V

Experimental synthesis of ancestral proteins to test biological hypotheses

CHAPTER 16

Using ancestral gene resurrection to unravel the evolution of protein function

Joseph W. Thornton and Jamie T. Bridgham

16.1 Introduction

In the century and a half since Darwin, the central goal of evolutionary biology has been to provide historical explanations for the diversity of species and their myriad adaptations. The recent advent of molecular biology and genomics presents us with a new kind of biodiversity that is equally astonishing: thousands of genes in every genome, each with specific, exquisitely tuned functions. How this functional biodiversity of genes and proteins evolved is arguably the central question in molecular evolution.

Most work to date on the evolution of gene function has relied on statistical methods to infer process from patterns in present-day sequence data (for overviews, see Li, 1997; Page and Holmes, 1998). Many valuable insights have emerged from this approach, but the hypotheses that have been generated remain for the most part empirically untested. Recently, however, advances in phylogenetics and DNA-synthesis techniques have made it possible to infer the sequences of ancestral genes and then synthesize and express them in the laboratory. As a result, hypotheses about the functions of ancient genes—and the mechanistic basis for their evolution—can now be empirically tested using the reductionist power of experimental molecular biology.

In this chapter, we review our use of ancestral gene resurrection to understand how the members of a biologically crucial gene family, the steroid hormone receptors, evolved their diverse and highly specific functions. We also discuss some methodological questions and concerns—particularly related to uncertainty in the reconstruction of ancestral sequences—and point to potential future directions for the budding field of ancestral gene resurrection.

16.2 The evolution of molecular interactions

Virtually everything that living cells do is regulated by specific molecular interactions, such as those between enzymes and substrates, receptors and ligands, and transcription factors and their DNA-binding sites. Genomic diversity is also largely due to the diversity of molecular interactions: members of most gene families have a core conserved function (such as DNA binding or a specific mode of catalysis) but have diversified by changing their specific binding partners. Despite the biological importance of specific molecular interactions, however, there has been very limited work, theoretical or empirical, to understand the general dynamics by which they evolve (see Fryxell, 1996; Aharoni *et al.*, 2005; Haag and Molla, 2005; Bridgham *et al.*, 2006).

Tightly integrated molecular partnerships also exemplify an important and largely unresolved evolutionary issue: the evolution of complexity. The classic model for the evolution of complex systems is that they result from a gradual process of elaboration and optimization under the influence of selection. This model is well supported for some complex structures, such as metazoan eyes: the presence of eyes of intermediate complexity in

a variety of taxa indicates that more complex eyes evolved gradually (and repeatedly) from a primitive light-sensing organ (Futuyma, 1998). It is not clear, however, how this model can explain the evolution of tightly integrated molecular systems, in which the function of each part depends on its interaction with the other parts. Simultaneous emergence of more than one element by mutational processes is unlikely, so it is not apparent how selection can drive the evolution of any part or the system as a whole. What, for example, is the selection pressure that drives the evolution of a new hormone if there is not already a receptor to transduce its signal? Conversely, what is the function of a new receptor if there is not already a hormone for it to receive?

Darwin was well aware of this puzzle. He wrote in *The Origin of Species,* "If it could be demonstrated that any complex organ existed, which could not possibly have been formed by numerous, successive, slight modifications, my theory would absolutely break down" (Darwin, 1859). He also recognized that for many present-day complex systems, it would be difficult to reconstruct the stepwise process by which they evolved: "In order to discover the early transitional grades through which the organ has passed, we should have to look to very ancient ancestral forms, long since become extinct." This is a particular problem for studying the evolution of molecular complex systems, the ancestral forms of which, unlike those of morphological features, are not preserved as fossils.

The advent of ancestral gene resurrection provides a way to study the ancestral forms of molecules that would otherwise be scientifically inaccessible. We can now resurrect and characterize the functions of ancient genes, including those that participate in specific interactions. Thus we can begin to unravel the events by which tightly integrated molecular complexes emerged by stepwise Darwinian processes.

16.3 Steroid hormones and their receptors

To understand the evolution of molecular complexity, we study a specific model system: the tight functional interactions between steroid hormones and their intracellular receptors. Steroid hormone receptors are ligand-regulated transcription factors. They are activated by contact with specific steroid hormones, such as testosterone, estradiol, progesterone, cortisol, and aldosterone. These hormones are produced in the gonads or adrenal/interrenal glands through a pathway of enzymatic modifications beginning with cholesterol (Figure 16.1). Humans have six steroid receptors (SRs): two for estrogens (ERα and ERβ) and one each for testosterone and other androgens (AR), progestins (PR), glucocorticoids (GR), and mineralocorticoids (MR). The classic effects of steroid hormones include control of secondary sexual differentiation, and reproductive function in females (estrogens and progestins) and males (androgens), response to stress (glucocorticoids), and maintenance of osmotic homeostasis (aldosterone).

Mechanistically, SRs are molecular mediators (Gronemeyer *et al.,* 2004). In the absence of the hormone, SRs are typically in the cytosol. Steroid hormones are hydrophobic, so they cross the cell membrane by diffusion. Each hormone binds with extraordinary affinity and specificity to a receptor. Hormone-binding triggers a change in the receptor's conformation that allows it to dimerize, translocate to the nucleus, and bind tightly to specific response elements in the nucleus; short DNA sequences in the control region of target genes. The receptor then attracts coactivator proteins that modify chromatin, attract elements of the basal transcription complex, or otherwise increase transcription of the target gene (Figure 16.1). SRs have a conserved modular structure, consisting of a highly conserved DNA-binding domain (DBD), which recognizes and binds to response elements, and a moderately conserved ligand-binding domain (LBD), which binds to the hormone (the ligand) and contains the hormone-activated transcriptional activation function. Receptors also contain a poorly conserved flexible hinge region, which orients the DBD and LBD relative to each other, and a non-conserved N-terminal domain, which contains an autonomous transcriptional function. The DBD and LBD are functionally separable from the rest of the sequence, allowing the construction of chimeric proteins that combine

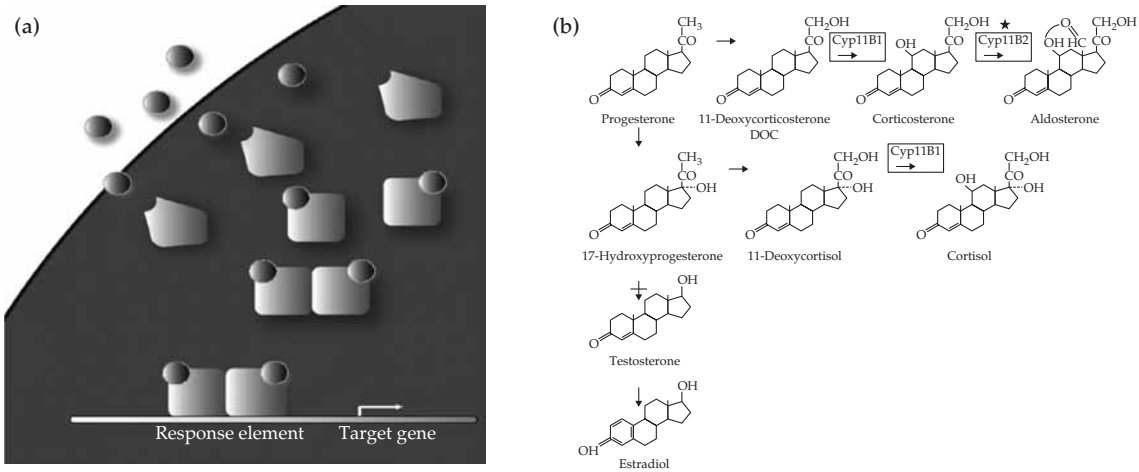


Figure 16.1 Steroid hormone and receptor biology. (a) Simplified mechanism of hormone receptor action. Steroid hormones (dark balls) are hydrophobic, so they cross cell membranes by diffusion. In the cytosol, a hormone molecule binds tightly to a specific receptor, conferring a conformational change that allows the receptor to dimerize, enter the nucleus, and bind to specific response elements in the promoters of target genes. The hormone-bound SR attracts coactivator proteins that increase expression of the target gene. (b) Steroid synthesis pathway. All major steroid hormones are produced in a pathway of enzyme-mediated modifications, beginning with a progesterone precursor. Reactions catalyzed by one of these enzymes, cytochrome P450–11B, are boxed. Only in tetrapods can this protein also catalyze the hydroxylation of corticosterone to produce aldosterone (star).

the functions of one protein's DBD with those of another protein's LBD or activation domain (Green and Chambon, 1987).

