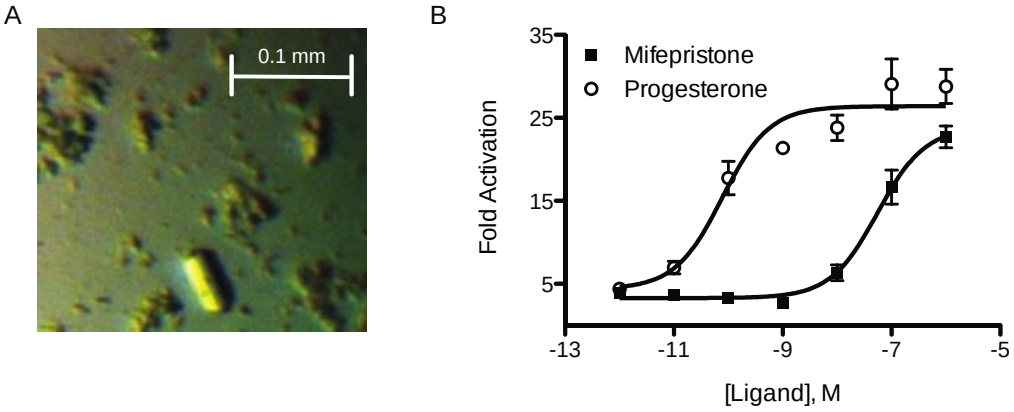


Figure 4.1: Crystals of the ancSR2–progesterone–mifepristone complex and in vitro activation data.

A. The ancSR2–progesterone–mifepristone ternary complex crystals measured approximately 50 x 20 x 20 microns. B. In luciferase reporter assays, ancSR2 is strongly activated by mifepristone (EC_{50} = 56 ± 1.2 nM) as well as progesterone (EC_{50} = 78 ± 1.7 pM).

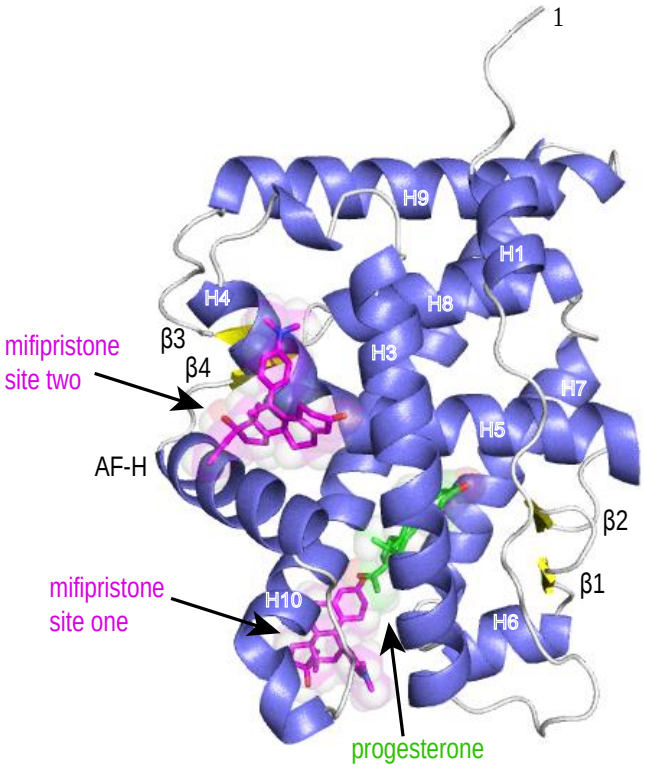


Figure 4.2: Overall structure of the ancSR2–progesterone–mifepristone complex.

Overall structure of the ancSR2 LBD with bound progesterone and mifepristone shown as green and magenta, respectively with oxygens, colored red. Helices are blue, β-sheets are yellow, loops are white. Figures were generated in PyMol (Schödenger, LLC).

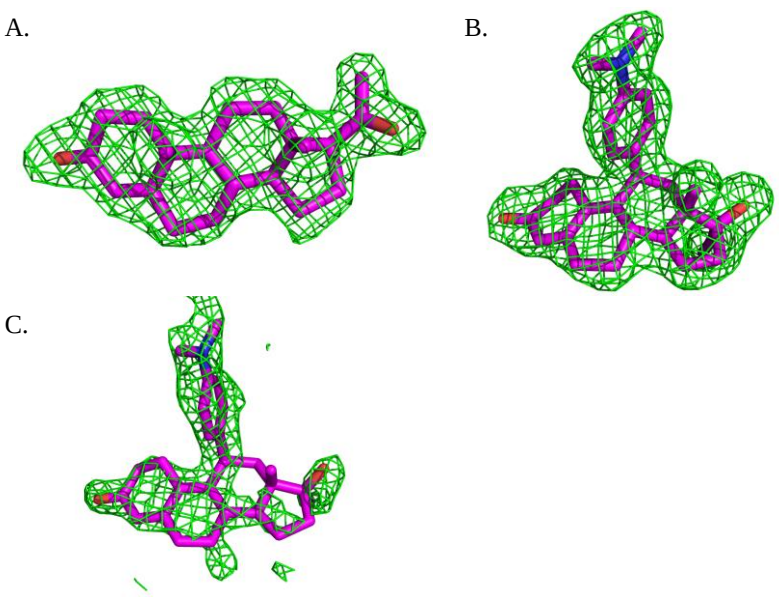


Figure 4.3: Omit maps of bound ligands.

F_o-F_c electron density (green) contoured to 2 σ showing evidence for bound ligand. Omit maps were generated by removal of the ligand from the structure and running 3 cycles of gradient energy minimization and B-factor optimization in PHENIX (version dev-1423) to minimize model bias. A. Electron density within the LBP corresponding to the volume of progesterone. B. Electron density at the base of the receptor corresponding to the volume of mifepristone (site-one). C. Electron density at the coactivator cleft corresponding to the volume of mifepristone (site-two).

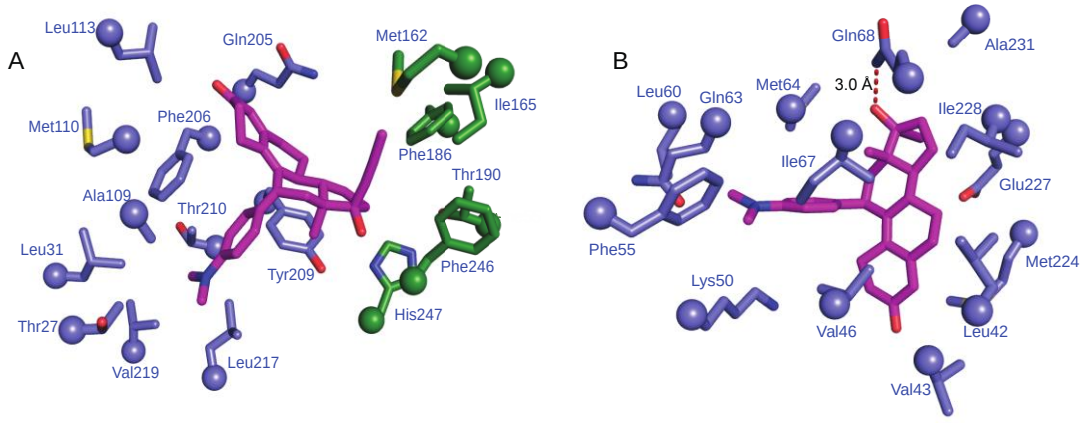


Figure 4.4: Mifepristone binding site interactions.

AncSR2 is shown in slate blue; mifepristone is shown in magenta (oxygens, red; nitrogens, blue). Residues within 4.2 Å of the ligand are shown. A. Site-one mifepristone interacts with both the monomer in the asymmetric unit as well as residues in a symmetry mate (forest green). B. Site-two mifepristone interacts with residues in the ancSR2 coactivator cleft.

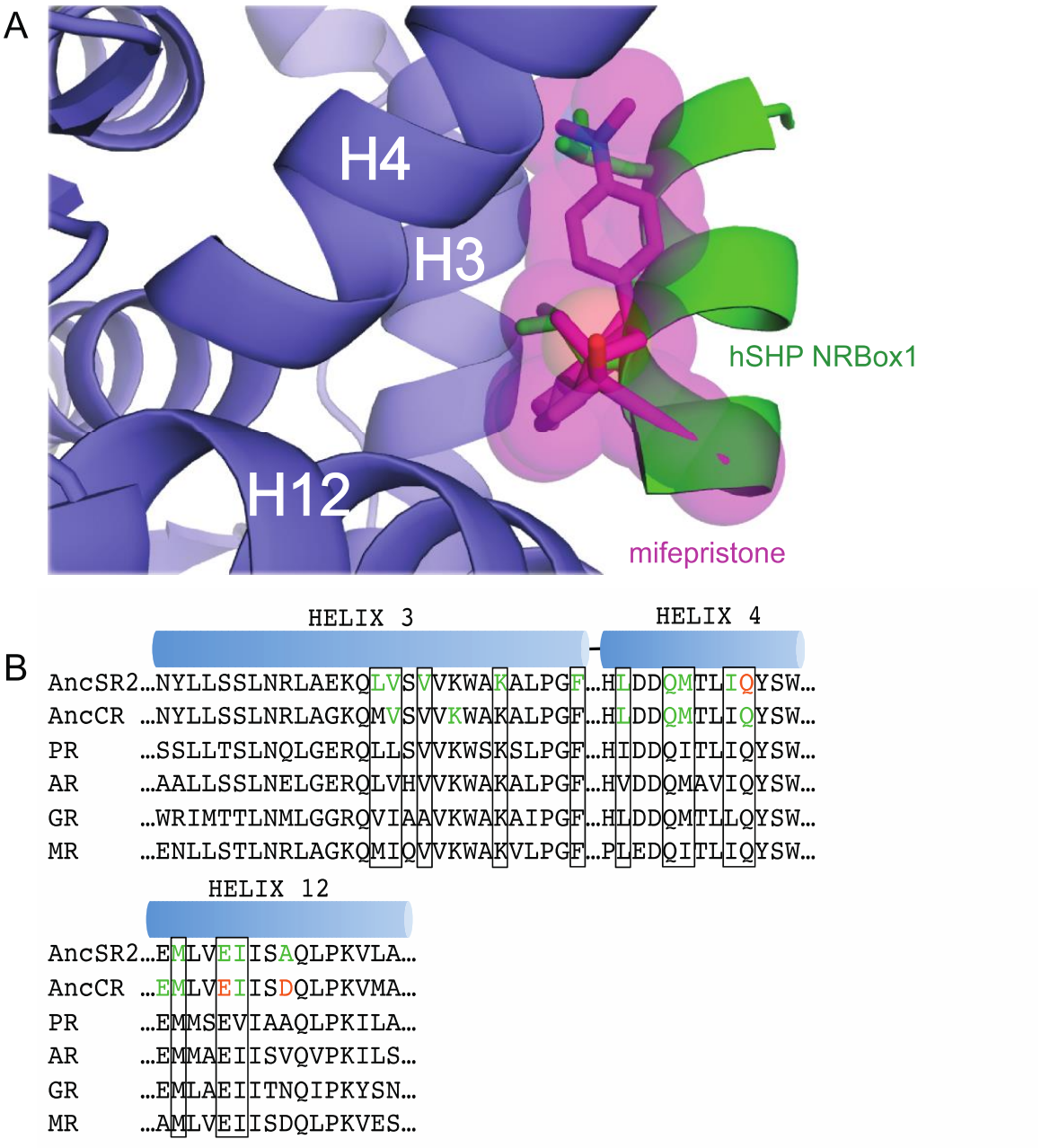


Figure 4.5: Mifepristone occupies the coactivator protein space.

A. ancSR2–progesterone–mifepristone (protein, slate blue; mifepristone, magenta) was overlaid with ancestral corticoid receptor–deoxycorticosterone–hSHP NRBox1 (PDB accession code: 2Q3Y) complex. Mifepristone occupies the same space as the hSHP NRBox1 peptide (green, leucine side chains shown as sticks). B. *Sequence alignment of ancestral and extant 3-ketosteroid*

receptor coactivator binding clefts. Sequence alignments of the ancSR2, ancCR, Progesterone Receptor (PR), Androgen Receptor (AR), Glucocorticoid Receptor (GR), and Mineralocorticoid Receptor (MR) coactivator binding clefts. Hydrophobic interactions (green) and hydrogen bonds (red) are shown for the interaction between AncSR2–mifepristone site-two and ancCR–SHP. Conserved residues across the SR lineage are indicated by a black box.

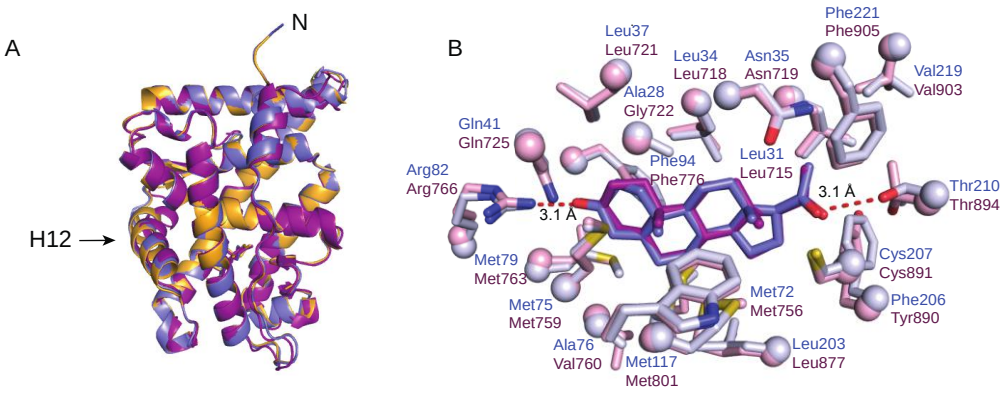


Figure 4.6: Global alignment of progesterone-bound steroid receptors.

A cartoon representation of the ancSR2–progesterone–mifepristone (slate blue), ancSR2–progesterone (orange, PDB accession code: 4FN9), and progesterone receptor–progesterone (purple, PDB accession code 1A28) complexes overlay with high overall structural similarity. B. *Ligand adopts an identical position and conformation in two progesterone-bound receptor structures*. The ancSR2–progesterone–mifepristone (ligand – slate blue, receptor – light blue) and progesterone receptor–progesterone (ligand – magenta, receptor – light pink, PDB accession code 1A28) structures show identical positioning and conformation of the ligand within the ligand binding pocket. Progesterone makes hydrogen bonds with Arg82 and Thr210 in both structures (dashed red lines).

