MECHANISMS OF EVOLUTION BY GENE DUPLICATION: THE ORIGINS OF CORTICOSTEROID SIGNALING

by

SEAN MICHAEL CARROLL

A DISSERTATION

Presented to the Department of Biology
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

September 2009

University of Oregon Graduate School

Confirmation of Approval and Acceptance of Dissertation prepared by:

Sean Carroll

Title:

"Mechanisms of Evolution by Gene Duplication: The Origins of Corticosteroid Signaling"

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Biology by:

Patrick Phillips, Chairperson, Biology
Joseph Thornton, Advisor, Biology
William Cresko, Member, Biology
John Postlethwait, Member, Biology
Kenneth Prehoda, Outside Member, Chemistry

and Richard Linton, Vice President for Research and Graduate Studies/Dean of the Graduate School for the University of Oregon.

September 5, 2009

Original approval signatures are on file with the Graduate School and the University of Oregon Libraries.

An Abstract of the Dissertation of

Sean Michael Carroll for the degree of Doctor of Philosophy

in the Department of Biology to be taken September 2009

Title: MECHANISMS OF EVOLUTION BY GENE DUPLICATION: THE ORIGINS OF CORTICOSTEROID SIGNALING

Approved: _____

Dr. Joseph W. Thornton, Advisor

Gene duplication underlies the evolution of many protein functions and is a known stimulus for molecular innovation. Many models exist to explain the maintenance of duplicate genes in the genome and the dynamics that drive the evolution of novel protein functions; few if any of these models, however, incorporate knowledge of how protein structures and functions actually evolve. A growing body of work on the historical mechanisms of molecular evolution and the ways in which proteins evolve in the lab has provided profound insights into the ways in which proteins respond to mutation, selection, and drift. Evolutionary models of duplicate gene evolution could greatly benefit from the knowledge gained from these mechanistic studies of protein evolution.

My dissertation seeks to address this gap in knowledge by reconstructing the process by which novel steroid signaling pathways evolved after gene duplication. I focus specifically on a class of hormones called corticosteroids – critical regulators of the stress

response, metabolism, and immunity – and the mineralocorticoid and glucocorticoid receptors that mediate the steroid response. Both the enzymes that synthesize corticosteroids and the hormone receptors are the result of ancient gene duplication events, and I make use of methods in phylogenetics, molecular biology, and structural biology to reconstruct the mechanisms and dynamics by which they evolved.

This dissertation comprises three separate but complementary studies that illuminate the origins of corticosteroid signaling. In the first project, I show how lineage-specific steroid signaling arose in elasmobranchs as a novel hormone exploited the structural promiscuity of preexistent receptors. Next, I describe how degenerative and stabilizing mutations defined the divergence of the glucocorticoid receptor after gene duplication. And finally, I use phylogenetic and functional analyses to reconstruct the origins of corticosteroid synthesis with the duplication of enzymes in the steroid synthesis pathway. Together, I provide a comprehensive reconstruction of the evolution of corticosteroid signaling. This work also highlights specific evolutionary mechanisms – molecular exploitation, structural and functional promiscuity, degenerative mutations, and stabilizing mutations – that could drive the evolution of novel protein functions after gene duplication.

This dissertation includes both previously published and unpublished co-authored materials.

CURRICULUM VITAE

NAME OF AUTHOR: Sean Michael Carroll

PLACE OF BIRTH: Havre de Grace, Maryland

DATE OF BIRTH: April 27, 1981

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, Oregon
McDaniel College, Westminster, Maryland

DEGREES AWARDED:

Doctor of Philosophy in Biology, 2009, University of Oregon
Bachelor of Arts in Biology, 2003, McDaniel College

AREAS OF SPECIAL INTEREST:

Molecular Evolution
Structural Biology

PROFESSIONAL EXPERIENCE:

Graduate Teaching Fellow, Department of Biology, University of Oregon, 2003-2004

Intern, Department of Biology, University of South Carolina, 2002

Intern, Osiris Therapeutics, Baltimore, Maryland, 2001

GRANTS, AWARDS AND HONORS:

National Science Foundation Integrative Graduate Education and Research Traineeship in Evolution, Development, and Genomics, 2005-2008

Deutscher Akademischer Austausch Dienst: German Academic Exchange Service, 2005 (Awarded but subsequently declined)

National Institutes of Health Developmental Biology Training Grant, University of Oregon, 2004

Graduation with Distinction, *magna cum laude*, McDaniel College, 2003

National Science Foundation Research Experiences for Undergraduates Summer Fellowship, University of South Carolina, 2002

PUBLICATIONS:

Carroll, S. M., Bridgham, J. T., and Thornton, J. W. (2008). Evolution of hormone signaling in elasmobranchs by exploitation of promiscuous receptors. Mol Biol Evol *25*, 2643-2652.

Bridgham, J. T., Carroll, S. M., Thornton, J. W. (2006). Evolution of hormone-receptor complexity by molecular exploitation. Science *312*, 97-101.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

PROTEIN EVOLUTION BY GENE DUPLICATION

Gene duplication is a widely accepted source of molecular innovation, particularly for the evolution of novel protein functions. Many models exist to explain the mechanisms that drive the maintenance and functional divergence of gene duplicates, but few, if any, attempt to incorporate a growing body of work on structural and biophysical mechanisms of protein evolution. Are there universal changes in protein structure that underlie the functional diversification of proteins after gene duplication? The purpose of this review is to synthesize our knowledge of evolution by gene duplication with empirical analyses of protein structure to provide a more comprehensive understanding of the mechanisms driving the evolution of novel protein functions.

The central goal of evolutionary biology is to provide a historical explanation for the diversity and complexity of life. Just as Darwin's "principle of descent with modification" (Darwin, 1964) explains the origin of species and their exquisitely optimized forms, there are similar explanations for the patterns and processes of evolution at the molecular level. Proteins occupy a special place in molecular evolution as the workhorses of the cell; the interactions of proteins with nucleic acids, metabolites,

and the environment drive virtually all cellular functions and underlie phenotypic diversity and complexity. Therefore a central goal of molecular evolution is to understand how proteins arise and diversify in function.

Properties of Protein Structure and Function

Protein biology was in the midst of a "great biochemical revolution" (Dickerson, 2005) in the 1950's and '60's, as the molecular world was revealed in the light of X-ray crystallography. No longer were the structures of DNA and proteins shrouded in mystery; they could now be visualized in terms of discrete macromolecular structures. As a result, a wealth of research emerged to explain the relationship between a protein's structure, sequence, and function, and how they evolve. Early comparisons of proteins, particularly the vertebrate hemoglobins, revealed many features of protein evolution that are relevant today. Species differences in protein function were discovered to evolve from a common structural scaffold; for example, the lineage-specific adaptations of bird, fish, and crocodile hemoglobins are accomplished by one or a few mutations that result in only minor changes to the protein fold (Perutz, 1983; Powers, 1972). Parts of some proteins were discovered to be more "dispensable" than others for function (Anfinsen, 1959), in what was a preface to the neutral theory of molecular evolution (Kimura, 1968). Changes in some regions had little effect on function and were more tolerant to mutation, while others were found to be essential and highly conserved between species. It was also recognized that protein sequences and structures harbor an account of their evolutionary past. Zuckerkandl and Pauling (Zuckerkandl and Pauling, 1965) deemed DNA and

protein sequences "documents of evolutionary history," recording aspects of evolution such as relationships to other sequences (phylogeny), the branches on which sequence changes occurred, the rate of substitution, and even the sequences of long-extinct ancestral molecules. Similarly, conserved macromolecular structures suggested common descent of proteins from speciation and gene duplication events. This is best illustrated in the classic papers by Perutz (Perutz et al., 1960) and Kendrew (Kendrew et al., 1960) on the respective structures of human hemoglobin and whale myoglobin proteins. These and other works within the hemoglobins (Ingram, 1961; Pauling and Zuckerkandl, 1963) were the first to highlight the role of gene duplication as important for the evolution of novel protein functions.

Ohno's seminal work solidified the importance of gene duplication in evolution, and today there are three major ways – gene conservation, subfunctionalization, and neofunctionalization – by which duplicate genes are preserved (Ohno, 1970; Piatigorsky and Wistow, 1991; Hughes, 1994; Force et al., 1999). This review focuses primarily on the evolution of novel gene functions by neofunctionalization; the different fates of gene duplicates and the population genetic models for their preservation are the subject of numerous other reviews (Hahn, 2009; Lynch and Katju, 2004; Conant and Wolfe, 2008; Roth et al., 2007; Zhang, 2003). Ohno theorized that after gene duplication, one copy is freed of its ancestral function and is allowed to explore previously "forbidden" mutations in a period of relaxed or absent selection. Most often this mutation accumulation quickly results in the degeneration and nonfunctionalization of the duplicate loci. In rare instances, however, the mutational drift will stumble upon a novel gene function that

confers a fitness advantage and leads to the preservation of the duplicate loci. There are several different models describing the mutational mechanisms of neofunctionalization, and these are reviewed elsewhere (Hahn, 2009). Here, I discuss how knowledge of protein evolution that has emerged since the time of Ohno and could enrich our understanding of neofunctionalization as a mechanism for the preservation duplicate genes and the subsequent evolution of novel protein functions. I avoid new terminology or statements that further confuse an already heavily researched topic; my goal is to show that common processes of protein evolution can increase the likelihood that duplicate genes are preserved and evolve new functions.

Modern Concepts of Protein Function and Evolution

Recent advances in our knowledge of protein functional evolution have come from two major disciplines. The first includes work that seeks to empirically test hypotheses on the historical mechanisms of protein evolution by employing the manipulative experiments of molecular biology (Golding and Dean, 1998; Dean and Thornton, 2007). Highlights of the so-called "functional synthesis" include the molecular basis of evolved pesticide resistance (Newcomb et al., 1997), the adaptive landscape of an ancient switch in coenzyme use (Miller et al., 2006), and the diversification of opsins to varying light conditions (Yokoyama et al., 1999; Shi et al., 2001; Yokoyama, 2002; Yokoyama et al., 2008). These case studies offer insights into historical processes but are limited to the end-result of a single evolutionary "experiment" or outcome, and cannot directly address things such as the repeatability or predictability of evolution (Gould,

1989; Lenormand et al., 2009; Weinreich et al., 2006; Blount et al., 2008). However, this approach is well complemented by protein engineers and experimental evolutionists that can observe protein evolution in real time. Directed evolution, in particular, most often seeks to evolve an enzyme to acquire some new desired function through repeated rounds of mutation and selection (Arnold et al., 2001; Bershtein and Tawfik, 2008a). The researcher can monitor and control the strength and nature of selection, the pool of variation to select upon, the properties of evolutionary intermediates, and the precise mutational trajectories taken in one or any number of evolutionary experiments. These experiments offer more precise measurements of protein evolution but lack the complexity of natural systems. Collectively, mechanistic studies of historical evolution and directed evolution experiments have illuminated such things as the evolutionary potential of promiscuous protein functions, how proteins evolve in a period of relaxed selection, and the fundamental importance of protein stability. Evolutionary models of neofunctionalization could greatly benefit from this knowledge.

The Exploitation of Fortuitous Protein Promiscuity

Shortly after Ohno outlined the fate of duplicate genes (Ohno, 1970), functional promiscuity emerged as an important force in the evolution of novel protein functions. Jensen (Jensen, 1976) championed the importance of protein promiscuity: "fortuitously formed compounds that happened to be useful would have conferred a selective advantage, thereby providing a basis for increased and more specific production of that compound (by gene duplication and specialization by mutation)." In its broadest sense,

promiscuity is merely a lack of specificity in protein functions. Proteins evolve under purifying selection to maintain a native function, but in many instances this is accompanied by one or more low-level and nonspecific functions that are of no selective consequence. This phenomenon is most often attributed to enzymes, but promiscuity has been observed in many different types of proteins across the tree of life (Copley, 2003; Khersonsky et al., 2006; O'Brien and Herschlag, 1999; D'Ari and Casadesús, 1998). A recent genome-wide survey of *E. coli* found that promiscuity is quite common: approximately 20% of gene knockouts are rescued by the promiscuous functions of typically unrelated genes (Patrick et al., 2007). These mutants were rescued by a large sink of latent and fortuitous protein functions (Patrick and Matsumura, 2008; Patrick et al., 2007; Jensen, 1976) that is potentially even larger when genes are exchanged between *E. coli* strains or other bacteria (Hendrickson, 2009; Touchon et al., 2009; Touchon et al., 2009).

Arguably, the exploitation of promiscuous activities blurs the lines between the three outcomes of duplicate genes. Promiscuous activities can arise after gene duplication and legitimately drive neofunctionalization or, alternatively, they can be present before duplication and selectively neutral. The exploitation of ancestral promiscuity is "new" in the sense that these functions are newly exposed to selection, but the activities might have arisen before or after gene duplication. Both theory (Jensen, 1976; Bergthorsson et al., 2007) and empirical studies (Patrick and Matsumura, 2008; McLoughlin and Copley, 2008) suggest that these low-level promiscuous functions could be exposed to positive selection if the protein (and its native function) are overexpressed. Duplicate proteins

would amplify promiscuous activities and could lead to the "gene conservation" fate of duplicates (Bergthorsson et al., 2007; Ohno, 1970); subsequent mutations optimize the functions of each paralog, to the detriment of the side-activity (Bergthorsson et al., 2007; Jensen, 1976). Selection upon promiscuous activities can also set the stage for subfunctionalization after gene duplication. McLoughlin and Copley (McLoughlin and Copley, 2008) devised an experimental system in which both the native and promiscuous functions of an enzyme are suddenly necessary for survival. They effectively evolve a situation of "gene sharing" (Jensen, 1976; Piatigorsky and Wistow, 1991; Hughes, 1994) with noticable tradeoffs between the two functions. This multifunctional enzyme is a model precursor for subfunctionalization by the partitioning and optmization of functions between duplicates (Piatigorsky and Wistow, 1991; Hughes, 1994). In short, protein promiscuity could expand the potential adaptive pathways to a new function and increase the likelihood that duplicate genes are preserved.

Historical and real-time studies of protein evolution have illuminated ways in which protein promiscuity increases evolvability. In some protein engineering experiments, both the starting and end products of the selection process are specialists for a particular function, and yet analyses of the mutational intermediates often find that evolution proceeded through a generalist protein capable of multiple interactions (Matsumura and Ellington, 2001; Fasan et al., 2008). These results might suggest that there is always a functional tradeoff between the native and evolved functions, but there are many examples in which no such tradeoff exists (reviewed in Khersonsky et al., 2006). The dynamic nature of functional promiscuity has been captured in the lab by

allowing proteins to neutrally diverge from their ancestral sequence. Experiments that evolved an enzyme with random mutations plus purifying selection for the native function found that the variants accrued a number of novel promiscuous activities in the course of their divergence (Bloom et al., 2007b; Amitai et al., 2007). These and other studies suggest that sequence divergence permits the exploration of an often-vast neutral network of promiscuous activities (Amitai et al., 2007; Bloom and Arnold, 2009; Peisajovich and Tawfik, 2007). The exploitation of promiscuous activities by selection is then highly contingent on evolutionary history (Patrick and Matsumura, 2008; Patrick et al., 2007) and the current sequence space occupied by the protein; latent promiscuous functions could suddenly become adaptive after a shift in the environment or genome.

The exploitation of ancestral promiscuity for novel functions is elegantly demonstrated in the origins of steroid signaling. In humans and other tetrapods, the mineralocorticoid receptor (MR) is activated by aldosterone to regulate blood pressure and salt balance, whereas the glucocorticoid receptor (GR) is activated by cortisol and regulates various immune, metabolic, and stress responses. Phylogenetic analyses showed that MR and GR descend from an ancient gene duplication, which begs the question: how did the specific interactions of MR with aldosterone and GR with cortisol arise? Using ancestral state reconstruction, Bridgham *et al.* (Bridgham et al., 2006) found that the resurrected last common ancestor of MR and GR, named AncCR, was activated by a variety of hormones, including those such as cortisol and aldosterone that had yet to evolve. But why would a receptor be preadapted for hormones that had yet to appear? The authors suggest that the AncCR's affinity is merely a byproduct of constraints

imposed by structurally similar, more ancient endogenous hormones. The determination of AncCR's protein structure revealed a common binding mode for structurally analogous hormones (Ortlund et al., 2007), and suggested that empty "nooks" in the hormone-binding pocket fortuitously fit hormones that evolved millions of years later (Carroll et al., 2008). The structural promiscuity that descended from AncCR facilitated the recruitment of MR by aldosterone upon the later synthesis of the hormone in tetrapods (Bridgham et al., 2006). Similar instances of novel molecules recruiting preexistent molecules for new functional roles have been found elsewhere in steroid receptors (Thornton, 2001; Carroll et al., 2008) and other in proteins (Cai et al., 2007; Krasowski et al., 2005; Cardoso et al., 2007).

Protein Functions in Response to Mutation

Neutral processes dominate most of protein evolution and play a major role in the various outcomes of gene duplication. The neutral theory predicts that most sequence diversity is the result of random fixations of mutation while purifying selection maintains protein function (Kimura, 1983). Ohno believed that after gene duplication, one copy is "free to accumulate a series of forbidden mutations" in a period of mutational drift (Ohno, 1970). The end-result of random mutations in a period of relaxed or absent selection is most often nonfunctionalization, but these mutations can sometimes play a creative role in the preservation of duplicate genes. So how tolerant are proteins to random mutational drift? Are there certain properties that make one protein more robust to mutation than others? This could be particularly relevant to the period of "forbidden"

mutations after gene duplication. If proteins crumble after only a few mutations then the opportunity for a new or optimized function to arise is indeed extremely small. Historical and engineered studies of protein evolution offer empirical tests of a protein's response to periods of mutation accumulation with and without selection.

Typically, periods of "undirected" sequence evolution are unnecessary in directed evolution. Starting with a protein and a native function (A), selection for activity (B) usually proceeds by a simple uphill climb to the novel function (Tracewell and Arnold, 2009). However an alternative approach, and one that is particularly intriguing for studies of historical evolution, is to select for new functions from libraries created by so called "mutational drift" (Gupta and Tawfik, 2008); this term can be misleading because sequences can be evolved with random mutations and no selection (drift), or random mutations plus purifying selection for the native protein function (sequence divergence). Typically, successive rounds of mutation and selection on the native function occur before screening the variants for novel protein functions. Protein populations generated using this technique have been found to display different properties than those under constant selection, including greater protein stability and an increased opportunity for new functions to arise (Bershtein et al., 2008; Bloom et al., 2007b; Bloom et al., 2007a; Bloom et al., 2007).

A wealth of directed evolution experiments have used enzyme activity as a surrogate for fitness to classify the distribution of the effects of random mutation. These data (\reviewed in\ (Bloom and Arnold, 2009)) were generated from a broad sampling of functionally diverse enzymes and give very similar results – most mutations are

selectively neutral, a smaller proportion are deleterious, and a tiny fraction, typically half to one-hundredth of a percent, are selectively advantageous for a novel protein function. Structural analyses have shown, in accordance with the neutral theory (Kimura and Ota, 1974) and earlier predictions (Anfinsen, 1959), that neutral mutations are not equally distributed in the protein structure. Of 920 "tolerated" mutations to a DNA repair enzyme (AAG), mutations near the active site and DNA-binding surface are underrepresented or absent while those on the surface and in protein loops are enriched (Guo et al., 2004). These results suggest that proteins are initially very tolerant to random mutations, at least in some parts of the protein structure.

Repeated rounds of random mutagenesis shed light on the robustness of proteins during prolonged mutation accumulation experiments. Bershtein *et al.* (Bershtein et al., 2006) subjected *TEM-1* β-lactamase to a prolonged drift – upwards of 20 mutations per gene – in the absence of purifying selection, and subsequently measured the proportion of enzyme variants that continued to confer antibiotic resistance (fitness). When presented with a high selective pressure (large dose of antibiotic), fitness decreased rapidly with mutational load, so that only half of the variants confer resistance with just two mutations. Given a lower dose of antibiotic, protein fitness is initially quite robust to mutations; however, in later rounds of mutagenesis, the fitness decrease steepens and is indicative of negative epistasis. The authors conclude that robustness and epistasis are linked in proteins and determined by a common parameter – protein stability. In all, these results show that the distribution of fitness effects of protein mutations can be dependent on the presence or strength of selection, and likely the specific protein architecture. These

factors could dominate the evolutionary dynamics of gene duplicates towards nonfunctionalization versus other fates (Bershtein and Tawfik, 2008b).

Degenerative mutations accrued in weak or absent selection are known to play a creative role in certain models of subfunctionalization (Force et al., 1999); a recent study presents a case in which these mutations are the likely mechanism for neofunctionalization. Steroid receptors most often function as hormone activated transcription factors, possessing a largely modular domain architecture to coordinate multiple subfunctions such as binding to DNA or to hormone (Beato, 1989). Amphioxus possesses two steroid receptors that evolved with the duplication of an ancestral estrogen-activated receptor (Bridgham et al., 2008). One of these paralogs has retained an ancestral-like function, whereas the other lost its subfunction to bind hormone. The degenerative mutations in this copy created a novel receptor that competes for the same DNA-binding elements as its paralog, effectively functioning as a transcriptional repressor. These results demonstrate that degenerative evolution after gene duplication may be a creative force in and of itself, significantly increasing the likelihood of neofunctionalization.

The Importance of Protein Stability

Long before knowledge of three-dimensional protein structures or even the nature of the polypeptide chain, scientists recognized the importance of a properly folded protein state. Outside of normal physiological conditions – such as changes in pH, solvent, or temperature – proteins lose their native folded state and, with it, their function (Nelson, 2003). Protein denaturation, however, is readily reversible when the unfolded

chain is returned to its physiological norm. Anfinsen showed in his experiments that renaturation is guided solely by the biophysical properties encoded in the protein's sequence (Anfinsen, 1973), and a return to the native state is accomplished by finding the most thermodynamically stable protein conformation. These early studies of protein folding and stability inspired a wealth of research on the role that stability plays in protein evolution, and how protein stability itself evolves.

Research has shown that protein stability is exquisitely adapted to an organism's thermal environment. Enzymes from psychrophilic organisms are optimally active at cold temperatures (Russell et al., 1998; Davail et al., 1994), while enzymes from thermophilic organisms function at high temperatures (Singleton and Amelunxen, 1973); these adaptations have been recreated in the lab by selecting enzymes for increased thermal tolerance (Counago et al., 2006). The importance of thermal adaptation for protein function suggests that stability is an important and evolutionary labile property of proteins (Puigbò et al., 2008). Remarkably, this tight relationship between protein activity and thermostability has illuminated the thermal environment of reconstructed ancestral proteins that existed some billions of years ago (Gaucher et al., 2008; Gaucher et al., 2003).

So why aren't proteins selected to be overly stable? From a biophysical perspective, there might be tradeoffs between the catalytic activity and thermostability of thermal adapted enzymes (Arnold et al., 2001). Enzymes adapted to high temperatures are thought to have more stabilizing interactions to make their structures rigid, while those adapted to cold should be more flexible to allow for conformational changes

associated with catalytic activities. But an alternate theory is that, for most proteins, marginal stability arises from passive evolutionary mechanisms. Taverna and Goldstein (Taverna and Goldstein, 2002) suggest that proteins that meet the minimum stability requirements will predominate simply because of the vastness of sequence space and the rare chance that highly stable variants would arise by chance. Many random mutations are inherently destabilizing, through a variety of mechanisms (Matthews, 1987; Matthews, 1993). Tokuriki *et al.* (Tokuriki et al., 2007) compute the distribution of mutational effects on protein stability for 21 different proteins and find that, regardless of protein architecture, most mutations are inherently destabilizing.

A protein's stability plays a major role in governing its tolerance to mutation. Given the above assertions that proteins are only marginally stable and many mutations are destabilizing, a protein would quickly unfold and lose activity in the absence of selection. A protein's robustness to mutation is largely governed by a stability threshold, below which the protein becomes unstable and loses activity (Bershtein et al., 2006; Bloom et al., 2007). The protein's distance from this threshold, however, is dynamic and evolvable. The stability threshold can be relaxed by decreasing selection for proper function (Bershtein et al., 2006), by gaining one or more stabilizing mutations that increase the distance from the threshold (Bloom et al., 2005; Bloom et al., 2006), or even by more global mechanisms such as increasing the expression of chaperone proteins (Tokuriki and Tawfik, 2009). Mutations that increase stability are called "global suppressors," because they allow a greater number of destabilizing mutations to accumulate than would otherwise be allowed (Bershtein and Tawfik, 2008a; Mitraki et

al., 1991; Shortle and Lin, 1985). Strikingly, proteins with global suppressor mutations have an increased likelihood for novel functions to arise (Bershtein et al., 2008; Bershtein and Tawfik, 2008a; Bloom et al., 2006), and destabilizing mutations that confer novel functions are often balanced by changes to increase stability (Tokuriki et al., 2008). Increased protein stability can expand the range of promiscuous activities (Huang and Palzkill, 1997; Petrosino et al., 1998) by several proposed mechanisms (Tokuriki and Tawfik, 2009; Wroe et al., 2007). These results have profound implications for protein evolution after gene duplication – factors such as relaxed purifying selection or global suppressor mutations that relieve the effects of deleterious mutations increase the probability that beneficial mutations will arise.