The goal of research in our laboratory is to describe the specific mechanisms and dynamics by which new hormone–receptor and receptor–DNA relationships evolved. SRs have several characteristics that make them very suitable for ancestral gene resurrection. First, they form a monophyletic group within a larger superfamily of genes, so phylogenetic methods can be used to reconstruct their proliferation and divergence from a common ancestral protein (Thornton and DeSalle, 2000). Second, there are efficient, well-established molecular assays for determining the intrinsic functions of SRs, which can be used to characterize ancient receptors resurrected in the laboratory. Finally, there is an extensive database on the sequences, structures, and functions of extant receptors, providing a rich context for interpreting reconstructed ancestral sequences.

In this chapter, we review how we have used ancestral gene resurrection to address two evolutionary questions. First, how did the specific interactions of GR and MR with glucocorticoids and aldosterone evolve? Second, how were the PR's and AR's partnerships with progesterone and testosterone established. Our ultimate goal is to determine the evolutionary dynamics and molecular mechanisms by which all of the receptor-specific functions of this important gene family evolved.

16.4 Evolution of corticoid receptor specificity

MR and GR are sister receptors that descend from a gene duplication deep in the vertebrate lineage (Thornton, 2001) and now have distinct signaling functions. GR is specifically activated by the stress hormone cortisol in most vertebrates to regulate metabolism, inflammation, and immunity (Bentley, 1998). MR is activated by aldosterone to control electrolyte homeostasis and other processes (Bentley, 1998; Farman and Rafestin-Oblin, 2001). MR can also be activated by cortisol, although the presence of a cortisol-clearing enzyme in most MR-expressing tissues makes the receptor a largely aldosterone-specific factor (Farman and Rafestin-Oblin, 2001).

Aldosterone has only been detected in tetrapods, so it has long been assumed that the GR, which is

insensitive to aldosterone, retains the ancestral functions, with the MR's affinity for aldosterone being derived (Baker, 2001). We sought to test this hypothesis and determine the mechanistic basis for the evolution of GR/MR specificity by resurrecting the ancestral corticoid receptor (AncCR): the ancient protein from which GR and MR descend by gene duplication. Our work on this ancient gene was first presented in Bridgham *et al.* (2006), which provides further details on methods and results.

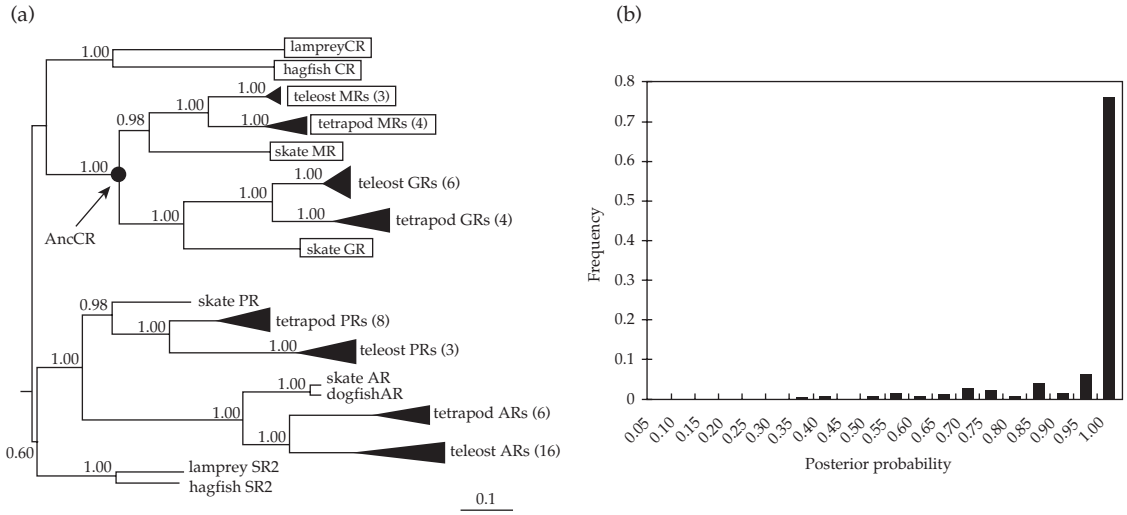
The first requirement for ancestral gene resurrection is an ample data-set of sequences from extant taxa. The accuracy with which an ancestral sequence is inferred depends strongly on the length of the branches that descend from the ancestral node and, to a lesser extent, on the accuracy of the phylogeny itself. If the node is surrounded by long branches on which the majority of phylogenetic information has been erased by subsequent substitutions, accurately reconstructing the ancestral state becomes very difficult (Zhang and Nei, 1997). A large number of SR sequences from tetrapods and teleosts were publicly available, but few sequences were known from agnathans and elasmobranches—the basal lineages that diverged from other vertebrates just before and after the node represented by AncCR. To improve the robustness of the sequence database, we used degenerate PCR and rapid amplification of cDNA ends (RACE) to isolate corticoid receptors from two jawless fishes—the lamprey *Petromyzon marinus* and the hagfish *Myxine glutinosa*—and an elasmobranch, the skate *Raja erinacea*. We recovered a single unduplicated corticoid receptor from the lamprey and hagfish, and clear orthologs of both the GR and MR from the skate (Figure 16.2a).

The second requirement for ancestral sequence reconstruction is a well-corroborated phylogenetic tree. Based on an alignment of 60 broadly sampled SR protein sequences—including the new sequences from basal vertebrates—we used maximum parsimony, Bayesian Markov chain Monte Carlo (BMCMC), and maximum likelihood to determine the phylogeny of corticoid receptors and their outgroups. For BMCMC, we integrated over numerous protein evolutionary models and found that the JTT + gamma model was supported with 100% posterior probability; we therefore used this model for maximum-likelihood analysis. Maximum likelihood, maximum parsimony, and BMCMC all recovered the same phylogeny, increasing confidence in the result (Kolaczkowski and Thornton, 2004; Thornton and Kolaczkowski, 2005). We found that the node that represents AncCR is extremely well supported, with posterior probability and bootstrap confidence measures equal to 1.0; this is particularly important, because ancestral reconstruction is generally robust to errors in the topology except at the node being reconstructed (Zhang and Nei, 1997). The majority of other nodes on the tree were also inferred with high confidence.

The tree (Figure 16.2a) indicates that the duplication of AncCR, which produced separate GR and MR lineages, occurred after the divergence of jawless vertebrates but before the split of cartilaginous from bony fish. AncCR therefore represents the unduplicated corticoid receptor gene, which existed in the genome of the last common ancestor of agnathans and jawed vertebrates, about 450–470 million years ago.

16.5 Resurrecting the AncCR

Using this phylogeny as a scaffold for phylogenetic inference, we next inferred the protein sequence of the AncCR using the maximum-likelihood-based method of (Yang *et al.*, 1995). The analysis assumed the JTT + gamma model of protein evolution, which was strongly supported by our Bayesian analysis of numerous models. The parameter values of the model, such as branch lengths and the shape parameter for among-site rate variation, were estimated by maximum likelihood. We focused on the LBD sequence, because this is the functional domain that confers ligand specificity. The AncCR LBD protein sequence was inferred with high support: the mean posterior probability was 94% per site, and two-thirds of sites had posterior probabilities of more than 99%. A small number of sites were ambiguously reconstructed, however, with alternative states that had non-trivial probability, an issue to which we will return below (Figure 16.2b).