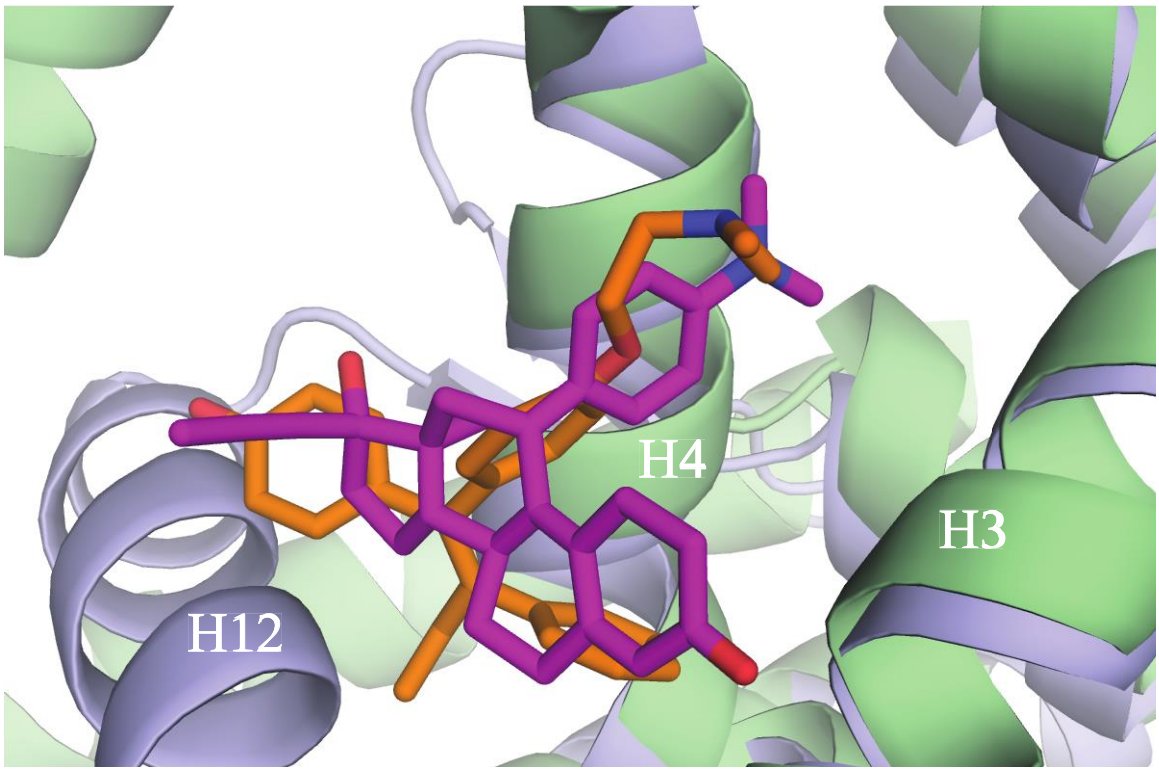


Figure 4.7: Mifepristone and 4-hydroxytamoxifen show similar binding modes to the steroid receptor coactivator binding cleft.

Alignment of the ancSR2–progesterone–mifepristone crystal structure and the ERβ–tamoxifen structure shows both mifepristone and 4-hydroxytamoxifen bound to the coactivator cleft. ancSR2 – slate blue, mifepristone – magenta, ERβ – light green, tamoxifen – orange.

References

1. Bridgham JT, Eick GN, Larroux C, Deshpande K, Harms MJ, et al. (2010) Protein evolution by molecular tinkering: diversification of the nuclear receptor superfamily from a ligand-dependent ancestor. PLoS Biol 8.

2. Nagy L, Schwabe JW (2004) Mechanism of the nuclear receptor molecular switch. Trends Biochem Sci 29: 317-324.

3. O'Malley BW, Tsai MJ (1992) Molecular pathways of steroid receptor action. Biol Reprod 46: 163-167.

4. McKenna NJ, O'Malley BW (2002) Minireview: nuclear receptor coactivators--an update. Endocrinology 143: 2461-2465.

5. McKenna NJ, O'Malley BW (2002) Combinatorial control of gene expression by nuclear receptors and coregulators. Cell 108: 465-474.

6. Nakao R, Haji M, Yanase T, Ogo A, Takayanagi R, et al. (1992) A single amino acid substitution (Met786----Val) in the steroid-binding domain of human androgen receptor leads to complete androgen insensitivity syndrome. J Clin Endocrinol Metab 74: 1152-1157.

7. Imasaki K, Hasegawa T, Okabe T, Sakai Y, Haji M, et al. (1994) Single amino acid substitution (840Arg-->His) in the hormone-binding domain of the androgen receptor leads to incomplete androgen insensitivity syndrome associated with a thermolabile androgen receptor. Eur J Endocrinol 130: 569-574.

8. Quigley CA, De Bellis A, Marschke KB, el-Awady MK, Wilson EM, et al. (1995) Androgen receptor defects: historical, clinical, and molecular perspectives. Endocr Rev 16: 271-321.

9. Nagpal S, Saunders M, Kastner P, Durand B, Nakshatri H, et al. (1992) Promoter context- and response element-dependent specificity of the transcriptional activation and modulating functions of retinoic acid receptors. Cell 70: 1007-1019.

10. Rodriguez AL, Tamrazi A, Collins ML, Katzenellenbogen JA (2004) Design, synthesis, and in vitro biological evaluation of small molecule inhibitors of estrogen receptor alpha coactivator binding. J Med Chem 47: 600-611.

11. Harms MJ, Eick GN, Goswami D, Colucci JK, Griffin PR, et al. (2013) Biophysical mechanisms for large-effect mutations in the evolution of steroid hormone receptors. Proc Natl Acad Sci U S A.

12. Bridgham JT, Ortlund EA, Thornton JW (2009) An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. Nature 461: 515-519.

13. Ortlund EA, Bridgham JT, Redinbo MR, Thornton JW (2007) Crystal structure of an ancient protein: evolution by conformational epistasis. Science 317: 1544-1548.

14. Thornton JW (2004) Resurrecting ancient genes: experimental analysis of extinct molecules. Nat Rev Genet 5: 366-375.

15. Harms MJ, Thornton JW (2010) Analyzing protein structure and function using ancestral gene reconstruction. Curr Opin Struct Biol 20: 360-366.

16. Kohn JA, Deshpande K, Ortlund EA (2012) Deciphering modern glucocorticoid cross-pharmacology using ancestral corticosteroid receptors. J Biol Chem 287: 16267-16275.

17. Eick GN, Colucci JK, Harms MJ, Ortlund EA, Thornton JW (2012) Evolution of minimal specificity and promiscuity in steroid hormone receptors. PLoS Genet 8: e1003072.

18. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. Macromolecular Crystallography, Pt A 276: 307-326.

19. Pereira de Jesus-Tran K, Cote PL, Cantin L, Blanchet J, Labrie F, et al. (2006) Comparison of crystal structures of human androgen receptor ligand-binding domain complexed with various agonists reveals molecular determinants responsible for binding affinity. Protein Sci 15: 987-999.

20. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66: 213-221.

21. Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, et al. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 66: 12-21.

22. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, et al. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res 35: W375-383.

23. Heikinheimo O, Kekkonen R, Lahteenmaki P (2003) The pharmacokinetics of mifepristone in humans reveal insights into differential mechanisms of antiprogestin action. Contraception 68: 421-426.

24. Song LN, Coghlan M, Gelmann EP (2004) Antiandrogen effects of mifepristone on coactivator and corepressor interactions with the androgen receptor. Mol Endocrinol 18: 70-85.

25. Nieman LK, Chrousos GP, Kellner C, Spitz IM, Nisula BC, et al. (1985) Successful treatment of Cushing's syndrome with the glucocorticoid antagonist RU 486. J Clin Endocrinol Metab 61: 536-540.

26. Hodgson MC, Astapova I, Cheng S, Lee LJ, Verhoeven MC, et al. (2005) The androgen receptor recruits nuclear receptor CoRepressor (N-CoR) in the presence of mifepristone via its N and C termini revealing a novel molecular mechanism for androgen receptor antagonists. J Biol Chem 280: 6511-6519.

27. Cadepond F, Ulmann A, Baulieu EE (1997) RU486 (mifepristone): mechanisms of action and clinical uses. Annu Rev Med 48: 129-156.

28. Zhang S, Jonklaas J, Danielsen M (2007) The glucocorticoid agonist activities of mifepristone (RU486) and progesterone are dependent on glucocorticoid receptor levels but not on EC50 values. Steroids 72: 600-608.

29. Boehm MF, McClurg MR, Pathirana C, Mangelsdorf D, White SK, et al. (1994) Synthesis of high specific activity [3H]-9-cis-retinoic acid and its application for identifying retinoids with unusual binding properties. J Med Chem 37: 408-414.

30. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. J Mol Biol 372: 774-797.

31. Williams SP, Sigler PB (1998) Atomic structure of progesterone complexed with its receptor. Nature 393: 392-396.

32. Carroll SM, Ortlund EA, Thornton JW (2011) Mechanisms for the evolution of a derived function in the ancestral glucocorticoid receptor. PLoS Genet 7: e1002117.

33. Bledsoe RK, Madauss KP, Holt JA, Apolito CJ, Lambert MH, et al. (2005) A ligand-mediated hydrogen bond network required for the activation of the mineralocorticoid receptor. J Biol Chem 280: 31283-31293.

34. Pratt WB, Toft DO (1997) Steroid receptor interactions with heat shock protein and immunophilin chaperones. Endocr Rev 18: 306-360.

35. Watkins RE, Wisely GB, Moore LB, Collins JL, Lambert MH, et al. (2001) The human nuclear xenobiotic receptor PXR: structural determinants of directed promiscuity. Science 292: 2329-2333.

36. Patch JA, Barron AE (2003) Helical peptoid mimics of magainin-2 amide. J Am Chem Soc 125: 12092-12093.

37. Wang Y, Chirgadze NY, Briggs SL, Khan S, Jensen EV, et al. (2006) A second binding site for hydroxytamoxifen within the coactivator-binding groove of estrogen receptor beta. Proc Natl Acad Sci U S A 103: 9908-9911.

CHAPTER 5: EXPRESSION, PURIFICATION, AND CRYSTALLIZATION OF THE ANCESTRAL ANDROGEN RECEPTOR-DHT COMPLEX

Jennifer K. Colucci[1] and Eric A. Ortlund[1d]

[1]*Department of Biochemistry and Winship Cancer Institute, Emory University School of Medicine, Atlanta, GA 30033, USA*

J. K. Colucci, E. A. Ortlund, Expression, purification and crystallization of the ancestral androgen receptor-DHT complex. *Acta crystallographica. Section F, Structural biology and crystallization communications* **69**, 994 (Sep, 2013).

The family of steroid receptors evolved from an estrogen-sensitive ancestor, through a promiscuous 3-keto steroid receptor, to be sensitve to a wide array of steroids. In order to understand the structural mechanisms governing response to androgenic steroids, the ancestral androgen receptor was resurrected (ancAR1) for structural and biochemical analysis. In this work, I cloned, expressed, purified, and crystallized the ancAR1 in complex with its cognate ligand, 5α-dihydrotesterone (DHT) and a fragment of a coregulator protein, transcription initial factor 2 (Tif2). This work was previously published in Acta crystallographica Section F: Structural biology and crystallization communications.

---

[d] Conceived and designed the experiments: JKC EAO. Performed the experiments: JKC. Analyzed the data: JKC EAO. Wrote the paper: JKC EAO.

Abstract

Steroid receptors (SRs) are a closely related family of ligand-dependent nuclear receptors that mediate the transcription of genes critical for development, reproduction and immunity. SR dysregulation has been implicated in cancer, inflammatory diseases and metabolic disorders. SRs bind their cognate hormone ligand with exquisite specificity, offering a unique system to study the evolution of molecular recognition. The SR family evolved from an estrogen-sensitive ancestor and diverged to become sensitive to progestagens, corticoids and, most recently, androgens. To understand the structural mechanisms driving the evolution of androgen responsiveness, the ancestral androgen receptor (ancAR1) was crystallized in complex with 5α-dihydrotestosterone (DHT) and a fragment of the transcriptional mediator/intermediary factor 2 (Tif2). Crystals diffracted to 2.1 Å resolution and the resulting structure will permit a direct comparison with its progestagen-sensitive ancestor, ancestral steroid receptor 2 (ancSR2).

Introduction

The androgen receptor (AR) is a member of the steroid receptor (SR) family of transcription factors, which play a major role in regulating vertebrate biology. AR responds to androgens, such as testosterone and 5α–dihydrotestosterone (DHT), to regulate genes central to male sexual development, immunity, and behavior (*52-55*). Given the widely prevalent role of androgens in normal physiology, AR signaling has been implicated in a number of diseases including cancer, cardiovascular defects, metabolic disorders, Alzheimer's, and Androgen Insensitivity Syndrome (AIS) (*52, 56, 57*). AR plays a particularly malicious role in prostate cancer by driving gene expression to fuel cell growth in both an androgen-dependent and androgen independent manner (*58, 59*).

AR displays the typical modular SR domain architecture with a N-terminal activation function 1 domain, a DNA binding domain, a short linker region, and a ligand binding domain (LBD). Without ligand, AR is unstable and resides in the cytoplasm complexed to chaperones (*60*). Upon binding to a high-affinity ligand such as DHT, the hormone-receptor complex translocates to the nucleus where it binds to coregulatory proteins such as the Tif2 to regulate target gene expression (*61*). This simple paradigm, where a small lipophilic ligand regulates complex gene programs, requires AR to recognize androgens with a high degree of specificity and exclude interaction with very similar steroids such as estrogens, progestagens, and corticosteroids. This exquisite sensitivity to androgens arose during early vertebrate evolution with the appearance the first AR (some 450 mya) following the duplication and subsequence divergence from a progestagen-activated ancestor (*17, 62*).

To gain insight into the fundamental processes governing the evolution of androgen specificity we initiated biophysical studies on the ancestral AR, ancAR1, the sequence of which was inferred using a well-described technique termed ancestral gene resurrection (to be

published) (*20, 63, 64*). We have cloned, expressed, purified, and crystallized ancAR1 in complex with the high affinity ligand DHT and a fragment of the transcriptional activator Tif2 which binds to the activated conformation of SRs. Diffraction data were collected to 2.1 Å.

Materials and Methods

Reagents.

Chemicals were purchased from Sigma (St Louis, Missouri, USA) or Fisher (Hampton, New Hampshire, USA). DHT was purchased from Toronto Research Chemicals (Toronto, Canada). The vector for His-tagged TEV protease was a gift from David Waugh (National Cancer Institute). The pLIC_MBP vector was a gift from John Sondek (UNC, Chapel Hill). The ancAR1 LBD was resurrected using well established protocols and was kindly provided by Dr. Joseph Thornton (University of Oregon, USA). The peptide corresponding to the nuclear receptor coactivator box 3 from human Tif2 was synthesized by RS Synthesis (Louisville, Kentucky, USA).

Cloning.