Mechanistic Studies of Protein Functional Evolution

Empirical studies of historical and engineered protein evolution have illuminated several properties that can contribute to our understanding of how novel functions evolve by gene duplication. First is the possibility that proteins are only as specific as they need to be. Functional promiscuity is a common byproduct of native protein functions, present in many different gene families across the tree of life. As long as promiscuous activities do not interfere with normal cellular functions, they are allowed to persist and drift with sequence divergence. Occasionally, though, this untapped and fortuitous promiscuity can be exploited for new functional roles by a variety of mechanisms. Promiscuous activities that arise and are selected upon before duplication (gene sharing) will lead to subfunctionalization (Ohno, 1970; Hughes, 1994; Piatigorsky and Wistow, 1991); those

that arise before or after duplication but are only exposed to selection after could lead to the appearance of neofunctionalization (Ohno, 1970), or in actuality, preservation by gene conservation (Bergthorsson et al., 2007). The exploitation of fortuitous functional promiscuity is not well explained by the canonical fates of duplicate genes.

Real-time studies of protein divergence allow us to monitor changes in native and promiscuous protein functions, protein structure, and stability as they evolve with mutation and varying degrees selection. These conditions simulate changes in functional constraints brought about by gene duplication; those gene functions that are more robust to mutation should have an increased likelihood of discovering creative mutations, and a slightly decreased chance that mutations will be destructive. These degenerative mutations, however, play creative roles in forms of subfunctionalization (Force et al., 1999) and as described above, neofunctionalization as well (Bridgham et al., 2008). Thus empirical studies of protein evolution have expanded our ideas of how proteins change in response to mutation and varying degrees of selection.

Finally, protein stability has emerged as a global determinant of protein structure and function and is in itself and evolvable character. Proteins are optimized by selection to be only as stable as is necessary to properly fold and function. Marginal protein stability could arise partly due to tradeoffs between flexibility and activity or because many mutations are inherently destabilizing. Mutations that increase stability, however, can have profound effects on subsequent protein evolution. These stabilizing and so-called "global suppressor" (Shortle and Lin, 1985) mutations allow a greater number of destabilizing and previously "forbidden" (Ohno, 1970) mutations to accrue than would

normally be allowed. Suppressor mutations fortuitously expand the neutral networks from which promiscuous or native functions arise, increase the chance of neofunctionalization, and extend the time to nonfunctionalization.

Ohno's highly prescient synthesis forever placed gene duplication as one of the major creative forces in evolution (Ohno, 1970). Since this time, the evolutionary consequences of gene duplication have been found to include adaptation, speciation, mutational robustness, evolvability, and complexity, amongst others (Postlethwait et al., 2004; Conant and Wolfe, 2006; Cooper et al., 2007; Escriva et al., 2006; Soltis and Soltis, 1999; Teichmann and Babu, 2004; Wagner et al., 2003; Hannay et al., 2008; Lynch and Conery, 2003; Hittinger and Carroll, 2007; Panopoulou and Poustka, 2005). Gene duplication has arguably sparked the greatest interest in the evolution of novel gene functions by neofunctionalization (Ohno, 1970; Hahn, 2009). Variations in the models of neofunctionalization have sparked much debate, partly because of a lack of evidence on how protein functions actually evolve; historical and engineered studies of protein evolution can now illuminate these processes. This knowledge certainly will not revolutionize the ideas founded by Ohno, but could help us to delineate the important variables of neofunctionalization – when the novel function arises; whether mutations are fixed by selection or drift; whether the mutations optimize, create, or destroy ancestral functions; whether trajectories are repeatable and predictable or highly contingent; and the effect of changing selective regimes on evolution. Studies of the specific mechanisms by which proteins evolve could contribute to a greater understanding of how novel genes

and gene networks evolve. My dissertation provides one such as example in the evolution of corticosteroid signaling.

STEROID SIGNALING AS A MODEL SYSTEM FOR STUDIES OF PROTEIN EVOLUTION

Steroid receptors and their hormones are an excellent system in which to study the dynamics and mechanisms of protein evolution. Steroid receptors (SRs) are ligand-activated transcription factors and molecular mediators: hormone binding induces a change in the receptor structure that causes cascade of interactions with chaperone proteins, DNA response elements, and coregulatory proteins that ultimately results in the regulation of gene expression in target tissues (Beato, 1989; Beato et al., 1995). Steroid hormones are produced by a series of enzymatic reactions from a cholesterol precursor and are typically classified by their site of origin – sex steroids such as progestins, androgens, and estrogens are synthesized predominantly in the gonads; corticosteroids are made in the adrenal cortex. Steroid signaling regulates many critical facets of life such as metabolism, immunity, reproduction, development, and the stress response (Bentley, 1998). The ability of this system to regulate widely different biological responses is due in large part to the specificity of a particular receptor for one or only a few steroids ligands. It is of great interest then to understand how new receptors and hormones evolve and are assembled into novel receptor-hormone interactions.

SRs originated from a series of gene and whole-genome duplications followed by the evolution of distinct hormone responses. Only recently were the evolutionary relationships between these receptors delineated. Humans possess six SRs: the estrogen receptors alpha (ERα) and beta (ERβ), progesterone receptor (PR), androgen receptor (AR), glucocorticoid receptor (GR), and mineralocorticoid receptor (MR). The diversity of receptors evolved from a single estrogen receptor-like ancestor over 600 million years ago (Thornton, 2001; Thornton et al., 2003). Evolutionary offshoots in the process of SR duplication and divergence are preserved across the tree of life: some invertebrates have a single (estrogen) receptor (Thornton et al., 2003; Keay et al., 2006; Keay and Thornton, 2009), others diverged after the first gene duplication and have only two receptors (Bridgham et al., 2008; Paris et al., 2008), basal vertebrates like lamprey have three (Thornton, 2001), and jawed vertebrates typically have six (Bridgham et al., 2006). Some lineages have experienced additional rounds of duplication and possess additional SR copies (Greenwood et al., 2003; Bury et al., 2003; Wu et al., 2003). Much less is known about the origins of steroid hormones, but work in this dissertation and elsewhere suggests that the duplication and divergence of steroid-synthesizing enzymes can drive the evolution of novel classes of steroids (Nelson, 1998; Nelson, 1999; Markov et al., 2008; Markov et al., 2009).

The protein architecture of SRs makes them particularly amenable to molecular analyses in the lab. SRs have a modular domain architecture, consisting of highly conserved DNA-binding (DBD) and ligand-binding domains, and variable hinge, N-terminal, and C-terminal regions (Kumar and Thompson, 1999; Kumar and Thompson,

2005). The LBD alone is sufficient to bind hormone and mediate ligand-dependent transcriptional activation, allowing the specific functions of the LBD to be cloned into reporter constructs and assayed *in vitro*. Because of their biomedical importance, a wealth of structural and functional data exists for human and other receptor LBDs, including a large number of polymorphisms, disease states, and engineered proteins. X-ray crystallography has revealed the structural basis of ligand-dependent transcriptional activation in the receptor LBD. The activation function helix (AF-H) acts as a molecular switch for transcriptional activation (Nagy and Schwabe, 2004). In the absence of ligand, AF-H is extended or in an "off" position; the binding of a steroid agonist, however, stabilizes the receptor structure and allows the AF-H to nestle against the receptor surface. This repositioning of AF-H creates a novel protein-protein interaction surface that allows coregulatory proteins to bind and recruit the transcriptional machinery to upregulate the expression of target genes. Empirical knowledge of how SRs bind to and are activated by hormone allows us to more easily dissect the molecular basis of evolutionary changes in receptor-hormone interactions.

A combination of techniques from phylogenetics, molecular, and structural biology permit highly mechanistic studies of SR evolution. This work is part of an emerging "functional synthesis" that seeks to understand the processes of molecular evolution at their most fundamental level (Dean and Thornton, 2007). Molecular phylogenetics provides an evolutionary framework on which to describe the relative timing of SR gene duplications, to explain the observed diversity of SR functions, and to infer on which branches functional changes occurred (Thornton, 2001; Bridgham et al.,

2006). The phylogeny also serves as a template for the reconstruction of ancestral states. This technique is the closest thing that we have to molecular fossils or a time machine: ancestral protein sequences are inferred by statistical models of the evolutionary process, the ancestor is "resurrected" via gene synthesis, and this ancestor is cloned and characterized with the same methods used for extant receptors (Thornton et al., 2003). Ancestral reconstruction allows us to empirically characterize the functions of extinct genes, to track the branches on which amino acid substitutions occurred, and to use manipulative experiments to determine the precise mechanism of functional change. My dissertation employs this highly mechanistic approach to study the evolution of corticosteroid signaling.

The Origins of Corticosteroid Signaling

My work focuses specifically on corticosteroid hormones and their receptors the MR and GR, to understand how new hormones and receptors evolve and are recruited into novel molecular partnerships. Corticosteroids are typically classified as either mineralocorticoids or glucocorticoids based on their respective abilities to activate MR or GR. In humans, MR is activated predominantly by aldosterone for the regulation of salt balance and blood pressure; GR is activated by cortsiol to control a variety of immune, metabolic, and stress responses. There are, however, lineage-specific receptor-hormone interactions that have evolved in several many species (Sturm et al., 2005; Idler and Truscott, 1966). So how can we explain the observed diversity and complexity of corticosteroid signaling?

GR and MR evolved with the duplication of a single ancestral receptor over 440 million years ago (Thornton, 2001; Bridgham et al., 2006). Using methods in ancestral state reconstruction and gene resurrection (Thornton et al., 2003), our lab previously demonstrated that the last common GR/MR ancestor functioned much like extant MRs – highly sensitive to a variety of hormones – and that GR functions are derived (Bridgham et al., 2006; Ortlund et al., 2007). After duplication, GRs evolved two major changes in their response to ligand: 1) reduced sensitivity to all hormones, as is found in the GR of an elasmobranchian cartilaginous fish (Carroll et al., 2008), and on top of this, 2) cortisol-specific activation, as is present in teleost and tetrapod receptors (Bridgham et al., 2006). The latter event is well studied: the historical mutations, their effects on protein structure-function, and ancient mutational trajectories have all been elucidated in the evolution of cortisol specificity (Bridgham et al., 2006; Ortlund et al., 2007). Considerably less is known about earlier periods of corticosteroid receptor evolution.

My dissertation seeks to reconstruct the evolutionary mechanisms and dynamics that gave rise to corticosteroid signaling pathways. I use a combination of techniques from phylogenetics, molecular biology, and structural biology to reveal the evolutionary origins of hormones, their receptors, and their functional interactions. This dissertation comprises three separate but complementary studies that elucidate the origins of corticosteroid signaling. In the first project, I show how lineage-specific receptor-hormone interactions evolved by exploiting the structural promiscuity of receptors. This work was originally published in the journal *Molecular Biology and Evolution* (vol. 25,

iss. 12, pp. 2643-52, 2008). It was co-authored with Jamie T. Bridgham, who assisted with materials and experimental design, and Joseph W. Thornton, who assisted with experimental design and edited the manuscript. Next, I study how degenerative and stabilizing mutations defined the functional divergence of GR after gene duplication. This work is in preparation for publication and is co-authored with Eric A. Ortlund, who assisted with experiments and experimental design, and Joseph W. Thornton, who assisted with experimental design and edited the manuscript. Finally, I reconstruct the molecular mechanisms of gene duplication that gave rise to the enzyme necessary for the synthesis of corticosteroid hormones. This work is in preparation for publication and is co-authored with Joseph W. Thornton, who assisted with experimental design and edited the manuscript. Together, this dissertation describes how seemingly complex and integrated facets of corticosteroid signaling can evolve by the simple and stepwise process of gene duplication and divergence.

BRIDGE TO CHAPTER II

The interactions of corticosteroid hormones with their receptors represent a simple, pairwise molecular interaction. These interactions are the basis for the molecular networks that drive virtually all cellular processes and underlie phenotypic diversity and complexity. To understand how molecular interactions evolve, I reconstructed the evolution corticosteroid signaling in elasmobranchs, a subclass of cartilaginous fishes. I show that novel receptor-hormone interactions can evolve by molecular exploitation, and that they are facilitated by the structural promiscuity of receptors.

CHAPTER II

EVOLUTION OF HORMONE SIGNALING IN ELASMOBRANCHS BY EXPLOITATION OF PROMISCUOUS RECEPTORS

This work was originally published in the journal *Molecular Biology and Evolution* (vol. 25, iss. 12, pp. 2643-52, 2008). It was co-authored with Jamie T. Bridgham, who assisted with materials and experimental design, and Joseph W. Thornton, who assisted with experimental design and edited the manuscript.

INTRODUCTION

Virtually all cellular functions are driven by specific interactions among biomolecules, such as enzymes and their substrates, transcription factors and their DNA binding sites, and receptors and their ligands. Despite extensive "top-down" work on the global structure of molecular interaction networks (Barabasi and Oltvai, 2004; Cork and Purugganan, 2004; Teichmann and Babu, 2004; Wilkins, 2005), only limited knowledge is available concerning the specific mechanisms and dynamics by which the molecular interactions that constitute these networks evolve (Zhu et al., 2005; Bridgham et al.,

2006; Prud'homme et al., 2006; Hittinger and Carroll, 2007; McGregor et al., 2007; Ortlund et al., 2007).

Corticosteroid hormones and their receptors provide an elegant model of molecular interactions, and the existence of a lineage-specific corticosteroid in the taxon *Elasmobranchii* offers the opportunity to investigate how new hormone-receptor interactions evolve. Corticosteroids are produced in the adrenal/interrenal gland through an enzyme-mediated biosynthetic pathway and secreted into the blood. The classic actions of corticosteroids are mediated by intracellular corticosteroid receptors (CRs), members of the steroid hormone receptor family that also includes receptors for androgens, progestins, and estrogens (Thornton, 2001). These proteins are ligand-activated transcription factors: upon binding to their hormone partners with high specificity and affinity, the receptor changes conformation, dimerizes, binds to specific DNA response elements, and attracts coregulator proteins, resulting in increased expression of nearby target genes (Beato et al., 1995). Like other steroid receptors, CRs have a modular domain structure, consisting of functionally autonomous conserved domains for DNA-binding and for ligand-activated transcription, as well as a nonconserved hinge and an N-terminal domain with additional ligand-independent transcriptional regulatory functions (Kumar and Thompson, 2005). Most vertebrates possess two paralogous CRs, a glucocorticoid receptor (GR) and mineralocorticoid receptor (MR), which arose from a single ancestral CR in a large-scale gene duplication some 470-440 million years ago (Thornton, 2001; Bridgham et al., 2006). In bony vertebrates, GR is activated by cortisol and regulates metabolism, immunity, and the

stress response; MR controls electrolyte homeostasis and blood pressure and is activated primarily by aldosterone in tetrapods and 11-deoxycorticosterone in teleosts (Bentley, 1998; Sturm et al., 2005). In addition to their classic receptor-mediated effects, some steroids also trigger rapid "non-classical" effects through other mechanisms (Thomas et al., 2005; Thomas et al., 2004; Prossnitz et al., 2007; Harvey et al., 2008).

A better understanding of steroid hormone-receptor interactions in basal vertebrates would help illuminate how GR, MR, and the physiological processes they regulate have evolved (Baker et al., 2007; Bury and Sturm, 2007). Some forty years ago it was discovered that the plasma of elasmobranchs –sharks, skates, and rays – contains very high levels of the corticosteroid 1α-hydroxycorticosterone (1α-B), which is not known to be produced in any other taxon (Idler and Truscott, 1966). 1α-B is synthesized in the interrenal gland in copious amounts relative to other endogenous hormones (Idler and Truscott, 1967; Truscott and Idler, 1972). It differs from other corticosteroids by the addition of a hydroxyl group at the C_1 position of the steroid backbone. The enzyme activity that drives this unusual 1α-hydroxylation reaction is present and functional in *in vitro* preparations of elasmobranch steroidogenic tissues (Truscott and Idler, 1968), but the underlying gene has not been identified. The molecular basis for 1α-B action in elasmobranchs, particularly its potential interactions with MR and GR, remains uncharacterized. It has been speculated that 1α-B functions as a dual ligand in both glucocorticoid and mineralocorticoid signaling pathways (Gelsleichter and Musick, 1999; Nunez and Trant, 1999; Nunez et al., 2006; Manire et al., 2007). 1α-B clearly has a mineralocorticoid-like role in the regulation of salt and osmolyte balance (Hazon and

Henderson, 1984; Armour et al., 1993), but the evidence for its glucocorticoid-like effects are more speculative. 1α-B levels have been shown to increase under stress and reduced osmolarity (Hazon and Henderson, 1984; Armour et al., 1993; Manire et al., 2007). Administration of glucocorticoids to elasmobranchs stimulates the regulation of carbohydrate metabolism (Patent, 1970), cartilage growth (Gelsleichter and Musick, 1999), and immune responses (Walsh et al., 2002). Like the glucocorticoids of other vertebrates, synthesis of 1α-B is regulated by andrenocorticotropic hormone and angiotensin II from the hypothalamic-pituitary adrenal axis (Honn and Chavin, 1978; Hazon and Henderson, 1985; O'Toole et al., 1990; Nunez and Vedeckis, 2002). Although cortisol and other recognized glucocorticoids are absent or present at very low concentrations in elasmobranchs, 1α-B circulates at extremely high levels in elasmobranchs (Truscott and Idler, 1972; Kime, 1977; Armour et al., 1993), just as glucocorticoids do in most other vertebrates.

Here we characterize the functional interactions of 1α-B with the GR and MR of an elasmobranch, the little skate (*Leucoraja erinacea*), and reconstruct how these receptor-hormone interactions evolved. We combine molecular functional assays, ancestral gene resurrection (Thornton, 2004), and analysis of protein structure to determine the functions of GR and MR with respect to this hormone and to characterize how the lineage-specific partnership of 1α-B with its receptors evolved.

METHODS

Isolation and Reconstruction of Corticosteroid Receptors.

Skate (*Leucoraja erinacea*) and hagfish (*Myxine glutinosa*) receptor ligand-binding domains (LBDs) were amplified using degenerate PCR and RACE from liver cDNA (Bridgham et al., 2006). Lamprey (*Petromyzon marinus*) CR cDNA was amplified similarly from a cDNA library (Thornton, 2001). Teleost (*Astatotilapia burtoni*) GR2a and MR were provided by R. Fernald, the human MR by R. Evans, and the human GR by B. Darimont.

Reconstruction and synthesis of ancestral receptors was performed as described in (Bridgham et al., 2006; Ortlund et al., 2007). Briefly, ancestral protein sequences were inferred using maximum likelihood (ML) phylogenetic reconstruction and a large dataset of steroid and related receptors. Nucleic acid sequences coding for the ligand binding domain were synthesized *de novo*. Ambiguously reconstructed sites defined as having an alternate amino acid state with a posterior probability greater than 0.20 or as having a different maximum likelihood state when reconstructed on any tree in the 95% credible set collected by Bayesian Markov Chain Monte Carlo analysis. Alternative states were introduced singly into the maximum likelihood sequence using site-directed mutagenesis (QuikChange II, Stratagene).

Receptor Activation.

LBDs (with hinge and CTE) of extant receptors were amplified using high-fidelity PCR and cloned into pSG5-GAL4DBD (gift of D. Furlow). Ancestral receptor LBDs (with CTE) were cloned into pSG5-GAL4DBD with a human GR hinge region. CHO-K1 cells were grown in 96-well plates in phenol red free α-MEM plus 10% dextran-charcoal-stripped fetal bovine serum (Hyclone), then transfected with 1 ng of receptor LBD, 100 ng of pFRluc reporter, and 0.1 ng of normalization vector phRLtk using Lipofectamine and Plus Reagents (Invitrogen). After 4 hours incubation, cells were treated with fresh medium, allowed to recover overnight, then treated in triplicate with hormone or vehicle control (ethanol) for 24 hours. Reporter expression was measured assayed using Dual-Glo (Promega) and expressed as the ratio of firefly luciferase to *Renilla* luciferase. Dose-response relationships were analyzed using Prism4 software (GraphPad). Receptors were considered unresponsive if they displayed <2-fold activation at >1 uM hormone. 1α-B was synthesized and provided by J. Rimoldi, University of Mississippi.

Quantitative PCR.

Expression of MR and GR was measured in various organs of *L. erinacea* provided by J. St. George, Boston University. Total RNA was isolated using RNeasy (Qiagen), and cDNA prepared using QuantiTect Reverse Transcription (Qiagen). Skeletal and cardiac muscle samples were digested with proteinase K (Qiagen) to improve yield. Primers were designed to amplify the ligand-dependent activation (AF-2) domain of MR

and GR LBDs. Housekeeping gene glyceraldehyde 3-phosphodehydrogenase (GAPDH GenBank DQ382343) was isolated using degenerate PCR and used as an internal reference for normalization. Q-PCR was performed on an ABI Prism 7900 HT with 1x Quantitect SYBR Green PCR Master Mix (Qiagen), 0.3 μM each primer, and skate cDNA. Cycling was as follows: 95°/15', [94°/15", 54°/30", 72°/30"] x 39 cycles, followed by melting curve analysis from 65°-95°. Reactions were run in triplicate and fluorescence detected during extension. Primer efficiencies (E_{MR} = 1.96, E_{GR} = 1.98, and E_{Gapdh} = 1.97) were determined by standard curve analyses (Simon, 2003) of serially diluted and linearized skate MR and GR in GAL4-DBD-pSG5 and skate Gapdh gene in pCR2.1 (Invitrogen).

In Silico Structural Analysis.

A model of 1α-B was constructed and energy minimized using ChemOffice Ultra (CambridgeSoft) and imported into MacPyMOL (Delano Scientific LLC) along with the crystal structure of AncCR/DOC (PDB 2Q3Y). 1α-B was aligned to DOC at the C_{10} carbon with the steroid backbones oriented in the same plane.

RESULTS

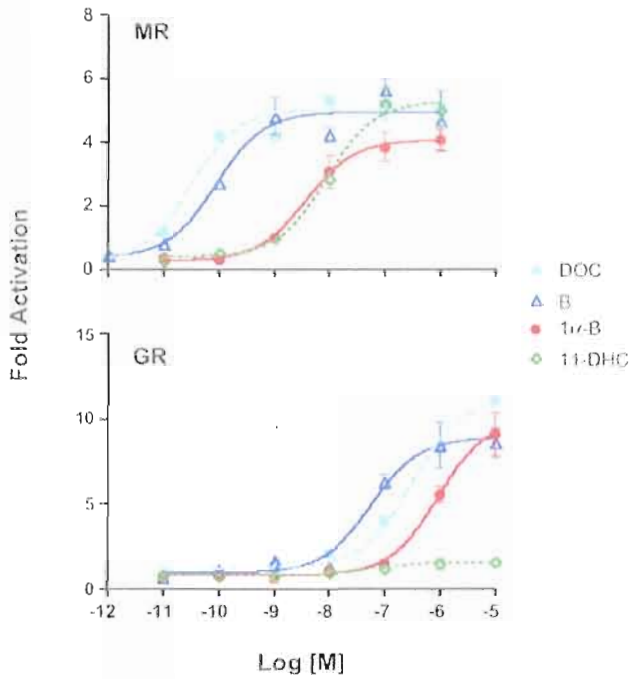
High- and low-sensitivity receptors for 1α-*B*

To determine the intrinsic hormone sensitivity of skate GR and MR, we expressed fusion constructs of receptor ligand-binding domains (LBDs) with the Gal4 DNA-binding domain (DBD). We evaluated transcription of a UAS-driven luciferase reporter gene in response to the predominant elasmobranch hormone 1α-B and several other steroid hormones found at lesser concentrations in elasmobranchs, including 11-deoxycorticosterone (DOC), corticosterone (B), and 11-dehydrocorticosterone (11-DHC) (Truscott and Idler, 1972). We also examined several mammalian corticosteroids not present in elasmobranchs, including aldosterone (Aldo), cortisol (F), and 11-deoxycortisol (DOF).

We found that both skate MR and GR LBDs are activated by 1α-B and other corticosteroid hormones. MR is a high-sensitivity receptor, activating transcription in response to low nanomolar concentrations of all hormones examined, including 1α-B. GR, in contrast, is activated by the same hormones, but requires 2 to 4 orders of magnitude higher concentrations to achieve half-maximal activation (Fig. 2.1). Based on these results, endogenous concentrations of 1α-B and possibly other steroids in elasmobranchs are expected to activate the MR, but only concentrations of 1α-B are likely to be high enough to activate the GR (see Discussion). The difference in quantitative sensitivity but not hormone specificity between GR and MR in

elasmobranchs contrasts with the situation in humans and other bony vertebrates, in which GR and MR have distinct hormone preferences.