(c)

	LBP	LBD
SkateMR	0.95	0.80
AstatotilapiaMR	0.95	0.78
SkateGR	0.95	0.76
HumanMR	0.95	0.75
HagfishCR	0.86	0.63
AstatotilapiaGRl	0.82	0.70
LampreyCR	0.77	0.66
HumanGR	0.77	0.66

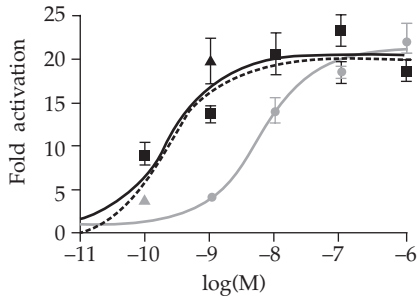


Figure 16.2 Resurrection of the ancestral corticoid receptor (CR). (a) Phylogeny of the steroid receptors. The gene family tree of 59 steroid and related receptor amino acid sequences was inferred using maximum likelihood (ML), Bayesian Markov Chain Monte Carlo (BMCMC), and maximum parsimony. ML branch lengths and BMCMC posterior probabilities for major nodes are shown. Parentheses, number of sequences in each clade. The ancestral corticoid receptor (AncCR) that we reconstructed is marked. Boxes, aldosterone-activated receptors. For details see Bridgham *et al.* 2006. (b) Distribution of the posterior probabilities of the most likely amino acid at each site in the 232-amino acid ligand binding domain of AncCR. (c) Sequence similarity of AncCR to extant corticoid receptors. Receptors that are activated by aldosterone are shown in bold. LBD, entire ligand-binding domain. LBP, ligand-binding pocket, consisting of sites that are known from structural studies to make contact with and coordinate binding of the hormone. (d) The ancestral corticoid receptor is activated by aldosterone. Increase in activation of a luciferase reporter gene by the resurrected AncCR-LBD is shown for increasing doses of aldosterone (black line, squares), cortisol (gray line, circles), and 11-deoxycorticosterone (dashed line, triangles). Fold-activation is relative to activation of the reporter in the absence of hormone.

Preliminary analysis led us to the hypothesis that the AncCR LBD would have MR-like functions, for two reasons. First, its sequence is most similar to the aldosterone-activated receptors: it differs from them by only one residue in the ligand-binding pocket but is considerably less similar to the aldosterone-insensitive GRs (Figure 16.2c). Second, when we functionally characterized the ligand sensitivity of extant receptors, we found that all the receptors from all the basal vertebrates were activated by very low doses of aldosterone, cortisol, and 11-deoxycorticosterone (DOC; Figure 16.3a and b). They are similar in this respect to MRs of tetrapods and teleosts (Hellal-Levy *et al.*, 1999; Greenwood *et al.*, 2003; Sturm *et al.*, 2005). The only receptors we found to be insensitive to aldosterone were the GRs of tetrapods and teleosts. The most parsimonious scenario

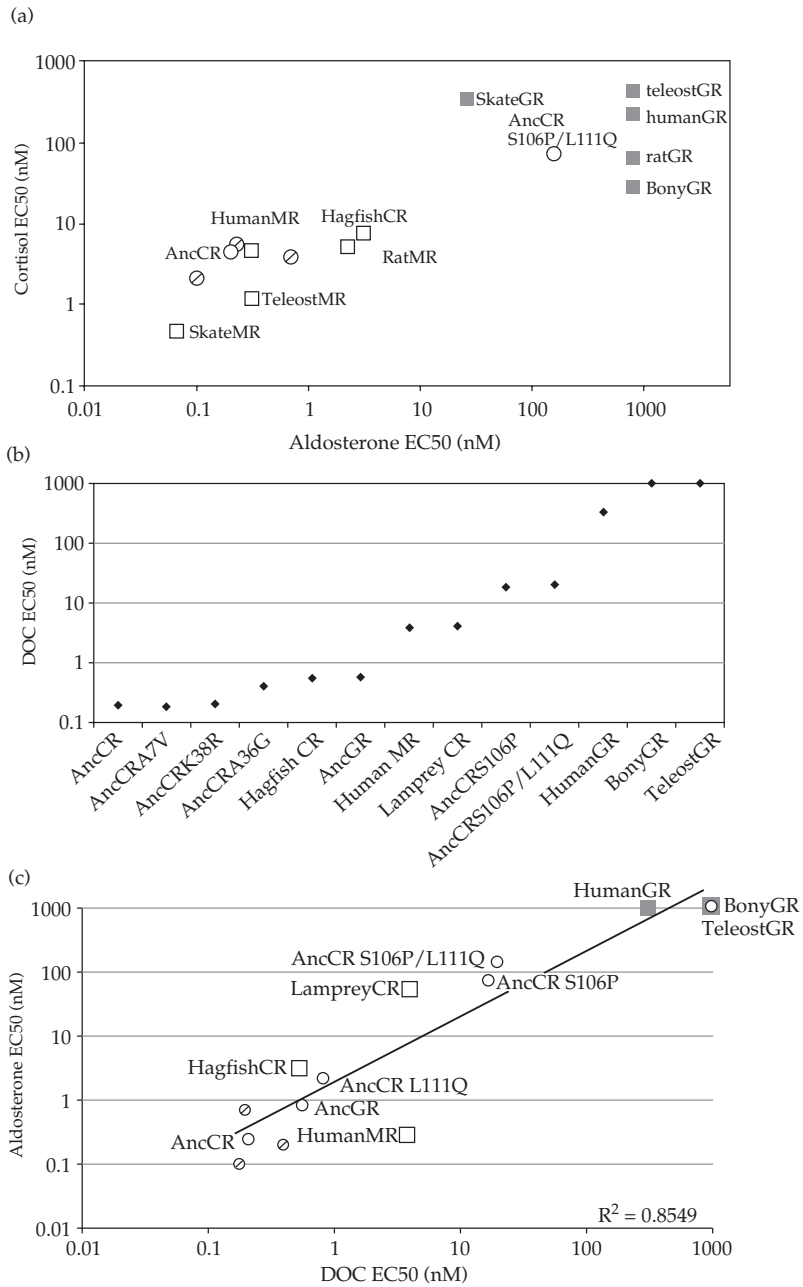


Figure 16.3 Hormone activation of extant and ancestral receptors. (a) Receptor sensitivity to aldosterone and cortisol. Each point represents the sensitivity of one receptor to these two hormones, expressed as the EC50—the concentration (nanomolar, nM) that is required to achieve half-maximal activation of a luciferase reporter gene; a lower EC50 indicates greater sensitivity. White squares are extant MRs and CRs; gray boxes are GRs. White circles are resurrected ancestral receptors; slashed circles are three variant versions that incorporate alternative reconstructions due to phylogenetic uncertainty. The AncCR containing two substitutions that recapitulate the evolution of the GR-like phenotype is also shown. (b) Sensitivity of extant and ancestral receptors to deoxycorticosterone (DOC), the putative ancestral ligand that is present in all vertebrates. (c) Correlation of receptor aldosterone sensitivity with DOC sensitivity. The linear regression shows that variation in a receptor's sensitivity to DOC predicts 85% of variation in aldosterone sensitivity.

a priori is therefore that AncCR was aldosterone-sensitive: the alternative hypothesis would require aldosterone activation to be gained independently in the lineages leading to the agnathan CRs, to the elasmobranch GR, and to the MRs of elasmobranches, teleosts, and tetrapods, all in the absence of the hormone—a most unlikely possibility (Figure 16.2a).

To test this hypothesis, we inferred a cDNA sequence that would code for the AncCR LBD sequence, and optimized it for expression in cultured cells using a standard table of mammalian codon bias. We had this cDNA synthesized commercially, an approach we find practical, accurate, and reasonably priced. We then subcloned the AncCR LBD cDNA into a vector for high-level expression in a fusion protein, and transfected that construct into mammalian Chinese hamster ovary (CHO) K1 cells. CHO-K1 cells are a standard system for characterizing vertebrate receptors for two reasons: they do not express any of their own SRs—so they provide a low-noise cellular background for functional assays—and they contain all the conserved accessory factors for SRs from a wide variety of taxa to function properly.