The ancAR1 LBD (residues 1–250) was cloned into pLIC_MBP, which contains a hexahistidine tag followed by the maltose-binding protein (MBP) and a Tobacco etch virus (TEV) protease site N-terminal to the protein. The forward cloning primer used was 50-TACTTCCAATCCAATGCGGCGATCGCCATTCCCATTTTCC-30; the reverse cloning primer used was 50-TTATCCACTTCCAAT- GCGCTAGTTTAAACTTACTGC-30. The sequence of ancAR1 is IPIFLSVLQSIEPEVVYAGYDNTQPDTSASLLTSLNELGERQLVRVVKWAKALPGFRNLH VDDQMTLIQYSWMGVMVFAMGWRSYKNVNSRMLYFAPDLVFNEQRMQKSTMYNLC VRMRHLSQEFVWLQVTQEEFLCMKALLLFSIIPVEGLKNQKYFDELRMNYIKELDRVISF QGKNPTSSSQRFYQLTKLLDSLQPIVRKLHQFTFDLFVQSQSLSVEFPEMMSEIISAQVPKI LAGMVKPLLFHKQ. The crystallization construct contains the residues SNA C-terminal to the receptor as a relic from the TEV protease cleavage site.

Expression and Purification.

AncAR1 LBD was expressed as a 6xHis-MBP fusion protein in Escherichia coli BL21(DE3) cells. Cultures (1.0 L in Terrific Broth) were grown to an OD600 of 0.8 and induced with a final concentration of 400 mM IPTG and 50 mM DHT at 291 K overnight. Cell mass was collected by centrifugation at 4000 rev min^{-1} for 20 min, resuspended in 150 mM NaCl, 20 mM Tris–HCl pH 7.4, 5% glycerol, 25 mM imidazole, 0.1% PMSF and lysed using sonication on ice. AncAR1-MBP was initially purified using Ni^{2+}-affinity chromatography (HisTrap column, GE Healthcare). Fractions containing ancAR1-MBP were identified by denaturing polyacrylamide gel electrophoresis (SDS–PAGE), pooled and dialyzed against 150 mM NaCl, 20 mM Tris–HCl pH 7.4, 5% glycerol, 1 mg TEV protease. Following TEV protease cleavage, the tagged MBP was removed by an additional Ni2+-affinity column. The flowthrough containing ancAR1 LBD was concentrated using an Amicon Ultra 10K centrifugal filter device (Millipore), concentrated to 3.3 mg ml^{-1} and dialyzed against 150 mM sodium chloride, 20 mM Tris–HCl pH 7.4, 5% glycerol. The final purity of the ancAR1 LBD was assessed using SDS–PAGE (Figure 5.1).

Crystallization and Data Collection.

Prior to crystallization, an additional 50 mM DHT was added to the ancAR1–DHT complex to ensure full occupancy of DHT in the ligand-binding pocket. Additionally, 500 mM of a peptide derived from human Tif2, corresponding to the nuclear receptor coactivator box 3 (740-KENALLRYLLDKDD-753), was added to the receptor– ligand complex for crystallization, yielding a 5:1 molar ratio of peptide:ancAR1. Crystallization trials were performed using sitting-drop vapor diffusion, mixing 0.2 ml of the protein sample with an equal volume of screening solution and equilibrating against 60 ml screening solution in the reservoir. Initial screening was performed using 480 conditions from the commercially available kits The JCSG+ Suite, The PEGs Suite, The Nucleix Suite, The Classics Lite Suite and The AmSO4 Suite (Qiagen). Positive

hits were obtained using The PEGs Suite condition A11 [0.1 M MES pH 6.5, 25% (w/v) PEG 1000], The JCSG+ Suite condition A2 (0.1 M trisodium citrate pH 5.5, 20% PEG 3000) and The JCSG+ Suite condition G9 (0.1 M KCN, 30% PEG MME 2000). These hits were further expanded to generate diffraction-quality crystals.

Crystals of the ternary ancAR1 LBD – DHT – TIF2 complex were grown by hanging drop vapor diffusion at 22 °C from solutions containing 1.0 μL of protein at 3.3 mg mL^{-1} protein and 1.0 μL of the following crystallant 20 % PEG 1000, 0.3 M MES pH 6.5. Crystals were cryoprotected by transient soaking in crystallant containing 20 % glycerol and were flash-cooled in liquid N_2 at 100 K. Two hundred and five frames were collected at 0.5 ° oscillation. Data to 2.1 Å resolution were collected at the South East Regional Collaborative Access Team (SER-CAT) 22-BM at the Advanced Photon Source (APS) at Argonne National Laboratory in Chicago, IL using a wavelength of 0.97 Å, and were processed and scaled with HKL2000 (Table 4.1) (*65*). Data processing revealed that crystals of the ancAR1-DHT complex grew in the $P4_32_12$ space group.

Results and Discussion

In order to understand the evolution of ligand specificity in steroid hormone nuclear receptors, we cloned, overexpressed and purified ancAR1 in complex with its most potent ligand, DHT, and a fragment of the human coactivator Tif2. A denaturing SDS–PAGE gel shows a pure receptor–ligand complex with no contaminating bands (Figure 5.1). The calculated molecular weight of the protein is 29,498 Da.

Crystals were grown by equilibrating a 1.0 µl protein solution and 1.0 µl mother liquor using hanging drop vapor diffusion (Figure 5.2). The receptor crystallized as long rods in a solution of 20% PEG 1000 and 0.3 M MES pH 6.5. Crystals formed in the $P4_32_12$ space group and diffracted to 2.1 Å (Figure 5.3). The Matthews coefficient (V_M) is 2.96 $Å^3$ Dal^{-1} with one monomer in the asymmetric unit, corresponding to a solvent content of 58.4 % (*66*).

Solving the crystal structure of this ancient receptor-ligand complex is a critical step towards understanding the structural and biophysical changes that occurred in the AR lineage to develop sensitivity for androgenic compounds. This structure will permit direct structural comparison with the progestagen-activated ancestral steroid receptor 2 (*67*) and will guide future functional studies identifying the residues that were responsible for the functional shift from 17-acetyl steroid to 17-hydroxyl (androgen) responsiveness. Understanding how ligand recognition can be evolved, harvested, and exploited is essential to the progression of protein engineering, drug design and discovery, and a true comprehension of our molecular history.

Acknowledgements

We would like to thank Dr. Geeta Eick and Dr. Joseph W. Thornton (University of Oregon, USA) for providing the ancAR1 gene. Data were collected at Southeast Regional Collaborative Access Team (SER-CAT) 22-BM beamline at the Advanced Photon Source, Argonne National Laboratory. Supporting institutions may be found at http://www.ser-cat.org/members.html. This research is supported by R01 GM081592 (Joseph Thornton PI, EAO coPI), start-up funds from Emory University (EAO) and JKC's American Heart Association pre-doctoral fellowship (10PRE3530007). Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. W-31-109-Eng-38.

Figures

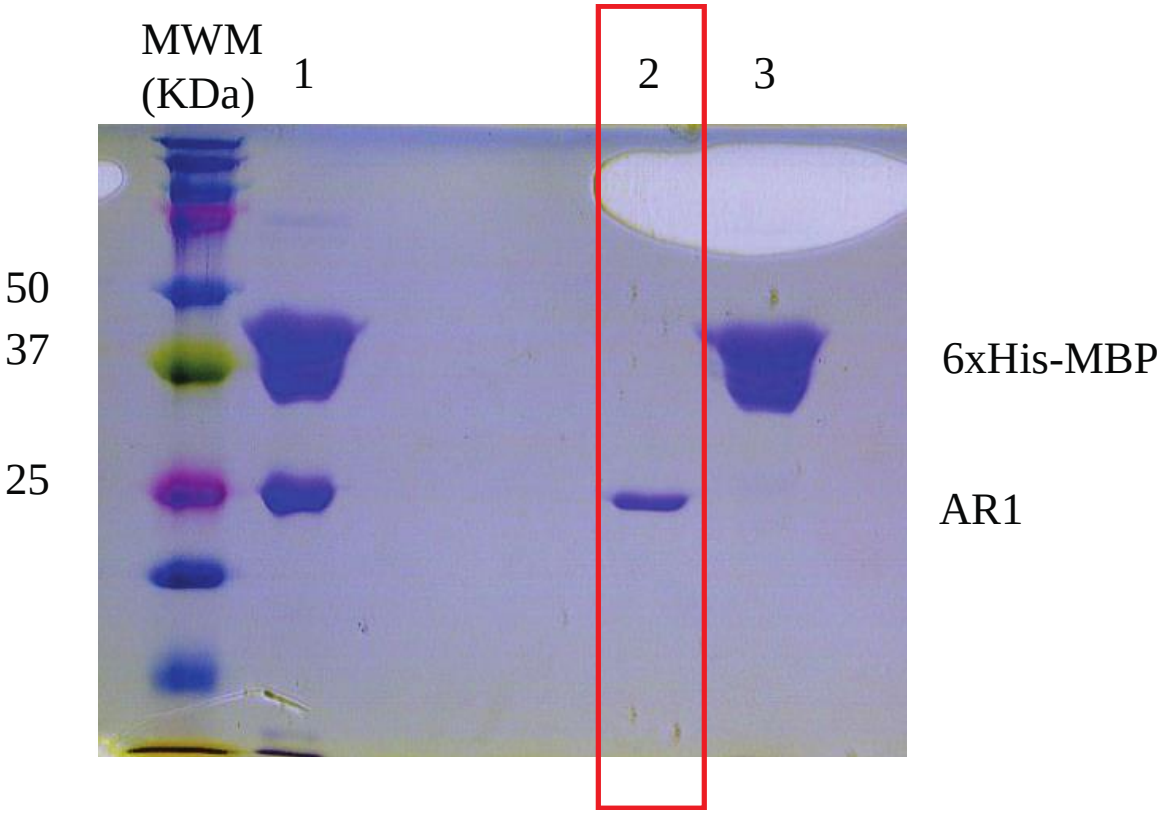


Figure 5.1: Following a series of affinity columns, ancAR1–DHT was purified to homogeneity.

Lane 1, TEV protease-cleaved ancAR1-MBP fusion protein. Lane 2, purified ancAR1. Lane 3, cleaved MBP. Lane M contains molecular-weight markers (labeled in kDa).

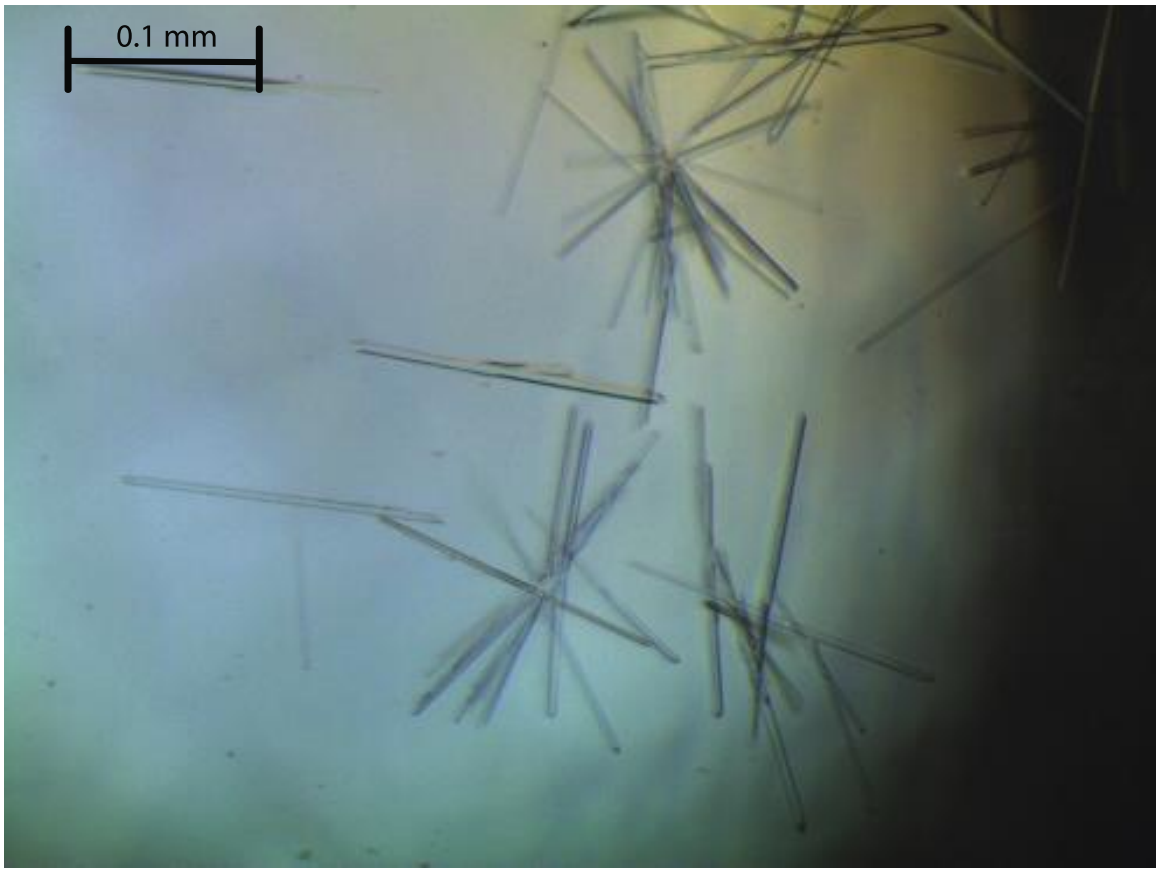


Figure 5.2: Crystals of ancAR1–DHT.

The long rod-shaped crystals are approximately 100– 200 mm in length. The crystals grew in 20% PEG 1000, 0.3 M MES pH 6.5.