A



B

	EC_{50} (nM)	
Hormone	Skate MR	Skate GR
B	0.09	58.40
DOC	0.03	305.57
1α-B	3.84	947.35
11-DHC	8.81	NA
Aldo	0.07	11.00
F	1.04	138.92
DOF	21.93	161.55

Figure 2.1. The skate MR and GR are high- and low-sensitivity receptors, respectively, for 1α-B and other corticosteroids. (**A**) Ligand-dependent transcriptional activity of receptor ligand-binding domains was determined in the presence of increasing concentrations of various hormones using a luciferase reporter gene assay. Fold activation is the reporter activity of hormone-treated samples divided by vehicle-only control; points show the mean of three replicates plus SEM. Four hormones found in elasmobranchs were tested: 11-deoxycorticosterone (DOC), corticosterone (B), 1α-hydroxycorticosterone (1α-B), and 11-dehydrocorticosterone (11-DHC). (**B**) Skate GR and MR hormone sensitivity. The concentration of hormone required for half-maximal reporter activation (EC_{50}) of skate MR and GR is shown in nanomolar (nM). NA, no activation, defined as EC_{50} > 1mM of hormone. Aldo, aldosterone; F, cortisol; DOF, 11-deoxycortisol.

MR and GR are Ubiquitously Coexpressed

To determine whether differences in gene expression could be important for generating distinct tissue-specific gluco- or mineralocorticoid responses, we used quantitative PCR to measure GR and MR transcripts across skate tissues. We found that both MR and GR are ubiquitously expressed, with little variation in the relative quantities of the two transcripts. GR and MR transcript levels, normalized to expression of the housekeeping gene Gapdh, varied by less than an order of magnitude among tissues, except for in skeletal muscle, where Gapdh levels were very high (Fig. 2.2A); a similar pattern was observed when transcripts were not normalized to Gapdh (not shown). The ratio of MR to GR expression, which is not affected by differences in Gapdh levels, varied by a factor of less than two among tissues (Figs. 2.2B).

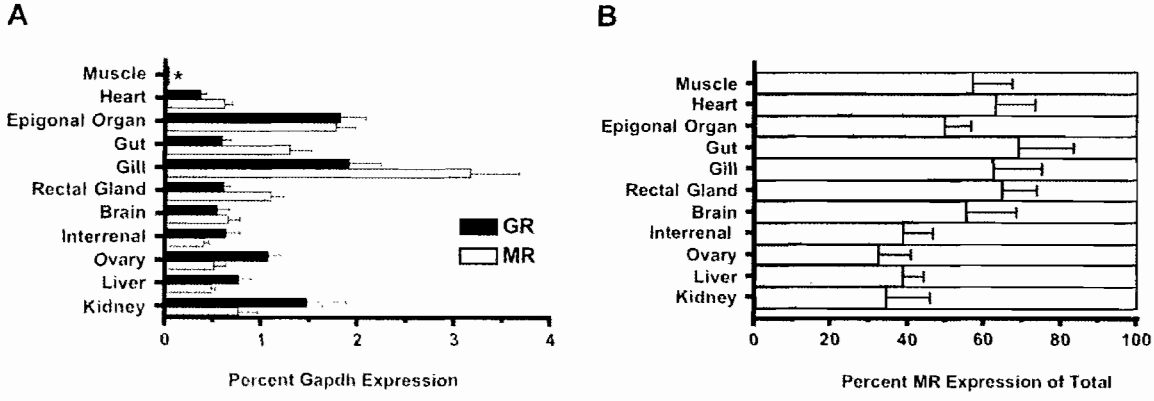


Figure 2.2. MR and GR are widely coexpressed in skate. Abundance of MR, GR, and Gapdh mRNAs were determined in various tissues of adult skate using Q-PCR. (**A**) The mean expression level and SEM was calculated for MR and GR from triplicate reactions and normalized to Gapdh. Asterisk denotes a tissue-specific increase in Gapdh expression that results in a decrease in normalized MR and GR expression. (**B**) Relative MR and GR expression in each tissue. Gray bars (with SEM) show MR expression as percent of total (MR plus GR) expression.

1α-B Activates CRs from other vertebrates

We next sought to determine how the lineage-specific partnership of 1α-B with corticosteroid receptors was assembled during evolution. We began by determining whether corticosteroid receptors from other species are also sensitive to 1α-B. MRs from bony vertebrates – both tetrapods (*Homo sapiens*) and teleosts (*Astatotilapia burtoni*) – activated reporter gene transcription in the presence of sub-micromolar concentrations of 1α-B (Fig. 2.3, Table 2.1), despite the hormone's absence from lineages other than elasmobranchs. The CR of an agnathan – the Atlantic hagfish (*Myxine glutinosa*), which possesses a single unduplicated gene orthologous to both GR and MR – had similar 1α-B sensitivity. Only the GRs of bony vertebrates – which are activated only by 17-hydroxylated corticosteroids like cortisol (Bridgham et al., 2006) -- and the CR of the sea lamprey (*Petromyzon marinus*) were insensitive to 1α-B.

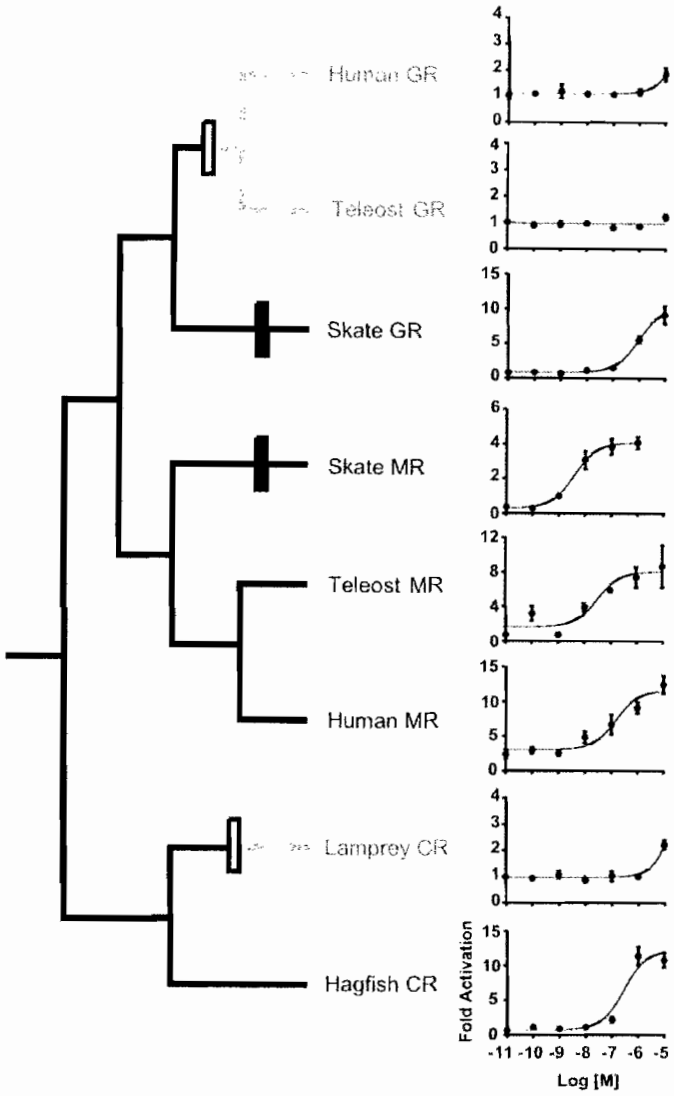


Figure 2.3. GR and MR sensitivity to 1α-B is taxonomically widespread. Sensitivity of corticosteroid receptors from a variety of vertebrates to 1α-B was assessed using a luciferase reporter assay. Black and gray branches show receptors sensitive and insensitive to 1α-B, respectively. Open boxes mark evolutionary loss of ancestral activation by 1α-B in the most parsimonious scenario. Black boxes denote the evolution of 1α-B synthesis in elasmobranchs.

Table 2.1. Sensitivity of extant and ancestral receptors to 1α-B. Sensitivity is presented as the EC50 – the concentration at which half-maximal reporter gene activation (EC_{50}) is achieved.

Receptor	EC50 (nM)	Genbank ID	Species
Skate MR	3.8	ABD46745	Leucoraja erinacea
AncCR	20.6	ABD46748	Ancestral reconstruction
AncGR1	25.1	ABU96169	Ancestral reconstruction
Teleost MR	32.5	AAM27890	Astatotilapia burtoni
Human MR	153.1	NP_000892	Homo sapiens
Hagfish CR	288.1	ADB46742	Myxine glutinosa
Lamprey CR	No activation	AAK20929	Petromyzon marinus
AncGR2	No activation	ABU96170	Ancestral reconstruction
Teleost GR2	No activation	AAM27888	Astatotilapia burtoni
Human GR	No activation	NP_000167	Homo sapiens
Skate GR	No activation	ABD46744	Leucoraja erinacea

Ancient receptor sensitivity to 1α-B

The widespread sensitivity of vertebrate corticosteroid receptors to 1a-B suggests an ancient origin of 1α-B responsiveness. To test the hypothesis that receptor sensitivity to 1α-B predates the elasmobranch-specific emergence of the hormone, we resurrected and functionally characterized three ancient corticosteroid receptors as they existed in ancestral vertebrates. Specifically, we used a large database of extant receptor sequences and used phylogenetic techniques to infer the maximum likelihood amino acid sequences of three ancient receptors: the ancestral corticoid receptor (AncCR, the unduplicated

ancestral gene from which extant MRs and GRs descend by gene duplication), GR in the last common ancestor of all jawed vertebrates (AncGR1), and GR from the last common ancestor of all bony vertebrates (AncGR2), after their split from cartilaginous fishes (Fig. 4). We then synthesized DNAs coding for the ligand-binding domains of these reconstructed proteins, expressed them, and characterized their functions using a reporter transcription assay in cell culture (Bridgham et al., 2006; Ortlund et al., 2007).

As predicted, we found that the most ancient receptors -- AncCR and AncGR1 -- are extremely sensitive to 1α-B, activating transcription with EC_{50}s of ~20 nM. AncGR2, in contrast, was unresponsive, as expected based on the lack of sensitivity to 1α-B in its descendants, the GRs of tetrapods and teleosts (Fig. 2.4, Table 2.1). To determine whether the 1α-B sensitivity of AncCR might be an artifact of uncertainty in the inference of the ancestral sequence, we identified sites that were ambiguously reconstructed (defined as having an alternative amino acid state with posterior probability > 0.20). In all cases but five, the alternate state is found in other 1α-B-activated receptors and is therefore not sufficient to abolish sensitivity to that hormone. Introducing each of these five alternate states into the AncCR by site-directed mutagenesis had no effect on ligand-activation (Table 2.2). Among sites that make contact with the ligand in the AncCR crystal structure (Ortlund et al., 2007), only one was ambiguously reconstructed; introducing this alternate state into AncCR had no effect on sensitivity to 1α-B (Table 2.2). We conclude that AncCR's response to 1α-B is not an artifact of uncertainty in the ancestral reconstruction.

To determine whether AncCR's sensitivity to 1α-B may be due to error in the phylogeny on which the ancestral reconstruction is based, we inferred the maximum likelihood sequence of AncCR on each of the 467 trees in the 95% credible set from a large Bayesian analysis. At only one sequence site did the ancestral reconstructions differ among trees. We introduced the alternate state at this site (A7V) into AncCR by mutagenesis and found it had no effect on sensitivity to 1α-B (Table 2.2).

We conclude that AncCR and AncGR1 were sensitive to 1α-B, and this ancient sensitivity was retained in most of the lineages descending from those ancestors, including the MR and GR of elasmobranchs. After the divergence of bony from cartilaginous fishes, the GRs of bony vertebrates subsequently lost 1α-B sensitivity, during the same period in which the receptor became cortisol-specific. This result indicates that corticosteroid receptors were capable of being activated by 1α-B many millions of years before synthesis of the hormone itself evolved in the elasmobranch lineage.

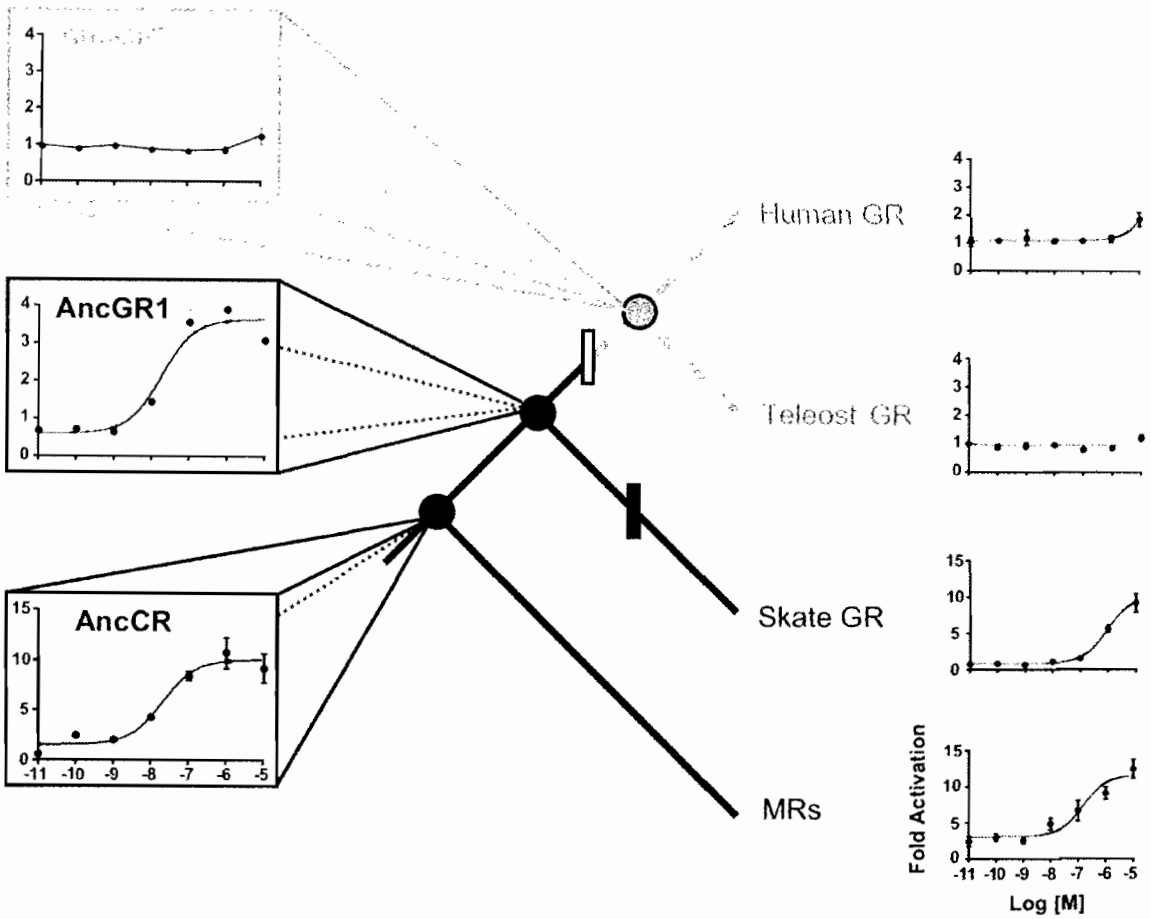


Figure 2.4. Corticosteroid receptor sensitivity to 1α-B predates the evolution of 1α-B synthesis. Three ancestral receptors from early vertebrates (circled nodes) were "resurrected" by ancestral sequence reconstruction and gene synthesis. Graphs depict luciferase reporter activity in the presence of increasing 1α-B concentrations, relative to vehicle-only control. Dose-response curves for selected extant receptors are also shown. Black and gray circles denote ancestral receptors sensitive and insensitive, respectively to 1a-B. White rectangle shows loss of 1α-B activation; black rectangle, origin of 1α-B synthesis. AncCR represents the unduplicated ancestral gene from which GR and MR descend; AncGR1 is GR in the last common ancestor of jawed vertebrates; AncGR2 is GR in the last common ancestor of bony vewrtebrates.

Table 2.2. AncCR's sensitivity to 1a-B is robust to statistical and phylogenetic uncertainty. Alternate reconstructions of ancestral states were introduced into the maximum likelihood reconstruction of AncCR by site-directed mutagenesis, and their sensitivity to 1α-B determined with a reporter gene assay.

Receptor	EC50 (nM)[a]	Source of uncertainty[b]	Receptors insensitive to 1a-B with alternate state[c]
AncCR	20.6	-	-
AncCR A7V	23.3	Phylogenetic	Pma CR
AncCR A36G	3.0	Stochastic – LBP	-
AncCR S20T	10.5	Stochastic – extant	Abu GR; Pma CR
AncCR K38R	12.3	Stochastic – extant	Has GR
AncCR S76A	4.5	Stochastic – extant	Hsa GR; Pma CR
AncCR V137A	9.8	Stochastic – extant	PmaCR
AncCR V224A	8.1	Stochastic – extant	Abu GR; Pma CR

[a] Sensitivity to 1a-B is reported as the concentration required for half-maximal activation of a luciferase reporter gene.

[b]Phylogenetic uncertainty refers to maximum likelihood reconstructions that differ among trees in the 95% credible set. Stochastic uncertainty refers to non-optimal reconstructions with posterior probability > 0.2 on the maximum likelihood phylogeny. LBP, sites in the ligand binding pocket. Extant, alternate reconstruction is present in one or more receptors insensitive to 1a-B.

[c]Abu, *Astatotilapia burtoni*; Hsa, *Homo sapiens*; Pma, *Petromyzon marinus*.

Structural Analysis of 1α*-B Docked in the AncCR*

To determine why ancient receptors were activated by 1α-B, we examined the previously solved crystal structure of AncCR (Ortlund et al., 2007). We hypothesized that AncCR was structurally "preadapted" to bind 1α-B because of that hormone's similarity to DOC, an ancient hormone that is the putative ancestral ligand for AncCR

(Ortlund et al., 2007). We generated a structural model of 1α-B docked into the ligand-binding domain of the AncCR with DOC. 1α-B differs from DOC only by the presence of hydroxyl groups at the C_1 and C_{11} positions. The structure shows that AncCR's ligand pocket contains ample room to accommodate the 11-hydroxyl of 1α-B; the receptor's previously identified ability to bind cortisol and aldosterone, which also carry the 11-hydroxyl, without conformational adjustment (Ortlund et al., 2007). The ligand pocket also contains unoccupied space in the alpha-plane above the C_1 carbon (Fig. 2.5A); in the model of AncCR-with 1α-B, this space is occupied by and adequate to accommodate the 1α-hydroxyl (Fig. 2.5B). A slight hydrophobic clash of this hydroxyl with AncCR's Leu32 and Phe92, however, is the likely cause of the receptor's slightly reduced sensitivity to 1α-B compared to other ligands that lack the 1α-hydroxyl. That the receptor retains nanomolar sensitivity to 1α-B, however, indicates that minor adjustments of the receptor backbone or side chain rotamers are sufficient to relieve the clash.

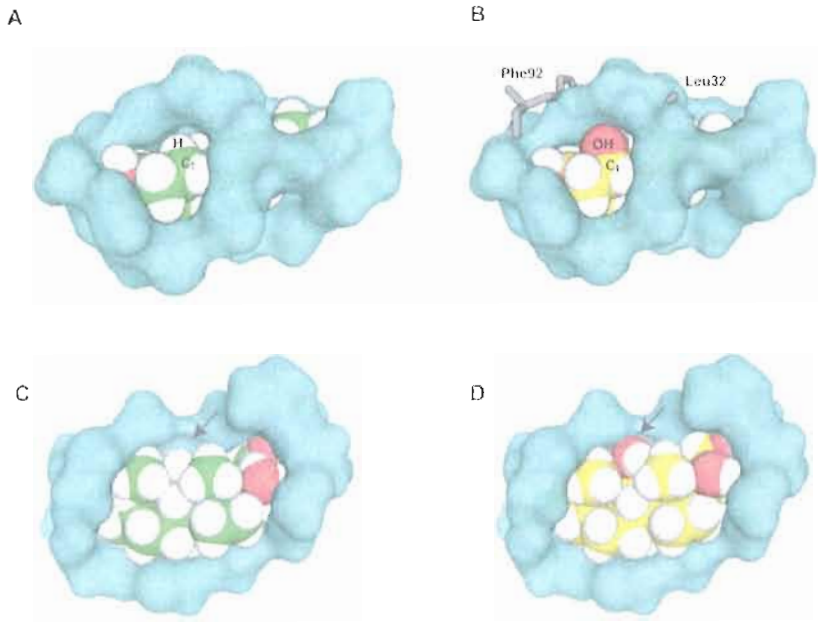


Figure 2.5. The structural basis for receptor promiscuity. The crystal structure of AncCR's ligand pocket in complex with DOC (green, panels A and C) and an *in silico* model of AncCR with 1a-B (yellow, panels B and D) are shown from two perspectives (A,B and C,D). Electron density of atoms within 4 angstroms of ligand is shown as a teal surface; some atoms from the foreground surface have been excluded to aid viewing. Red, oxygen atoms; white, hydrogen atoms. In panel A, empty space above the C1 carbon of DOC is apparent (black arrow), which can accommodate the added hydroxyl of 1a –B (panel B, labeled OH). In panel C, empty space (arrow) above C11 and its hydrogen (labeled H) accommodates the added hydroxyl of 1a-B (panel D).

DISCUSSION

Our analysis indicates that the lineage-specific partnership of 1α-B with MR and GR in elasmobranchs evolved due to molecular exploitation – recruitment of older molecules, which previously had different functions, into new functional relationships (Thornton, 2001; Bridgham et al., 2006). Both the widespread taxonomic distribution of 1a-B sensitivity and experimental analysis of resurrected ancestral receptors indicate that the ancestral gene from which both GR and MR evolved was already sensitive to 1α-B,

millions of years before synthesis of the hormone itself evolved (Fig. 2.6). Sensitivity to 1α-B has been retained to the present in numerous descendants of the AncCR, including those in numerous species that lack 1α-B, such as mammals, in which administration of exogenous 1α-B elicits a strong mineralocorticoid response (Idler et al., 1967). Our results indicate that the partnership of 1α-B with its receptors in elasmobranchs evolved when a newly synthesized hormone recruited pre-existing receptors, which previously had different ligands, into new functional partnerships. This evolutionary dynamic is similar to that previously observed for the interaction between aldosterone and the MR of tetrapods (Bridgham et al., 2006). Other steroid hormone/receptor interactions evolved when intermediates in the biosynthesis of ligands for more ancient receptors were recruited into partnerships with newly duplicated receptors (Thornton, 2001; Thornton et al., 2003). Several other apparent examples of evolutionary recruitment of ancient molecules into new partnerships have recently been described (Krasowski et al., 2005; Cai et al., 2007; Cardoso et al., 2007), suggesting that the evolution of specific molecular interactions by molecular exploitation may be a dominant theme in the emergence of biological systems.

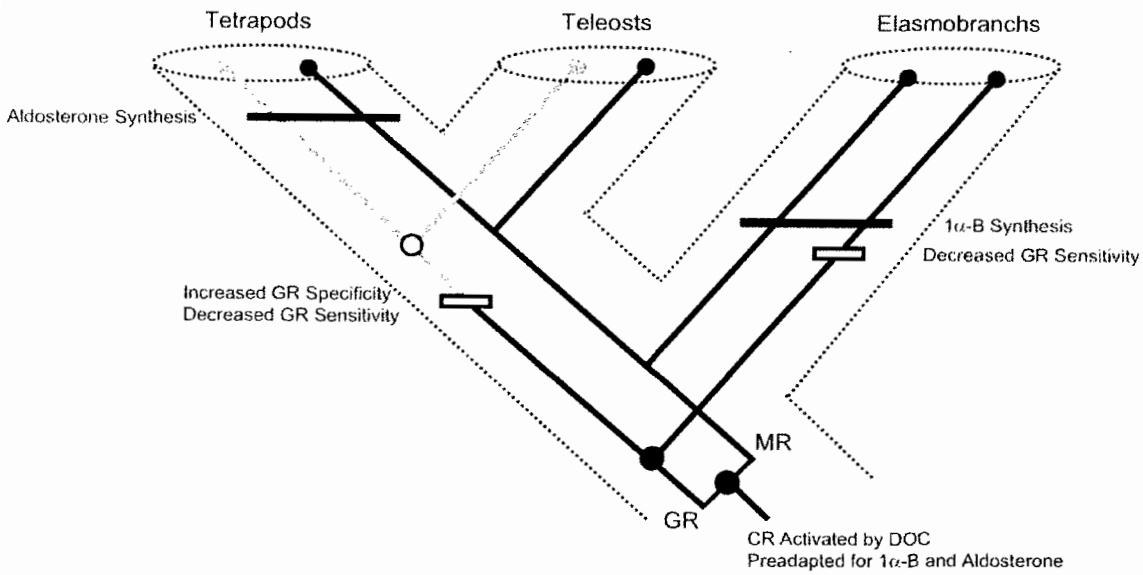


Figure 2.6. Receptor exploitation in the evolution of corticosteroid signaling. The phylogeny traces the evolution of receptor genes and their ligand sensitivity (black and gray lines, sensitive and insensitive to 1a-B, respectively) within the divergence of major vertebrate lineages (funnels). Closed circles at nodes on the phylogeny indicate ancestral receptors that were resurrected and functionally characterized. Open boxes mark functional shifts in the GR lineage. The origin of aldosterone and 1α-B synthesis is shown by black bars in the tetrapod and elasmobranch lineages, respectively.