We used a luciferase reporter-gene assay to determine the AncCR LBD's responsiveness to various corticosteroid hormones. We found that AncCR is a very sensitive and specific aldosterone receptor, activating transcription 20-fold at subnanomolar concentrations. Like the extant CRs and MRs it is also activated by low doses of DOC and, to a lesser extent, cortisol (Figures 16.2d and 16.3). These results corroborated our hypothesis that the ancestral CR had MR like functions and that aldosterone sensitivity is far more ancient than previously assumed.

16.6 Robustness to uncertainty

Ancestral reconstruction using maximum likelihood will converge on the true ancestral states with increasing confidence as the amount of data at each site increases, if the correct tree and evolutionary model are used. As the number of available sequences related to the ancestor increases, uncertainty about the ancestral state declines; the posterior probability of the maximum-likelihood state at every site approaches 1.0, and the probability of error approaches zero.

In reality, however, the number of available sequences that provide phylogenetic information about an ancestor of interest is always finite, so the maximum-likelihood reconstruction of the ancestral sequence will usually be uncertain. That is, at some sites there will be more than one possible state with posterior probability greater than zero. In most cases, the maximum likelihood estimate will be the true ancestral state, but at some sites the ancestral amino acid may be one with a lower likelihood, resulting in erroneous reconstruction.

There are three potential causes of error in ancestral sequence reconstruction. The first is stochastic error. If a 1000-site protein is reconstructed with 0.99 posterior probability at every site, 1% of all sites—or 10 residues overall—in the maximum-likelihood sequence are expected to have the incorrect state. For a site to be reconstructed erroneously, the probability that the state pattern observed in extant sequences would evolve from the true ancestral state must be lower than the probability that the state pattern would evolve from a different (untrue) state. That is, a low-probability set of evolutionary events must have taken place instead of more-likely scenarios; this sort of error will occur, for example, if two sister sequences that descend from their ancestor on short branches share state i, but this is due to two convergent substitutions from ancestral state j rather than the higher-probability scenario of conservation from state i in the ancestor. Low-probability events do occur, albeit with low frequency; over a large number of sites, states with suboptimal likelihoods will be the true state in a few cases. The probability that the maximum-likelihood state is erroneous at a site is, of course, inversely proportional to the posterior probability of that state. Sites that have alternative reconstructions with non-trivial posterior probabilities are therefore the ones at which the true state is most likely to be different from the maximum-likelihood state.

The second and third potential cause of erroneous states arise from the fact that maximum likelihood calculates the probability of ancestral states conditionally on a phylogenetic tree and

evolutionary model. If the tree or the model assumed are incorrect, calculated likelihoods will not accurately reflect the actual likelihoods of the possible ancestral states, and the inferred maximum-likelihood sequence may not be the sequence with the highest true likelihood. Because experimental results concerning the functions of a reconstructed ancestral gene are only as good as the inferred sequence on which they rely, it is important to explore whether the maximum-likelihood sequence and its functions are robust to potential errors in sequence reconstruction induced by these factors.

We characterized the robustness of the AncCR sequence to error in several ways. First, we characterized the robustness of AncCR's aldosterone response to stochastic error. We examined all positions that were ambiguously reconstructed on the maximum-likelihood tree, defined as having an alternate state with a posterior probability of more than 0.20. In all cases but one, the alternate state is found in other aldosterone-activated receptors and is therefore not sufficient to abolish aldosterone sensitivity. We introduced the one exception into AncCR using site-directed mutagenesis; it had no effect on ligand activation (Figure 16.3a). We also used structural information to identify sites likely to be of functional importance. We examined sites that make contact with the ligand in the MR crystal structure (Fagart *et al.*, 2005) and found that only one of these was ambiguously reconstructed. The mutagenized AncCR with the alternate state remained activated by very low aldosterone concentrations (Figure 16.3a). These data indicate that AncCR's aldosterone sensitivity is not likely to be due to stochastic errors in the maximum-likelihood reconstruction of the protein sequence.

Second, we examined the effect of uncertainty about the phylogenetic tree on the AncCR sequence and its inferred function. Although the maximum-likelihood tree was generally well supported, there were a few nodes at which alternative topologies were not ruled out. To address the possibility that different trees might change the inferred ancestral sequence and its function, we used BMCMC to collect a large sample of plausible trees (posterior probability > 0.0002), and the reconstructed the AncCR sequence independently on all 467 trees in the 95% credible set. The ancestral sequence on every tree was identical to that on the maximum-likelihood tree at every site but one, which had an alternate state on some trees. When this state was introduced into the reconstructed protein, AncCR became even more sensitive to aldosterone (Figure 16.3a). This result indicates the AncCR's aldosterone sensitivity is not likely to be an artifact of assuming an incorrect tree.

As for model violation, it is possible that the JTT + gamma model we used is erroneous, and violation of this model's assumptions could have produced some inaccurate ancestral states. In our Bayesian analysis, we integrated over a large number of available models and found that the JTT + gamma model was supported with 100% posterior probability; that is, alternative models therefore had vanishingly small probability due to a very poor fit to the data. We therefore did not explore alternative reconstructions using these other models, because the weighted posterior probability of the ancestral sequence states that would be inferred using these models is approximately zero. This does not mean, however, that JTT + gamma represents the true evolutionary model; it implies only that it is the best of the protein evolutionary models currently implemented. The real evolutionary process is almost always more complex than these models, but no methods are available for determining whether the best-fit model is in fact good enough to predict with high accuracy (Thornton and Kolaczkowski, 2005). Thus we cannot rule out the possibility that model violation could have introduced some errors into our reconstruction. We can say only this: based on our Bayesian analysis, the ancestral sequence we inferred—and our experimental results—are robust to ambiguity in the choice of models from among those currently available.

16.7 The evolution of the MR–aldosterone interaction

The aldosterone-sensitivity of AncCR is surprising, because aldosterone is a relatively recent,

tetrapod-specific hormone. Aldosterone has been reliably detected in tetrapods and lungfish (Bentley, 1998), but is absent from teleosts (Jiang *et al.,* 1998), elasmobranchs (Simpson and Wright, 1970; Nunez and Trant, 1999), and agnathans (Bridgham *et al.,* 2006). Prior work has shown that aldosterone's emergence is due to evolutionary modification of a key enzyme in the steroidogenic pathway (Figure 16.1b)—cytochrome P450–11B (Cyp11B). The ancestral function of Cyp11B is to catalyze the 11-hydroxylation of DOC in the synthesis of glucocorticoids, a function present in all jawed vertebrates. Only in tetrapods has this enzyme evolved the additional capacity to hydroxylate corticosterone at the 18-position to produce aldosterone (Nonaka *et al.,* 1995; Jiang *et al.,* 1998; Bulow and Bernhardt, 2002). This novel catalytic function appended aldosterone as a new terminal hormone at the end of the more ancient glucocorticoid synthesis pathway (Figure 16.1b).

Our data, together with this knowledge about the evolution of the steroid-synthesis pathway, indicate that the sensitivity of corticoid receptors to aldosterone is more ancient than the hormone itself. The aldosterone-sensitive AncCR existed in an organism that almost certainly did not produce aldosterone, just as present-day teleosts, elasmobranches, and agnathans contain MRs and CRs that are activated by aldosterone, despite the documented absence of the hormone from these taxa. AncCR must have been regulated by a different ligand; one candidate is DOC. DOC is clearly ancient: it is known to be produced by agnathans (Weisbart and Youson, 1977), and tetrapods, teleosts, and elasmobranchs all make it as an intermediate in the synthesis of other corticosteroids (Figure 16.1b), Our experiments show that AncCR is extremely sensitive to DOC, as are both agnathan CRs (Figure 16.3a). DOC is also an effective activator of the human MR (Hellal-Levy *et al.,* 1999) and may be the physiological MR ligand in teleosts (Sturm *et al.,* 2005).