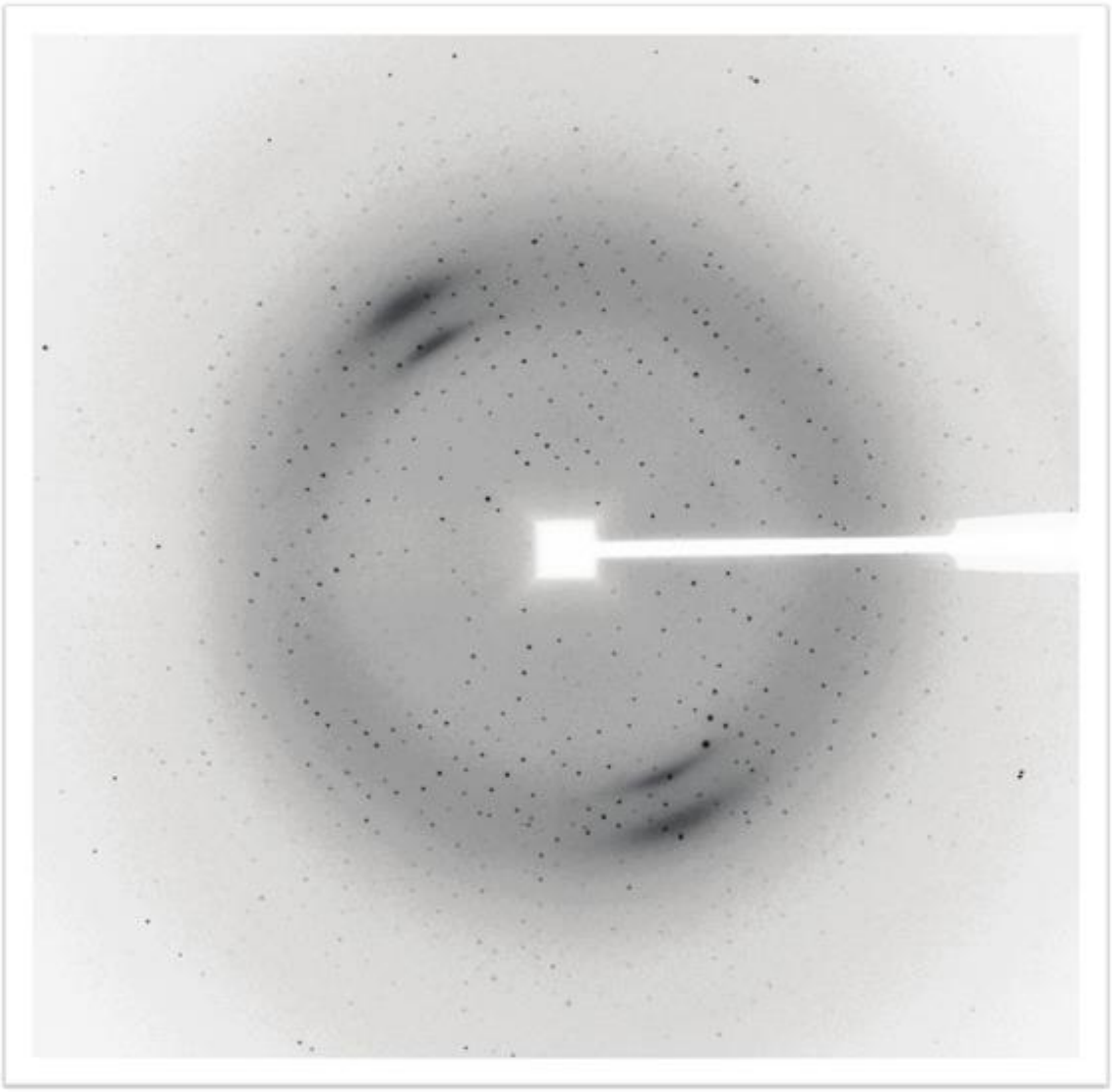


Figure 5.3: Diffraction image of an ancAR1–DHT crystal.

The detector edge corresponds to 2.15 Å resolution.

Table 5.1: Diffraction and processing statistics for ancAR1-DHT crystal

	ancAR1-DHT
Resolution (Å)	2.10 (29.40 – 2.10)*
Space Group	$P4_32_12$
Unit Cell Dimensions (Å, °)	a=b=68.9, c=147.3
	α=β=γ=90
No. of Reflections	172781
R^a_{sym} (%)	10.9 (43.7)
Completeness (%)	99.9 (100)
Ave. Redundancy	8.0 (7.3)
I/σ	17.4 (4.3)
Mosaicity (°)	0.45

[a] $R_{sym} = \Sigma|I- \langle I\rangle|/ \Sigma|I|$, where I is the observed intensity and <I> is the average intensity of several symmetry-related observations.

*Data in parentheses represent highest shell

Table 5.1: Data collection statistics for AncAR1-DHT-Tif2.

References

1. B. Alberts, *Molecular biology of the cell*. (Garland Science, New York, ed. 4th, 2002), pp. xxxiv, 1548 p.
2. R. P. Bhattacharyya, A. Remenyi, B. J. Yeh, W. A. Lim, Domains, motifs, and scaffolds: the role of modular interactions in the evolution and wiring of cell signaling circuits. *Annual review of biochemistry* **75**, 655 (2006).
3. R. Rohs *et al.*, Origins of specificity in protein-DNA recognition. *Annual review of biochemistry* **79**, 233 (2010).
4. R. J. Lefkowitz, The superfamily of heptahelical receptors. *Nature cell biology* **2**, E133 (Jul, 2000).
5. S. D. Copley, Enzymes with extra talents: moonlighting functions and catalytic promiscuity. *Current opinion in chemical biology* **7**, 265 (Apr, 2003).
6. R. A. Jensen, Enzyme recruitment in evolution of new function. *Annual review of microbiology* **30**, 409 (1976).
7. P. J. O'Brien, D. Herschlag, Catalytic promiscuity and the evolution of new enzymatic activities. *Chemistry & biology* **6**, R91 (Apr, 1999).
8. O. Khersonsky, D. S. Tawfik, Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annual review of biochemistry* **79**, 471 (2010).
9. D. S. Tawfik, Messy biology and the origins of evolutionary innovations. *Nature chemical biology* **6**, 692 (Oct, 2010).
10. Y. Yoshikuni, T. E. Ferrin, J. D. Keasling, Designed divergent evolution of enzyme function. *Nature* **440**, 1078 (Apr 20, 2006).
11. D. A. Liberles, M. D. Tisdell, J. A. Grahnen, Binding constraints on the evolution of enzymes and signalling proteins: the important role of negative pleiotropy. *Proceedings. Biological sciences / The Royal Society* **278**, 1930 (Jul 7, 2011).
12. S. Bershtein, K. Goldin, D. S. Tawfik, Intense neutral drifts yield robust and evolvable consensus proteins. *Journal of molecular biology* **379**, 1029 (Jun 20, 2008).
13. J. D. Bloom, P. A. Romero, Z. Lu, F. H. Arnold, Neutral genetic drift can alter promiscuous protein functions, potentially aiding functional evolution. *Biology direct* **2**, 17 (2007).
14. J. D. Bloom, F. H. Arnold, In the light of directed evolution: pathways of adaptive protein evolution. *Proceedings of the National Academy of Sciences of the United States of America* **106 Suppl 1**, 9995 (Jun 16, 2009).

15. M. J. Keiser *et al.*, Predicting new molecular targets for known drugs. *Nature* **462**, 175 (Nov 12, 2009).
16. H. Gronemeyer, J. A. Gustafsson, V. Laudet, Principles for modulation of the nuclear receptor superfamily. *Nature reviews. Drug discovery* **3**, 950 (Nov, 2004).
17. G. N. Eick, J. W. Thornton, Evolution of steroid receptors from an estrogen-sensitive ancestral receptor. *Molecular and cellular endocrinology* **334**, 31 (Mar 1, 2011).
18. J. A. Katzenellenbogen, The structural pervasiveness of estrogenic activity. *Environmental health perspectives* **103 Suppl 7**, 99 (Oct, 1995).
19. D. O. Norris, J. A. Carr, *Endocrine disruption : biological basis for health effects in wildlife and humans*. (Oxford University Press, New York, 2006), pp. xiv, 477 p.
20. J. W. Thornton, Resurrecting ancient genes: experimental analysis of extinct molecules. *Nature reviews. Genetics* **5**, 366 (May, 2004).
21. D. A. Liberles, *Ancestral sequence reconstruction*. (Oxford University Press, Oxford ; New York, 2007), pp. xiii, 252 p.
22. J. W. Thornton, E. Need, D. Crews, Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling. *Science* **301**, 1714 (Sep 19, 2003).
23. J. T. Bridgham, J. E. Brown, A. Rodriguez-Mari, J. M. Catchen, J. W. Thornton, Evolution of a new function by degenerative mutation in cephalochordate steroid receptors. *PLoS genetics* **4**, e1000191 (2008).
24. D. S. Geller *et al.*, Activating mineralocorticoid receptor mutation in hypertension exacerbated by pregnancy. *Science* **289**, 119 (Jul 7, 2000).
25. J. Veldscholte *et al.*, The androgen receptor in LNCaP cells contains a mutation in the ligand binding domain which affects steroid binding characteristics and response to antiandrogens. *The Journal of steroid biochemistry and molecular biology* **41**, 665 (Mar, 1992).
26. X. Y. Zhao *et al.*, Glucocorticoids can promote androgen-independent growth of prostate cancer cells through a mutated androgen receptor. *Nature medicine* **6**, 703 (Jun, 2000).
27. J. W. Thornton, Evolution of vertebrate steroid receptors from an ancestral estrogen receptor by ligand exploitation and serial genome expansions. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 5671 (May 8, 2001).
28. T. Mizuta, K. Asahina, M. Suzuki, K. Kubokawa, In vitro conversion of sex steroids and expression of sex steroidogenic enzyme genes in amphioxus ovary. *Journal of experimental zoology. Part A, Ecological genetics and physiology* **309**, 83 (Mar 1, 2008).

29. A. D'Aniello *et al.*, Occurrence of sex steroid hormones and their binding proteins in Octopus vulgaris lam. *Biochemical and biophysical research communications* **227**, 782 (Oct 23, 1996).
30. D. A. Close, S. S. Yun, S. D. McCormick, A. J. Wildbill, W. Li, 11-deoxycortisol is a corticosteroid hormone in the lamprey. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 13942 (Aug 3, 2010).
31. D. M. Taverna, R. A. Goldstein, Why are proteins so robust to site mutations? *Journal of molecular biology* **315**, 479 (Jan 18, 2002).
32. M. J. Ryan, J. H. Fox, W. Wilczynski, A. S. Rand, Sexual selection for sensory exploitation in the frog Physalaemus pustulosus. *Nature* **343**, 66 (Jan 4, 1990).
33. W. Wickler, *Mimicry in plants and animals*. World university library (McGraw-Hill, New York,, 1968), pp. 253 p.
34. D. P. Edwards, The roles of tolerance in the evolution, maintenance and breakdown of mutualism. *Die Naturwissenschaften* **96**, 1137 (Oct, 2009).
35. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**, 1792 (2004).
36. S. Guindon, O. Gascuel, A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic biology* **52**, 696 (Oct, 2003).
37. M. Anisimova, O. Gascuel, Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Systematic biology* **55**, 539 (Aug, 2006).
38. Z. Yang, S. Kumar, M. Nei, A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics* **141**, 1641 (Dec, 1995).
39. Z. Yang, PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer applications in the biosciences : CABIOS* **13**, 555 (Oct, 1997).
40. V. Hanson-Smith, B. Kolaczkowski, J. W. Thornton, Robustness of ancestral sequence reconstruction to phylogenetic uncertainty. *Molecular biology and evolution* **27**, 1988 (Sep, 2010).
41. D. Picard, K. R. Yamamoto, Two signals mediate hormone-dependent nuclear localization of the glucocorticoid receptor. *The EMBO journal* **6**, 3333 (Nov, 1987).
42. J. Keay, J. T. Bridgham, J. W. Thornton, The Octopus vulgaris estrogen receptor is a constitutive transcriptional activator: evolutionary and functional implications. *Endocrinology* **147**, 3861 (Aug, 2006).
43. J. T. Bridgham, S. M. Carroll, J. W. Thornton, Evolution of hormone-receptor complexity by molecular exploitation. *Science* **312**, 97 (Apr 7, 2006).

44. L. Clinckemalie, D. Vanderschueren, S. Boonen, F. Claessens, The hinge region in androgen receptor control. *Molecular and cellular endocrinology* **358**, 1 (Jul 6, 2012).
45. J. T. Bridgham *et al.*, Protein evolution by molecular tinkering: diversification of the nuclear receptor superfamily from a ligand-dependent ancestor. *PLoS biology* **8**, (2010).
46. P. D. Adams *et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography* **66**, 213 (Feb, 2010).
47. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta crystallographica. Section D, Biological crystallography* **60**, 2126 (Dec, 2004).
48. N. Eswar, D. Eramian, B. Webb, M. Y. Shen, A. Sali, Protein structure modeling with MODELLER. *Methods in molecular biology* **426**, 145 (2008).
49. S. C. Lovell *et al.*, Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins* **50**, 437 (Feb 15, 2003).
50. G. J. Kleywegt, T. A. Jones, Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta crystallographica. Section D, Biological crystallography* **50**, 178 (Mar 1, 1994).
51. M. J. Harms *et al.*, Biophysical mechanisms for large-effect mutations in the evolution of steroid hormone receptors. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 11475 (Jul 9, 2013).
52. Y. Li, K. Izumi, H. Miyamoto, The Role of the Androgen Receptor in the Development and Progression of Bladder Cancer. *Jpn J Clin Oncol* **42**, 569 (Jul, 2012).
53. P. M. Holterhus, Molecular androgen memory in sex development. *Pediatric endocrinology reviews : PER* **9 Suppl 1**, 515 (Sep, 2011).
54. K. A. Walters, U. Simanainen, D. J. Handelsman, Molecular insights into androgen actions in male and female reproductive function from androgen receptor knockout models. *Human reproduction update* **16**, 543 (Sep-Oct, 2010).
55. R. S. Wang, S. Yeh, C. R. Tzeng, C. Chang, Androgen receptor roles in spermatogenesis and fertility: lessons from testicular cell-specific androgen receptor knockout mice. *Endocrine reviews* **30**, 119 (Apr, 2009).
56. T. E. Hickey, J. L. L. Robinson, J. S. Carroll, W. D. Tilley, The Androgen Receptor in Breast Tissues: Growth Inhibitor, Tumor Suppressor, Oncogene? *Molecular endocrinology* **26**, 1252 (Aug, 2012).
57. I. A. Hughes, R. Werner, T. Bunch, O. Hiort, Androgen Insensitivity Syndrome. *Semin Reprod Med* **30**, 432 (Sep, 2012).

58. S. M. Green, E. A. Mostaghel, P. S. Nelson, Androgen action and metabolism in prostate cancer. *Molecular and cellular endocrinology* **360**, 3 (Sep 5, 2012).

59. L. Tamburrino *et al.*, Androgen receptor (AR) expression in prostate cancer and progression of the tumor: Lessons from cell lines, animal models and human specimens. *Steroids* **77**, 996 (Aug, 2012).

60. Y. Fang, A. E. Fliss, D. M. Robins, A. J. Caplan, Hsp90 regulates androgen receptor hormone binding affinity in vivo. *The Journal of biological chemistry* **271**, 28697 (Nov 8, 1996).

61. L. Nagy, J. W. Schwabe, Mechanism of the nuclear receptor molecular switch. *Trends in biochemical sciences* **29**, 317 (Jun, 2004).

62. Y. Ogino, H. Katoh, S. Kuraku, G. Yamada, Evolutionary history and functional characterization of androgen receptor genes in jawed vertebrates. *Endocrinology* **150**, 5415 (Dec, 2009).

63. M. J. Harms, J. W. Thornton, Analyzing protein structure and function using ancestral gene reconstruction. *Current opinion in structural biology* **20**, 360 (Jun, 2010).

64. E. A. Ortlund, J. T. Bridgham, M. R. Redinbo, J. W. Thornton, Crystal structure of an ancient protein: evolution by conformational epistasis. *Science* **317**, 1544 (Sep 14, 2007).

65. Z. Otwinowski, W. Minor, Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol* **276**, 307 (1997).

66. B. W. Matthews, Solvent content of protein crystals. *Journal of molecular biology* **33**, 491 (Apr 28, 1968).

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

68. B. W. O'Malley, M. J. Tsai, Molecular pathways of steroid receptor action. *Biology of reproduction* **46**, 163 (Feb, 1992).

69. N. J. McKenna, B. W. O'Malley, Minireview: nuclear receptor coactivators--an update. *Endocrinology* **143**, 2461 (Jul, 2002).

70. N. J. McKenna, B. W. O'Malley, Combinatorial control of gene expression by nuclear receptors and coregulators. *Cell* **108**, 465 (Feb 22, 2002).

