Robustness of Reconstructed Ancestral Protein Functions to Statistical Uncertainty

Geeta N. Eick,[1,2] Jamie T. Bridgham,[1] Douglas P. Anderson,[1,3] Michael J. Harms,[1,3] and Joseph W. Thornton*,[4]

[1]Institute of Ecology & Evolutionary Biology, University of Oregon, Eugene, OR

[2]Department of Anthropology, University of Oregon, Eugene, OR

[3]Institute of Molecular Biology, University of Oregon, Eugene, OR

[4]Department of Ecology & Evolution and Department of Human Genetics, University of Chicago, Chicago, IL

*Corresponding author: E-mail: joet1@uchicago.edu

Associate editor: Michael S. Rosenberg

Abstract

Hypotheses about the functions of ancient proteins and the effects of historical mutations on them are often tested using ancestral protein reconstruction (APR)—phylogenetic inference of ancestral sequences followed by synthesis and experimental characterization. Usually, some sequence sites are ambiguously reconstructed, with two or more statistically plausible states. The extent to which the inferred functions and mutational effects are robust to uncertainty about the ancestral sequence has not been studied systematically. To address this issue, we reconstructed ancestral proteins in three domain families that have different functions, architectures, and degrees of uncertainty; we then experimentally characterized the functional robustness of these proteins when uncertainty was incorporated using several approaches, including sampling amino acid states from the posterior distribution at each site and incorporating the alternative amino acid state at every ambiguous site in the sequence into a single "worst plausible case" protein. In every case, qualitative conclusions about the ancestral proteins' functions and the effects of key historical mutations were robust to sequence uncertainty, with similar functions observed even when scores of alternate amino acids were incorporated. There was some variation in quantitative descriptors of function among plausible sequences, suggesting that experimentally characterizing robustness is particularly important when quantitative estimates of ancient biochemical parameters are desired. The worst plausible case method appears to provide an efficient strategy for characterizing the functional robustness of ancestral proteins to large amounts of sequence uncertainty. Sampling from the posterior distribution sometimes produced artifactually nonfunctional proteins for sequences reconstructed with substantial ambiguity.

Key words: ancestral protein reconstruction, ancestral sequence reconstruction, protein evolution.

Introduction

Ancestral protein reconstruction (APR)—phylogenetic inference of ancient protein sequences, followed by gene synthesis, expression, and experimental characterization—has become a widely used strategy to experimentally test hypotheses about the functional and biochemical properties of ancient proteins (Jermann et al. 1995; Chandrasekharan et al. 1996; Chang et al. 2002; Thornton et al. 2003; Thomson et al. 2005; Gaucher et al. 2008; Hobbs et al. 2012; Akanuma et al. 2013; Bar-Rogovsky et al. 2013; Risso et al. 2013; Williams et al. 2013; Boucher et al. 2014; Akanuma et al. 2015; Bickelmann et al. 2015; Carrigan et al. 2015; Clifton and Jackson 2016; Devamani et al. 2016; Steindel et al. 2016). APR has also been used to experimentally determine the effects of specific historical changes in protein sequence on the properties of ancient proteins by introducing mutations that recapitulate ancient sequence substitutions into reconstructed ancestral proteins (Zhang and Rosenberg 2002; Bridgham et al. 2006; Kaiser et al. 2007; Ortlund et al. 2007; Yokoyama et al. 2008; Lynch et al. 2011; Finnigan et al. 2012; Harms et al. 2013; Smith et al. 2013; Wilson et al. 2015). By combining explicit reconstruction of history with experimental analysis of molecular properties, APR can bring the decisive inference style of molecular biology to questions about the mechanisms and dynamics by which proteins evolved.

The advantages of APR, however, depend upon the reliability of the inferred ancient sequences. Most APR studies have reconstructed ancestral sequences using the maximum likelihood (ML) approach, which yields a single best estimate of the ancestral sequence (Yang et al. 1995; Pupko et al. 2000). Beginning with an alignment of present-day protein sequences and a phylogeny of those sequences, ML-APR uses a probabilistic Markov model of the process of sequence evolution to calculate the likelihood of every possible ancestral state at every site in the protein sequence for any internal node of interest on the phylogeny. The likelihood of an ancestral amino acid state at some site is defined as the probability that all the states at this site found in present-day proteins at the tips of the tree would have evolved given that ancestral state, the phylogeny, and the model. The

Open Access

posterior probability (PP) of that state can then be expressed as the likelihood of that state (weighted by its equilibrium frequency, which serves as the prior) divided by the sum of the prior-weighted likelihoods of all 20 possible states at that site. The maximum likelihood (ML) estimate of the ancestral state (or, strictly speaking, the maximum a posteriori estimate) is the one with the highest prior-weighted likelihood. The ML estimate of the ancestral sequence is the string of ML states at all sites; it represents the ancestral sequence that maximizes the conditional probability that all observed extant sequences in the alignment would have evolved.

The ML sequence is the best point estimate of the true ancestral sequence, but it is seldom inferred with certainty. In virtually all real-world cases, the reconstructed ML sequence contains some ambiguously inferred sites. At such sites, the PP of the ML state is less than 1, and less likely but still plausible alternative states exist. Uncertainty typically arises when the pattern of amino acids across the tips of the phylogeny requires numerous independent changes of state given any of the plausible ancestral amino acids and these scenarios have probabilities that are not dramatically different from each other given the tree and its branch lengths.