Aldosterone differs structurally from DOC only by the presence of 18-keto and 11-hydroxyl groups; our experiments show that neither of these moieties affect activation of the ancestral or extant CRs and MRs. In all the receptors we examined, both ancestral and extant, there is a very strong correlation between sensitivity to aldosterone and sensitivity to DOC (Figure 16.3c). Whatever the precise identity of the ancestral ligand, AncCR's sensitivity to aldosterone, like that of the CRs and MRs in species that do not produce the hormone, must be due to its similarity to the endogenous steroids that are the receptor's natural ligands.

These results demonstrate how the aldosterone–MR partnership in tetrapods evolved in a stepwise, Darwinian fashion. Our data show that the receptor's affinity for aldosterone preceded the appearance of the hormone; AncCR's sensitivity to aldosterone was a structural byproduct of the receptor's affinity for its natural hormone. In this sense, the receptor was ''preadapted'' to bind aldosterone when the hormone appeared much later due to modification of the steroidogenic pathway. We call this dynamic molecular exploitation, because it involves a newly evolved molecule—the hormone, in this case—recruiting into a new interaction a more ancient molecule that was previously constrained by selection for an entirely different function.

16.8 The mechanistic basis for GR evolution

One of the most exciting new applications of ancestral gene resurrection is in determining the mechanistic basis for the evolution of gene function. By resurrecting multiple ancestral nodes on a tree it is possible to determine experimentally when a novel function evolved. Candidate amino acid positions can then be identified as those residues that changed state on the same branch on which the new function emerged; the hypothesis is that these substitutions represent the mechanistic basis for the evolution of the new function. These candidate substitutions can then be introduced into the resurrected ancestral gene to determine their actual effect on the function.

We have used this strategy to study the evolution of the GR's derived functions (Bridgham *et al.,* 2006). Our findings concerning the function of the AncCR indicate that the specific MR–aldosterone partnership in tetrapods is due to the loss of the

ancestor's aldosterone-sensitivity in the lineage leading to the GRs of bony vertebrates. To determine when this shift occurred, we resurrected two additional ancestral receptors on the tree, both of which existed after the duplication of AncCR: the GR in the ancestor of all jawed vertebrates (AncGR) and the GR in the ancestor of bony vertebrates (BonyGR). Using transcriptional reporter assays, we showed that AncGR is indeed activated by aldosterone, cortisol, and DOC, but the more recent BonyGR displays the full GR-like phenotype, with no aldosterone response and a reduced sensitivity to cortisol (Figure 16.4a). This result indicates that aldosterone sensitivity was lost from the GRs after the elasmobranch divergence but before the tetrapod/teleost split.

To understand the mechanistic basis for this functional shift, we combined ancestral gene resurrection with mutagenesis to identify the specific amino acids responsible for the GR's loss of aldosterone response. We identified candidate substitutions by finding amino acid changes that are phylogenetically and functionally diagnostic, defined as having occurred on the branch where aldosterone sensitivity was lost, with one state conserved in all the aldosterone-activated receptors and a different state in all aldosterone-insensitive GRs. Although it is possible that functionally crucial sites that changed on the key branch may not have been conserved since then in extant sequences, we reasoned that the residues most important to the functions of the GR and MR would probably be constrained by selection, making them reasonable first candidates.

There were four diagnostic changes that met these criteria. To test their functional importance, we introduced each substitution singly and in combination into the AncCR by site-directed mutagenesis. We determined experimentally whether they were capable of producing the GR-like phenotype: loss of aldosterone sensitivity, with moderate cortisol sensitivity maintained. The combination of S106P and L111Q conferred a GR-like function, reducing the receptor's sensitivity to aldosterone by three orders of magnitude while retaining moderate sensitivity to cortisol and DOC (Figures 16.3a and 16.4b). None of the other mutants tested showed this pattern (Bridgham *et al.*, 2006).

Knowing that multiple substitutions were required to yield the GR function, we sought to determine the order in which these substitutions are likely to have occurred. To reconstruct the trajectory of GR evolution, we introduced each replacement in isolation and found that both are required to yield the full GR phenotype. The L111Q mutation alone radically reduced activation by all the ligands tested (Figure 16.4b). S106P strongly impaired both aldosterone and cortisol sensitivity, but this receptor retained significant DOC sensitivity, suggesting a neutral phenotype with regard to the likely ancestral ligand. In the S106P background, L111Q restores the cortisol response to a level characteristic of extant GRs, while further reducing aldosterone response and leaving DOC activation more or less unchanged (Figure 16.4b).

These results indicate that a mutational path beginning with S106P followed by L111Q converts the ancestor to the modern GR phenotype via a functional intermediate step and is therefore the most likely evolutionary scenario (Maynard-Smith, 1970). This result also points to strong intragenic epistasis—that is, the effect of the L111Q substitution depends on the state at site 106—and indicates that that the order of substitutions strongly constraints the potential evolutionary paths that the protein may take through sequence space. The results we observed for these two substitutions suggest that the ancestral sequence evolved along a neutral ridge through sequence space (Gavrilets, 2004), bypassing the nonfunctional valley represented by L111Q alone.

To illuminate how these substitutions altered the evolving GR's response to ligand, we compared the crystal structures of the human GR and MR and found that the two substitutions cooperate to maintain cortisol activation despite the loss of aldosterone sensitivity (Figure 16.4c; see also Fagart *et al.*, 2005; Li *et al.*, 2005). The substitution of proline for serine at the position corresponding to 106 in the AncCR introduces a kink in a loop between two helices; this kink pulls one of the adjacent helices forward, changing the shape of the ligand pocket in a way that partially destabilizes the binding of hormones. As a result, activation by aldosterone is radically reduced, and