71. R. Nakao *et al.*, A single amino acid substitution (Met786----Val) in the steroid-binding domain of human androgen receptor leads to complete androgen insensitivity syndrome. *The Journal of clinical endocrinology and metabolism* **74**, 1152 (May, 1992).

72. K. Imasaki *et al.*, Single amino acid substitution (840Arg-->His) in the hormone-binding domain of the androgen receptor leads to incomplete androgen insensitivity syndrome associated with a thermolabile androgen receptor. *European journal of endocrinology / European Federation of Endocrine Societies* **130**, 569 (Jun, 1994).

73. C. A. Quigley *et al.*, Androgen receptor defects: historical, clinical, and molecular perspectives. *Endocrine reviews* **16**, 271 (Jun, 1995).

74. S. Nagpal *et al.*, Promoter context- and response element-dependent specificity of the transcriptional activation and modulating functions of retinoic acid receptors. *Cell* **70**, 1007 (Sep 18, 1992).

75. A. L. Rodriguez, A. Tamrazi, M. L. Collins, J. A. Katzenellenbogen, Design, synthesis, and in vitro biological evaluation of small molecule inhibitors of estrogen receptor alpha coactivator binding. *Journal of medicinal chemistry* **47**, 600 (Jan 29, 2004).

76. M. J. Harms *et al.*, Biophysical mechanisms for large-effect mutations in the evolution of steroid hormone receptors. *Proceedings of the National Academy of Sciences of the United States of America*, (Jun 24, 2013).

77. J. T. Bridgham, E. A. Ortlund, J. W. Thornton, An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature* **461**, 515 (Sep 24, 2009).

78. J. A. Kohn, K. Deshpande, E. A. Ortlund, Deciphering modern glucocorticoid cross-pharmacology using ancestral corticosteroid receptors. *The Journal of biological chemistry* **287**, 16267 (May 11, 2012).

79. K. Pereira de Jesus-Tran *et al.*, Comparison of crystal structures of human androgen receptor ligand-binding domain complexed with various agonists reveals molecular determinants responsible for binding affinity. *Protein science : a publication of the Protein Society* **15**, 987 (May, 2006).

80. V. B. Chen *et al.*, MolProbity: all-atom structure validation for macromolecular crystallography. *Acta crystallographica. Section D, Biological crystallography* **66**, 12 (Jan, 2010).

81. I. W. Davis *et al.*, MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**, W375 (Jul, 2007).

82. O. Heikinheimo, R. Kekkonen, P. Lahteenmaki, The pharmacokinetics of mifepristone in humans reveal insights into differential mechanisms of antiprogestin action. *Contraception* **68**, 421 (Dec, 2003).

83. L. N. Song, M. Coghlan, E. P. Gelmann, Antiandrogen effects of mifepristone on coactivator and corepressor interactions with the androgen receptor. *Molecular endocrinology* **18**, 70 (Jan, 2004).
84. L. K. Nieman *et al.*, Successful treatment of Cushing's syndrome with the glucocorticoid antagonist RU 486. *The Journal of clinical endocrinology and metabolism* **61**, 536 (Sep, 1985).
85. M. C. Hodgson *et al.*, The androgen receptor recruits nuclear receptor CoRepressor (N-CoR) in the presence of mifepristone via its N and C termini revealing a novel molecular mechanism for androgen receptor antagonists. *The Journal of biological chemistry* **280**, 6511 (Feb 25, 2005).
86. F. Cadepond, A. Ulmann, E. E. Baulieu, RU486 (mifepristone): mechanisms of action and clinical uses. *Annual review of medicine* **48**, 129 (1997).
87. S. Zhang, J. Jonklaas, M. Danielsen, The glucocorticoid agonist activities of mifepristone (RU486) and progesterone are dependent on glucocorticoid receptor levels but not on EC50 values. *Steroids* **72**, 600 (Jun, 2007).
88. M. F. Boehm *et al.*, Synthesis of high specific activity [3H]-9-cis-retinoic acid and its application for identifying retinoids with unusual binding properties. *Journal of medicinal chemistry* **37**, 408 (Feb 4, 1994).
89. E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state. *Journal of molecular biology* **372**, 774 (Sep 21, 2007).
90. S. P. Williams, P. B. Sigler, Atomic structure of progesterone complexed with its receptor. *Nature* **393**, 392 (May 28, 1998).
91. S. M. Carroll, E. A. Ortlund, J. W. Thornton, Mechanisms for the evolution of a derived function in the ancestral glucocorticoid receptor. *PLoS genetics* **7**, e1002117 (Jun, 2011).
92. R. K. Bledsoe *et al.*, A ligand-mediated hydrogen bond network required for the activation of the mineralocorticoid receptor. *The Journal of biological chemistry* **280**, 31283 (Sep 2, 2005).
93. W. B. Pratt, D. O. Toft, Steroid receptor interactions with heat shock protein and immunophilin chaperones. *Endocrine reviews* **18**, 306 (Jun, 1997).
94. R. E. Watkins *et al.*, The human nuclear xenobiotic receptor PXR: structural determinants of directed promiscuity. *Science* **292**, 2329 (Jun 22, 2001).
95. J. A. Patch, A. E. Barron, Helical peptoid mimics of magainin-2 amide. *Journal of the American Chemical Society* **125**, 12092 (Oct 8, 2003).

96. Y. Wang *et al.*, A second binding site for hydroxytamoxifen within the coactivator-binding groove of estrogen receptor beta. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 9908 (Jun 27, 2006).

CHAPTER 6: BEYOND MINIMAL SPECIFICITY: EVOLVING THE ABILITY TO DISCRIMINATE AMONG DIVERSE 3-KETOSTEROIDS

Jennifer K. Colucci[1], Geeta N. Eick[2], Joseph W. Thornton[2], and Eric A. Ortlund[1e]

[1]*Department of Biochemistry and Winship Cancer Institute, Emory University School of Medicine, Atlanta, GA 30033, USA*

[2]*Institute of Ecology and Evolution, University of Oregon, Eugene, Oregon, USA*

[3]*Howard Hughes Medical Institute, Eugene, Oregon, USA.*

Steroids differ primarily at their A- and D-rings; two classes of SRs have evolved to recognize these differences. Previously, we have investigated the mechanisms behind A-ring recognition (Chapter 3). However, here we investigate the structural mechanism driving the evolution of diverse functional group recognition on the D-ring. Using the previously-published structure of ancSR2 and a homology model of ancSR1, I identified three candidate amino acids that I believed might contribute to D-ring recognition and designed mutations to test this theory. Later, I analyzed data from a collaborating lab that tested the activity of ancSR2 with these three amino acids mutated and various steroids. This work has not yet been published.

[e] Conceived and designed the experiments: JKC GNE JWT EAO. Performed the experiments: JKC GNE. Analyzed the data: JKC GNE JWT EAO. Wrote the paper: JKC EAO.

Abstract

Protein function is determined by the interplay between a protein and its binding partners. These interactions allow the cell to respond to extracellular cues, such as steroid hormones, and respond accordingly. Understanding how specific molecular interactions evolve and the structural changes that drive this process is crucial to develop a complete understanding of hormone signaling pathways. Here, we use the closely related protein family of steroid receptors (SRs) to investigate the evolution of ligand specificity. Each SR has evolved to bind with high specificity to its cognate ligand with limited cross reactivity. The seminal SR (ancSR1) is responsive to estrogenic steroids, while the second-oldest SR (ancSR2) is responsive to nonaromatic 3-ketosteroids. These two steroid classes differ in two regions of their molecular scaffold, their A- and D-rings. We have previously shown the mechanism by which A-ring specificity evolved. Here, we use biochemistry, ancestral gene resurrection, and structural biology to identify and characterize the historical sequence changes that permit recognition of substituents found on the D-ring of nonaromatized steroids.

Introduction

Cellular functions are determined by the interplay between proteins and their binding partners. The evolution of these dynamic interactions can arise by many different processes. Recent work has shown that epistatic interactions, molecular exploitation, and molecular frustration have played large roles in the evolution of a sub-family of nuclear receptors [1-3]. These evolutionary processes, in concert, have built a closely related assemblage of decidedly specific receptor-ligand pairs.

Steroid receptors (SRs) play a critical role in metazoan physiology controlling the transcription of genes central to development, reproduction, homeostasis, and immunity [4]. SRs have a modular domain architecture consisting of an unstructured N terminal domain (NTD), a DNA binding domain, a small hinge region, and a ligand binding domain (LBD) [5, 6]. The ligand binding domains of these receptors respond to small lipophilic molecules that diffuse into the cell, triggering a translocation of the receptor-ligand complex from the cytoplasm to the nucleus [4]. Once in the nucleus, the DNA binding domain (DBD) interacts with target genes and recruits transcriptional coregulators to modulate transcriptional activity [4].

There are five extant steroid receptors in humans: the estrogen, progesterone, androgen, glucocorticoid, and mineralocorticoid receptors. These receptors respond to estrogen, progesterone, testosterone, cortisol, and aldosterone, respectively (Figure 6.1). Discrimination between steroids occurs primarily at two distal ends of the hormone, specifically the A- and D-rings (Figure 6.1). Estrogenic steroids have an aromatized A-ring with a small hydroxyl group at the carbon 3 and carbon 17 positions of the steroid. All other steroids have a non-aromatized A-ring with a keto group at the carbon 3 position. These nonaromatic three-keto steroids can have a variety of substituents at the seventeen position of the ring.

All steroid receptors evolved from a common ancestor, ancestral steroid receptor 1 (ancSR1), which was activated by estrogenic compounds (Figure 6.1) [7]. After a gene duplication event, the second copy of the gene, ancestral steroid receptor 2 (ancSR2), evolved to recognize an array of steroids with a nonaromatic A-ring and a three-keto substituent [8].

Previous work has shown that a single amino acid substitution in the ancSR1 LBD, E41Q, switched hormone preference from the aromatic steroid estrogen to 3-ketosteroids [3]. An additional amino acid substitution, L75M, further diminishes estrogen sensitivity by adding an additional unfulfilled weak hydrogen bond above the aromatic A-ring [3]. These forward substitutions create an ensemble of states with unfulfilled hydrogen bond potential when an aromatic steroid is bound. The affinity of ancSR1 E41Q/L75M for 3-ketosteroids is not as strong as that of the wild type ancSR2 [9]; therefore, there must be additional amino acid changes along the evolutionary trajectory that aided the sensitivity of ancSR2 for non-aromatic 3-ketosteroids with a 17-acetyl moiety. The homology model of ancSR1 and the crystal structure ancSR2 (PDB accession code: 4FN9) show an unfilled cavity at the distal end of the steroid, permitting steroids with a larger D-ring substituent to fit into the ligand-binding pocket (LBP) [9]. The cavity volume of the ancSR1-estrogen complex is 494.7 $Å^3$; while the cavity volume of the AncSR2-progesterone complex is 624.8 $Å^3$. This is due to an increase in the cavity near the D-ring, permitting steroids with bulky carbon 17 substituents to bind. Thus, the transition in D-ring sensitivity from ancSR1 to ancSR2 is facilitated by molecular exploitation of preexisting space and hydrogen bonding capacity within this pocket to permit the recognition of carbon 17 functional groups.

While the ancSR2 is competent to bind varied carbon 17 substituents, it most strongly prefers the hydroxyacetyl group present on corticosteroids (EC_{50} = ~1.0 nM) followed by that of progestagens, which contain an acetyl group at this position (EC_{50} = ~10 nM) [3]. AncSR1 lacks any activation by corticosteroids and progestagens [3]. The mode of corticosteroid activation in

nonaromatic 3-ketosteroid receptors has been well documented and is the result of an activating hydrogen bond network [1, 10], which is conserved in the ancSR2-DOC complex [8]. AncSR2 is also strongly activated by progestagens, which are nonaromatic 3-keto steroids with a smaller 17-acetyl group [3].

Through evolutionary, structural, and biophysical studies we have achieved a detailed understanding of the molecular forces that permitted the recognition of the nonaromatized A-ring in progestagens, corticoids, and androgens. However, it is unclear how the ancSR2 evolved a preference for the bulky D-ring substituents that differentiate these steroids from each other. We utilized resurrected ancestral proteins, structural biology, and biochemical methods to determine the molecular mechanism that drove evolution of 17-acetyl specificity. We analyzed a homology model of ancSR1 and the crystal structure of ancSR2 to identify substitutions that were critical for D-ring substituent recognition. We then examined all possible combinations of these residues to trace out an evolutionary pathway in which 17-hydroxyl sensitivity was diminished and 17-acetyl sensitivity was gained. This further extends our understanding of the structural and mechanistic forces that drove the evolution of 3-keto steroid specificity in steroid hormone receptors.

Materials and Methods

Reagents.

Chemicals were purchased from Sigma (St. Louis, MO) or Fisher (Hampton, NH). The vector for His tagged TEV was a gift from David Waugh (NCI, VA). pLIC_MBP and was a gift from John Sondek (UNC, Chapel Hill).

Structural Analysis.

The structures of the ancSR2 – progesterone and ancSR2 – DOC (PDB accession codes 4FN9 and 4FNE, respectively) were used for structural analysis [8]. The ancSR1 homology model was generated by manually threading the ancSR1 sequence onto human ERα (PDB accession code: 1ERE; 62 % sequence identity) using COOT v0.9 [11]. Side chains were manually adjusted to the most common rotamer and the resulting model was subjected to three rounds of gradient energy minimization in Phenix to idealize geometry [11, 12]. Cavity volumes were calculated using CastP [13]. Figures were generated with PyMol v1.5.0.1 (Schrodinger, LLC).

Mutagenesis.

Mutagenesis to reverse historical substitutions was performed using QuikChange (Stratagene) and verified by DNA sequencing.

Reporter activation assays.

AncSR2 LBD and ancSR1 LBD were cloned into the Gal4-DBD-pSG5 vector; 31 amino acids of the GR hinge containing the nuclear localization signal-1 were inserted between the Gal4 DBD and LBD to ensure nuclear localization and conformational independence of the two domains [8]. The hormone-dependent transcriptional activity of resurrected ancestral receptors and their variants was assayed using a luciferase reporter system. Briefly, CHO-K1 cells were grown in 96-well plates and transfected with 1.0 ng of receptor plasmid, 100.0 ng of a UAS-driven firefly luciferase reporter (pFRluc), and 0.1 ng of the constitutive phRLtk Renilla luciferase reporter plasmid, using Lipofectamine and Plus Reagent in OPTIMEM (Invitrogen). After 4 hours, transfection medium was replaced with phenol-red-free αMEM supplemented with 10% dextran-charcoal stripped FBS (Hyclone). After overnight recovery, cells were incubated in triplicate with the hormone of interest from 10^{-12} to 10^{-5} M for 24 hours, then assayed using Dual-Glo luciferase (Promega). Firefly luciferase activity was normalized to Renilla luciferase activity. Luminescence was read using a Synergy 4 microplate reader (BioTek). Dose-response relationships were

estimated using nonlinear regression in Prism4 software (GraphPad Software, Inc.); fold increase in activation was calculated relative to vehicle-only (ethanol) control.