Molecular exploitation is greatly facilitated by untapped promiscuity in ancient proteins, which allows them to accommodate new partners that are minor variants of their original binding partners. 1α-B is identical to the likely ancestral ligand DOC, except for additional oxygen atoms at the C_1 and C_{11} positions. Our structural analysis of AncCR indicates that the ancestral ligand pocket had ample room to accommodate these modifications in 1α-B, when the hormone evolved later in the elasmobranchs. AncCR's promiscuous response to steroids that had not yet evolved contrasts with its unresponsiveness to the more ancient androgens, estrogens, or progestins, which differ from the corticosteroids at other key locations on the steroid backbone (Bridgham et al.,

2006). AncCR therefore appears to have been only as specific as it needed to be: it excluded other endogenously produced hormones that would inappropriately activate the receptor, but not other potential ligands whose synthesis had not yet evolved. When steroids that fortuitously fit the pocket appeared later, this promiscuity set the stage for the evolution of novel receptor-ligand partnerships with physiological or developmental functions. Structural promiscuity – a ligand pocket that contains extra space or flexibility, allowing to accommodate a range of ligands -- is a common protein feature and is likely to have facilitated the evolution of new functions in both receptors and enzymes (Brzozowski et al., 1997; Bledsoe et al., 2002; Pereira de Jesus-Tran et al., 2006; Nettles et al., 2007; Suino-Powell et al., 2008; James and Tawfik, 2001; Copley, 2003; Khersonsky et al., 2006; Bloom et al., 2007b).

Our results shed light on the functional roles of corticosteroids and their receptors in elasmobranchs. The ligand binding domains of skate MR and GR differ little in hormone specificity but radically in sensitivity: MR is activated by much lower corticosteroid concentrations than are required to activate GR. Analysis of serum steroid concentrations in a wide variety of elasmobranch species have indicated that 1α-B levels are very high – up to several hundred nanomolar (Truscott and Idler, 1968; Kime, 1977) – and that corticosterone, DOC, and 11-DHC are present at orders of magnitude lower concentrations (Idler and Truscott, 1967; Truscott and Idler, 1972; Kime, 1977; Hazon and Henderson, 1984; Armour et al., 1993; Rasmussen and Crow, 1993; Snelson et al., 1997). Our results indicate that skate MR has corticosteroid sensitivity similar to that of tetrapod and teleost MRs (Bridgham et al., 2006), whereas the skate GR's sensitivity is

orders of magnitude lower. Although *in vitro* reporter assays using fusion proteins are not precise predictors of *in vivo* dose-response relationships, physiological 1α-B levels, at least under some conditions, are likely to be well beyond those necessary to activate skate MR; it is possible that DOC and corticosterone may also serve as MR ligands. Skate GR, in contrast, is likely to be activated only by elevated concentrations of 1α-B (and possibly corticosterone) such as those that occur under conditions of stress (Manire et al., 2007). Our findings suggest the hypothesis that low corticosteroid levels regulate the MR to control osmolarity, whereas stress-induced increases in the same hormones regulate the GR to control the stress response. Like the elasmobranch GR, the GR of bony vertebrates is a low-sensitivity receptor that activates the stress response only when cortisol levels are very high (Bentley, 1998); unlike the elasmobranch GRs, however, GRs in bony vertebrates evolved novel specificity as well, so that they are insensitive to even high doses of mineralocorticoids (Bridgham et al., 2006). Other factors, such as post-translational modification and the availability of coregulators, may also contribute to differences in GR and and MR function.

MR and GR are ubiquitously coexpressed, at roughly equal proportions, in a wide variety of skate tissues. This is largely consistent with patterns of expression found in teleosts (Bury et al., 2003; Greenwood et al., 2003; Sturm et al., 2005) and tetrapods (de Castro et al., 1996; Le Menuet et al., 2000). These data provide no evidence for subfunctionalization of expression domains as important in the maintenance of the GR and MR after duplication. We cannot rule out the possibility, however, that GR and MR

may be differentially expressed on a finer scale, in specific cell types or groups of cells or at specific stages.

The interaction of corticosteroid receptors, their ligands, and the DNA response elements they regulate represent a relatively simple but physiologically essential molecular network. Gene duplication provides raw material for the genesis and expansion of such networks (Wagner, 1994; Teichmann and Babu, 2004), but duplication alone is not enough to make a network more elaborate or generate a new one. To increase complexity, functional interactions must diversify, either by the evolution of new interactions or the partitioning of ancestral functions among new network nodes generated by duplication. We have shown that the corticosteroid hormone/receptor network was elaborated when receptor gene duplication and divergence, together with the extension of a biosynthetic pathway, produced new receptors and hormones that were slight structural and functional modifications of their ancestral forms. The receptors' promiscuity set the stage for these new hormones to be integrated into the existing receptor signaling network. Subsequently, divergence of the two sister receptors – leading to reduced sensitivity in the GR – resulted in markedly different quantitative sensitivities to corticosteroid hormones and the possibility of distinct network responses to varying levels of hormone. These observations indicate that the complex molecular networks that drive physiological process can evolve by mechanisms as simple as molecular exploitation and the partial degradation of function after gene duplication.

BRIDGE TO CHAPTER III

These results lead to the question of how the degradation of GR functions evolved. Because the skate GR is the only known receptor of its kind, reduced sensitivity could have arisen within the cartilaginous fishes or in the divergence of GR after gene duplication. The latter offers an intriguing hypothesis – that GR functions might have evolved by high-frequency degenerative mutations after gene duplication. The GR lineage could have begun the 'road to nonfunctionalization,' but was preserved when degenerative mutations created a low sensitivity receptor that was recruited for novel signaling roles. Chapter III seeks to test this hypothesis and reconstruct the changes in GR sequence, structure, and function after duplication of AncCR. Using a combination of ancestral gene resurrection, site-directed mutagenesis, and X-ray crystallography, I found that the first GR ancestor was in fact a low sensitivity receptor. Reduced sensitivity evolved by a number of mutations that destabilize the receptor structure, but are buffered by an additional change that increases global receptor stability. These results suggest that the neofunctionalization of GR was driven by high-probability degenerative mutations that created a novel function through the partial degradation of ancestral structures.

CHAPTER III

DEGENERATIVE AND STABILIZING MUTATIONS IN THE ORIGIN OF THE GLUCOCORTICOID RECEPTOR

This work is in preparation for publication and is co-authored with Eric A. Ortlund, who assisted with experiments and experimental design, and Joseph W. Thornton, who assisted with experimental design and edited the manuscript.

INTRODUCTION

Gene duplication provides the raw genetic material upon which new protein functions can evolve. Ohno theorized that after gene duplication, one copy is freed from purifying selection and allowed to accumulate previously "forbidden" mutations that, in very rare instances, lead to novel gene functions (Ohno, 1970). In most cases, however, these mutations are degenerative, resulting in a loss of function, and the duplicate gene is lost (Demuth and Hahn, 2009). Degenerative mutations can sometimes play a constructive role in duplicate gene preservation if the gene contains multiple, independent subfunctions. In the duplication-degeneration-complementation (DDC) model (Force et al., 1999), the various subfunctions of a single ancestral gene are partitioned between

duplicate loci by complementary degenerative mutations; duplicates are preserved because each is necessary to recapitulate the full complement of ancestral subfunctions. This process, however, merely allots ancestral gene functions between daughter genes and does not directly result in the evolution of novel functions. Recent work by Bridgham *et al.* (Bridgham et al., 2008) suggests that highly probable loss-of-function mutations can in fact be the mechanism for neofunctionalization. In that case, degenerative mutations in one duplicate copy knocked out a single modular protein subfunction and created a novel competitive repressor of its paralog, suggesting that degenerative evolution after gene duplication can be a creative force in and of itself, significantly increasing the likelihood of neofunctionalization by potentially loss-of-function mutations. However, the prevalence of this process in the evolution of duplicate genes and gene functions is not yet known.

We hypothesized that degenerative mutational processes may have played a creative role in the preservation and divergence of the glucocorticoid receptor. The glucocorticoid receptor (GR) functions as a ligand-activated transcription factor, relaying the signal of a far-off secreted hormone to the local regulation of gene expression in target tissues (Beato et al., 1995). In humans, the GR is activated by cortisol to regulate facets of immunity, metabolism, and the stress response (Bentley, 1998). GR and its paralog the mineralocorticoid receptor (MR) arose by the duplication of a single ancestral receptor over 440 million years ago (Thornton, 2001; Bridgham et al., 2006). Using methods in the phylogenetic reconstruction of ancestral proteins (Thornton et al., 2003), our group previously demonstrated that the last common GR/MR ancestor functioned

much like extant MRs – highly sensitive to a variety of corticosteroid hormones – and that GR functions are derived. After duplication, GRs evolved two major changes in their response to ligand: 1) reduced sensitivity to all hormones, as is present in all known GRs, and on top of this, 2) cortisol-specific activation, as is present in teleost and tetrapod receptors (Bridgham et al., 2006). The latter event is well studied: the historical mutations, their effects on protein structure-function, and ancient mutational trajectories have all been elucidated in the evolution of cortisol specificity (Bridgham et al., 2006; Ortlund et al., 2007). The loss of GR sensitivity, however, has been largely unexplored and could reflect the partial degeneration of ancestral receptor structures and functions.

The purpose of this study was to determine precisely when and how reduced GR sensitivity evolved. Because all known GRs have reduced sensitivity, we hypothesized that the very first GR also functioned as a low sensitivity receptor, and sought to test this hypothesis using ancestral state reconstruction and gene resurrection (Thornton, 2004). The protein sequence of the first GR ancestor - AncGR1.1 - was reconstructed and "resurrected," its sensitivity to hormone was assessed using a reporter gene assay, and its protein structure was solved using X-ray crystallography. We show that after gene duplication, the GR lineage evolved by the partial degradation of a highly sensitive ancestor to become a novel low sensitivity receptor for corticosteroid signaling.

METHODS

Receptor Isolation.

The little skate (*Leucoraja erinacea*) GR ligand-binding domain (LBD) was isolated previously using degenerate PCR and RACE with liver cDNA (Bridgham et al., 2006). The skate GR protein sequence was used in a tblastn search of the elephant shark genome (http://esharkgenome.imcb.a-star.edu.sg/) to identify the GR LBD; gene-specific primers were designed to amplify the elephant shark GR LBD coding sequence. All other Chondrichthyan GR LBDs were isolated by hemi-degenerate PCR from cDNA using a degenerate primer in the GR DNA-binding domain (DBD) in combination with a gene-specific primer for a ~ 25 bp sequence conserved in the 3'-UTR of the elephant shark and skate (5'-TCATATGCACTACATATGGTTTACAGA-3'). High-fidelity PCR was used to amplify GR LBDs from five cartilaginous fish species: elephant shark (*Callorhincus milii*), Atlantic sharpnose shark (*Rhizoprionodon terraenovae*), brownbanded bambooshark (*Chiloscyllium punctatum*), small-spotted catshark (*Scyliorhinus canicula*), and Atlantic stingray (*Dasyatis sabina*). Template cDNA for PCR was graciously provided by B. Venkatesh (*C. milii* and *C. punctatum*) and S. Nunez (*R. terraenovae*, *S. canicula*, and *D. sabina*).

Phylogenetic Analysis.

The conserved DNA- and ligand-binding domains of 100 steroid receptor protein sequences were aligned using Clustal X (Larkin et al., 2007). Maximum likelihood

phylogenetics was performed using PhyML_aLRT (Anisimova and Gascuel, 2006) assuming the Jones model of evolution (Jones et al., 1992) and a four-category discrete gamma distribution of among-site rate variation, with the shape parameter estimated from the data; this model was previously shown to be highly supported, with 100% posterior probability, when this and other models are compared in a Bayesian analysis (Bridgham et al., 2006). Support at nodes was calculated as the approximate likelihood ratio (LR) of the maximum likelihood tree versus the next best tree that does not contain that node (Anisimova and Gascuel, 2006).

Ancestral Sequence Reconstruction and Gene Resurrection.

The maximum likelihood tree topology differed from previously published SR phylogenies (Bridgham et al., 2006) with respect to the placement of jawless fish receptors; to account for this discrepancy in trees, ancestral receptor sequences were reconstructed over both the experimental and published (Bridgham et al., 2006) trees weighted by their probability. Ancestral states were inferred using the Ancestral Reconstruction Tool [Victor Hanson-Smith, unpublished] given the above alignment matrix and the JTT model. For any ancestor relevant to our study, no site in the inferred sequence possessed amino acid states that differed between trees. A nucleic acid sequence coding for the LBD of the last common ancestor of all GRs (AncGR1.1) and optimized for expression in mammalian cells was synthesized *de novo* (Genscript, Piscataway, NJ) and characterized as described below.

Tests of Evolutionary Rates.

Rates of protein evolution were analyzed in HyPhy (Pond et al., 2005). A likelihood ratio test (LRT) was used to compare the relative branch lengths from the last common GR/MR ancestor (AncCR) to the last common ancestors of all GRs (AncGR1.1) or MRs (AncMR1). The likelihood was calculated given the alignment, maximum likelihood tree, and the JTT model. The alternate and null hypotheses of the LRT optimized branch lengths given the full (unconstrained) likelihood model or one in which the MR and GR branches post-gene duplication were constrained to have the same length. The likelihood ratio of alternate and null models was determined and a p-value calculated using a chi-squared distribution with one degree of freedom.

Receptor Characterization.

LBDs were cloned as fusion proteins into a pSG5-Gal4DBD expression vector (gift of D. Furlow) and cotransfected using Lipofectamine and Plus Reagents (Invitrogen, Carlsbad, CA) with a UAS-driven luciferase reporter gene (pFRluc) into mammalian cell culture (CHO-K1) and grown in phenol red-free α-MEM plus 10% dextran-charcoal-stripped fetal bovine serum (Hyclone, Logan, UT). Cells were incubated with transfection reagents for four hours, after which they were treated with fresh medium; after recovery, cells were treated in triplicate with hormone or vehicle control, and incubated overnight. Reporter expression was measured using Dual-Glo (Promega, Madison, WI) and dose-

response relationships analyzed using Prism4 (GraphPad, La Jolla, CA). Site-directed mutagenesis was carried out using QuickChange II (Stratagene, La Jolla, CA) and clones verified by DNA sequencing.

Protein Growth, Purification, and Crystallography.

AncGR1.1 was subcloned into a pMCSG7-MBP-His expression vector, transformed into BL21 (DE3) pLysS cells, and grown to an OD600 of 0.8-1.0. Cultures were induced with 0.1 uM IPTG plus 50 uM of the steroid 11-deoxycorticosterone (DOC) and grown overnight at 16° C. Purification of AncGR1.1 was performed using nickel affinity chromatography and a sizing column. Pure AncGR1.1 was concentrated to 3.7 mg/mL and dialyzed into a crystallization buffer consisting of 20 mM Tris, pH 6.5, 150 mM NaCl, 5% glycerol, 50 uM CHAPS, and 50 uM hormone (11-deoxycorticosterone).

Multiple sparse matrix screens were set with AncGR1.1 protein using a Phoenix crystallization robot (Art Robbins Instruments, Sunnyvale, CA); hits formed at 22° C from the Salt Rx screen (Hampton Research, Aliso Viejo, CA). Crystals were optimized at 22° C in hanging drop diffusion plates with: 2.5-2.8 M sodium acetate trihydrate, pH 7.0, 0.1 M BIS-TRIS propane, pH 7.0, and a small peptide designed from the TIF2 Box3 steroid receptor coactivator protein. Crystals were soaked in a cryoprotectant solution containing 20% glycerol and flash-frozen in liquid nitrogen. Data was collected to 1.95 Å resolution at the South East Regional Collaborative Access Team (SER-CAT) at the

Advanced Photon Source (Argonne National Laboratory). Initial phasing of the AncGR1.1 plus DOC structure was determined using molecular replacement of the AncCR with DOC (2Q3Y); refinement of the structure was carried out using COOT version 0.5 (Emsley and Cowtan, 2004) and CCP4i (http://www.ccp4.ac.uk/pjb.html). The root mean square deviation (rmsd), a measure of the overall similarity between protein backbones, was calculated using CaspR (Claude et al., 2004).

RESULTS

GR Phylogenetics, Evolutionary Rate Variation, and Reconstructed Ancestral States

To increase taxon sampling for ancestral reconstruction we isolated additional GRs sampled from throughout the cartilaginous fishes. The GR phylogeny and nodes at which ancestral receptor sequences were reconstructed are shown in Fig. 3.1. This phylogeny was used to test for differences in evolutionary rates after duplication and to reconstruct ancestral receptor states.

To determine whether novel GR functions evolved by increased rates of protein evolution after gene duplication, we sought to compare the relative branch lengths of the GR and MR lineages stemming from the AncCR. The classic model of neofunctionalization (Ohno, 1970) predicts that the duplicate lineage acquiring a novel function should evolve quickly with the rapid fixation of beneficial mutations, while the ancestral functions are maintained by purifying selection in the paralogous copy and this lineage continues to evolve slowly. For phylogenies such as ours with thorough taxon

sampling, we can test for differences in the rate of duplicate gene evolution on branches that cover the same amount of time. We utilized a likelihood-based test to assess the relative branch lengths of the GR and MR lineages from the duplication of AncCR to the divergence of cartilaginous fishes, so that any differences in branch length would reflect unequal rates of substitution after gene duplication. The AncCR:AncGR1.1 branch is noticeably longer (0.123) than the corresponding AncCR:AncMR1 branch (0.041) (Fig. 3.1), but we found only a marginally non-significant difference between the rates of divergence after gene duplication ($p = 0.06$).

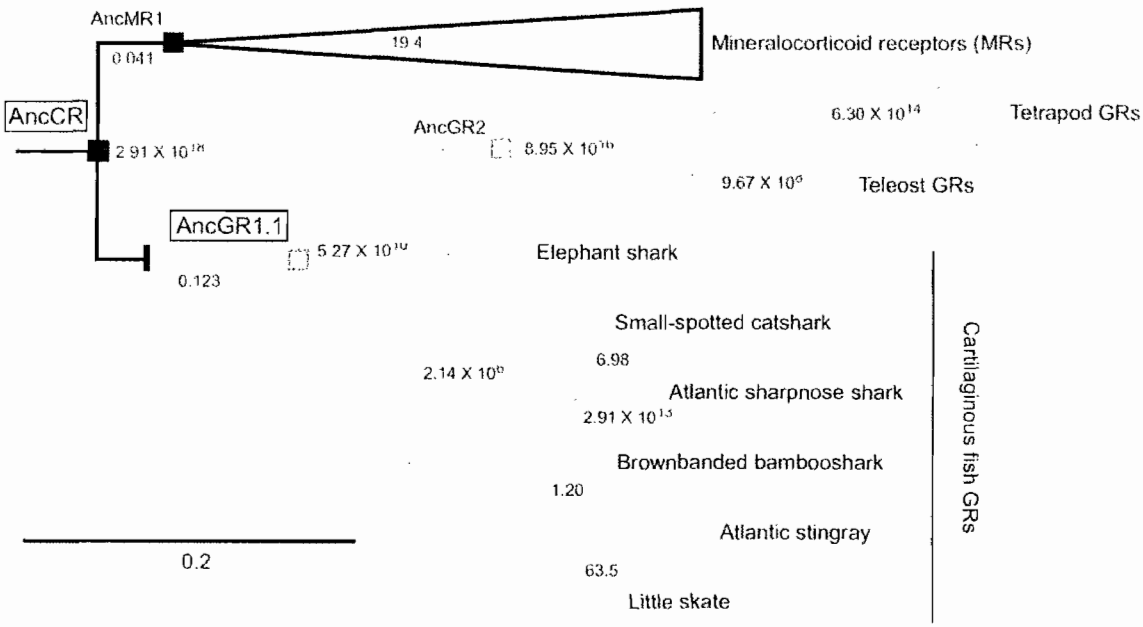


Figure 3.1. Schematic of the GR clade from the maximum likelihood phylogeny. The likelihood ratio and the names of relevant ancestral sequences are shown at major nodes; branch lengths following the duplication of AncCR are also displayed. AncCR, the last common ancestor of all MRs and GRs; AncGR1.1, the GR ancestor of cartilaginous and bony fishes; AncGR2, the GR ancestor of ray- and lobe-finned fishes; AncMR1, the MR ancestor of cartilaginous and bony fishes. Black, high sensitivity receptors; gray, low sensitivity receptors; black dash, the evolution of reduced sensitivity. The AncCR:AncMR1 and AncCR:AncGR1.1 branch lengths were analyzed using a likelihood-based test of relative rates.

The increased taxon sampling of this study was found to affect reconstructed sequences and to increase overall confidence in the reconstruction. The sequence of AncGR1.1 is 7% different from a prior reconstruction of this node (AncGR1) that was inferred using only the skate GR from all cartilaginous fishes. AncGR1.1 has a higher posterior probability averaged across sites (0.951) than AncGR1 (0.930), a greater number of sites reconstructed with 100% posterior probability, and fewer sites with statistical uncertainty. The reconstructed sequences of AncCR and AncGR2 were affected by taxon sampling to a much lesser extent.

AncGR1.1 and its Descendents are Low Sensitivity GRs

AncGR1.1 and its descendant GRs in cartilaginous fishes are all low sensitivity receptors in our reporter gene assay, activated only by high doses of hormone. AncGR1.1 requires greater than nanomolar concentrations of the hormones DOC and B for half-maximal activation, and close to micromolar concentrations of 1α-B and 11-DHC (Fig. 3.2A and Table 3.1). Low sensitivity is conserved in cartilaginous fish GRs (Fig. 3B and Table 3.1), though 11-DHC did not activate (defined as < 2-fold activation for EC_{50} > 1 μM of hormone) and the Atlantic stingray GR was unresponsive with all hormones. Low sensitivity receptors require much higher doses of hormone for activation than high sensitivity receptors such as AncCR; for all ligands tested, AncGR1.1 is approximately 25- to 530-fold less responsive to hormone than AncCR. The sensitivity of AncGR1.1 is also slightly reduced relative to AncGR1, receptors that differ only in the taxon sampling

used for ancestral reconstruction. The low sensitivity of AncGR1.1 strongly suggests that a reduced response to hormone evolved soon after the duplication of AncCR.

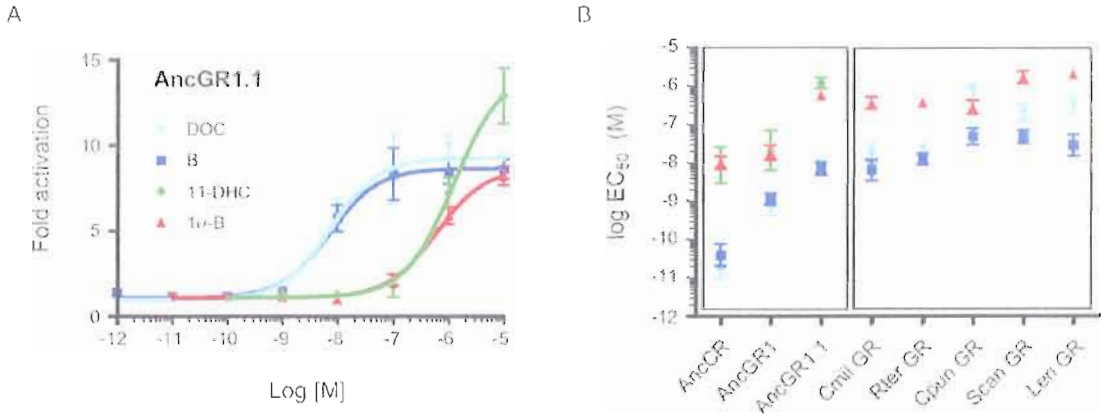


Figure 3.2. AncGR1.1 and its descendent GRs in cartilaginous fishes are low sensitivity receptors. **A)** Ligand-dependent transcriptional activation of the AncGR1.1 was measured in the presence of increasing concentrations of hormone using a luciferase reporter gene assay. Points are the mean of triplicate reactions with the standard error of the mean. Four hormones were tested: 11-deoxycorticosterone (DOC), corticosterone (B), 11-dehydrocorticosterone (11-DHC), and 1α-hydroxycorticosterone (1α-B). **B)** High and low sensitivity receptors. Dose-response curves were calculated for a number of receptor-hormone pairs and plotted as the log effective concentration for half maximal activation (log EC_{50}) in molar with standard error. Larger values, low sensitivity; smaller values, high sensitivity. Hormones are the same as in (A); 11-DHC did not activate cartilaginous fish GRs in our assay (defined as < 2-fold activation for EC_{50} > 1 μM of hormone) and is not shown. Receptors are grouped by reconstructed ancestors or extant cartilaginous fish GRs. AncGR1 is a prior reconstruction of AncGR1.1; cartilaginous fish species are listed Table 3.1.

Table 3.1. Sensitivity of extant and ancestral receptors to hormone. Dose-response curves were calculated for receptors in the presence of hormone; the data is plotted as the log EC_{50} plus standard error.