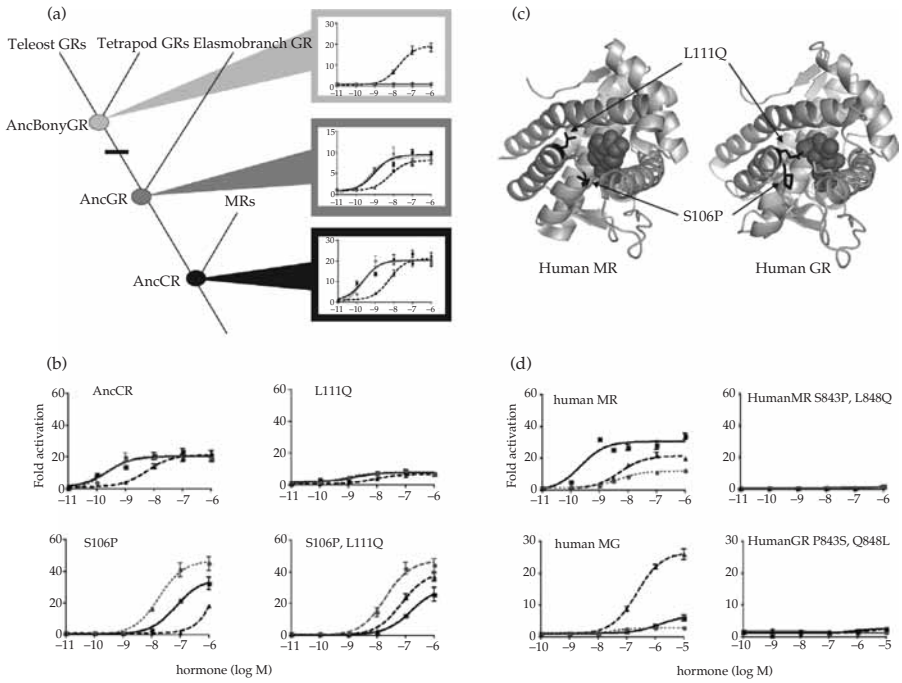


Figure 16.4 Evolution of the GR phenotype. (a) Resurrection of a series of ancestral receptors determines when aldosterone sensitivity was lost from the GR lineage. Ancestral receptors at the three nodes indicated were reconstructed by maximum likelihood, synthesized and functionally characterized in a luciferase reporter assay with increasing doses of aldosterone (solid black line), DOC (gray), and cortisol (dotted black). The GR in the ancestor of all jawed vertebrates retains the aldosterone-sensitive phenotype of AncCR. In contrast, the GR in the ancestor of all bony vertebrates (BonyGR) has the full GR-like phenotype, with response to aldosterone and DOC abolished and moderate sensitivity to cortisol retained. These results indicate that the GR-phenotype emerged on the branch marked with the black bar. (b) Identification of substitutions crucial to the emergence of GR-like function. Four substitutions occurred on the branch where GR function evolved (black bar in panel (a) and were then conserved, with one state in all aldosterone-insensitive receptors and another in all aldosterone-activated receptors. These were introduced singly and in combination into the ancestral background. One two-fold mutant (S106P/L111Q) had a largely GR-like phenotype, with aldosterone sensitivity reduced by three orders of magnitude and moderate cortisol sensitivity retained. Large arrows show the evolutionary trajectory to the double mutant through a functional intermediate step; small arrow, trajectory involving a nonfunctional intermediate. (c) Structural basis for the functional shift caused by the two-fold substitution. S106P and L111Q are plotted on the structures of the human MR and the human GR with their ligands. S106P introduces a kink that excludes aldosterone and moves L111Q into a position where it can form a hydrogen-bond with the unique 17-hydroxyl group of cortisol, stabilizing binding of cortisol in the GR-like structure. (d) The effects of substitutions on function depend on the ancestral sequence background. When introduced into the AncCR sequence, S106P and L111Q together confer the GR-like phenotype, but when the same substitutions are introduced into the extant human MR, they do not have the same effect. The reverse substitutions in the extant human GR background are not capable of reversing the GR-like phenotype to restore the ancestral function.

cortisol—which was a weaker activator to begin with—becomes a completely ineffective ligand. The kink, however, also brings site 111 into a position where it is close to the 17-position on the ligand; the substitution of a polar glutamine for the hydrophobic leucine then forms a hydrogen bond with the 17-hydroxyl, which is only found in cortisol. This bond stabilizes cortisol binding and restores activation by cortisol, compensating for the general reduction in ligand sensitivity induced by S106P. Our findings therefore indicate that the aldosterone specificity of MR evolved by a simple and conserved structural mechanism: two crucial replacements in the GRs that change the general architecture of the binding pocket and then compensate by creating a new stabilizing interaction with one specific ligand. The effect is to wipe out the ancestral sensitivity to aldosterone and retain a moderate response to cortisol.

In summary, our findings demonstrate that the MR–aldosterone partnership evolved in a stepwise fashion consistent with Darwinian theory, but the

functions being selected for changed over time. AncCR's sensitivity to aldosterone was present before the hormone, a byproduct of selective constraints on the receptor for activation by the native ligand, just as the agnathan CRS elasmobranch and teleost MRs can be activated by aldosterone despite its absence from those organisms. AncCR and its descendant genes were structurally preadapted for activation by aldosterone when that hormone evolved millions of years later. After the duplication that produced GR and MR, only two substitutions in the GR lineage were required to abolish aldosterone sensitivity and yield two receptors with distinct hormone-response profiles. The evolution of an MR that could be independently regulated by aldosterone enabled a more specific endocrine response, because it allowed electrolyte homeostasis to be controlled without also triggering the GR stress response, and vice versa.

16.9 Ancestral gene resurrection for studying structure–function relationships

Many structure–function studies seek to determine the role of individual residues in producing specific protein functions. Candidate residues are often identified by sequence comparisons and then introduced into extant proteins to test the hypothesis that they determine some aspect of function. If intragenic epistasis is important, however, then the effect of a substitution may depend on the sequence at other sites, however; the only way to reliably determine whether a substitution was crucial for producing a new function is to introduce it into the ancestral background in which it originally occurred.

Understanding the ligand specificity of GR and MR is a ripe goal in molecular endocrinology. These receptors play key roles in numerous diseases, so they are prime drug targets. Their partially overlapping specificities to many synthetic and natural ligands results in unwanted side effects, however. Better insight into the structural basis of the GR's and MR's ligand-binding functions would help guide efforts to design receptor-specific agonists and antagonists.

Having identified S106P and L111Q as phylogenetically and functionally diagnostic substitutions and then verified their ability to produce the GR-like phenotype when introduced into the ancestral sequence, we sought to determine whether these same two substitutions are sufficient to yield a GR-like function when introduced into an extant MR. We constructed a double-mutant humanMR with the homologous S843P and L848Q substitutions. We found that in this background these two substitutions do not produce a cortisol-specific receptor as they do in AncCR; rather, they render the human MR completely unresponsive to cortisol, aldosterone, and DOC (Figure 16.4d). Conversely, we tested whether introducing the ancestral states at these positions into the human GR could restore aldosterone sensitivity as they do in AncCR; however, the P637S/Q642L double substitution in the human GR background also abolished all activation by the receptor (Figure 16.4d).

These data indicate that the functional effect of evolutionarily important residues depends crucially on the amino acids present at other positions in the protein. Our findings show that if we are to understand the mechanistic basis for the evolution of protein functions we must use the ancestral sequence as the substrate for hypothesis testing. Introducing putatively important sequence changes into extant proteins, a common practice, does not reliably affect function in the same way as when the same substitutions occur in the ancestral sequence background.

16.10 Sex-steroid evolution

Our first work in ancestral gene resurrection helped reveal how the evolution of AR's interaction with testosterone and PR's partnership with progesterone evolved. This research involved the resurrection of the common ancestor of the entire SR gene family, or AncSR1 for short. We began by inferring the optimal phylogeny of a large database of SRs and closely related nuclear receptors by both parsimony and BMCMC; both methods found nearly identical trees, and most of the nodes on the tree were strongly supported (Figure 16.5a Thornton, 2001; Thornton *et al.*, 2003). The tree indicated that AncSR1 existed before the divergence of protostomes from deuterostomes, some