Results

Comparison of estrogen versus progesterone recognition in the ligand binding pocket

To identify the amino acid substitutions that conferred full activation by steroids with differential carbon 17 substituents we analyzed the X-ray crystal structure of the ancestral SR2-Progesterone complex (PDB accession code: 4FN9) and created a homology model of the ancSR1-estrogen complex [8]. Comparison of the two models revealed three substitutions (A76i/L203g/F206h; ancSR2 residues in uppercase, ancSR1 residues in lower case) as candidates for enabling D-ring specific interaction. We hypothesize that reversing these amino acids in ancSR2 could recapitulate 17-hydroxyl steroid activation.

The ancSR1-estrogen complex shows the edge of the steroid along carbons 14 and 15 interacting with helix ten in the binding pocket (Figure 6.2). The position of estrogen in the LBP of ancSR1 facilitates a key hydrogen bond between the 17-hydroxyl and a conserved histidine (h206) [14]. Isoleucine 76 of ancSR1 facilitates additional hydrophobic interactions with carbons six and seven of estrogen, stabilizing the base of the steroid. Glycine 203 allows tight main chain contact with carbons fourteen and fifteen of estrogen, facilitating a key hydrogen bond interaction with h206. Progesterone, however, rotates 9.5 ° counterclockwise - with respect to estrogen - within the binding pocket of ancSR2. This rotation is primarily driven by the presence of L203 beneath the D-ring, which drives the steroid toward helix 3; the positioning of progesterone enables hydrogen bonding between the steroidal oxygen 20 and threonine 210 of ancSR2. This rotated binding mode is also conserved in the binding of corticosteroids, as shown in a previously published crystal structure [9]. The cavity volumes of the ancSR1-estrogen and ancSR2-progesterone complexes are 494.7 $Å^3$and 624.8 $Å^3$, respectively, which allows room for the larger carbon seventeen substituent [13]. This finding agrees with prior work that shows that receptors that are activated by steroids with larger substituents at carbon seventeen have larger cavity volumes [15].

Which amino acid substitutions facilitate recognition of bulky carbon 17 substituents?

Three amino acid reversions in ancSR2 increase 17-hydroxy steroid preference 1200 %. To test the role of these three amino acids (positions 76, 203, and 206) on D-ring recognition, we performed luciferase reporter assays with WT ancSR1, WT ancSR2 and mutant receptors that are specific combinations of ancSR1 and ancSR2. To ensure binding variances result from only differences localized to the steroid's D-ring, we used two synthetic ligands, 19-Nortestosterone (NorT) and 19-Norprogesterone (NorP), which differ only at the carbon 17 position. NorT has a hydroxyl group (similar to that of estrogen) and NorP has an acetyl group (similar to that of progesterone). As we hypothesized, the presence of the ancestral states at positions 76, 203, and 206 have a very low NorT/NorP EC_{50} ratio indicating preferential binding of the smaller steroid, NorT (Figure 6.3, Table 6.1). The presence of the evolutionarily derived states at all three positions increases the NorT/NorP value to 1200 % of the ancestral states.

The ability to tolerate functional changes in D-ring recognition is contingent upon additional epistatic interactions. There are several states with mixed populations of ancestral and derived amino acids that lead to increased EC_{50} values and decreased maximal activation. States containing the ancestral i76 and the derived L203 show very low ligand binding activity with both NorT and NorP. Modeling these two residues into the binding pocket of ancSR2 reveals a steric clash between the side chains of i76 and L203 for four of the six rotamers of isoleucine (Figure 6.4). To maintain activation throughout its evolution, ancSR1 likely accumulated the permissive i76A substitution first, followed by the g203L substitution (Figure 6.5). A mix of the derived and ancestral states of A76 and g203 shows a large increase in the EC_{50}s for both NorP and NorT. This state lacks the supporting hydrophobic interactions along the steroids' carbons four, six, seven, fourteen, and fifteen that act as a driving force for steroid binding. Leucine 203 is

incompatible with a ligand lying flat against H10 in the LBP due to a steric clash with the ligand. Thus, the g203L substitution repositions the ligand in the binding pocket by a rotation toward H3, as seen in the ancSR2-progesterone and ancSR2-DOC crystal structures [9]. This repositioning places the carbon 17 substituent in a cavity with exploitable hydrogen bonding capacity in both the ancSR1 and ancSR2 states. WT ancSR2 retains the ability to be activated by NorT, likely through rotation of the ligand and exploitation of the hydrogen bonding capability of T210.

A shift to NorT preference over NorP in ancSR2 g203L/h206F/i76A was not achieved. This could stem from the contributions of the nonaromatic A-ring in NorT anchoring the ligand in a position where it is hindered from rotating toward H3 and thus fulfilling its optimum hydrogenbond capabilities. It is possible that there are additional changes between the ancSR1 and ancSR2 that contribute to D-ring selectivity; addition of these reverse mutations into ancSR2 may serve to recapitulate 17-hydroxyl preference.

Discussion

AncSR2 ligand preference evolved through a series of discrete steps. Large effect mutations in ancSR2, such as q41E [3] likely occurred first, yet epistasis shaped robust 17-acetyl activation by requiring three additional substitutions g203L/h206F/i76A (Figure 6.5). The i76A switch was required in order to tolerate the remaining two substitutions; these had the net effect of repositioning the steroid in the pocket to confer tighter binding and greater activation by 17-acetyl steroids. The evolution of 17-acetyl specificity was co-opted from the increased cavity volume near carbons thirteen and seventeen and available hydrogen bond capability. This exploitation of existing capabilities is not a new notion in molecular evolution theory [2, 16-18]. The mineralocorticoid receptor's affinity for aldosterone was also a by-product of evolution long before aldosterone had evolved [19].

Earlier work has shown that in SR evolution, epistatic interactions played a major role in guiding the evolution of ligand specificity in corticoid receptors [1]. Seemingly neutral mutations within the LBP became non-neutral (permissive) in the context of function-shifting substitutions. These substitutions altered protein-ligand and intra-protein interactions to shape ligand selectivity. Thus, the results seen here mimic an important evolutionary theme of epistasis steering molecular evolution.

The modification of the ligand position within the LBP seen in the ancSR1 to ancSR2 transition is a phenomenon also seen in another steroid binding pocket. Elephants have low levels of progesterone, and have evolved to use a unique steroid, 5α–dihydroprogesterone, which has low affinity for the human progesterone receptor [20]. However, in the elephant progestin receptor (ePR), a single glycine to alanine substitution repositions the ligand within the binding pocket in order to increase affinity for elephant's unique gestagin [20].

Pairing the D-ring substitutions discussed here with the A-ring substitutions previously published [3], we can fully recapitulate the two-pronged shift in ligand-contacting residues necessary for differential ligand activation. Now that the earliest evolutionary shifts from

estrogenic to nonaromatic 17-acetyl steroids and the evolution of corticoid specificity have been described, the evolution of androgen specificity remains an interesting and highly relevant topic. The androgen receptor (AR) is a unique 3-ketosteroid receptor in that its ligand's carbon 17 substituent reverts back to a hydroxyl, like that of estrogen (Figure 6.1). We envision that a similar approach to those used here could help to understand the evolution of the acquisition of 17-hydroxyl specificity in nonaromatic 3-ketosteroid receptors. The interaction of AR with its ligand has been widely discussed in the literature due to the implication of the receptor in testicular and prostate cancers and Androgen Insensitivity Syndrome. A deeper understanding AR's response to its ligand could help to build a molecular framework for synthetic selective AR modulators.

Figures

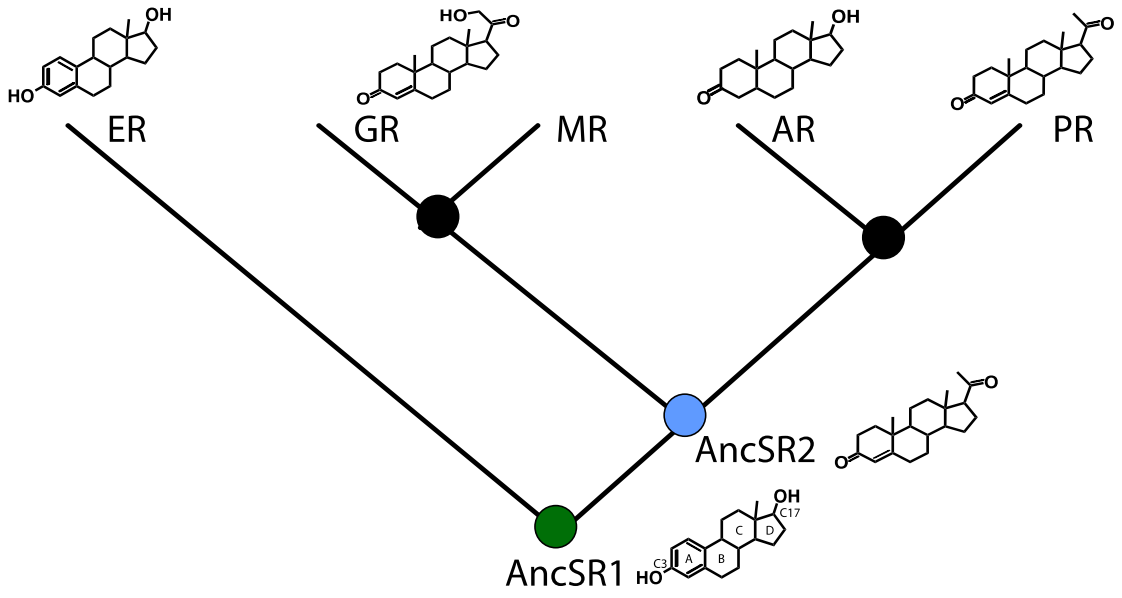


Figure 6.1: Phylogeny of the Steroid Receptor lineage.

This reduced cladogram of SR expansion shows the interrelationship of each of the SRs and their cognate ligands. The endogenous ligand for each receptor is shown at the respective node or branch tip. The ancSR1 and estrogen receptor (ER) respond to estrogen. Both the glucocorticoid receptor (GR) and mineralocorticoid receptor (MR) are responsive to deoxycorticosterone (DOC). The ancSR2 and progesterone receptor (PR) are responsive to progesterone. The androgen receptor (AR) is responsive to dihydrotestosterone (DHT). Node representing ancSR1 (green) and ancSR2 (blue) are indicated with filled circles. Carbons 3 and 17 and the A-, B-, C-, and D-rings are labeled.

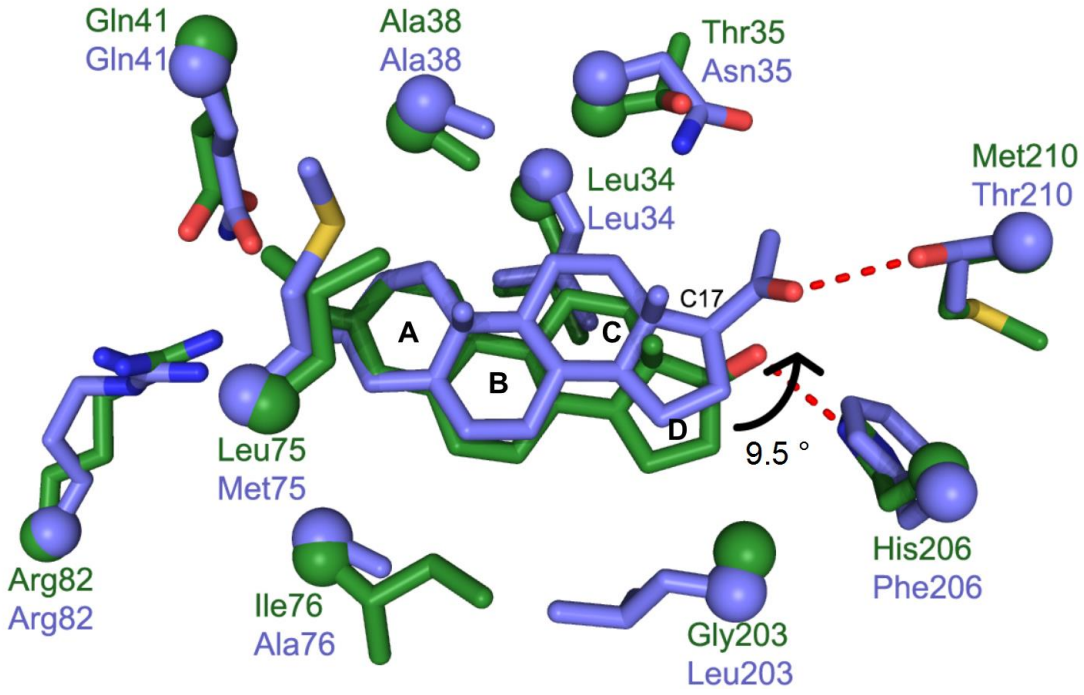


Figure 6.2: Rotation of the 17-acetyl ligand in the binding pocket allows for exploitation of pre-existing hydrogen bond capacity.

Ligand binding pocket of the ancSR1-Estrogen homology model (green) aligned with the ancSR2-Progesterone crystal structure (blue) (4FN9). The carbon 3 position of the steroids overlap, while the distal end of the steroid is free to rotate within the pocket; arrow indicates the direction of rotation. The carbonyl is available to make a hydrogen bond with either T210 or N35.

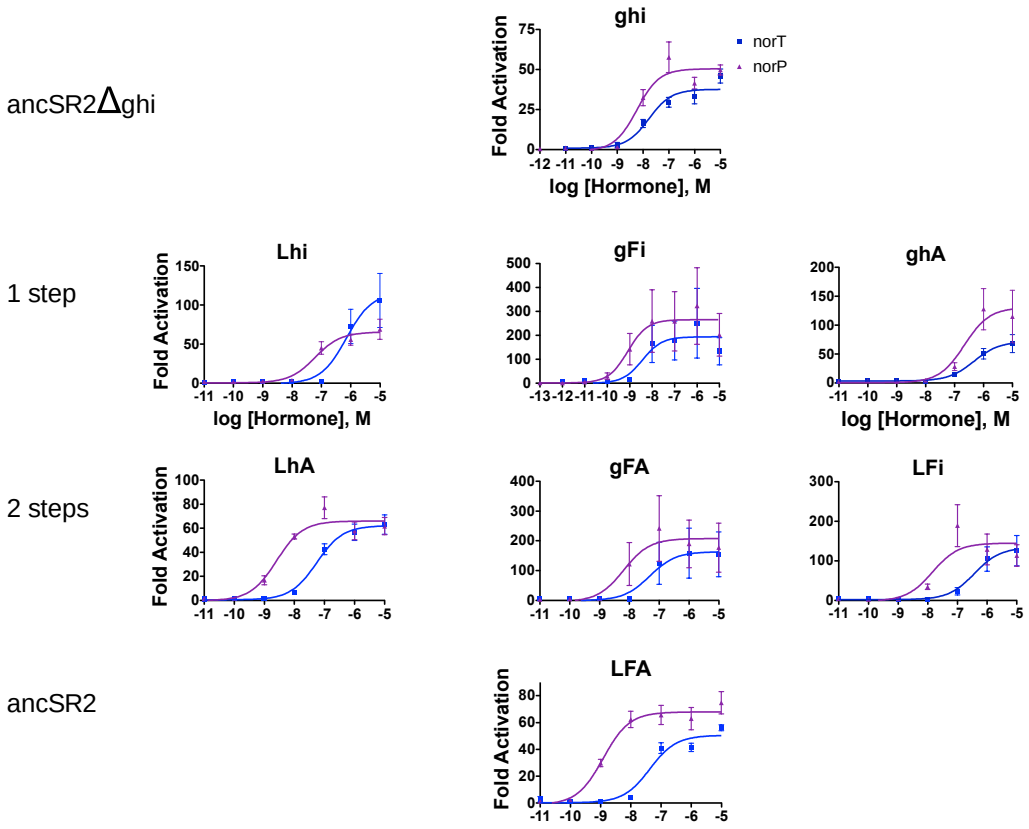


Figure 6.3: Forward Evolution of D-ring residues increased preference for 17-acetyl ligands.