	DOC		B		11-DHC		1α-B	
Receptor	log EC_{50}	SE	log EC_{50}	SE	log EC_{50}	SE	log EC_{50}	SE
AncCR	-10.870	0.34	-10.410	0.29	-7.315	0.17	-7.992	0.16
AncGR1	-9.182	0.20	-8.944	0.15	-7.304	0.12	-7.735	0.19
AncGR1.1	-8.144	0.20	-8.151	0.16	-5.924	0.14	-6.232	0.09
Cmil GR	-7.761	0.25	-8.196	0.28	NA	-	-6.438	0.16
Rter GR	-7.623	0.12	-7.912	0.16	NA	-	-6.433	0.12
Cpun GR	-6.143	0.16	-7.320	0.21	NA	-	-6.552	0.18
Scan GR	-6.702	0.24	-7.322	0.17	NA	-	-5.767	0.16
Leri GR	-6.515	0.21	-7.548	0.28	NA	-	-5.700	0.13
Dsab GR	NA	-	NA	-	NA	-	NA	-

[a] AncGR1, prior reconstruction of AncGR1.1; Cmil, elephant shark (*Callorhincus milii*); Rter, Atlantic sharpnose shark (*Rhizoprionodon terraenovae*); Cpun, brownbanded bambooshark (*Chiloscyllium punctatum*); Scan, small-spotted catshark (*Scyliorhinus canicula*); Leri, little skate (*Leucoraja erinacea*); Dsab, Atlantic stingray (*Dasyatis sabina*).

[b] NA, no activation, defined as < 2-fold activation for EC_{50} > 1 μM of hormone

The molecular basis of decreased hormone sensitivity

To dissect the mechanism of reduced hormone sensitivity we examined the historical substitutions that occurred on the branch from AncCR to AncGR1.1. Out of 36 changes on this branch, we chose a set of candidate substitutions that were hypothesized to cause the reduced hormone sensitivity of AncGR1.1. Seventeen of these substitutions were already tested in AncGR1 and did not reduce receptor sensitivity, leaving 19 remaining candidates. Of these, only 7 cause a radical change in amino acid properties

and are therefore predicted to strongly affect receptor function. To test whether any of these 7 candidate substitutions recapitulate the loss of GR sensitivity, we introduced each historical change into the maximum likelihood (ML) AncCR background using site-directed mutagenesis and tested the effect on hormone sensitivity in the reporter gene assay with increasing concentrations of DOC.

Of seven candidate substitutions, two markedly reduced AncCR's sensitivity to hormone. The V43A and R116H single mutants reduced the EC_{50} of DOC by over two orders of magnitude to near AncGR1.1-like values (Fig. 3.3 and Table 3.2). Other mutations caused slight reductions in AncCR sensitivity (Table 3.2), but the V43A and R116H substitutions are clearly those with the greatest effect. When combined, V43A and R116H produce a nearly unresponsive receptor that required over 10^5 times the dose of hormone required to activate the ML AncCR, a value orders of magnitude higher than required to activate AncGR1.1 or any of the extant GRs. Both V43A and R116H occurred on the branch to AncGR1.1, suggesting that additional substitutions buffer the loss of sensitivity conferred by these mutations.

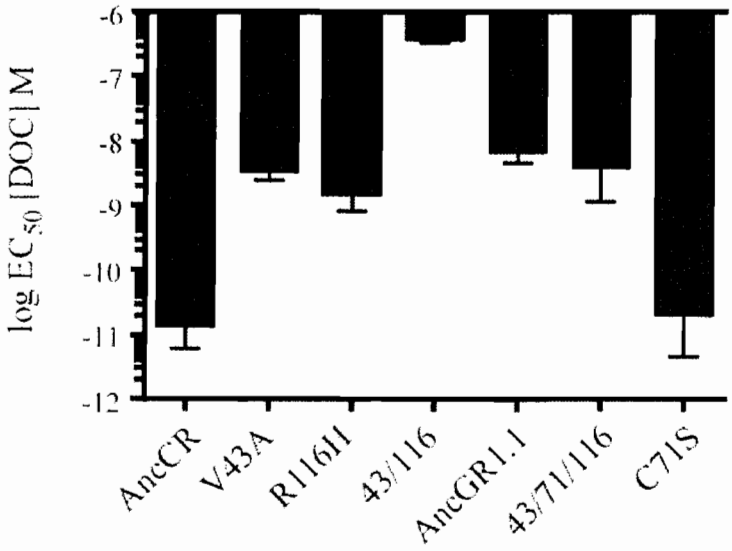


Figure 3.3. Three historical substitutions are sufficient to reduce AncCR sensitivity. Shown is the log EC_{50} of molar 11-deoxycorticosterone (DOC) with a high sensitivity receptor (AncCR), a low sensitivity receptor (AncGR1.1), and single or combination AncCR mutants. The V43A and R116H substitutions were identified from candidate sites; C71S was identified from the structure of AncGR1.1.

Table 3.2. Two historical GR substitutions markedly reduce AncCR sensitivity. Candidate substitutions were introduced into AncCR by site-directed mutagenesis and their effects on receptor sensitivity measured as the change in log EC_{50} of molar 11-deoxycorticosterone (DOC) from the maximum likelihood (ML) reconstruction of AncCR.

Receptor	log EC_{50}	SE	Change in Sensitivity (Fold)
AncCR ML	-10.870	0.34	-
L1M	-10.250	0.30	- 4.2
V43A	-8.443	0.15	- 267.3
K83Q	-10.330	0.41	- 3.5
A107S	-10.350	0.24	- 3.3
Q113K	-9.773	0.43	- 12.5

Table 3.2. (continued).

Receptor	log EC_{50}	SE	Change in Sensitivity (Fold)
R116H	-8.812	0.28	- 114.3
Q211E	-10.170	0.43	- 5.0
43/116	-6.418	0.06	- 28313.9
AncGR1.1 ML	-8.144	0.20	- 532.1
43/71/116	-8.387	0.54	- 304.1
C71S	-10.690	0.64	- 1.5

[a] All mutants are in the AncCR; AncGR1.1 is shown as a low sensitivity receptor

[b] Change in Sensitivity (Fold) is the ± fold-difference in the EC_{50} of molar DOC of mutant and ML receptors

The Crystal Structure of AncGR1.1 Complexed with DOC

To reveal the structural basis of reduced GR sensitivity and identify other important substitutions, we used X-ray crystallography to solve the protein structure of AncGR1.1 complexed with ligand. To compare the respective structures of high versus low sensitivity receptors, we aligned AncCR and AncGR1.1; as expected, AncGR1.1 shares a high degree of structural similarity with AncCR. The main chain conformations of these receptors are structurally very conserved (Fig. 3.4A), with an rmsd of only 0.658 Å. Side chains within 4 Å of DOC are conserved, except for A36G, which alters a ligand-contacting residue. The reversion of this site to its ancestral state in AncGR1.1 (G36A) did not increase sensitivity; contrarily, it reduced responsiveness another 5-fold (data not shown). Because there are no major structural differences in the ligand-binding pocket or

the protein backbone, we focused on evolutionary changes in protein structure induced by the site 43 and 116 substitutions to reveal the mechanism of decreased sensitivity.

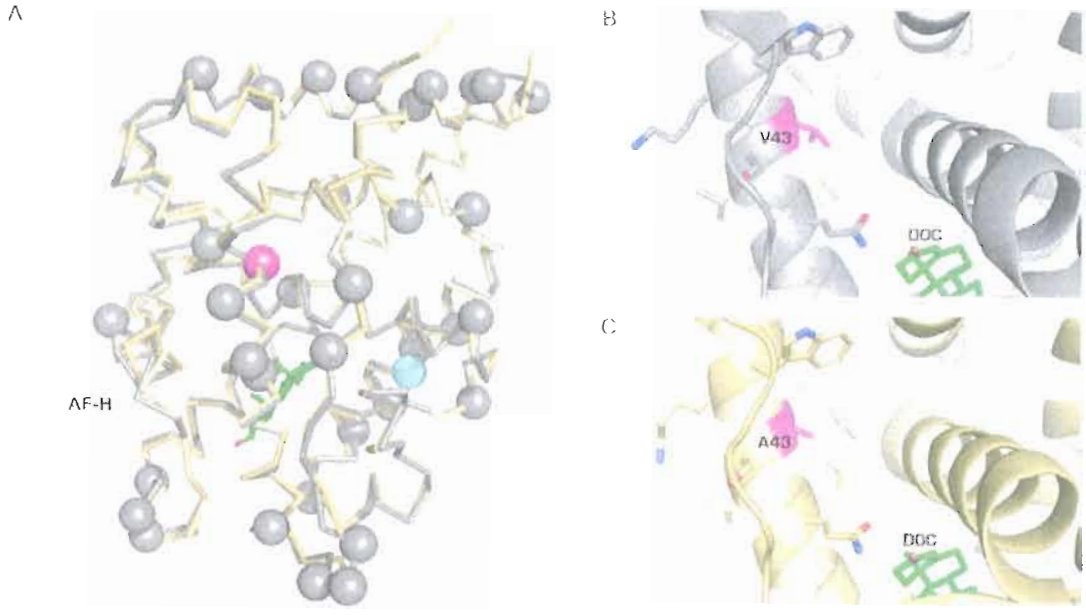


Figure 3.4. The solved crystal structure of AncGR1.1 in complex with 11-deoxycorticosterone (DOC) compared to the previously solved structure of AncCR with DOC (2Q3Y). **A)** AncCR and AncGR1.1 are highly structurally conserved despite 36 substitutions between the two receptors. Gray, AncCR; yellow, AncGR1.1; spheres, sequence differences; magenta sphere, position 43; cyan sphere, position 116; green, DOC; AF-H, activation function helix. **B)** and **C)** show the structural environment around position 43. The substitution of V43 in AncCR (**B**) for A43 in AncGR1.1 (**C**) is thought to weaken the hydrophobic interactions of the surrounding helices.

Mutations that Reduce Sensitivity Cause a Loss of Favorable Interactions

Structural analysis of the V43A substitution suggested a loss of hydrophobic interactions from AncCR to AncGR1.1. V43 in AncCR (Fig. 3.4B) helps to stabilize helix 3 (H3) against H5 and the ligand. The V43A substitution results in a smaller side chain and weaker hydrophobic interactions in the AncGR1.1 structure (Fig. 3.4C). The

loss of contacts with V43A is likely to destabilize the packing of H3 against the receptor and affect ligand contacting residues and/or the hydrophobic cleft. The hydrophobic cleft is an important coactivator docking site on the receptor surface formed at the interface of H3, H5, and the activation function helix (H12 or AF-H) (Feng et al., 1998); these coactivators bridge the receptor-ligand complex to the transcriptional machinery and are necessary for receptor function. Destabilizing mutations around the cleft would likely decrease coactivator binding and weaken ligand-dependent transcriptional activation.

Position 116 is antipodal to 43 on the receptor structure and a hub in an extensive and stabilizing hydrogen bond network. In AncCR, R116 is part of a large network of h-bonds running from the C-terminal loop of H5 through the inner-facing side chains of H7 and into H6 (Fig. 3.5A). This network is greatly reduced in AncGR1.1 due to the R116H substitution and possibly another substitution at the neighboring site 113 (Fig. 3.5B). H116 has a much-reduced ability to form hydrogen bonds, and could be repelled by the like positively charged side chain of K113; the latter site forms a novel salt bridge to E100, further rearranging the ancestral h-bond network. Indeed, Q113K was identified in the original group of seven candidate substitutions from AncCR to AncGR1.1 and had the third greatest effect on sensitivity after the changes at sites 43 and 116 (Table 3.2).

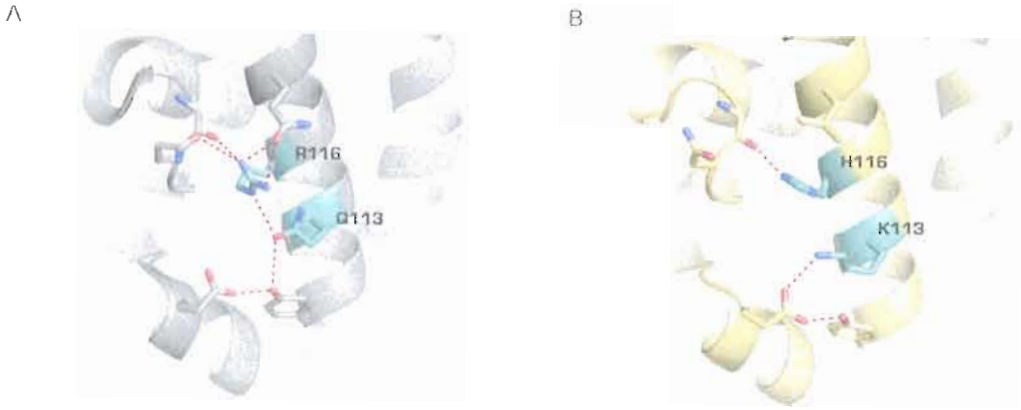


Figure 3.5. AncGR1.1 lost hydrogen bonds in the region of position 116. **A)** R116 in AncCR is part of an extensive and stabilizing hydrogen bond network. **B)** This h-bond network is greatly reduced in AncGR1.1 with an R116H substitution and a neighboring Q113K substitution that rearranges the ancestral pattern of h-bonds.

The loss of hydrogen bonds due to substitutions at sites 116 and 113, as well as the loss of hydrophobic interactions at position 43, suggests that the novel function of AncGR1.1, its reduced sensitivity, evolved because of the partial degeneration of ancestral structures and functions. Our results indicated that other substitutions build favorable interactions in the AncGR1.1 to buffer mutations at sites 43 and 116; we returned to the AncGR1.1 structure to identify these sites.

C71S Adds Core Molecular Bonding and Suppresses Desensitizing Mutations

By analyzing the positions of AncCR:AncGR1.1 changes on the protein structure, C71S emerged as a candidate historical substitution for new favorable interactions in AncGR1.1. C71S creates in AncGR1.1 a new water-mediated hydrogen bond network

with L139 and T143 that stabilizes the receptor core (Fig. 3.6); C71 in AncCR cannot form this structure. We introduced S71 into AncCR and found that this mutation (AncCR_C71S) has absolutely no effect on receptor sensitivity (Fig. 3.3 and Table 3.2). In combination with the destabilizing mutations (AncCR_V43A, C71S, R116H), C71S recapitulates the evolution of AncGR1.1 sensitivity by buffering the effects of mutations at sites 43 and 116.

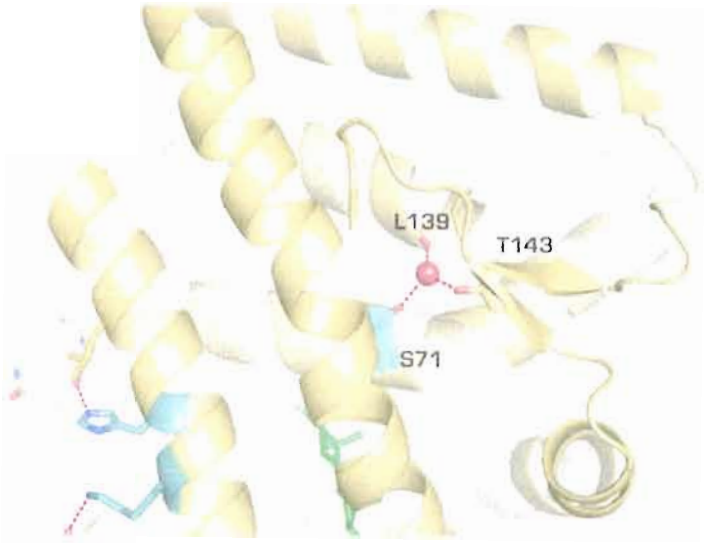


Figure 3.6. A C71S substitution from AncCR to AncGR1.1 adds favorable hydrogen bonds and buffers the effects of other mutations. S71 creates a novel water-mediated hydrogen bond network with L139 and T143, potentially stabilizing the core of the AncGR1.1 structure.

Statistical uncertainty in reconstructed AncGR1.1 sites

To determine whether reduced sensitivity is an artifact of chance errors in the reconstruction, we assessed plausible alternate reconstructions of AncGR1.1 sites. We examined those sites that are most likely to change AncGR1.1 function: a) those shown

in this study to be important for receptor sensitivity, and b) those positions with plausible alternate states (PP > 0.20) that are present in one or more high sensitivity receptors and therefore might increase receptor sensitivity. Alternate reconstructions were introduced into AncGR1.1 using mutagenesis and receptor sensitivity was measured as the response to DOC.

None of the alternate reconstructions in the AncGR1.1 sequence undermine the conclusion that the ancestral GR was a low-sensitivity receptor. All substitutions have virtually no effect on receptor sensitivity (Table 3.3), except for H116Q. This alternate reconstruction produces a modest (7-fold) increase in the response of AncGR1.1 to DOC but is still approximately 80-fold less responsive to hormone than AncCR. Further, the structural basis for reduced sensitivity does not change with the alternate reconstruction: both H116 and Q116 would result in a loss of hydrogen bonds from the highly sensitive AncCR state (R116). These results indicate that the reduced sensitivity of AncGR1.1 is robust to uncertainty in the sequence reconstruction, as is the biophysical mechanism that caused reduced sensitivity

Table 3.3. Statistical uncertainty in the reconstruction of functionally important AncGR1.1 sites. Plausible alternate reconstructions of AncGR1.1 were introduced into the maximum likelihood (ML) sequence and assayed for their effects on the EC_{50} of molar 11-deoxycorticosterone (DOC). Sites are those identified from the reduced sensitivity of AncCR or from plausible alternate AncGR1.1 states (defined as > 0.20 PP with the state present in high sensitivity receptors).

Receptor	PP (Alt)	log EC50	SE	Change in Sensitivity (Fold)
AncGR1.1 ML	-	-8.144	0.20	-
T25S	0.27	-8.245	0.32	+ 1.3
A43V*	0.45	-8.041	0.17	- 1.3
E110D	0.30	-8.196	0.25	+ 1.1
K113Q*	0.28	-7.993	0.21	- 1.4
H116Q*	0.28	-8.972	0.22	+ 6.7
T214S	0.24	-8.159	0.18	+ 1.0
T214A	0.22	-8.137	0.28	- 1.0

* Statistical uncertainty in sites important for sensitivity.

[a] All mutants are in the AncCR; AncGR1.1 is shown as a low sensitivity receptor

[b] Change in Sensitivity (Fold) is the ± fold-difference in the EC_{50} of molar DOC of mutant and ML receptors

[c] PP (Alt) is the posterior probability of plausible alternate states

DISCUSSION

The Evolution of GR Functions by Gene Duplication

Our work suggests that GR arose by the partial degradation of structures and functions present in a highly sensitive ancestor. After duplication of the ancestral

corticoid receptor, mutations in the GR lineage resulted in the loss of hydrophobic interactions, hydrogen bonds, and ligand-binding residues. The degradation of these structures created in the GR ancestor a reduced sensitivity to hormone, where up to two orders of magnitude more ligand is needed for transcriptional activation. Several single mutations are sufficient to recapitulate the reduction of hormone sensitivity after gene duplication, but their combined effects are too much and result in a largely nonfunctional receptor. Receptor function is maintained by an additional substitution that increases core receptor stability and buffers the degenerative effects of other mutations. Additional substitutions then optimized the functions of the GR ancestor as a low sensitivity corticosteroid receptor.

The common ancestor of all jawed vertebrates, in which AncGR1.1 existed, possessed MR and GR that functioned similarly to the respective high and low sensitivity receptors of cartilaginous fishes. In these animals, MR is activated by baseline corticosteroid levels, while GR is activated only by the highest physiological doses of these same hormones (Carroll et al., 2008). Our results demonstrate that the last common ancestor of all GRs was a low sensitivity receptor, and that reduced sensitivity is conserved in the GRs of cartilaginous fishes. The other descendents of AncGR1.1 – the GRs of teleosts and tetrapods – acquired increased specificity to cortisol (Bridgham et al., 2006). Other innovations in corticosteroid-receptor signaling have evolved in multiple different lineages (Bury and Sturm, 2007; Prunet et al., 2006; Keightley et al., 1998).

The Structural Basis of Decreased Receptor Sensitivity

After duplication of AncCR, a combination of potentially stabilizing and destabilizing mutations generated the reduced hormone sensitivity of AncGR1.1. Thermodynamic stability is a global determinant of protein function and is in itself an evolvable character. Periods of random mutation accumulation can quickly degrade protein structures and, past a certain stability threshold, cause the protein to unfold and lose activity (Bershtein et al., 2006; Bloom et al., 2005; Bloom and Arnold, 2009; Bloom et al., 2007). Conversely, stabilizing mutations are thought to buffer the protein beyond the stability threshold and allow a greater number of otherwise deleterious mutations to accrue (Shortle and Lin, 1985; Mitraki et al., 1991; Huang and Palzkill, 1997; Bershtein and Tawfik, 2008a). These so-called "global suppressor" mutations (Shortle and Lin, 1985) might facilitate the preservation of duplicate genes. Periods of relaxed selection and random mutation accumulation are an important part of evolution by neofunctionalization but most often result in nonfunctionalization and gene loss (Ohno, 1970). Suppressor mutations would allow a greater number of mutations to accrue than would otherwise be tolerable, effectively increasing the likelihood of neofunctionalization and decreasing the chances of nonfunctionalization. Here, multiple destabilizing mutations were buffered by the added stability of C71S to generate a novel low sensitivity receptor that was integrated into corticosteroid signaling pathways. Given the complexity of GR functions and its modular protein domain architecture, it is unknown whether the GR locus was preserved after duplication by mutations conferring decreased sensitivity, or mutations to other protein domains.

Site 71 was first identified as a determinant of protein stability in the recombinant expression of human GR (Bledsoe et al., 2002); in fact, all crystal structures of corticosteroid receptors with agonist bound have had to engineer a serine at site 71 to express soluble protein (Kauppi et al., 2003; Bledsoe et al., 2005; Li et al., 2005; Ortlund et al., 2007). We found that this exact experiment was performed over 440 MYA, where a C71S substitution increased the stability of AncGR1.1. Given its neutrality in AncCR, the C71S substitution may have been a chance permissive mutation for decreased sensitivity (Ortlund et al., 2007) or arisen later in GR evolution as a compensatory mutation for the loss of stability caused by other degenerative mutations (Depristo et al., 2005).

Accuracy of Ancestral State Reconstructions

This study displays the power of ancestral sequence reconstruction for mechanistic studies of molecular evolution, but our analyses demonstrate that measures should be taken to assess the accuracy of inferred ancestral states. We found that the choice or depth of taxon sampling can affect the identity and statistical support of reconstructed sequences. Reconstructions at the same node – AncGR1 and AncGR1.1 – have slightly different sequences and functions due only to the number of cartilaginous fish GRs used in the analysis. Opportunities to increase taxon sampling and break up long phylogenetic branches should increase the accuracy of maximum likelihood reconstructions, as has been found in reconstructions by parsimony (Salisbury and Kim, 2001). Equally important is an assessment of the quality of the inferred sequence. Three

of four sites important to the decreased sensitivity of GRs are ambiguously reconstructed in AncGR1.1; fortunately, alternate states at these sites have little effect on receptor sensitivity. Nonetheless, alternate reconstructions could potentially alter the function of reconstructed ancestral states in other model systems and skew the evolutionary inferences drawn from these results. Finally, we show that evolutionary changes to protein structure can allow substitutions that are predicted by the likelihood model to be low probability mutations.

New Functions can arise by the Partial Degeneration of Ancestral Functions

Our study highlights a creative role for partial loss-of-function mutations in the evolution of gene functions. This is similar to the process described by Bridgham *et al.* (2008), where after gene duplication, complete loss-of-function mutations in one modular protein domain of one duplicate can create a competitive repressor of the fully functional paralog. The degradation of GR, however, presents a slightly different mechanism in the partial loss of activity within a modular function – the receptor ligand-binding domain - that created a novel differential response to hormone levels. Steroid signaling relies on very precise molecular cues, and even slight changes in receptor sensitivity could have noticeable effects on the biological response (Simons, 2006). After duplication of AncCR, the MR ligand-binding domain retained ancestral sensitivity, whereas the evolution of reduced sensitivity created a novel regulator of steroid signaling in the GR lineage. These examples demonstrate how mutations that impair native protein functions can drive the evolution of novel functional roles.

BRIDGE TO CHAPTER IV

Despite the wealth of knowledge on steroid receptor evolution, very little is known about the origins of the hormones themselves. It is thought that corticosteroids arose by the duplication of enzymes involved in the synthesis of more ancient hormones such as sex steroids, but a rigorous test of this hypothesis has yet to be done. In Chapter IV, I reconstruct both the timing and the mechanism of the gene duplication that gave rise to the *Cyp21* enzyme that drives corticosteroid synthesis. Using genomic analyses, molecular phylogenetics, and analyses of conserved synteny, I show that the gene duplication that gave rise to *Cyp21* occurred prior to the divergence of vertebrates, and by a complex set of mutational events.

CHAPTER IV

THE EVOLUTION OF CYP21 BY GENE DUPLICATION AND CHROMOSOMAL RELOCATION

This work is in preparation for publication and is co-authored with Joseph W. Thornton, who assisted with experimental design and edited the manuscript.