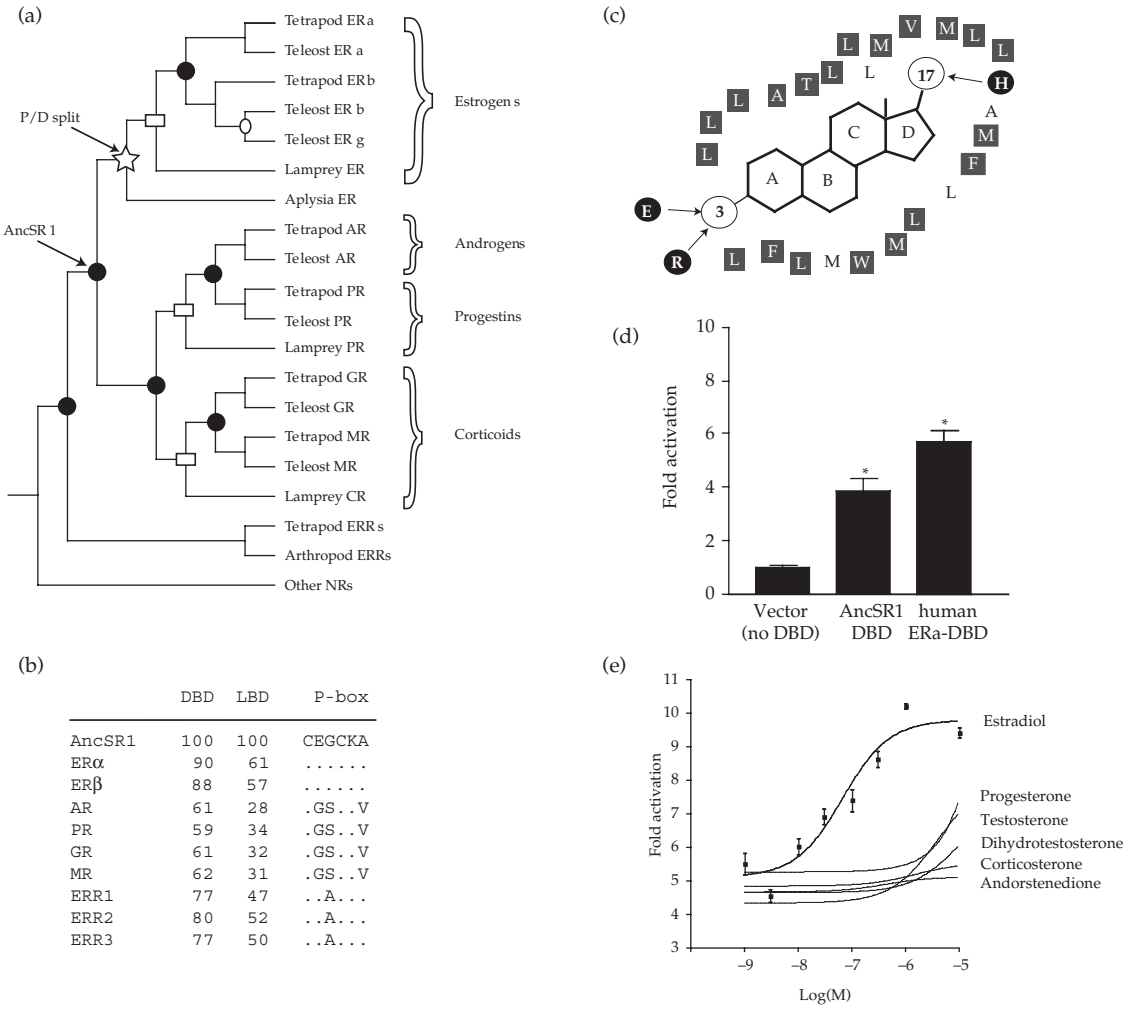


Figure 16.5 The ancestral steroid hormone receptor functioned as an estrogen receptor (see Thornton 2001 and Thornton *et al.* 2003 for details). (a) Phylogeny of steroid hormone receptors. Dark circles represent gene duplications; star, speciation event that split protostomes from deuterostomes; rectangles, speciation of jawed from jawless vertebrates. AncSR1 is the ancestral gene from which all extant steroid receptors descend. ERRs and other NRs are members of other nuclear receptor families. (b) Sequence similarity of the inferred AncSR1 protein sequence to human steroid and related receptors. Percent amino acid identity of AncSR1 is shown for the DNA-binding domain and the ligand-binding domain. The sequence in the P-box, a region of the DBD that confers specificity for response elements, is also shown. Dots indicate residues identical to AncSR1. (c) The ligand-binding pocket of AncSR1 is almost identical to that of the human estrogen receptor. Residues lining the pocket of AncSR1 are shown in relation to the steroid ligand, based on crystallography of several extant receptors. Gray shaded residues are identical to those in the human ERα. Circled residues interact with steroid moieties at the 3- and 17- positions and confer specificity for the various steroid hormones. (d) AncSR1 activates transcription from estrogen response elements. Activation of an ERE-driven luciferase gene is shown for the DBDs of the resurrected ancestral receptor and the human ERα. Asterisks, significantly different from control, $P<0.01$. (e) AncSR1 activates transcription in response to estrogen but not other hormones. Activation of a luciferase reporter by the ligand-binding domain of the resurrected ancestral receptor is shown over a range of concentrations of several steroid hormones.

600 million years ago, a result consistent with the discovery of ER gene sequences in mollusks (Thornton *et al.*, 2003; Keay *et al.*, 2006). We then inferred the ancestral sequences of the DBDs and LBDs by maximum likelihood, assuming the JTT + gamma model, which was again strongly supported in the Bayesian analysis.

We examined the AncSR1 sequence in light of known structure–function relationships to predict its function, and we hypothesized that it was likely

to function like an ER. First, AncSR1 was far more similar to the extant ERs (90% in the DBD and 85% in the residues that line the ligand-binding pocket of the LBD) than it was to the other SRs (at most 62 and 34%, respectively; Figure 16.5b). More specifically, there are six critical residues—three in the LBD and three in the DBD—that are known from crystallographic and mutagenesis studies to discriminate between estrogens and the other steroid hormones and between estrogen-response elements and the elements recognized by the other receptors. Every one of these critical sites in the ancestral protein contained the ER-specific residues, and all were reconstructed with high posterior probability (Figure 16.5b and c).

To test the hypothesis that AncSR1 had ER-like functions, we synthesized cDNA sequences that code for the inferred ancestral protein's functional domains and cloned them into vectors for high-level expression in CHO-K1 cells. Using a luciferase reporter assay, we showed that the AncSR1 DBD activated transcription of genes flanked by estrogen-response elements, to which other SRs do not bind, almost as effectively as modern-day ERs do (Figure 16.5d). This result corroborates the hypothesis that the DNA-binding functions of the ERs represent the conserved ancestral state, and those of the AR, PR, GR, and MR are derived.

The LBD activated transcription in a dose-dependent manner in the presence of low levels of estrogens but was completely insensitive to other steroid hormones (Figure 16.5e). The AncSR1 LBD was less effective than the human ER LBD, activating transcription to a lower level and requiring somewhat higher amounts of estrogens to achieve maximal activation; nevertheless, the specificity of AncSR1 to estrogens was very high. Further, in a ligand-binding assay, the receptor specifically bound radiolabeled estrogens, again with lower affinity than present-day ERs (Thornton *et al.*, 2003).

Together, these data indicated that the ancestor of the entire SR family functioned as a specific ER, activating transcription of genes flanked by estrogen-response elements when (and only when) estrogens are present. The functions of the other members of the SR family—including sensitivity to other hormones, such as testosterone and progesterone—are therefore derived characteristics that emerged after receptor gene duplications and sequence divergence (Thornton *et al.*, 2003). As for the somewhat impaired functions of the AncSR1 LBD compared to the extant ERs, there are two possible interpretations. First, the true ancestral receptor may have been a suboptimal ER, the functions of which were optimized during evolution of the ER lineage. A more likely interpretation in our view is that errors in the inferred sequence of AncSR1 LBD may have impaired the receptor's functions in our hands. Because LBDs are less conserved than DBDs and AncSR1 is such an ancient protein, the AncSR1 LBD was inferred with considerably lower mean posterior probability confidence than either AncSR1 DBD or the AncCR LBD (see Thornton *et al.*, 2003). Accordingly, the expected number of incorrect amino acids in the reconstructed sequence due to stochastic error is higher. Each such mistake has the potential to compromise function, just as nonsynonymous mutations in the sequence would be expected to do. Whatever the explanation for the quantitatively reduced function of the AncSR1, however, it is clear that this ancient protein responded to estrogens, and that sensitivity to the other steroids is a derived function that emerged in lineage leading to the AR, PR, GR, and MR.

In light of the synthesis pathway for producing steroid hormones, our results on the functions of AncSR1 shed light on how the novel interactions of PR and AR with their ligands evolved. Estrogen is the terminal hormone produced in a pathway that uses progesterone and testosterone as intermediates (Figure 16.1b); estrogen synthesis through this pathway appears to be extremely ancient, as all three steroids are present in vertebrates and invertebrates, such as mollusks (D'Aniello *et al.*, 1996; Di Cosmo *et al.*, 2001; Zhu *et al.*, 2003). Our experiments on AncSR1 indicates that that the last hormone to be produced was the first one to serve as an SR ligand. Before progesterone and testosterone served as SR ligands, then, they must have been present as intermediates in the production of estrogen. After the duplication of SR1, duplicated receptors evolved increased affinity for these steroids, turning what had

been biochemical stepping stones into *bona fide* hormones.

These results indicate that the interactions of AR and PR with their ligands also evolved by molecular exploitation. In contrast to the MR–aldosterone partnership, the steroids in this case were present before the receptors evolved. For the AR and PR, duplicated receptors diverged in sequence and recruited older ligands, which had previously served an entirely different function, into a novel signaling partnership.