Dose response curves of ancSR2 WT and mutants upon treatment with NorP (purple trianges) and NorT (blue squares). Graphs show fold activation [Firefly luciferase/ Renilla luciferase (FFL/RL)] vs. log[hormone] (M). Values can be seen in Table 6.1. Lower case letters represent ancestral states; capital letters represent derived states (A76i/L203g/F206h; ancSR2 residues in uppercase, ancSR1 residues in lower case) . N=3. *Data generated by Geeta N. Eick, Thornton lab*.

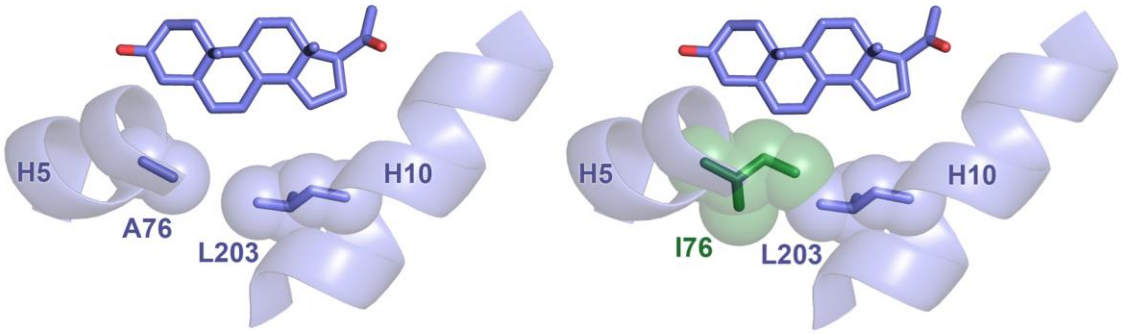


Figure 6.4: Epistatic interactions shaped ancSR2 evolution.

A close-up of the WT ancSR2 (blue) LBP bound to progesterone. Ligands and side chains shown as sticks, helices shown as cartoons; derived state - slate blue; ancestral state - green; oxygens – red; van der Waals surfaces shown for side chains. **A**. Residues A76 and L203 are 4.1 Å apart and do not clash. **B**. The ancestral i76 is modeled into the ancSR2 backbone. i76 and L203 are 1.9 Å apart and show a clear steric clash. Lower case letters represent ancestral states; capital letters represent derived states.

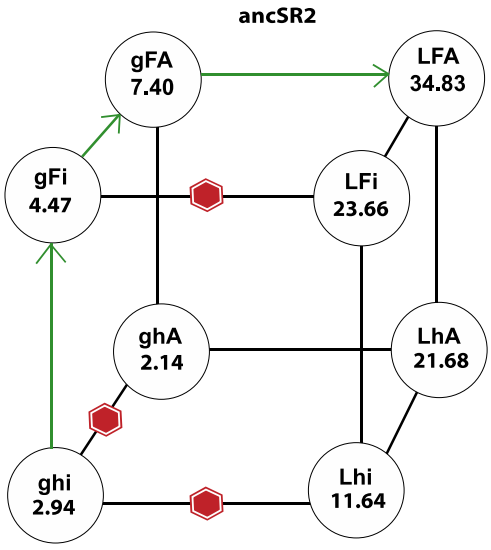


Figure 6.5: Evolutionary pathway to the evolution of 17-acetyl recognition.

Each vertex of the cube represents a receptor variant; letters are residues at positions 203/206/76. Red hexagons denote unlikely evolutionary trajectories. Values within the circles are NorT/NorP EC_{50} ratios defined in Table 6.1. Lower case letters represent ancestral states; capital letters represent derived states.

	Hormone Sensitivity (EC_{50}), nM		
ancSR2	NorT	NorP	NorT/NorP
ghi	16.9	5.8	2.9
ghA	466.7	218.3	2.1
gFi	3.7	0.8	4.5
Lhi	711.2	61.1	11.6
gFA	48.0	6.5	7.4
LhA	55.2	2.6	21.7
LFi	342.8	14.5	23.7
LFA	18.5	0.5	34.8

Table 6.1: Hormone sensitivity of WT ancSR2 and mutants.

In luciferase reporter assays, seven mutants of ancSR2 and the WT were treated with NorP and NorT. Table shows EC_{50} values (in nM). The ratio of NorT/NorP shows the fold preference for NorP. Lower case letters represent ancestral states; capital letters represent derived states. *Data generated by Geeta N. Eick.*

References

1. Bridgham, J.T., E.A. Ortlund, and J.W. Thornton, *An epistatic ratchet constrains the direction of glucocorticoid receptor evolution.* Nature, 2009. **461**(7263): p. 515-9.
2. Bridgham, J.T., S.M. Carroll, and J.W. Thornton, *Evolution of hormone-receptor complexity by molecular exploitation.* Science, 2006. **312**(5770): p. 97-101.
3. Harms, M.J., et al., *Biophysical mechanisms for large-effect mutations in the evolution of steroid hormone receptors.* Proc Natl Acad Sci U S A, 2013. **110**(28): p. 11475-80.
4. Nagy, L. and J.W. Schwabe, *Mechanism of the nuclear receptor molecular switch.* Trends Biochem Sci, 2004. **29**(6): p. 317-24.
5. Wurtz, J.M., et al., *A canonical structure for the ligand-binding domain of nuclear receptors.* Nat Struct Biol, 1996. **3**(2): p. 206.
6. Govindan, M.V., et al., *Cloning of the human glucocorticoid receptor cDNA.* Nucleic Acids Res, 1985. **13**(23): p. 8293-304.
7. Eick, G.N. and J.W. Thornton, *Evolution of steroid receptors from an estrogen-sensitive ancestral receptor.* Mol Cell Endocrinol, 2011. **334**(1-2): p. 31-8.
8. Volleth, M. and G. Eick, *Chromosome evolution in bats as revealed by FISH: the ongoing search for the ancestral chiropteran karyotype.* Cytogenet Genome Res, 2012. **137**(2-4): p. 165-73.
9. Eick, G.N., et al., *Evolution of minimal specificity and promiscuity in steroid hormone receptors.* PLoS Genet, 2012. **8**(11): p. e1003072.
10. Bledsoe, R.K., et al., *A ligand-mediated hydrogen bond network required for the activation of the mineralocorticoid receptor.* J Biol Chem, 2005. **280**(35): p. 31283-93.
11. Emsley, P., et al., *Features and development of Coot.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.
12. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.
13. Dundas, J., et al., *CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues.* Nucleic Acids Res, 2006. **34**(Web Server issue): p. W116-8.
14. Tanenbaum, D.M., et al., *Crystallographic comparison of the estrogen and progesterone receptor's ligand binding domains.* Proc Natl Acad Sci U S A, 1998. **95**(11): p. 5998-6003.

15. Madauss, K.P., et al., *Progesterone receptor ligand binding pocket flexibility: crystal structures of the norethindrone and mometasone furoate complexes.* J Med Chem, 2004. **47**(13): p. 3381-7.
16. Gompel, N., et al., *Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila.* Nature, 2005. **433**(7025): p. 481-7.
17. Piatigorsky, J., *Multifunctional lens crystallins and corneal enzymes. More than meets the eye.* Ann N Y Acad Sci, 1998. **842**: p. 7-15.
18. Ryan, M.J., *Sexual selection, receiver biases, and the evolution of sex differences.* Science, 1998. **281**(5385): p. 1999-2003.
19. Ortlund, E.A., et al., *Crystal structure of an ancient protein: evolution by conformational epistasis.* Science, 2007. **317**(5844): p. 1544-8.
20. Wierer, M., et al., *A single glycine-alanine exchange directs ligand specificity of the elephant progestin receptor.* PLoS One, 2012. **7**(11): p. e50350.

CHAPTER 7: DISCUSSION

These collective works take a holistic approach towards understanding the evolutionary mechanisms that generate protein-ligand specificity. These range from sequence prediction and gene synthesis to *in vitro* and *in vivo* assays and structural biology to connect sequence changes to the functional shifts that created SR-signaling pathways.

How did the differences in ligand specificity between the ERs and naSRs evolve?

Humans have two phylogenetic classes of SRs (NR3A1 and NR3C1-4), which correspond to the chemical classes of endogenous ligands that activate each receptor's LBD. The first class (NR3A1) contains the estrogen receptors (ERs), whose endogenous ligands are 18-carbon steroids with an aromatized A-ring and a hydroxyl attached to carbon 3 on the steroid skeleton (Figure 2.1A). The second class (NR3C1-4) contains the nonaromatized steroid receptors (naSRs) – includes receptors for androgens (androgen receptor), progestagens (progesterone receptor), glucocorticoids (glucocorticoid receptor), and mineralocorticoids (mineralocorticoid receptor). These ligands all contain a nonaromatized A-ring, an additional methyl at carbon 19, and, in most cases, a ketone at carbon 3 (Figure 2.1A). We sought to understand the basis for the evolution of ligand specificity between ERs and naSR receptors.

Our findings, viewed in the context of the ancient pathway for steroid hormone synthesis, suggest that some hormone-receptor pairs were assembled during evolution by a process of molecular exploitation, where molecules with a different ancient function were recruited into new signaling partnerships after gene duplication and/or functional divergence [1, 2]. That the ancient ancSR1 was specific for estrogens implies that progestagens and androgens, which are intermediates in the synthesis of estrogens (Figure 2.1A), existed before steroid receptors evolved to transduce their signals. When ancSR2 and its descendants evolved the capacity to be activated by nonaromatized steroids, these intermediary biochemical stepping stones in estrogen synthesis were recruited into new, *bona fide* signaling partnerships.

We concluded that an inversion of ligand specificity for endogenous steroid hormones took place during the evolutionary interval between ancSR1 and ancSR2. This inversion of preference differs from a previously-seen phenomenon that shows evolution occurring via a narrowing of specificity from a promiscuous ancestor [3]. In a case where specificity narrows, the protein begins with broad specificity and its ancestors loses the ability to activate by a number of ligands. Here, we see that the ancestor lacks any of its predecessors' specificities and instead takes on novel function. After this inversion of function, the promiscuous responses of ancSR2 to nonaromatized steroids were differentially partitioned among its descendant lineages to yield the more specific PR, GR, MR, and AR.

This evolution of ligand specificity arose according to a principle of minimal specificity: at each point in time, SRs evolved to be specific enough to distinguish among the substances to which they were naturally exposed, but not more so [4]. Our findings provide an historical explanation for modern SRs' diverse sensitivities to natural and man-made substances, including the extant SRs susceptibility to synthetic endocrine disruptors [4]. They show that knowledge of history can contribute to predicting the ligands to which a modern protein will respond and indicate that promiscuity reflects the restricted capacity of natural selection to differentiate between perfect and "good enough." Structurally, the basis of the promiscuous responses of both ancSR1 and ancSR2 to non-target ligands were found to be due in large part to unfilled volume in the ligand binding pocket and untapped potential of polar side chains to form hydrogen bonds with polar atoms on the ligand [5, 6].

What are the mechanisms that dictate the ligand preferences of ERs and naSRs?

After identifying the sequence changes during the functional switch between ERs and naSRs, we sought to understand the biophysical mechanisms dictating this switch [4, 7] (see Chapters 2, 3). By combining ancestral reconstruction with studies of protein structure and

dynamics, we showed how two historical mutations remodeled the hydrogen-bonding network between hormone and SR, changing the dynamics of the complex in a ligand-specific way that radically shifted the receptor's hormone specificity.

Using molecular dynamics experiments to predict the molecular networks of the ligand-receptor pairs, we were able to interpret the effects of residue identity on the complex's energetic landscape (Chapter 3, Figure 3.3). The derived residues in ancSR2 established a new favorable interaction with 3-ketosteroids while simultaneously establishing a frustrated ensemble of suboptimal hydrogen bonding networks for aromatized steroids. This specifically excluded estrogens from the ancSR2 ligand binding pocket by introducing new interaction partners (i.e. the functional groups on amino acid side chains) that could not be simultaneously satisfied, given the position of the 3-hydroxyl on aromatized steroids.

Our findings shed light on the classic evolutionary debate concerning the size distribution of mutational effects during evolution [8, 9] and reveal the underlying causes of that distribution during the evolution of a biologically important new function. Two historical replacements were sufficient to drive a huge shift in SR ligand preference. Despite having no apparent effect on global protein structure and minor effects on the local biochemical properties of the ligand binding 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 hydrogen-bonding network when the ancestral ligand was present (Chapter 3, Figure 3.3).

Our observations underscore the fact that proteins evolve as complex systems with astronomical degrees of freedom and nonlinear relationships between sequence, biophysical properties, and function [10]. A protein-ligand complex samples a huge number of conformational microstates; just one or a very few mutations may cause states with different conformations and couplings to become energetically favorable [11]. This can magnify subtle changes in the biochemistry of a few amino acids into large perturbations in the protein's

behavior and, in turn, cause major evolutionary shifts in function. We also show that protein structure acts as more than a negative constraint on the freedom of protein sequences to evolve [12]: it also enables relatively small steps in sequence space to produce dramatic evolutionary shifts in a protein's biochemical behavior and biological functions.

How can ancestral proteins be used to understand modern pharmacology?

Steroid receptors are key regulators of metazoan physiology and are heavily targeted for pharmaceutical intervention. However, most currently approved SR-targeting pharmaceuticals have off-target pharmacology due to the high degree of structural similarity within the SR family. A better understanding of the SR-ligand relationship may enable the design of better pharmaceutical ligands. Given their increased stability, ancestral steroid receptors are ideal tools for the extensive mutagenesis required to build robust structure-function relationships for both endogenous and synthetic ligands [13, 14]. This same property makes them ideal tools to obtain crystal structures low affinity or weak SR modulators (i.e. lead compounds) that have been recalcitrant to crystallization. These tools have previously been successful in understanding the affinity of ancestral corticoid receptors for both endogenous ligands and modern pharmaceuticals [14].