INTRODUCTION

Corticosteroids are critical regulators of vertebrate immunity, glucose metabolism, and the stress response (Bentley, 1998). Corticosteroids are produced in the adrenal gland, released into the bloodstream, and received by the mineralocorticoid and glucocorticoid receptors, ligand-activated transcription factors that are expressed in target cells (Beato, 1989). The chemical synthesis of corticosteroids is driven by the 21-hydroxylase activity of the CYP21 enzyme, and mutations to this gene cause human disease in the form of congenital adrenal hyperplasia (White and Speiser, 2000). Despite their fundamental importance, very little is known about the evolutionary origins of corticosteroid and the CYP21 enzyme.

A reasonable hypothesis is that corticosteroids evolved by elaborating upon preexistent steroidogenic pathways. The synthesis of corticosteroids and other steroid hormones – including sex steroids such as progestins, androgens, and estrogens – is accomplished by successive enzymatic modifications to a cholesterol precursor. Progestins are the first products of steroid synthesis, after which the pathway bifurcates to produce either corticosteroids or, alternatively, androgens and finally estrogens (Fig. 4.1). The synthesis of sex steroids is quite ancient (Thornton, 2001; Thornton et al., 2003; Keay and Thornton, 2009), and corticosteroids are thought to be a recent evolutionary offshoot of this pathway. Both CYP21 and the enzyme CYP17 share certain progestin hormones (progesterone and 17-hydroxyprogesterone) as substrates for the synthesis of corticosteroids or sex steroids, respectively. It has been proposed that *Cyp21* arose by the duplication of a *Cyp17*-like ancestor that functioned in the synthesis of sex steroids, but a rigorous test of this hypothesis has yet to be done.

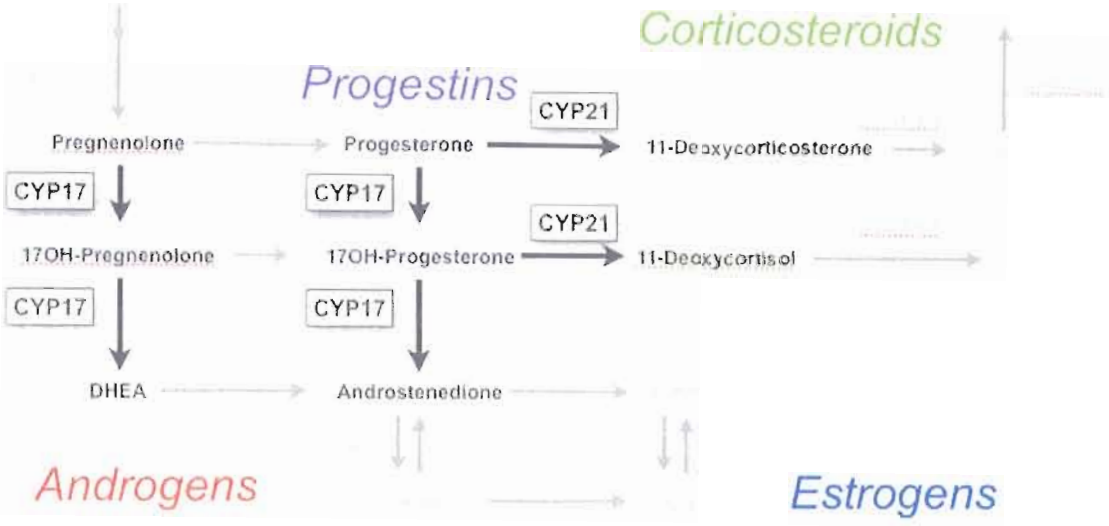


Figure 4.1. – The steroid synthesis pathway and major classes of hormone. Arrows indicate the conversion of one steroid to another by enzymes (boxed). Those reactions pertaining to CYP21 and CYP17 are shown in bold.

Studies of steroid biology in basal vertebrates and their closest outgroups offer insights into *Cyp21* evolution. Genomic and molecular analyses have shown that the steroidogenic pathway is largely conserved in amphioxus, a close chordate outgroup to vertebrates, minus *Cyp21* and other enzymes in corticosteroid synthesis (Castro et al., 2005; Mizuta et al., 2008; Mizuta and Kubokawa, 2007). Amphioxus possesses a single *Cyp17*-like gene (Mizuta and Kubokawa, 2007) but its true homology to *Cyp17* has been questioned (Markov et al., 2009) and its enzymatic functions have yet to be characterized. In another phylogenetically informative taxa, our lab has shown that lampreys – the most basal vertebrate ingroup – possess a single steroid receptor that is co-orthologous to the mineralocorticoid (MR) and glucocorticoid (GR) receptors of higher vertebrates and specific for corticosteroids (Thornton, 2001; Bridgham et al., 2006). The *in vivo* and *in*

vitro conversion of labeled steroid precursors to corticosteroids has been observed in these same animals (Weisbart and Youson, 1975; Weisbart and Youson, 1977; Weisbart et al., 1978). These data suggest that lamprey diverged after the *Cyp17/Cyp21* duplication and should possess copies of both genes, whereas basal craniates such as amphioxus diverged before the duplication and should have a single, co-orthologous copy of *Cyp17/Cyp21*.

To elucidate the timing of the *Cyp17/Cyp21* duplication, we sought to determine what *Cyp*s are present in the amphioxus and sea lamprey and to clearly discern their phylogenetic relationships. In addition, we sought to determine the molecular mechanism of gene duplication using analyses of syntenic conservation between the human *CYP17* and *CYP21* genomic loci. Our results show that a *Cyp17/Cyp21* ancestor duplicated just prior to the evolution of vertebrates by a complex set of mutational events.

METHODS

Isolation and Analysis of Cyp Sequences

The sequence of the published amphioxus (*B. belcheri*) *Cyp17*-like gene (Mizuta and Kubokawa, 2007) was used in a blastn search of the *B. floridae* genome (http://genome.jgi-psf.org/Brafl1/Brafl1.home.html) (Putnam et al., 2008). A predicted cDNA encoding *Cyp17* (BRAFLDRAFT_121141) was downloaded along with over 50 kb of the genomic scaffold (Bf_V2_122) on which BRAFLDRAFT_121141 was predicted. Primers were designed to the predicted start and stop codons using Geneious

v.4.7.4 (Biomatters LTD, Auckland, NZ). Amphioxus Cyp17 was amplified from ovary cDNA using high-fidelity PCR and cloned into the pCR2.1 sequencing vector (Invitrogen, Carlsbad, CA).

Lamprey Cyps were identified in the 5.9x preliminary assembly of the *P. marinus* genome (http://pre.ensembl.org/index.html) using a tblastn search of related *Cyp17* and *Cyp21* sequences. Sequences with the reciprocal best hits against the Genbank nr database were downloaded as the entire genome contig from the genome assembly. Contigs with hits to *Cyp17* were PMAR3:Contig38214:1:11273:1 and PMAR3:Contig9569:1:15185:1; those with hits to *Cyp21* were PMAR3:Contig36955:1:9414:1 and PMAR3:Contig5425:1:27435:1. The assembly of contigs and identification of putative start and stop codons were computed in Geneious. Lamprey *Cyp17* and *Cyp21* were PCR amplified from ovary cDNA and cloned into the pCR2.1 sequencing vector.

Cyp exon/intron boundaries were determined by aligning PCR-amplified cDNA sequences with the corresponding genomic contig or scaffold. The identification of transposable elements and low complexity DNA sequences was determined manually by blastx searchers of non-coding DNA, and by RepeatMasker (http://www.repeatmasker.org). Protein alignments and the presentation of sequence motifs was performed in Geneious.

Molecular Phylogenetics

The lamprey and amphioxus sequences were added to a large dataset of related *Cyp*s with *Cyp1a* and *Cyp1b* genes as the outgroup. Sequences were aligned using MUSCLE (Edgar, 2004). ProtTest2.1 (Abascal et al., 2005) was used to select the best evolutionary model for phylogenetics according to the Akaike Information Criterion framework (Posada and Buckley, 2004). Maximum likelihood phylogenetics was performed in PhyML_aLRT (Guindon and Gascuel, 2003; Anisimova and Gascuel, 2006) using the selected JTT model (Jones et al., 1992) assuming an estimated proportion of invariant sites plus a four-category gamma distribution with alpha estimated from the data. Statistical confidence at nodes was calculated as the χ^2 statistic or 1-p, where p is the estimated probability that the likelihood ratio support of a given node would occur by chance if that node is actually false (Anisimova and Gascuel, 2006).

Analyses of Chromosomal Location and Conserved Gene Order

Chromosome synteny – conservation of gene content and order – was identified using the Synteny Database (Catchen et al., 2009). Searches were conducted using the human, chicken, and amphioxus genomes with a window size of 100 genes. Briefly, this method uses reciprocal best hits from blast to identify paralogous and orthologous genes and their positions, and subsequently uses a sliding window to discover conserved gene order in nearby chromosomal regions. The Synteny Database was also used to examine

the reciprocal best hits of predicted genes on the amphioxus scaffold containing *Cyp17* (Bf_V2_122) with their putative orthologs in humans.

RESULTS

Identification of Amphioxus and Lamprey Cyps

A single gene similar to *Cyp17* was identified in amphioxus, and two genes similar to *Cyp17* and *Cyp21* were found in lamprey. We compared the predicted protein sequences of these *Cyp*s to those of human and eel and found that there is considerable sequence divergence between enzymes (Table 4.1). All sequences have very low amino acid similarity, but the amphioxus and lamprey *Cyp17* genes are most similar to the eel *Cyp17*, and lamprey *Cyp21* is most similar to eel *Cyp21.* All *Cyp* sequences show greater similarity in previously documented conserved regions, including the Ono-sequence (Ono et al., 1988), Ozols' tridecapeptide sequence (Ozols et al., 1981), and the heme binding site (data not shown).

Table 4.1. Percent of sequence identity between amphioxus, lamprey, zebrafish, and human Cyps. The percent of identical states between alignable sites are shown.

	Amphioxus Cyp17	Lamprey Cyp17	Lamprey Cyp21
Amphioxus Cyp17	-	37	32
Lamprey Cyp17	37	-	37
Eel Cyp17	39	52	36
Human Cyp17	34	42	32

Table 4.1. (continued).

	Amphioxus Cyp17	Lamprey Cyp17	Lamprey Cyp21
Lamprey Cyp21	32	37	-
Eel Cyp21	38	34	40
Human Cyp21	31	33	34

Maximum Likelihood Phylogenetics of Cyp Protein Sequences

The maximum likelihood phylogeny of the amphioxus, lamprey, and related *Cyp*s sequences places the timing of an ancient *Cyp17/Cyp21* gene duplication. Lamprey *Cyp17* and *Cyp21* are orthologous to their respective clades, and the single amphioxus *Cyp17*-like gene is co-orthologous to these genes. These relationships are strongly supported in the analysis: the placements of lamprey *Cyp17* and *Cyp21*, the *Cyp17/Cyp21* duplication, and the co-orthologous gene of amphioxus are all supported by a χ^2 statistic of 1.00. Despite high support at most nodes, there are several instances where the topology does not match the accepted species phylogeny, reflecting the rapid evolution of *Cyp* sequences in certain lineages (ie. mammalian *Cyp17*). Nonetheless, these results strongly suggest that: a) amphioxus has a single gene co-orthologous to *Cyp17* and *Cyp21*, b) lamprey has orthologs of both enzymes, and c) the duplication that gave rise to the vertebrates *Cyp17* and *Cyp21* lineages occurred between the divergence of amphioxus and lamprey.

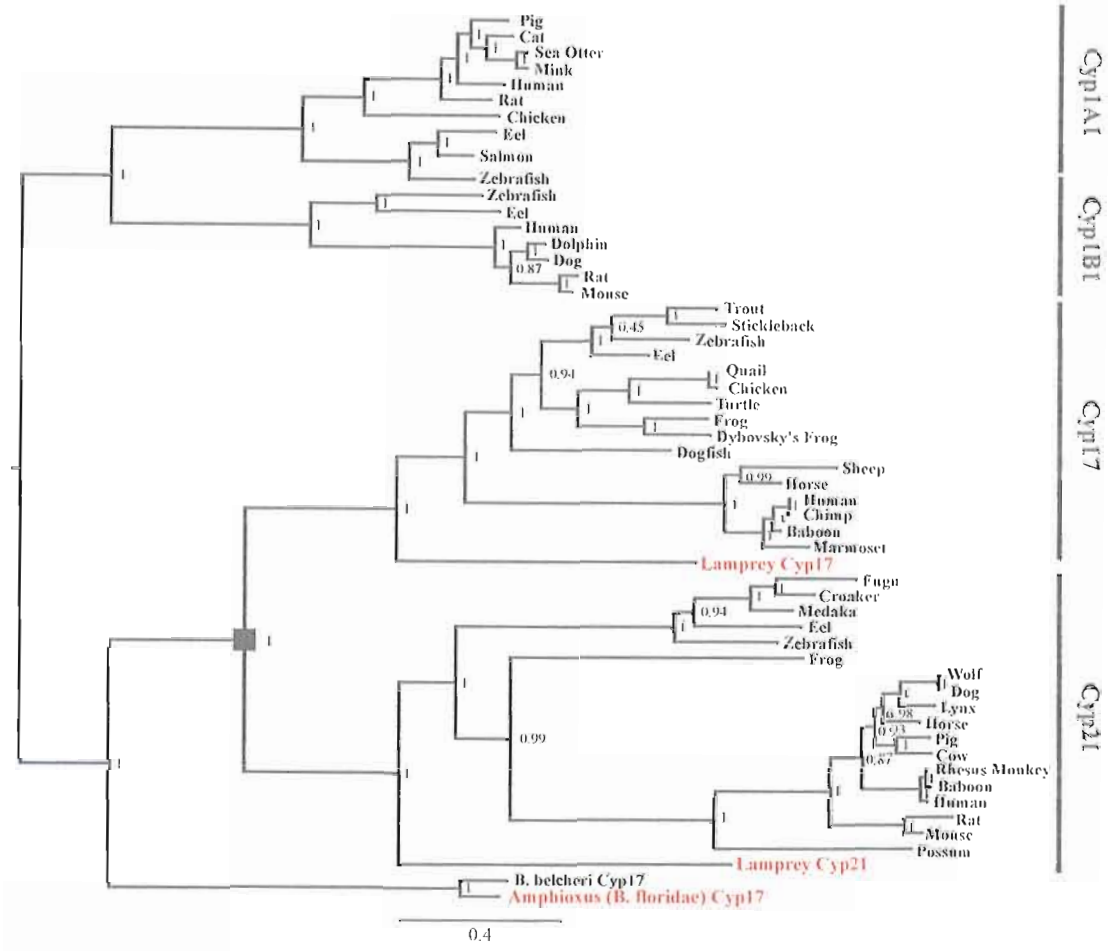


Figure 4.2. Maximum likelihood phylogeny of Cyp17 and Cyp21 proteins. Lamprey and amphioxus Cyps are marked in bold red. The Cyp17/Cyp21 gene duplication is marked with a box. Statistical accuracy at nodes was calculated as the χ^2 statistic or 1-p, where p is the estimated probability that the likelihood ratio support of a given node would occur by chance if that node is actually false (Anisimova and Gascuel, 2006).

The Molecular Mechanism of the Cyp17/Cyp21 Duplication

To determine whether *Cyp17* and *Cyp21* were generated by whole-genome duplication, we performed an automated search for syntenic conservation between the human chromosomes on which *CYP17* and *CYP21* reside using the Synteny Database (Catchen et al., 2009). Analyses of synteny – homologous genes conserved in duplicated

chromosomal regions – are useful for identifying large-scale duplication events such as whole-genome duplication. If generated by this mechanism, we expected to find conserved paralogous synteny between the genomic neighborhoods of the human *CYP17* and *CYP21* genes. Instead, we found a different pattern synteny that suggests *Cyp21* arose by a more complex set of mutational mechanisms. In the synteny analysis, human *CYP17* lies on chromosome 10 (position 10q24.3) and form a four-fold paralogous cluster with regions of chromosomes 2, 5, and 4 (Fig. 4.3A). The presence of four-fold paralogous clusters suggests that these regions were generated by two rounds of large-scale duplications from a single ancestral chromosome. *CYP21* lies on chromosome 6 (position 6p21.3) and forms a different tetra-paralogon with regions of chromosomes 1, 9, and 19 (Fig. 4.3B). Together, these non-overlapping four-fold paralogous clusters suggest that one or both of *Cyp17* and *Cyp21* relocated in the course of their evolution, moving into regions that were generated by two rounds of whole-genome duplication. Whether the *Cyp17*/*Cyp21* duplication was the result of duplicative transposition (and transfer) or whole-genome duplication followed by transposition is unknown.

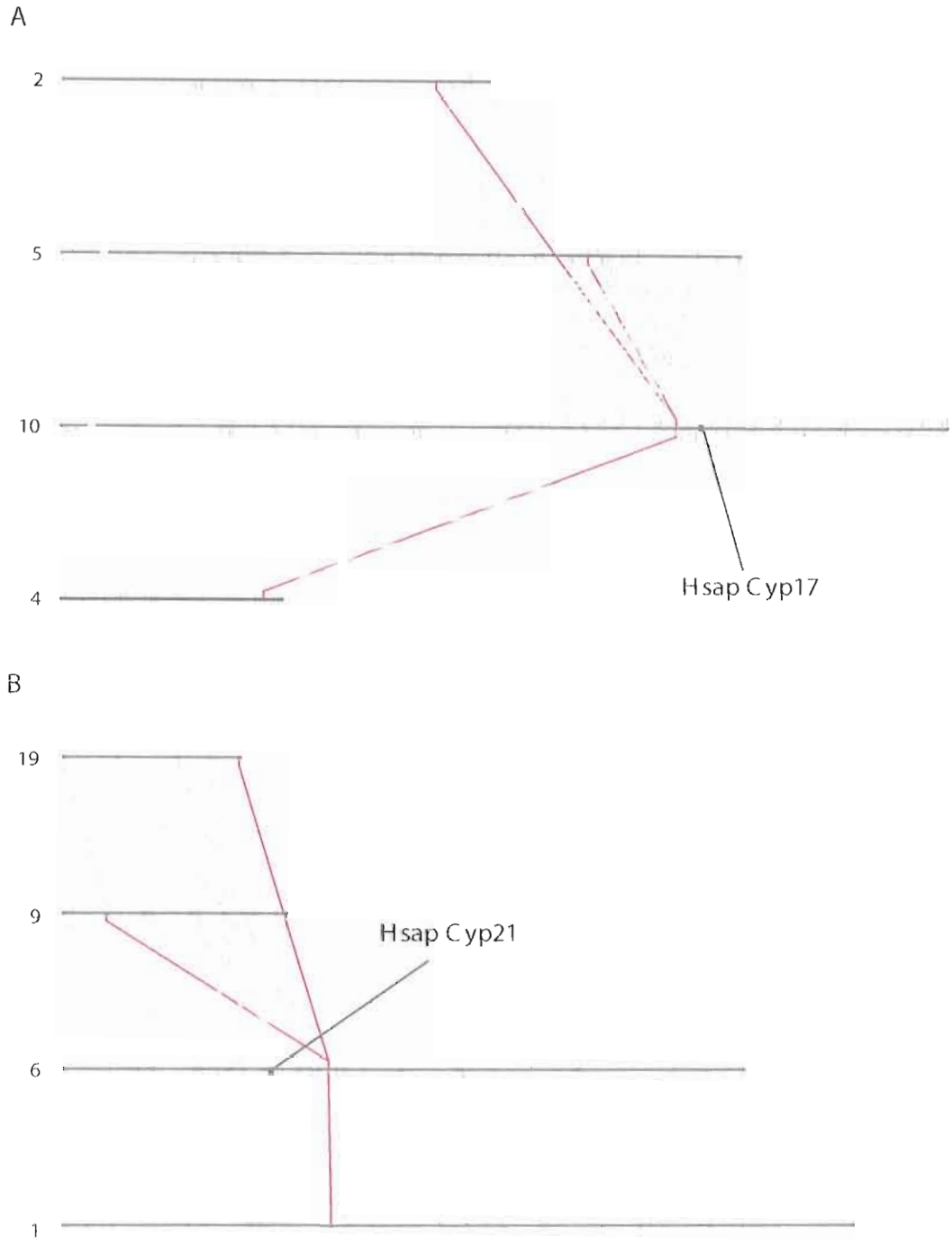


Figure 4.3. Conserved synteny in human chromosomal regions flanking Cyp17 and Cyp21. Shown are the paralogons in which human Cyp17 (A) and Cyp21 (B) reside, though syntenic clusters were only identified by neighboring genes. Human chromosomes are shown as darker and thicker horizontal lines; connecting gray lines show reciprocal best hits of paralogous human genes. The approximate positions of Cyp17 and Cyp21 are indicated. Red lines identify the relationships of four-fold paralogous genes: NFKB2 in (A) and RGL2 in (B).

To determine which paralog relocated after the *Cyp17/Cyp21* duplication, we sought to trace each tetra-paralogous chromosomal cluster in humans back to a single co-orthologous region in amphioxus using the Synteny Database (Catchen et al., 2009) and previously published synteny analyses. The *CYP21* cluster includes many loci in the major histocompatibility complex (MHC) and is one of the most well-studied examples of four-fold synteny, homologous to a single chromosomal segment in amphioxus (Castro et al., 2004; Abi-Rached et al., 2002). The amphioxus *Cyp17/Cyp21* co-ortholog is not present in this region, indicating that *Cyp21* relocated after genome duplication. The history of the *Cyp17* locus is less clear, as we were unable to trace the human *CYP17* cluster back to a single homologous region in amphioxus. The genomic neighborhood of the amphioxus co-ortholog was shown previously to have only weak synteny with human chromosomes 10, 8, 3, and X (Putnam et al., 2008). Collectively, these results do not allow us to discern the location of the ancestral *Cyp17/Cyp21* locus, and it is unclear exactly how many times *Cyp*s have relocated in the course of their evolution.

DISCUSSION

Corticosteroid Signaling Evolved from Preexistent Estrogen Signaling Pathways

Our work sheds light on the timing and mechanism of the gene duplication that gave rise to *Cyp21* and corticosteroid synthesis. We identified orthologs of Cyp17 and Cyp21 in the sea lamprey and, in contrast to other analyses (Mizuta and Kubokawa,

2007; Markov et al., 2009), a single gene that is co-orthologous to Cyp17 and Cyp21 in an amphioxus (*B. floridae*). This suggests that corticosteroids arose in vertebrates by the duplication of enzymes that previously functioned in the synthesis of sex steroids.

Given these results, we can reconstruct the timing of when corticosteroids and their receptors evolved. Amphioxus possesses only the receptors and hormones for estrogen signaling (Mizuta et al., 2008; Mizuta and Kubokawa, 2007; Bridgham et al., 2008); some time between the divergence of amphioxus and lamprey, primitive vertebrates evolved corticosteroid-sensitive receptors (Thornton, 2001) and the *Cyp21* enzyme necessary for corticosteroid synthesis. The order by which corticosteroids and their receptors evolved – either receptor-first or hormone-first – is not known; prior work, however, suggests that both are plausible scenarios for the evolution of novel receptor-hormone interactions (Thornton, 2001; Bridgham et al., 2006; Carroll et al., 2008).

There is some debate as to the antiquity of estrogen synthesis and the ancestral pathway from which the synthesis of corticosteroids arose. Work from our lab and others strongly suggests that estrogens and their receptors are quite ancient (Thornton, 2001; Keay and Thornton, 2009; Garcia-Alonso et al., 2006; D'Aniello et al., 1996; Di Cosmo et al., 2001; Di Cosmo et al., 2002), while others have suggested that these hormones evolved independently in vertebrate and invertebrate lineages (Markov et al., 2008; Markov et al., 2009). This latter hypothesis, however, is questionable due to inferences made from weakly supported *Cyp* phylogenies. Due to the high rates of sequence divergence between *Cyp* homologs, the validity of *Cyp* phylogenetic analyses should be carefully assessed. This study found very strong support for clades within the vertebrate

Cyp17 and *Cyp21* lineages that are incongruent with the established species phylogeny; similar false relationships are likely to form with greater phylogenetic distances. Because of these inherent difficulties, we believe that a combination of analyses – including molecular phylogenetics, genomic analyses, and molecular assays – are necessary for robust inferences of *Cyp* and steroid hormone evolution.

The Mechanism of the Cyp17/Cyp21 Duplication

There are four common molecular mechanisms of gene duplication, including: whole-genome duplication (polyploidization), duplicative transposition (non-allelic homologous recombination and non-homologous end joining), unequal crossing over (tandem duplication), and retrotransposition (Hahn, 2009; Hastings et al., 2009; Paques and Haber, 1999). The least likely mechanism for the *Cyp17/Cyp21* duplication is retroposition, as this model would require that the duplicate *Cyp* locus regained all of its intronic sequences after duplication. *Cyp17* and *Cyp21* may have evolved by tandem duplication, but this requires two separate mutational events (tandem followed by duplicative transposition) and there is no additional evidence supporting tandem duplication.