16.11 Future directions in ancestral gene resurrection

We see three major areas for advancement of ancestral gene resurrection. The first is improvements in the methods for phylogenetic inference of ancestral states. Existing evolutionary models do not capture all the dynamics of real evolutionary processes, because they assume a largely homogeneous evolutionary process across sites and lineages. Real sequences are subject to selection pressures that vary considerably among sites and lineages. For example, a site may be subject to strong constraints in one lineage but not in another, a phenomenon called heterotachy (Lopez *et al.*, 2002), which has been shown to undermine the accuracy of likelihood-based phylogenetic methods under some conditions (Kolaczkowski and Thornton, 2004). Several groups, including ours, are working on mixed models to improve performance in the face of heterotachy and other forms of heterogeneity. We need to know whether unincorporated heterogeneity reduces the accuracy of ancestral state inference, and—if it does—determine whether mixed models improve our ability to correctly infer ancestral sequences.

Second, ancestral resurrection can be used even more ambitiously for understanding the evolution of protein function—particularly in determining how specific sequence changes have led to the evolution of new functions. We see three particularly interesting possibilities in this area. The first is to expand on the preliminary work we have reported using site-directed mutagenesis on ancestral sequences to recapitulate evolutionary substitutions in their ancestral background and determine their effect on function. This technique, expanded in a high-throughput framework, could allow the adaptive landscape on which gene sequences have evolved to be characterized, allowing us deep insights into unresolved questions about the evolutionary role of epistasis, the relative importance of substitutions of large and small effect, the prevalence of compensatory and permissive mutations, and the reversibility and contingency of the evolutionary trajectories that actually took place. Characterizing this adaptive landscape will require determining the functions of the substitutions that occurred on critical branches of phylogenies by introducing all the possible combinations of phylogenetically diagnostic residues into ancestral backgrounds.

Another strategy that could yield great insights into the mechanistic basis for the evolution of gene function is to resolve the three-dimensional structures of ancestral proteins—particularly from multiple nodes on a tree—compare them to extant proteins. Together with ancestral mutagenesis experiments to test mechanistic hypotheses, this approach could provide the missing link for understanding *how* specific substitutions generated novel functions, and how protein structures have evolved over time.

A third area for further development is the experimental evolution of ancestral sequences. Microbial experimental evolution systems have proven to be extraordinarily powerful for understanding the nature of the evolutionary process: these systems allow evolution by natural selection to take place in large populations using multiple replicates under controlled laboratory conditions, and the evolutionary intermediates can be sampled at regular intervals, stored in the freezer, and characterized for their sequences, fitness, and functions. We are currently developing a system in which we can subject a recombinant SR to strong selection to evolve affinity for new ligands. The purpose is to understand the mechanisms and evolutionary dynamics by which receptors evolve new specificities. We plan to introduce resurrected ancestral receptors into this system to determine whether the trajectories of sequence evolution taken by receptors in evolving specificity for steroid hormones during real historical evolution

represent the only way for that evolutionary problem to be solved. This approach should also let us characterize the nature of the adaptive landscape on which receptors evolve, and the structural constraints that determine the trajectories of sequence evolution.

With these kinds of applications, ancestral gene resurrection will help us gain insight into some of the most difficult, important, and previously intractable problems in evolutionary biology. We have no doubt that others will think of additional ways in which this powerful new tool can advance our understanding of the processes by which genes and their myriad functions have evolved.

References

Aharoni, A., Gaidukov, L., Khersonsky, O., Gould, S.M., Roodveldt, C., and Tawfik, D.S. (2005) The 'evolvability' of promiscuous protein functions. *Nat. Genet.* **37**: 73–76.

Baker, M.E. (2001) Hydroxysteroid dehydrogenases: ancient and modern regulators of adrenal and sex steroid action. *Mol. Cell. Endocrinol.* **175**: 1–4.

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

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

Bulow, H.E. and Bernhardt, R. (2002) Analyses of the CYP11B gene family in the guinea pig suggest the existence of a primordial CYP11B gene with aldosterone synthase activity. *Eur. J. Biochem.* **269**; 3838–3846.

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.

Darwin, C. (1859) *On the Origin of the Species by Means of Natural Selection or the Preservation of Favoured Races in the Struggle for Life.* John Murray, London.

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.

Fagart, J., Huyet, J., Pinon, G.M., Rochel, M., Mayer, C., and Rafestin-Oblin, M.E. (2005) Crystal structure of a mutant mineralocorticoid receptor responsible for hypertension. *Nat. Struct. Mol. Biol.* **12**: 554–555.

Farman, N. and Rafestin-Oblin, M.E. (2001) Multiple aspects of mineralocorticoid selectivity. *Am. J. Physiol. Renal Physiol.* **280**: F181–F192.

Fryxell, K.J. (1996) The coevolution of gene family trees. *Trends Genet.* **12**: 364–369.

Futuyma, D.J. (1998) *Evolutionary Biology,* 3rd edn. Sinauer Associates, Sunderland, MA.

Gavrilets, S. (2004) *Fitness Landscapes and the Origin of Species*. Princeton University Press, Princeton, NJ.

Green, S. and Chambon, P. (1987) Oestradiol induction of a glucocorticoid-responsive gene by a chimaeric receptor. *Nature* **325**: 75–78.

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.

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

Haag, E.S. and Molla, M.N. (2005) Compensatory evolution of interacting gene products through multifunctional intermediates. *Evolution* **59**: 1620–1632.

Hellal-Levy, C., Couette, B., Fagart, J. Souque, A., Gomez-Sanchez, C., and Rafestin-Oblin, M.E. (1999) Specific hydroxylations determine selective corticosteroid recognition by human glucocorticoid and mineralocorticoid receptors. *FEBS Lett.* **464**: 9–13.

Jiang, J.Q., Young, G., Kobayashi, T., and Nagahama, Y. (1998) Eel (*Anguilla japonica*) testis 11beta-hydroxylase gene is expressed in interrenal tissue and its product lacks aldosterone synthesizing activity. *Mol. Cell. Endocrinol.* **146**: 207–211.

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.

Kolaczkowski, B. and Thornton, J.W. (2004) Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. *Nature* **431**: 980–984.

Li, W.-H. (1997) *Molecular Evolution*. Sinauer Associates, Sunderland, MA.

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.

Lopez, P., Casane, D., and Philippe, H. (2002) Heterotachy, an important process of protein evolution. *Mol. Biol. Evol.* **19**: 1–7.

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

Nonaka, Y., Takemori, H., Halder, S.K., Sun, T.J., Ohta, M., Hatano, O., Takakusu, A., and Okamoto, M. (1995) Frog cytochrome P-450 (11 beta,aldo), a single enzyme involved in the final steps of glucocorticoid and mineralocorticoid biosynthesis. *Eur. J. Biochem.* **229**: 249–256.

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

Page, R.D.M. and Holmes, E.C. (1998) *Molecular Evolution: a Phylogenetic Approach*. Blackwell Scientific, Oxford.

Simpson, T.H. and Wright, R.S. (1970) Synthesis of corticosteroids by the interrenal gland of selachian elasmobranch fish. *J. Endocrinol.* **46**: 261–268.

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.

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. USA* **98**: 5671–5676.

Thornton, J.W. and DeSalle, R. (2000) A new method to localize and test the significance of incongruence: detecting domain shuffling in the nuclear receptor superfamily. *Syst. Biol.* **49**: 183–201.

Thornton, J.W. and Kolaczkowski, B. (2005) No magic pill for phylogenetic error. *Trends Genet.* **21**: 310–311.

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

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.

Yang, Z., Kumar, S., and Nei, M. (1995) A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics* **141**: 1641–1650.

Zhang, J. and Nei, M. (1997) Accuracies of ancestral amino acid sequences inferred by the parsimony, likelihood, and distance methods. *J. Mol. Evol.* **44** (suppl. 1): S139–S146.

Zhu, W., Mantione, K., Jones, D., Salamon, E., Cho, J.J., Cadet, P., and Stefano, G.B. (2003) The presence of 17-beta estradiol in *Mytilus edulis* gonadal tissues: evidence for estradiol isoforms. *Neuro. Endocrinol. Lett.* **24**: 137–140.