To investigate the relationship between ancestral receptors and modern pharmaceutical ligands, we attempted to crystalize ancSR2 with the SR antagonist mifepristone. Surprisingly, we found mifepristone fortuitously docked at the protein surface, poised to interfere with coregulator binding. Recent attention has been given to generating pharmaceuticals that block the coregulator binding site in order to obstruct coregulator binding and achieve tissue-specific SR regulation independent of hormone binding [15, 16]. Mifepristone's interaction with the coactivator cleft of this SR suggests that it may be a useful molecular scaffold for further coactivator binding inhibitor (CBI) development.

These novel coactivator binding pharmaceuticals would be instrumental in the treatment of a range of diseases. Currently, there is a struggle to design effective peptidomimetic CBIs due to their inability to permeate the cell membranes [17]. Mifepristone is already well established as having effective extracellular to intracellular transport and thus shows strong potential as a pharmaceutical scaffold. Further work is needed to determine whether mifepristone or mifepristone derivatives are able to compete for the coactivator cleft in extant steroid receptors.

This structure is only the second to show a small molecule bound to the coactivator cleft of a steroid receptor. The first showed the anti-cancer drug 4-hydroxytamoxifen (HT) bound at the coactivator cleft of estrogen receptor beta (ERβ) (PDB accession code 2FSZ) [18]. Superposition of these two structures reveals that the ligands adopt nearly identical positions at the coactivator cleft (Figure 5.7). Both insert phenolic substituents into the H3/H4 gap and are held in place primarily by hydrophobic interactions. Future work should be devoted to investigating whether mifepristone has potential for acting as a scaffold for further CBI development.

How do epistatic interactions influence the evolution of ligand specificity?

We have shown that ancSR2 ligand preference evolved through a series of steps. Large effect mutations in ancSR2, such as Q41E [7] likely occurred first, yet epistasis shaped robust 17-acetyl activation by requiring three additional substitutions: G203L/H206F/I76A (Figure 6.5). An isoleucine 76 to alanine switch was required in order to tolerate the remaining two substitutions; these had the net effect of repositioning the steroid in the pocket to confer tighter binding and greater activation by 17-acetyl steroids. This evolution of 17-acetyl specificity was enabled by the increased cavity volume near carbons thirteen and seventeen and the available hydrogen bonding capability. This exploitation of existing hydrogen-bonding capabilities is not a new notion in the field of evolution of ligand specificity [19-22], and was recently shown to be a

primary component in the evolution of ancSR2 A-ring specificity [4]. In the case of the mineralocorticoid receptor (MR)-aldosterone relationship, MR's affinity for aldosterone was a by-product of evolution long before aldosterone had evolved [13].

Earlier work has shown that later in the SR evolutionary trajectory, during the functional specialization of CRs, epistatic interactions have guided the evolution of ligand specificity [23]. Seemingly neutral mutations within the ligand binding pocket became non-neutral (permissive) in the context of function-shifting substitutions. These substitutions altered protein-ligand and intra-protein interactions to shape ligand selectivity. Thus, these results mimic an important evolutionary theme whereby epistasis steers molecular evolution.

Composite discussion

Since protein-ligand pairs evolved deep in the past, we can only speculate on the evolutionary mechanism that created them. Unfortunately, there are no fossils available to shed light on this important process. Through sequence analyses, we have resurrected several of the genes representing the most likely common ancestors to extant steroid receptors (SRs). These fully-functional proteins are tools that allow us to study evolution and ligand specificity. The benefit of using these ancestral proteins to analyze the evolution of ligand specificity is that there are fewer amino acid differences between related receptors (demonstrating differential ligand activation) versus modern proteins. This allows us to focus on a smaller number of amino acids when examining the biophysical and structural mechanisms for altered hormone or drug recognition. In addition, these receptors are more tolerant to the structural perturbation introduced in mutagenesis studies. Therefore, they allow us to explore many more mutational combinations than would be possible with the modern receptors.

At its most fundamental, the sum of these studies provides an essential understanding of the dynamics of SR-ligand interactions. Since SRs play such a central role in metazoan physiology and their misregulation is implicated in cancers, inflammatory diseases, and cardiac diseases, it is essential to understand the means by which ligand – mediated SR responses are generated. SR function is easily modulated via pharmaceutical intervention due to the manipulability of their molecular switch [24].

The ligand binding domains of all SRs share a highly conserved three-dimensional structure due to their close evolutionary relationship (Figure 3.5). The natural ligands of all SRs, with the exception of the evolutionarily distant estrogen receptor, are 3-ketosteroids with small substituent variations. Cross-reactivity among these similar receptors causes the side effects seen with pharmaceuticals targeting these receptors. For example, most clinically-used synthetic progesterone receptor agonists, such as those used as birth control, can bind to MR, inhibiting its

function and causing hypertension and weight gain [25, 26]. Cross-reactivity of PR ligands with AR lead to decreased libido; cross-reactivity with the estrogen receptor lead to breast tenderness [25]. Thus far, structure-function studies of modern receptors have been ineffective in overcoming this complex problem. The mechanism underlying this promiscuity is derived from the fact that SRs share common structural features derived from their evolutionary relationship. Despite fifty years of intense effort, attempts to address this off-target pharmacology have proved unsuccessful. Through our efforts, we have identified the receptor – ligand interactions that are crucial for endogenous ligand recognition. Identification of these mechanisms of selectivity may allow us to build a conceptual framework for designing receptor-specific pharmaceuticals that lack off-target pharmacology.

While at its most basic, these studies show the structural, dynamic, and sequence changes that dictate ligand specificity; at its most ambitious, these data could be scaled up to shed insight on the mechanisms dictating ligand specificity in protein families with alternate folds. The Structural Classification of Proteins (SCOP) reports 1,393 unique protein folds (based on RMSD value) as of the year 2013. It is currently unknown whether or not the same principles that apply to ligand recognition for the SR fold (i.e. epistatic interactions working in concert to alter ligand affinity and activation, alteration of the protein's conformational dynamics upon ligand binding, change in protein stability depending on the identity of the bound ligand) can be extrapolated across protein-ligand evolutionary relationships for various protein folds. New research in molecular evolution will reveal the transferability of our findings. Alternatively, the null hypothesis would state that the rules learned here about dictating these evolutionary relationships are not portable and would need to be discovered for each of the nearly 1,400 known protein folds.

An ultimate goal of the protein engineering field is to generate the ability to design *de novo* protein-ligand interactions. This has been attempted several times with limited success. In one of the most successful examples, Baker et al. endeavored to develop novel enzymes to

catalyze reactions that are not carried out by natural biological catalysts [27]. However, the efficiency of these enzymes is still several orders of magnitude less than the "average enzyme" [27, 28], suggesting that even the best efforts at designing protein interactions are unsatisfactory. More recently, structure-based drug design has attempted to rationally design ligands for disease-contributing receptors. While these efforts have shown limited success [29], researchers believe that there is still "adequate room for the development of more sophisticated methodologies" [30].

Through providing a unified account of historical evolution at genetic, biophysical, and functional levels, this project has advanced our understanding of the evolutionary process. It has provided a mechanistic explanation that links genetic phenomena (such as epistasis and the distribution of large- and small-effect mutations) to classes of biophysical mechanisms, provided a rationale for why important evolutionary transitions took the genetic pathways that they did, and explained why extant proteins ended up with the architectures they now have. This project has also provided researchers with a model for how to connect evolutionary genetics to protein biochemistry using diverse techniques in a unified conceptual framework.

Future Directions

Can we completely recapitulate the switch in hormone selectivity from ancSR1 to ancSR2?

Thus far, the two manuscripts that have addressed the hormone switch from the ancSR1 to ancSR2 have been focused on the A- or D-ring substituent changes alone (see Chapters 3 & 5). This research showed that amino acid positions 41, and 75 are critical for dictating A-ring preference and positions 76, 203, and 206 are critical for dictating D-ring preference (16). This research relied on non-natural steraloids that differed from endogenous steroids at only their A- or D-rings. Pairing the two sets of residue changes (for a total of five exchanges) between ancSR1 and ancSR2 should theoretically yield a complete hormone switch. An experiment that introduced e41Q, l75M, i76A, g203L, and h206F into ancSR1 or Q41e, M75l, A76i, L203g, F206h into ancSR2 should alter preference to progestagens and estrogens, respectively. If this combination of

five amino acid exchanges is not sufficient to alter hormone preference to the levels at which they are in the WT receptor, then there are additional changes necessary that went unseen in our original analyses.

What factors contribute to the evolution of androgen specificity?

Corticoids and progestagens, the primary ligands of GR, MR, and PR, all have bulky substituents at their carbon 17 positions. The androgen receptor, despite its high sequence and structural similarities to GR, MR, and PR, differs significantly from these receptors in that it binds ligand with a small hydroxyl group at its carbon 17 position (Figure 6.1). This change in specificity represents a reversion back to a carbon seventeen substituent that was recognized by AR's ancestor, ancSR1, over 500 mya. In order to understand AR specificity, it is necessary to identify the changes that occurred in the AR lineage that excluded large carbon seventeen substituents in favor of a hydroxyl substituent. The completion of this study will provide a unified account of the historical evolution of steroid receptor-ligand interactions at the genetic, biophysical, and functional levels.

A comparison of the sequences as well as the structures of the ancAR1-DHT-Tif2 complex (Chapter 4) and the extant human AR – DHT complex (PDB accession code: 1T7T, [31]) reveal several promising targets, namely residues 880 and 887 (hAR numbering). Preliminary data suggest that these residues may be responsible for shifting ligand responsiveness from progestagens to androgens (data not shown) between ancSR3 and ancAR1, however the biophysical mechanism for this switch remains unknown. Using similar techniques as previous studies (luciferase reporter assays, fluorescence polarization, HDX, and molecular dynamics simulations) to analyze the wild type and mutant receptors, the mechanism of the ligand switch should be revealed.

References

1. Thornton, J.W., E. Need, and D. Crews, *Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling.* Science, 2003. **301**(5640): p. 1714-7.
2. Thornton, J.W., *Evolution of vertebrate steroid receptors from an ancestral estrogen receptor by ligand exploitation and serial genome expansions.* Proc Natl Acad Sci U S A, 2001. **98**(10): p. 5671-6.
3. Wouters, M.A., et al., *A despecialization step underlying evolution of a family of serine proteases.* Mol Cell, 2003. **12**(2): p. 343-54.
4. Eick, G.N., et al., *Evolution of minimal specificity and promiscuity in steroid hormone receptors.* PLoS Genet, 2012. **8**(11): p. e1003072.
5. Copley, S.D., *Enzymes with extra talents: moonlighting functions and catalytic promiscuity.* Curr Opin Chem Biol, 2003. **7**(2): p. 265-72.
6. Khersonsky, O. and D.S. Tawfik, *Enzyme promiscuity: a mechanistic and evolutionary perspective.* Annu Rev Biochem, 2010. **79**: p. 471-505.
7. Harms, M.J., et al., *Biophysical mechanisms for large-effect mutations in the evolution of steroid hormone receptors.* Proc Natl Acad Sci U S A, 2013. **110**(28): p. 11475-80.
8. Charlesworth, D. and B. Charlesworth, *Theoretical genetics of Batesian mimicry I. single-locus models.* J Theor Biol, 1975. **55**(2): p. 283-303.
9. Orr, H.A., *The genetic theory of adaptation: a brief history.* Nat Rev Genet, 2005. **6**(2): p. 119-27.
10. Alexander, P.A., et al., *A minimal sequence code for switching protein structure and function.* Proc Natl Acad Sci U S A, 2009. **106**(50): p. 21149-54.
11. Weber, G., *Energetics of ligand binding to proteins.* Adv Protein Chem, 1975. **29**: p. 1-83.
12. Worth, C.L., S. Gong, and T.L. Blundell, *Structural and functional constraints in the evolution of protein families.* Nat Rev Mol Cell Biol, 2009. **10**(10): p. 709-20.
13. Ortlund, E.A., et al., *Crystal structure of an ancient protein: evolution by conformational epistasis.* Science, 2007. **317**(5844): p. 1544-8.
14. Kohn, J.A., K. Deshpande, and E.A. Ortlund, *Deciphering modern glucocorticoid cross-pharmacology using ancestral corticosteroid receptors.* J Biol Chem, 2012. **287**(20): p. 16267-75.
15. Rodriguez, A.L., et al., *Design, synthesis, and in vitro biological evaluation of small molecule inhibitors of estrogen receptor alpha coactivator binding.* J Med Chem, 2004. **47**(3): p. 600-11.

16. Estebanez-Perpina, E., et al., *A surface on the androgen receptor that allosterically regulates coactivator binding.* Proc Natl Acad Sci U S A, 2007. **104**(41): p. 16074-9.
17. Patch, J.A. and A.E. Barron, *Helical peptoid mimics of magainin-2 amide.* J Am Chem Soc, 2003. **125**(40): p. 12092-3.
18. Wang, Y., et al., *A second binding site for hydroxytamoxifen within the coactivator-binding groove of estrogen receptor beta.* Proc Natl Acad Sci U S A, 2006. **103**(26): p. 9908-11.
19. Gompel, N., et al., *Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila.* Nature, 2005. **433**(7025): p. 481-7.
20. Piatigorsky, J., *Multifunctional lens crystallins and corneal enzymes. More than meets the eye.* Ann N Y Acad Sci, 1998. **842**: p. 7-15.
21. Ryan, M.J., *Sexual selection, receiver biases, and the evolution of sex differences.* Science, 1998. **281**(5385): p. 1999-2003.
22. Bridgham, J.T., S.M. Carroll, and J.W. Thornton, *Evolution of hormone-receptor complexity by molecular exploitation.* Science, 2006. **312**(5770): p. 97-101.
23. Bridgham, J.T., E.A. Ortlund, and J.W. Thornton, *An epistatic ratchet constrains the direction of glucocorticoid receptor evolution.* Nature, 2009. **461**(7263): p. 515-9.
24. Nagy, L. and J.W. Schwabe, *Mechanism of the nuclear receptor molecular switch.* Trends Biochem Sci, 2004. **29**(6): p. 317-24.
25. Madauss, K.P., E.L. Stewart, and S.P. Williams, *The evolution of progesterone receptor ligands.* Med Res Rev, 2007. **27**(3): p. 374-400.
26. Quinkler, M., et al., *Enzyme-mediated protection of the mineralocorticoid receptor against progesterone in the human kidney.* Mol Cell Endocrinol, 2001. **171**(1-2): p. 21-4.
27. Rothlisberger, D., et al., *Kemp elimination catalysts by computational enzyme design.* Nature, 2008. **453**(7192): p. 190-5.
28. Bar-Even, A., et al., *The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters.* Biochemistry, 2011. **50**(21): p. 4402-10.
29. Druker, B.J., *STI571 (Gleevec) as a paradigm for cancer therapy.* Trends Mol Med, 2002. **8**(4 Suppl): p. S14-8.
30. Mandal, S., M. Moudgil, and S.K. Mandal, *Rational drug design.* Eur J Pharmacol, 2009. **625**(1-3): p. 90-100.
31. Hur, E., et al., *Recognition and accommodation at the androgen receptor coactivator binding interface.* PLoS Biol, 2004. **2**(9): p. E274.