Several lines of evidence suggested that the duplication creating *Cyp17* and *Cyp21* was part of a large-scale and possibly whole-genome duplication event. First, the timing of the duplication – between amphioxus and lamprey – is consistent with hypotheses of polyploidization in vertebrate genome evolution. Almost forty years ago,

Ohno proposed that several rounds of whole-genome duplication occurred in some vertebrate ancestor(s) (Ohno, 1970); today, the "2R" hypothesis (Hughes, 1999) is generally accepted as one round of whole-genome duplication between amphioxus and lamprey, and another round after lamprey but before the split of jawed vertebrates (Holland et al., 1994). Second is that steroid receptors have expanded by *en bloc* duplications that are associated with the 2R. Amphioxus has two steroid receptor genes, lamprey has three (one copy was lost after duplication), and jawed vertebrates such as humans have six (Thornton, 2001; Bridgham et al., 2008; Bridgham et al., 2006); genome mapping analyses show that at least four of the human steroid receptors are syntenic with one another and evolved by the twice-duplication of a common ancestral chromosome (Thornton, 2001). Our synteny analyses clearly demonstrate that the chromosomes on which *CYP17* and *CYP21* reside are the products of two rounds of whole-genome duplication; the fact that these clusters are non-overlapping, however, means that one or more *Cyp*s have translocated in the course of their evolution and we cannot discern whether it was a whole-genome duplication that specifically gave rise to *Cyp17* and *Cyp21*. Nevertheless, the ancestral *Cyp17/Cyp21* locus was undoubtedly duplicated in this whole-genome duplication event, but the products of this duplication may have been lost in the course of evolution and the event that gave rise to *Cyp21* may have been driven by other molecular mechanisms.

Duplicative transposition is an alternative mechanism for the *Cyp17/Cyp21* duplication. We found many instances of transposon insertions (data not shown) in the genomic sequences of amphioxus and lamprey *Cyp*s, which are known to facilitate

recombination and duplicative transposition (Gray, 2000). The amphioxus *Cyp17* locus actually has an inverted, partial tandem duplicate that most likely evolved by non-allelic homologous recombination of flanking DNA transposon sequences. These transposable elements can induce chromosomal duplications and rearrangements through many different mechanisms and are enriched at the boundaries of segmental duplications in humans (Bailey et al., 2003) and flies (Fiston-Lavier et al., 2007). Transposon-mediated recombination could have duplicated the *Cyp17/Cyp21* ancestor and moved these genes to their current chromosomal locations in one mutational event. It is also possible that *Cyp21* arose by whole-genome duplication followed by duplicative transposition and the loss of the ancestral *Cyp21* locus. Because evolutionary divergence prevents us from "anchoring" the ancestral *Cyp* locus, the exact mechanism by which *Cyp21* evolved is unknown. Improved annotations of the amphioxus, lamprey, and other phylogenetically informative genomes would improve synteny analyses and help to delineate the mechanism of duplication.

Interestingly, the process of recombination associated with duplicative transposition persists at the human *CYP21* locus. Over 95% of *CYP21* disease state alleles are the result of recombination between the functional locus and a nearby pseudogene (*CYP21P*) that shares 98% sequence identity with its functional paralog (White et al., 1986; Higashi et al., 1986). Recombination causes deletions and gene conversions in *CYP21*, whereby nonfunctional mutations from the pseudogenic copy are recombined into the functional locus. The *CYP21* locus is particularly prone to these events because of its position in the MHC class III cluster, a region in which high rates of

recombination are maintained by balancing selection to generate allelic diversity in immune effector molecules (Klitz et al., 1986; Takahata et al., 1992). Thus, similar mutational processes of recombination might underlie the origins of *Cyp21* and the generation of disease state alleles.

In addition to facilitating gene duplication, transposable elements could have played a role in the origins of 21-hydroxylase activities in *Cyp21*. Both amphioxus and lamprey *Cyps* have acquired independent retroelement insertions in their intronic sequences, as well as variable exon/intron boundaries between genes (data not shown). Transposon insertions are normally deleterious to gene function, but they can sometimes serve as the substrate for novel coding sequence in a process called "exon recruitment" (Nekrutenko and Li, 2001; Kazazian, 2004). Retroposons can also contain their own splice sites and result in novel alternatively spliced transcripts (Makalowski et al., 1994), or copy and paste exons to new genomic locations (Moran et al., 1999). Mutational processes associated with the insertion of transposable elements might have equally if not more important than sequence substitutions to the divergence of *Cyp21* from a *Cyp17*-like ancestor. Functional analysis and characterization of sites at exon/intron boundaries would help to explore this hypothesis.

The Novelty of Corticosteroid Signaling

Corticosteroid signaling is a critical regulator of vertebrate physiology and a vertebrate-specific innovation. The players necessary for corticosteroid synthesis evolved

prior to the divergence of lamprey, as did the corticosteroid-sensitive receptors (Thornton, 2001), arising from preexistent components of estrogen signaling. Even the adrenal gland – the organ in which corticosteroids are produced – shares a common developmental origin with the gonads (Hatano et al., 1996), and adrenal-like cells persist in testicular tissue (Val et al., 2006). In the evolution of *Cyp21*, seemingly "novel" enzymatic functions arose by redirecting the hydroxylase activity of *Cyp17* to a new atom on the same steroid substrates (progesterone and 17-hydroxyprogesterone). The distance between the C17 and C21 carbons on which the respective activities of *Cyp17* and *Cyp21* are carried out is a mere 2.6 Å. This shows how new protein functions, and in fact entire pathways of endocrine signaling, can arise by subtle changes in protein structure and function.

CHAPTER V

CONCLUSIONS

Gene and whole-genome duplications provide the raw material for molecular innovation. My dissertation describes the mechanisms and dynamics by which gene duplication and divergence gave rise to corticosteroid signaling in vertebrates. Corticosteroid hormones and their receptors are an ideal system in which to study how molecular diversity and complexity evolve. A combination of techniques from evolutionary biology and traditional molecular analyses permit highly mechanistic studies of the ways in which novel hormones, receptors, and their interactions arise. My work shows that seemingly complex facets of corticosteroid evolution can evolve by simple processes that elaborate upon preexistent molecular functions.

Much like the exploitation of promiscuous enzyme activities, the MR and GR of the skate evolved lineage-specific functions from the promiscuous activities of their ancestor. The skate MR and GR were preadapted for signaling with 1α-B due to the structural promiscuity of AncCR. The sensitivity of AncCR to 1α-B is a byproduct of activation towards structurally similar, more ancient ligands. The AncCR is not necessarily a nonspecific or a generalist receptor; it is only as specific as it needs to be to activate with corticosteroids but does not distinguish much among them. After

duplication of AncCR and the synthesis of 1α-B in elasmobranchs, MR and GR were recruited for lineage-specific signaling functions. Thus, 1α-B signaling evolved as the fortuitous byproduct of structures adapted for preexistent functions.

Similarly, "novel" functions arose in the GR lineage by the partial degradation of ancestral structures and functions. I used ancestral sequence reconstruction and gene resurrection to show that the first GR ancestor functioned as a low sensitivity receptor, much like the little skate GR. After gene duplication, the GR lineage accrued mutations that reduced the stability and sensitivity of the receptor through the loss of hydrophobic interactions, hydrogen bonds, and direct ligand contacts. This loss, however, was buffered by an accompanying substitution to increase core receptor stability. Reduced sensitivity in GR created an added level of complexity in corticosteroid signaling – a low sensitivity receptor with a response distinct from that of MR or ancestral receptors. Other gain-of-function mutations likely occurred in other GR domains or non-coding regions, but its "novel" low sensitivity response to hormone arose by degenerative mutations after gene duplication.

Finally, I show that corticosteroid hormones evolved by gene duplication and elaborations upon existing steroid synthesis pathways. The duplication giving rise to the Cyp17 and Cyp21 lineages occurred just prior to the divergence of lamprey, making corticosteroids a vertebrate innovation. The timing of this duplication coincides with a round of whole-genome duplication proposed by the 2R hypothesis (Holland et al., 1994), but analyses of gene synteny suggest that an array of mutational mechanisms led to the duplication and translocation of the Cyp21 locus. Additional genomic and

molecular analyses of Cyps in basal vertebrates will help to elucidate the origins of corticosteroid synthesis.

The origin of corticosteroids and their receptors presents a model system by which novel protein functions arise by gene duplication. This system arose largely by the exploitation of preexistent forms: structural and functional promiscuity of the AncCR, the degenerative evolution of ancestral forms and functions in GR, and the elaboration of preexistent steroid synthesis pathways for the production of corticosteroids. The evolutionary processes underlying the origins of novel hormones, receptors, and interactions are not unlike those observed in other historical studies of evolution, or in protein engineering experiments in the lab. These are mechanisms by which protein functions evolve.

REFERENCES

Abascal, F., Zardoya, R., and Posada, D. (2005). ProtTest: selection of best-fit models of protein evolution. Bioinformatics *21*, 2104-2105.

Abi-Rached, L., Gilles, A., Shiina, T., Pontarotti, P., and Inoko, H. (2002). Evidence of en bloc duplication in vertebrate genomes. Nat Genet *31*, 100-105.

Amitai, G., Gupta, R., and Tawfik, D. S. (2007). Latent evolutionary potentials under the neutral mutational drift of an enzyme. HFSP J. *1*, 67-78.

Anfinsen, C. (1973). Principles that govern the folding of protein chains. Science *181*, 223-230.

Anfinsen, C. B. (1959). The Molecular Basis of Evolution (New York: Wiley).

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

Armour, K. J., O'Toole, L. B., and Hazon, N. (1993). The effect of dietary protein restriction on the secretory dynamics of 1 alpha-hydroxycorticosterone and urea in the dogfish, Scyliorhinus canicula: a possible role for 1 alpha-hydroxycorticosterone in sodium retention. J Endocrinol *138*, 275-282.

Arnold, F. H., Wintrode, P. L., Miyazaki, K., and Gershenson, A. (2001). How enzymes adapt: lessons from directed evolution. Trends Biochem Sci *26*, 100-106.

Bailey, J. A., Liu, G., and Eichler, E. E. (2003). An Alu transposition model for the origin and expansion of human segmental duplications. Am J Hum Genet *73*, 823-834.

Baker, M. E., Chandsawangbhuwana, C., and Ollikainen, N. (2007). Structural analysis of the evolution of steroid specificity in the mineralocorticoid and glucocorticoid receptors. BMC Evol Biol *7*, 24.

Barabasi, A. L., and Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. Nat Rev Genet *5*, 101-113.

Beato, M. (1989). Gene regulation by steroid hormones. Cell *56*, 335-344.

Beato, M., Herrlich, P., and Schutz, G. (1995). Steroid hormone receptors: many actors in search of a plot. Cell *83*, 851-857.

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

Bergthorsson, U., Andersson, D. I., and Roth, J. R. (2007). Ohno's dilemma: evolution of new genes under continuous selection. P Natl Acad Sci Usa *104*, 17004-17009.

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

Bershtein, S., Goldin, K., and Tawfik, D. S. (2008). Intense neutral drifts yield robust and evolvable consensus proteins. Journal of Molecular Biology *379*, 1029-1044.

Bershtein, S., and Tawfik, D. S. (2008a). Advances in laboratory evolution of enzymes. Current opinion in chemical biology *12*, 151-158.

Bershtein, S., and Tawfik, D. S. (2008b). Ohno's model revisited: measuring the frequency of potentially adaptive mutations under various mutational drifts. Mol Biol Evol *25*, 2311-2318.

Bledsoe, R. K., Madauss, K. P., Holt, J. A., Apolito, C. J., Lambert, M. H., Pearce, K. H., Stanley, T. B., Stewart, E. L., Trump, R. P., Willson, T. M., and Williams, S. P. (2005). A ligand-mediated hydrogen bond network required for the activation of the mineralocorticoid receptor. J Biol Chem *280*, 31283-31293.

Bledsoe, R. K., Montana, V. G., Stanley, T. B., Delves, C. J., Apolito, C. J., McKee, D. D., Consler, T. G., Parks, D. J., Stewart, E. L., Willson, T. M., Lambert, M. H., Moore, J. T., Pearce, K. H., and Xu, H. E. (2002). Crystal structure of the glucocorticoid receptor ligand binding domain reveals a novel mode of receptor dimerization and coactivator recognition. Cell *110*, 93-105.

Bloom, J. D., and Arnold, F. (2009). In the Light of Evolution III: Two Centuries of Darwin Sackler Colloquium: In the light of directed evolution: Pathways of adaptive protein evolution. Proc Natl Acad Sci U S A *106*, 9995-10000 .

Bloom, J. D., Arnold, F. H., and Wilke, C. (2007). Breaking proteins with mutations: threads and thresholds in evolution. Mol Syst Biol *3*, 1-2.

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

Bloom, J. D., Lu, Z., Chen, D., Raval, A., Venturelli, O. S., and Arnold, F. H. (2007a). Evolution favors protein mutational robustness in sufficiently large populations. BMC Biol *5*, 29.

Bloom, J. D., Romero, P. A., Lu, Z., and Arnold, F. H. (2007b). Neutral genetic drift can alter promiscuous protein functions, potentially aiding functional evolution. Biol Direct *2*, 17.

Bloom, J. D., Silberg, J. J., Wilke, C., Drummond, D. A., Adami, C., and Arnold, F. H. (2005). Thermodynamic prediction of protein neutrality. P Natl Acad Sci Usa *102*, 606-611.

Bloom, J. D., Raval, A., and Wilke, C. O. (2007). Thermodynamics of neutral protein evolution. Genetics *175*, 255.

Blount, Z. D., Borland, C. Z., and Lenski, R. E. (2008). Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli. P Natl Acad Sci Usa *105*, 7899-7906.

Bridgham, J. T., Carroll, S. M., and Thornton, J. W. (2006). Evolution of hormone-receptor complexity by molecular exploitation. Science *312*, 97-101.

Bridgham, J. T., Brown, J. E., RodrÌguez-MarÌ, A., Catchen, J. M., and Thornton, J. W. (2008). Evolution of a new function by degenerative mutation in cephalochordate steroid receptors. PLoS Genetics *4*, e1000191.

Brzozowski, A. M., Pike, A. C., Dauter, Z., Hubbard, R. E., Bonn, T., Engstrom, O., Ohman, L., Greene, G. L., Gustafsson, J. A., and Carlquist, M. (1997). Molecular basis of agonism and antagonism in the oestrogen receptor. Nature *389*, 753-758.

Bury, N. R., and Sturm, A. (2007). Evolution of the corticosteroid receptor signalling pathway in fish. Gen Comp Endocrinol *153*, 47-56.

Bury, N. R., Sturm, A., Le Rouzic, P., Lethimonier, C., Ducouret, B., Guiguen, Y., Robinson-Rechavi, M., Laudet, V., Rafestin-Oblin, M. E., and Prunet, P. (2003). Evidence for two distinct functional glucocorticoid receptors in teleost fish. J Mol Endocrinol *31*, 141-156.

Cai, S. Y., Xiong, L., Wray, C. G., Ballatori, N., and Boyer, J. L. (2007). The farnesoid X receptor FXRalpha/NR1H4 acquired ligand specificity for bile salts late in vertebrate evolution. Am J Physiol Regul Integr Comp Physiol *293*, R1400-9.

Cardoso, J. C., Vieira, F. A., Gomes, A. S., and Power, D. M. (2007). PACAP, VIP and their receptors in the metazoa: insights about the origin and evolution of the ligand-receptor pair. Peptides *28*, 1902-1919.

Carroll, S. M., Bridgham, J. T., and Thornton, J. W. (2008). Evolution of hormone signaling in elasmobranchs by exploitation of promiscuous receptors. Mol Biol Evol *25*, 2643-2652.

Castro, L. F., Furlong, R. F., and Holland, P. W. (2004). An antecedent of the MHC-linked genomic region in amphioxus. Immunogenetics *55*, 782-784.

Castro, L. F., Santos, M. M., and Reis-Henriques, M. A. (2005). The genomic environment around the Aromatase gene: evolutionary insights. BMC Evol Biol *5*, 43.

Catchen, J. M., Conery, J. S., and Postlethwait, J. H. (2009). Automated identification of conserved synteny after whole-genome duplication. Genome Res *19*, 1497-1505.

Claude, J. B., Suhre, K., Notredame, C., Claverie, J. M., and Abergel, C. (2004). CaspR: a web server for automated molecular replacement using homology modelling. Nucleic Acids Res *32*, W606-9.

Conant, G., and Wolfe, K. (2008). Turning a hobby into a job: How duplicated genes find new functions. Nat Rev Genet *9*, 938-950.

Conant, G. C., and Wolfe, K. H. (2006). Functional partitioning of yeast co-expression networks after genome duplication. PLoS Biol *4*, e109.

Cooper, G. M., Nickerson, D. A., and Eichler, E. E. (2007). Mutational and selective effects on copy-number variants in the human genome. Nat Genet *39*, S22-9.

Copley, S. D. (2003). Enzymes with extra talents: moonlighting functions and catalytic promiscuity. Curr Opin Chem Biol *7*, 265-272.

Cork, J. M., and Purugganan, M. D. (2004). The evolution of molecular genetic pathways and networks. Bioessays *26*, 479-484.

Counago, R., Chen, S., and Shamoo, Y. (2006). In vivo molecular evolution reveals biophysical origins of organismal fitness. Mol Cell *22*, 441-449.

D'Aniello, A., Di Cosmo, A., Di Cristo, C., Assisi, L., Botte, V., and Di Fiore, M. M. (1996). Occurrence of sex steroid hormones and their binding proteins in Octopus vulgaris lam. Biochem Biophys Res Commun *227*, 782-788.

D'Ari, R., and Casadesús, J. (1998). Underground metabolism. Bioessays *20*, 181-186.

Darwin, C. (1964). On the Origin of Species (Cambridge: Harvard University Press).

Davail, S., Feller, G., Narinx, E., and Gerday, C. (1994). Cold adaptation of proteins. Purification, characterization, and sequence of the heat-labile subtilisin from the antarctic psychrophile Bacillus TA41. Journal of Biological Chemistry *269*, 17448-17453.

de Castro, M., Elliot, S., Kino, T., Bamberger, C., Karl, M., Webster, E., and Chrousos, G. P. (1996). The non-ligand binding beta-isoform of the human glucocorticoid receptor (hGR beta): tissue levels, mechanism of action, and potential physiologic role. Mol Med *2*, 597-607.

Dean, A. M., and Thornton, J. W. (2007). Mechanistic approaches to the study of evolution: the functional synthesis. Nat Rev Genet *8*, 675-688.

Demuth, J. P., and Hahn, M. W. (2009). The life and death of gene families. Bioessays *31*, 29-39.

Depristo, M., Weinreich, D., and Hartl, D. L. (2005). Missense meanderings in sequence space: a biophysical view of protein evolution. Nat Rev Genet *6*, 678-687.

Di Cosmo, A., Di Cristo, C., and Paolucci, M. (2001). Sex steroid hormone fluctuations and morphological changes of the reproductive system of the female of Octopus vulgaris throughout the annual cycle. J Exp Zool *289*, 33-47.

Di Cosmo, A., Di Cristo, C., and Paolucci, M. (2002). A estradiol-17beta receptor in the reproductive system of the female of Octopus vulgaris: characterization and immunolocalization. Mol Reprod Dev *61*, 367-375.

Dickerson, R. E. (2005). Present At the Flood : How Structural Molecular Biology Came About (Sunderland, Mass: Sinauer Associates).

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res *32*, 1792-1797.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr *60*, 2126-2132.

Escriva, H., Bertrand, S., Germain, P., Robinson-Rechavi, M., Umbhauer, M., Cartry, J., Duffraisse, M., Holland, L., Gronemeyer, H., and Laudet, V. (2006). Neofunctionalization in vertebrates: the example of retinoic acid receptors. PLoS Genet *2*, e102.

Fasan, R., Meharenna, Y. T., Snow, C. D., Poulos, T. L., and Arnold, F. H. (2008). Evolutionary history of a specialized p450 propane monooxygenase. Journal of Molecular Biology *383*, 1069-1080.

Feng, W., Ribeiro, R. C., Wagner, R. L., Nguyen, H., Apriletti, J. W., Fletterick, R. J., Baxter, J. D., Kushner, P. J., and West, B. L. (1998). Hormone-dependent coactivator binding to a hydrophobic cleft on nuclear receptors. Science *280*, 1747-1749.

Fiston-Lavier, A. S., Anxolabehere, D., and Quesneville, H. (2007). A model of segmental duplication formation in Drosophila melanogaster. Genome Res *17*, 1458-1470.

Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. Genetics *151*, 1531-1545.

Garcia-Alonso, J., Hoeger, U., and Rebscher, N. (2006). Regulation of vitellogenesis in Nereis virens (Annelida: Polychaeta): effect of estradiol-17beta on eleocytes. Comp Biochem Physiol A Mol Integr Physiol *143*, 55-61.

Gaucher, E. A., Govindarajan, S., and Ganesh, O. K. (2008). Palaeotemperature trend for Precambrian life inferred from resurrected proteins. Nature *451*, 704-707.

Gaucher, E. A., Thomson, J. M., Burgan, M. F., and Benner, S. A. (2003). Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins. Nature *425*, 285-288.

Gelsleichter, J., and Musick, J. A. (1999). Effects of insulin-like growth factor-I, corticosterone, and 3,3', 5-tri-iodo-L-thyronine on glycosaminoglycan synthesis in vertebral cartilage of the clearnose skate, Raja eglanteria. J Exp Zool *284*, 549-556.

Golding, G. B., and Dean, A. M. (1998). The structural basis of molecular adaptation. Mol Biol Evol *15*, 355-369.

Gould, S. J. (1989). Wonderful life: The burgess shale and the nature of life (New York: Norton)

Gray, Y. H. (2000). It takes two transposons to tango: transposable-element-mediated chromosomal rearrangements. Trends Genet *16*, 461-468.

Greenwood, A. K., Butler, P. C., White, R. B., DeMarco, U., Pearce, D., and Fernald, R. D. (2003). Multiple corticosteroid receptors in a teleost fish: distinct sequences, expression patterns, and transcriptional activities. Endocrinology *144*, 4226-4236.

Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol *52*, 696-704.

Guo, H. H., Choe, J., and Loeb, L. A. (2004). Protein tolerance to random amino acid change. Proc Natl Acad Sci U S A *101*, 9205-9210.

Gupta, R., and Tawfik, D. S. (2008). Directed enzyme evolution via small and effective neutral drift libraries. Nat Meth *5*, 939-942.

Hahn, M. W. (2009). Distinguishing Among Evolutionary Models for the Maintenance of Gene Duplicates. J Hered *100*, 605-617.

Hannay, K., Marcotte, E. M., and Vogel, C. (2008). Buffering by gene duplicates: an analysis of molecular correlates and evolutionary conservation. BMC Genomics *9*, 609.

Harvey, B. J., Alzamora, R., Stubbs, A. K., Irnaten, M., McEneaney, V., and Thomas, W. (2008). Rapid responses to aldosterone in the kidney and colon. J Steroid Biochem Mol Biol *108*, 310-317.

Hastings, P. J., Lupski, J. R., Rosenberg, S. M., and Ira, G. (2009). Mechanisms of change in gene copy number. Nat Rev Genet *10*, 551-564.

Hatano, O., Takakusu, A., Nomura, M., and Morohashi, K. (1996). Identical origin of adrenal cortex and gonad revealed by expression profiles of Ad4BP/SF-1. Genes Cells *1*, 663-671.

Hazon, N., and Henderson, I. W. (1984). Secretory dynamics of 1 alpha-hydroxycorticosterone in the elasmobranch fish, Scyliorhinus canicula. J Endocrinol *103*, 205-211.

Hazon, N., and Henderson, I. W. (1985). Factors affecting the secretory dynamics of 1 alpha-hydroxycorticosterone in the dogfish, Scyliorhinus canicula. Gen Comp Endocrinol *59*, 50-55.

Hendrickson, H. (2009). Order and disorder during Escherichia coli divergence. PLoS Genet *5*, e1000335.

Higashi, Y., Yoshioka, H., Yamane, M., Gotoh, O., and Fujii-Kuriyama, Y. (1986). Complete nucleotide sequence of two steroid 21-hydroxylase genes tandemly arranged in human chromosome: a pseudogene and a genuine gene. Proc Natl Acad Sci U S A *83*, 2841-2845.

Hittinger, C. T., and Carroll, S. B. (2007). Gene duplication and the adaptive evolution of a classic genetic switch. Nature *449*, 677-681.

Holland, P. W., Garcia-Fernandez, J., Williams, N. A., and Sidow, A. (1994). Gene duplications and the origins of vertebrate development. Dev Suppl 125-133.

Honn, K. V., and Chavin, W. (1978). In vitro tropic action of ACTH upon adrenocortical cyclic nucleotide (cAMP, cGMP) production and corticosterone output in blue (Prionace glauca Linnaeus) and in mako (Isurus oxyrinchus Rafinesque) sharks. Gen Comp Endocrinol *36*, 161-169.

Huang, W., and Palzkill, T. (1997). A natural polymorphism in -lactamase is a global suppressor. Proceedings of the National Academy of Sciences *94(16)*, 8801-8806.

Hughes, A. L. (1994). The evolution of functionally novel proteins after gene duplication. Proc Biol Sci *256*, 119-124.

Hughes, A. L. (1999). Phylogenies of developmentally important proteins do not support the hypothesis of two rounds of genome duplication early in vertebrate history. J Mol Evol *48*, 565-576.

Idler, D. R., Freeman, H. C., and Truscott, B. (1967). Biological activity and protein-binding of 1 alpha-hydroxycorticosterone: an interrenal steroid in elasmobranch fish. Gen Comp Endocrinol *9*, 207-213.

Idler, D. R., and Truscott, B. (1966). 1-alpha-hydroxycorticosterone from cartilaginous fish: a new adrenal steroid in blood. Journal of the Fisheries Research Board of Canada *23*, 615-619.

Idler, D. R., and Truscott, B. (1967). 1-alpha-hydroxycorticosterone: synthesis in vitro and properties of an interrenal steroid in the blood of cartilaginous fish (genus Raja). Steroids *9*, 457-477.

Ingram, V. M. (1961). Gene evolution and the haemoglobins. Nature *189*, 704-708.

James, L. C., and Tawfik, D. S. (2001). Catalytic and binding poly-reactivities shared by two unrelated proteins: The potential role of promiscuity in enzyme evolution. Protein Sci *10*, 2600-2607.

Jensen, R. A. (1976). Enzyme recruitment in evolution of new function. Annu Rev Microbiol *30*, 409-425.

Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci *8*, 275-282.

Kauppi, B., Jakob, C., Farnegardh, M., Yang, J., Ahola, H., Alarcon, M., Calles, K., Engstrom, O., Harlan, J., Muchmore, S., Ramqvist, A. K., Thorell, S., Ohman, L., Greer, J., Gustafsson, J. A., Carlstedt-Duke, J., and Carlquist, M. (2003). The three-dimensional structures of antagonistic and agonistic forms of the glucocorticoid receptor ligand-binding domain: RU-486 induces a transconformation that leads to active antagonism. J Biol Chem *278*, 22748-22754.

Kazazian, H. H. J. (2004). Mobile elements: drivers of genome evolution. Science *303*, 1626-1632.

Keay, J., Bridgham, J. T., and Thornton, J. W. (2006). The Octopus vulgaris estrogen receptor is a constitutive transcriptional activator: evolutionary and functional implications. Endocrinology *147*, 3861-3869.

Keay, J., and Thornton, J. W. (2009). Hormone-activated estrogen receptors in annelid invertebrates: implications for evolution and endocrine disruption. Endocrinology *150*, 1731-1738.

Keightley, M. C., Curtis, A. J., Chu, S., and Fuller, P. J. (1998). Structural determinants of cortisol resistance in the guinea pig glucocorticoid receptor. Endocrinology *139*, 2479-2485.

Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Phillips, D. C., and Shore, V. C. (1960). Structure of myoglobin: A three-dimensional Fourier synthesis at 2 Å resolution. Nature *185*, 422-427.

Khersonsky, O., Roodveldt, C., and Tawfik, D. S. (2006). Enzyme promiscuity: evolutionary and mechanistic aspects. Curr Opin Chem Biol *10*, 498-508.

Kime, D. E. (1977). Measurement of 1alpha-hydroxycorticosterone and other corticosteroids in elasmobranch plasma by radioimmunoassay. Gen Comp Endocrinol *33*, 344-351.

Kimura, M. (1983). The neutral theory of molecular evolution (Cambridge, UK: Cambridge University Press).

Kimura, M. (1968). Evolutionary rate at the molecular level. Nature *217*, 624-626.

Kimura, M., and Ota, T. (1974). On some principles governing molecular evolution. Proc Natl Acad Sci U S A *71*, 2848-2852.

Klitz, W., Thomson, G., and Baur, M. P. (1986). Contrasting evolutionary histories among tightly linked HLA loci. Am J Hum Genet *39*, 340-349.

Krasowski, M. D., Yasuda, K., Hagey, L. R., and Schuetz, E. G. (2005). Evolution of the pregnane x receptor: adaptation to cross-species differences in biliary bile salts. Mol Endocrinol *19*, 1720-1739.

Kumar, R., and Thompson, E. B. (1999). The structure of the nuclear hormone receptors. Steroids *64*, 310-319.

Kumar, R., and Thompson, E. B. (2005). Gene regulation by the glucocorticoid receptor: structure:function relationship. J Steroid Biochem Mol Biol *94*, 383-394.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. Bioinformatics *23*, 2947-2948.

Le Menuet, D., Viengchareun, S., Penfornis, P., Walker, F., Zennaro, M. C., and Lombes, M. (2000). Targeted oncogenesis reveals a distinct tissue-specific utilization of alternative promoters of the human mineralocorticoid receptor gene in transgenic mice. J Biol Chem *275*, 7878-7886.

Lenormand, T., Roze, D., and Rousset, F. (2009). Stochasticity in evolution. Trends Ecol Evol (Amst) *24*, 157-165.

Li, Y., Suino, K., Daugherty, J., and Xu, H. E. (2005). Structural and biochemical mechanisms for the specificity of hormone binding and coactivator assembly by mineralocorticoid receptor. Mol Cell *19*, 367-380.

Lynch, M., and Conery, J. S. (2003). The origins of genome complexity. Science *302*, 1401-1404.

Lynch, M., and Katju, V. (2004). The altered evolutionary trajectories of gene duplicates. Trends Genet *20*, 544-549.

Makalowski, W., Mitchell, G. A., and Labuda, D. (1994). Alu sequences in the coding regions of mRNA: a source of protein variability. Trends Genet *10*, 188-193.

Manire, C. A., Rasmussen, L. E. L., Maruska, K. P., and Tricas, T. C. (2007). Sex, seasonal, and stress-related variations in elasmobranch corticosterone concentrations. Comp Biochem Phys A *148*, 926-935.

Markov, G. V., Paris, M., Bertrand, S., and Laudet, V. (2008). The evolution of the ligand/receptor couple: a long road from comparative endocrinology to comparative genomics. Mol Cell Endocrinol *293*, 5-16.

Markov, G. V., Tavares, R., Dauphin-Villemant, C., Demeneix, B. A., Baker, M. E., and Laudet, V. (2009). Independent elaboration of steroid hormone signaling pathways in metazoans. Proc Natl Acad Sci U S A *106*, 11913-11918.

Matsumura, I., and Ellington, A. D. (2001). In vitro evolution of beta-glucuronidase into a beta-galactosidase proceeds through non-specific intermediates. J Mol Biol *305*, 331-339.

Matthews, B. W. (1987). Genetic and structural analysis of the protein stability problem. Biochemistry *26*, 6885-6888.

Matthews, B. W. (1993). Structural and genetic analysis of protein stability. Annu Rev Biochem *62*, 139-160.

McGregor, A. P., Orgogozo, V., Delon, I., Zanet, J., Srinivasan, D. G., Payre, F., and Stern, D. L. (2007). Morphological evolution through multiple cis-regulatory mutations at a single gene. Nature *448*, 587-590.

McLoughlin, S. Y., and Copley, S. D. (2008). A compromise required by gene sharing enables survival: Implications for evolution of new enzyme activities. P Natl Acad Sci Usa *105*, 13497-13502.

Miller, S. P., Lunzer, M., and Dean, A. M. (2006). Direct demonstration of an adaptive constraint. Science *314*, 458-461.

Mitraki, A., Fane, B., Haase-Pettingell, C., Sturtevant, J., and King, J. (1991). Global suppression of protein folding defects and inclusion body formation. Science *253*, 54-58.

Mizuta, T., Asahina, K., Suzuki, M., and Kubokawa, K. (2008). In vitro conversion of sex steroids and expression of sex steroidogenic enzyme genes in amphioxus ovary. J Exp Zool Part A Ecol Genet Physiol *309*, 83-93.

Mizuta, T., and Kubokawa, K. (2007). Presence of sex steroids and cytochrome P450 (CYP) genes in amphioxus. Endocrinology *148*, 3554-3565.

Moran, J. V., DeBerardinis, R. J., and Kazazian, H. H. J. (1999). Exon shuffling by L1 retrotransposition. Science *283*, 1530-1534.

Nagy, L., and Schwabe, J. W. (2004). Mechanism of the nuclear receptor molecular switch. Trends Biochem Sci *29*, 317-324.

Nekrutenko, A., and Li, W. H. (2001). Transposable elements are found in a large number of human protein-coding genes. Trends Genet *17*, 619-621.

Nelson, D. R. (1998). Metazoan cytochrome P450 evolution. Comp Biochem Physiol C Pharmacol Toxicol Endocrinol *121*, 15-22.

Nelson, D. R. (1999). Cytochrome P450 and the individuality of species. Arch Biochem Biophys *369*, 1-10.

Nelson, P. (2003). Biological physics: Energy, information, life (New York: W.H. Freeman).

Nettles, K. W., Bruning, J. B., Gil, G., O'Neill, E. E., Nowak, J., Guo, Y., Kim, Y., DeSombre, E. R., Dilis, R., Hanson, R. N., Joachimiak, A., and Greene, G. L. (2007). Structural plasticity in the oestrogen receptor ligand-binding domain. EMBO Rep *8*, 563-568.

Newcomb, R. D., Campbell, P. M., Ollis, D. L., Cheah, E., Russell, R. J., and Oakeshott, J. G. (1997). A single amino acid substitution converts a carboxylesterase to an organophosphorus hydrolase and confers insecticide resistance on a blowfly. Proc Natl Acad Sci USA *94*, 7464-7468.

Nunez, B. S., Evans, A. N., Simpson, M. A., Wong, W. P., and Ip, Y. K. (2006). Characterization of cDNAs encoding cholesterol side chain cleavage and 3beta-hydroxysteroid dehydrogenase in the freshwater stingray Potamotrygon motoro. Comp Biochem Physiol B Biochem Mol Biol *145*, 306-317.

Nunez, B. S., and Vedeckis, W. V. (2002). Characterization of promoter 1B in the human glucocorticoid receptor gene. Mol Cell Endocrinol *189*, 191-199.

Nunez, S., and Trant, J. M. (1999). Regulation of interrenal gland steroidogenesis in the Atlantic stingray (Dasyatis sabina). J Exp Zool *284*, 517-525.

O'Brien, P. J., and Herschlag, D. (1999). Catalytic promiscuity and the evolution of new enzymatic activities. Chemistry & Biology *6*, R91-R105.

O'Toole, L. B., Armour, K. J., Decourt, C., Hazon, N., Lahlou, B., and Henderson, I. W. (1990). Secretory patterns of 1 alpha-hydroxycorticosterone in the isolated perifused interrenal gland of the dogfish, Scyliorhinus canicula. J Mol Endocrinol *5*, 55-60.

Ohno, S. (1970). Evolution By Gene Duplication (Berlin, New York: Springer-Verlag).

Ono, H., Iwasaki, M., Sakamoto, N., and Mizuno, S. (1988). cDNA cloning and sequence analysis of a chicken gene expressed during the gonadal development and homologous to mammalian cytochrome P-450c17. Gene *66*, 77-85.

Ortlund, E. A., Bridgham, J. T., Redinbo, M. R., and Thornton, J. W. (2007). Crystal structure of an ancient protein: Evolution by conformational epistasis. Science *317*, 1544-1548.

Ozols, J., Heinemann, F. S., and Johnson, E. F. (1981). Amino acid sequence of an analogous peptide from two forms of cytochrome P-450. J Biol Chem *256*, 11405-11408.

Panopoulou, G., and Poustka, A. J. (2005). Timing and mechanism of ancient vertebrate genome duplications -- the adventure of a hypothesis. Trends Genet *21*, 559-567.

Paques, F., and Haber, J. E. (1999). Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae. Microbiol Mol Biol Rev *63*, 349-404.

Paris, M., Pettersson, K., Schubert, M., Bertrand, S., Pongratz, I., Escriva, H., and Laudet, V. (2008). An amphioxus orthologue of the estrogen receptor that does not bind estradiol: insights into estrogen receptor evolution. BMC Evol Biol *8*, 219.

Patent, G. J. (1970). Comparison of some hormonal effects on carbohydrate metabolism in an elasmobranch (*Squalus acanthias*) and a holocephalan (*Hydrolagus colliei*). Gen Comp Endocrinol *14*, 215-242.

Patrick, W. M., and Matsumura, I. (2008). A study in molecular contingency: glutamine phosphoribosylpyrophosphate amidotransferase is a promiscuous and evolvable phosphoribosylanthranilate isomerase. Journal of Molecular Biology *377*, 323-336.

Patrick, W. M., Quandt, E. M., Swartzlander, D. B., and Matsumura, I. (2007). Multicopy suppression underpins metabolic evolvability. Mol Biol Evol *24*, 2716-2722.

Pauling, L., and Zuckerkandl, E. (1963). Chemical paleogenetics: molecular "restoration studies" of extinct forms of life. Acta Chem Scand *17*, S9-S16.

Peisajovich, S. G., and Tawfik, D. S. (2007). Protein engineers turned evolutionists. Nat Meth *4*, 991-994.

Pereira de Jesus-Tran, K., Cote, P. L., Cantin, L., Blanchet, J., Labrie, F., and Breton, R. (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.

Perutz, M. F., ROSSMANN, M. G., CULLIS, A. F., MUIRHEAD, H., WILL, G., and NORTH, A. C. (1960). Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-A. resolution, obtained by X-ray analysis. Nature *185*, 416-422.

Perutz, M. F. (1983). Species adaptation in a protein molecule. Mol Biol Evol *1*, 1-28.

Petrosino, J., Cantu, C. 3rd, and Palzkill, T. (1998). beta-Lactamases: protein evolution in real time. Trends Microbiol *6*, 323-327.

Piatigorsky, J., and Wistow, G. (1991). The recruitment of crystallins: new functions precede gene duplication. Science *252*, 1078-1079.

Pond, S. L., Frost, S. D., and Muse, S. V. (2005). HyPhy: hypothesis testing using phylogenies. Bioinformatics *21*, 676-679.

Posada, D., and Buckley, T. R. (2004). Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. Syst Biol *53*, 793-808.

Postlethwait, J., Amores, A., Cresko, W., Singer, A., and Yan, Y. L. (2004). Subfunction partitioning, the teleost radiation and the annotation of the human genome. Trends Genet *20*, 481-490.

Powers, D. A. (1972). Hemoglobin adaptation for fast and slow water habitats in sympatric catostomid fishes. Science *177*, 360-362.

Prossnitz, E. R., Arterburn, J. B., and Sklar, L. A. (2007). GPR30: A G protein-coupled receptor for estrogen. Mol Cell Endocrinol *265-266*, 138-142.

Prud'homme, B., Gompel, N., Rokas, A., Kassner, V. A., Williams, T. M., Yeh, S. D., True, J. R., and Carroll, S. B. (2006). Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. Nature *440*, 1050-1053.

Prunet, P., Sturm, A., and Milla, S. (2006). Multiple corticosteroid receptors in fish: from old ideas to new concepts. Gen Comp Endocrinol *147*, 17-23.

Puigbò, P., Pasamontes, A., and Garcia-Vallve, S. (2008). Gaining and losing the thermophilic adaptation in prokaryotes. Trends Genet *24*, 10-14.

Putnam, N. H., Butts, T., Ferrier, D. E., Furlong, R. F., Hellsten, U., Kawashima, T., Robinson-Rechavi, M., Shoguchi, E., Terry, A., Yu, J. K., Benito-Gutierrez, E. L., Dubchak, I., Garcia-Fernandez, J., Gibson-Brown, J. J., Grigoriev, I. V., Horton, A. C., de Jong, P. J., Jurka, J., Kapitonov, V. V., Kohara, Y., Kuroki, Y., Lindquist, E., Lucas, S., Osoegawa, K., Pennacchio, L. A., Salamov, A. A., Satou, Y., Sauka-Spengler, T., Schmutz, J., Shin-I, T., Toyoda, A., Bronner-Fraser, M., Fujiyama, A., Holland, L. Z., Holland, P. W., Satoh, N., and Rokhsar, D. S. (2008). The amphioxus genome and the evolution of the chordate karyotype. Nature *453*, 1064-1071.

Rasmussen, L. E. L., and Crow, G. L. (1993). Serum Corticosterone Concentrations in Immature Captive Whitetip Reef Sharks, Triaenodon Obesus. Journal of Experimental Zoology *267*, 283-287.

Roth, C., Rastogi, S., Arvestad, L., Dittmar, K., Light, S., Ekman, D., and Liberles, D. A. (2007). Evolution after gene duplication: models, mechanisms, sequences, systems, and organisms. J Exp Zool B Mol Dev Evol *308*, 58-73.

Russell, R. J., Gerike, U., Danson, M. J., Hough, D. W., and Taylor, G. L. (1998). Structural adaptations of the cold-active citrate synthase from an Antarctic bacterium. Structure *6*, 351-361.

Salisbury, B. A., and Kim, J. (2001). Ancestral state estimation and taxon sampling density. Syst Biol *50*, 557-564.

Shi, Y., Radlwimmer, F. B., and Yokoyama, S. (2001). Molecular genetics and the evolution of ultraviolet vision in vertebrates. Proc Natl Acad Sci U S A *98*, 11731-11736.

Shortle, D., and Lin, B. (1985). Genetic analysis of staphylococcal nuclease: identification of three intragenic" global" suppressors of nuclease-minus mutations. Genetics *110*, 539-555.

Simon, P. (2003). Q-Gene: processing quantitative real-time RT-PCR data. Bioinformatics *19*, 1439-1440.

Simons, S. S. J. (2006). How much is enough? Modulation of dose-response curve for steroid receptor-regulated gene expression by changing concentrations of transcription factor. Curr Top Med Chem *6*, 271-285.

Singleton, R., and Amelunxen, R. E. (1973). Proteins from thermophilic microorganisms. Microbiology and Molecular Biology Reviews *37*, 320-342.

Snelson, F. F., Jr., Rasmussen, L. E., Johnson, M. R., and Hess, D. L. (1997). Serum concentrations of steroid hormones during reproduction in the Atlantic stingray, Dasyatis sabina. Gen Comp Endocrinol *108*, 67-79.

Soltis, D. E., and Soltis, P. S. (1999). Polyploidy: recurrent formation and genome evolution. Trends Ecol Evol *14*, 348-352.

Sturm, A., Bury, N., Dengreville, L., Fagart, J., Flouriot, G., Rafestin-Oblin, M. E., and Prunet, P. (2005). 11-deoxycorticosterone is a potent agonist of the rainbow trout (Oncorhynchus mykiss) mineralocorticoid receptor. Endocrinology *146*, 47-55.

Suino-Powell, K., Xu, Y., Zhang, C., Tao, Y. G., Tolbert, W. D., Simons, S. S., Jr., and Xu, H. E. (2008). Doubling the size of the glucocorticoid receptor ligand binding pocket by deacylcortivazol. Mol Cell Biol *28*, 1915-1923.

Takahata, N., Satta, Y., and Klein, J. (1992). Polymorphism and balancing selection at major histocompatibility complex loci. Genetics *130*, 925-938.

Taverna, D. M., and Goldstein, R. A. (2002). Why are proteins marginally stable? Proteins *46*, 105-109.

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

Thomas, P., Pang, Y., Filardo, E. J., and Dong, J. (2005). Identity of an estrogen membrane receptor coupled to a G protein in human breast cancer cells. Endocrinology *146*, 624-632.

Thomas, P., Pang, Y., Zhu, Y., Detweiler, C., and Doughty, K. (2004). Multiple rapid progestin actions and progestin membrane receptor subtypes in fish. Steroids *69*, 567-573.

Thornton, J. W. (2001). Evolution of vertebrate steroid receptors from an ancestral estrogen receptor by ligand exploitation and serial genome expansions. Proc Natl Acad Sci U S A *98*, 5671-5676.

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

Thornton, J. W., Need, E., and Crews, D. (2003). Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling. Science *301*, 1714-1717.

Tokuriki, N., and Tawfik, D. S. (2009). Chaperonin overexpression promotes genetic variation and enzyme evolution. Nature *459*, 668-673.

Tokuriki, N., Stricher, F., Schymkowitz, J., Serrano, L., and Tawfik, D. S. (2007). The stability effects of protein mutations appear to be universally distributed. Journal of Molecular Biology *369*, 1318-1332.

Tokuriki, N., Stricher, F., Serrano, L., and Tawfik, D. S. (2008). How protein stability and new functions trade off. PLoS Comput Biol *4*, e1000002.

Tokuriki, N., and Tawfik, D. S. (2009). Protein dynamism and evolvability. Science *324*, 203-207.

Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., Bingen, E., Bonacorsi, S., Bouchier, C., Bouvet, O., Calteau, A., Chiapello, H., Clermont, O., Cruveiller, S., Danchin, A., Diard, M., Dossat, C., Karoui, M., Frapy, E., Garry, L., Ghigo, J., Gilles, A., Johnson, J., Le Bouguénec, C., Lescat, M., Mangenot, S., Martinez-Jéhanne, V., Matic, I., Nassif, X., Oztas, S., Petit, M., Pichon, C., Rouy, Z., Ruf, C., Schneider, D., Tourret, J., Vacherie, B., Vallenet, D., Médigue, C., Rocha, E. P., and Denamur, E. (2009). Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths. PLoS Genet *5*, e1000344.

Tracewell, C. A., and Arnold, F. H. (2009). Directed enzyme evolution: climbing fitness peaks one amino acid at a time. Current opinion in chemical biology *13*, 3-9.

Truscott, B., and Idler, D. R. (1968). The widespread occurrence of a corticosteroid 1-alpha-hydroxylase in the interrenals of elasmobranchii. J Endocrinol *40*, 515-526.

Truscott, B., and Idler, D. R. (1972). Corticosteroids in plasma of elasmobranchs. Comp Biochem Physiol A *42*, 41-50.

Val, P., Jeays-Ward, K., and Swain, A. (2006). Identification of a novel population of adrenal-like cells in the mammalian testis. Dev Biol *299*, 250-256.

Wagner, A. (1994). Evolution of gene networks by gene duplications: a mathematical model and its implications on genome organization. Proc Natl Acad Sci U S A *91*, 4387-4391.

Wagner, G. P., Amemiya, C., and Ruddle, F. (2003). Hox cluster duplications and the opportunity for evolutionary novelties. Proc Natl Acad Sci U S A *100*, 14603-14606.

Walsh, C. J., Wyffels, J. T., Bodine, A. B., and Luer, C. A. (2002). Dexamethasone-induced apoptosis in immune cells from peripheral circulation and lymphomyeloid tissues of juvenile clearnose skates, Raja eglanteria. Dev Comp Immunol *26*, 623-633.

Weinreich, D. M., Delaney, N. F., DePristo, M. A., and Hartl, D. L. (2006). Darwinian evolution can follow only very few mutational paths to fitter proteins. Science *312*, 111-114.

Weisbart, M., and Youson, J. H. (1975). Steroid formation in the larval and parasitic adult sea lamprey, Petromyzon marinus L. Gen Comp Endocrinol *27*, 517-526.

Weisbart, M., and Youson, J. H. (1977). In vivo formation of steroids from [1,2,6,7-3H]-progesterone by the sea lamprey, Petromyzon marinus L. J Steroid Biochem *8*, 1249-1252.

Weisbart, M., Youson, J. H., and Wiebe, J. P. (1978). Biochemical, histochemical, and ultrastructural analyses of presumed steroid-producing tissues in the sexually mature sea lamprey, Petromyzon marinus L. Gen Comp Endocrinol *34*, 26-37.

White, P. C., and Speiser, P. W. (2000). Congenital adrenal hyperplasia due to 21-hydroxylase deficiency. Endocr Rev *21*, 245-291.

White, P. C., New, M. I., and Dupont, B. (1986). Structure of human steroid 21-hydroxylase genes. Proc Natl Acad Sci U S A *83*, 5111-5115.

Wilkins, A. S. (2005). Recasting developmental evolution in terms of genetic pathway and network evolution...and the implications for comparative biology. Brain Res Bull *66*, 495-509.

Wroe, R., Chan, H. S., and Bornberg-Bauer, E. (2007). A structural model of latent evolutionary potentials underlying neutral networks in proteins. HFSP J. *1*, 79-87.

Wu, K. H., Tobias, M. L., Thornton, J. W., and Kelley, D. B. (2003). Estrogen receptors in Xenopus: duplicate genes, splice variants, and tissue-specific expression. Gen Comp Endocrinol *133*, 38-49.

Yokoyama, S., Tada, T., Zhang, H., and 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.

Yokoyama, S., Zhang, H., Radlwimmer, F. B., and Blow, N. S. (1999). Adaptive evolution of color vision of the Comoran coelacanth (Latimeria chalumnae). Proc Natl Acad Sci U S A *96*, 6279-6284.

Yokoyama, S. (2002). Molecular evolution of color vision in vertebrates. Gene *300*, 69-78.

Zhang, J. Z. (2003). Evolution by gene duplication: an update. Trends in Ecology & Evolution *18*, 292-298.

Zhu, G., Golding, G. B., and Dean, A. M. (2005). The selective cause of an ancient adaptation. Science *307*, 1279-1282.

Zuckerkandl, E., and Pauling, L. (1965). Molecules as documents of evolutionary history. J Theor Biol *8*, 357-366.