This uncertainty implies that although the ML sequence is the most accurate estimate of the true ancestral sequence given the data, model, and tree, it is likely to contain some erroneously reconstructed states. Consider a 200 amino acid ancestral protein containing 180 unambiguously reconstructed sites (PP = 1.0) and 20 sites at which two states are plausible with PPs of 0.8 and 0.2. The probability that the ML sequence is correct at every single site is $1^{180} \times 0.8^{20} = 1.2\%$, and the expected number of erroneous residues in the sequence is 4. The ML reconstruction sits at the center of a cloud of plausible alternative sequences, each of which contains alternative states at some of the ambiguously reconstructed sites; with increasing distance from the ML sequence, expected accuracy declines. The ML sequence's plausible immediate neighbors—the reconstructions containing the plausible alternative state at one site—each has only a 0.3% chance of being precisely correct and 4.6 expected errors. In the set of sequences containing two plausible alternative states, each sequence has a 0.07% chance of being correct and 5.2 expected errors, and so on. At the far edge of the cloud of plausible sequences, the sequence containing the alternate state at all 20 ambiguously reconstructed sites has only a 10^{-14} probability of being correct and is expected to contain 16 errors.

The goal of APR is to determine the ancestral protein's functional characteristics, not its precise sequence. Nevertheless, experiments to determine function depend upon inference of the sequence. Thus, a crucial question in any study using APR is the extent to which the properties of reconstructed ancestral proteins are robust to statistical uncertainty about their primary sequence. Most studies to date that have addressed this question have done so by generating variants of the ML ancestral sequence, each of which contains a plausible alternate amino acid at one of the ambiguously reconstructed sites (typically defined as amino acids with a posterior probability above some arbitrary but reasonable cutoff, such as 0.2). The experimental characterization is then repeated for each of these single-residue neighbors of the ML sequence (Yokoyama and Radlwimmer 2001; Zhang and Rosenberg 2002; Ugalde et al. 2004; Thomson et al. 2005; Bridgham et al. 2006; Goldschmidt et al. 2008; Carroll et al. 2011; Eick et al. 2012; Finnigan et al. 2012; Bickelmann et al. 2015). This approach is sufficient to determine the impact of each plausible alternate amino acid—and thus of uncertainty at each ambiguous site—in isolation on inferences about the ancestral protein's function. If ambiguously reconstructed sites interact epistatically, however, it is possible that the true ancestral sequence may have functions different from those of the ML ancestor and its immediate single-variant neighbors.

A second strategy, used in only a few cases, introduces all of the plausible alternate states into a single protein and then functionally characterizes this "worst plausible case" protein, which we refer to as the AltAll reconstruction. (Akanuma et al. 2013; Bridgham et al. 2014; McKeown et al. 2014; Anderson et al. 2016). The AltAll sequence contains more errors than any other plausible reconstruction and is typically much more different from the ML protein than the true ancestral protein is expected to be. Its characterization therefore represents a fairly conservative test of functional robustness to sequence uncertainty: if the ML and AltAll ancestors both lead to the same inference of the ancestral protein's function, it is assumed that the correct ancestral sequence—which most likely lies between these sequences, but much closer to the ML sequence—is likely to share the inferred function, as well. This approach addresses potential epistatic interactions among plausible alternative states. It also is practicable in cases in which the number of ambiguously reconstructed residues is so large that the strategy of creating all single-mutant neighbors is impractical.

A third strategy is to use Bayesian sampling to construct a set of sequences by choosing an amino acid state from the posterior probability distribution of ancestral states at each site. Several such sampled proteins are then constructed and assayed to provide some indication of the distribution of functions associated with the posterior probability distribution of sequences (Williams et al. 2006; Pollock and Chang 2007; Gaucher et al. 2008; Hobbs et al. 2012; Hart et al. 2014; Howard et al. 2014; Risso et al. 2014; Bickelmann et al. 2015). This strategy calculates the posterior probability distribution of ancestral sequence states given model parameters estimated from the data by maximum likelihood, so it represents an empirical Bayesian rather than fully Bayesian technique. The strategy is appealing in principle, but a concern is the possible production of nonfunctional proteins: the ensemble of all possible sequences contains far more very low-probability than high-probability sequences, particularly when ambiguity in the reconstruction is high, and if reconstruction errors are more likely to be functionally deleterious than beneficial, there may be a bias towards non-functional or poorly functioning proteins. A variation of this approach is to generate a large number of ancestral sequences by sampling from the posterior distribution, and then experimentally characterize reconstructions from this ensemble that have a

high likelihood of being correct (Eick et al. 2012). Such high-probability sampled proteins also typically combine large numbers of plausible alternate states—helping to address the issue of epistasis —but not as many as the highly conservative AltAll ancestors.

The AltAll and Bayesian approaches have been used in only a handful of cases, so it remains unclear how generally robust functional inferences using APR are to simultaneous incorporation of multiple alternate states. To better understand this issue, we applied the AltAll and Bayesian sampling strategies to reconstructed ancestral proteins in three different families of protein domains—guanylate kinase enzyme/protein interaction domains, steroid hormone receptor DNA-binding domains, and steroid hormone receptor ligand-binding/transactivation domains. These families have different kinds of functions and architectures, and they vary in the amount of ambiguity in their ancestral reconstructions. Each has been studied in previously published papers using ML-APR to characterize the functions of reconstructed ancestral proteins and to identify key historical mutations that, when introduced into the reconstructed ancestral sequence, recapitulate major shifts in protein function. For each family, we compared the functions of the AltAll and Bayesian ancestors with those of the ML ancestor to determine whether the published inferences are robust to incorporation of uncertainty at multiple sites using these strategies. We also compared the effects of introducing the key historical substitutions into the ML and AltAll ancestors to characterize the robustness of these inferences to uncertainty about the ancestral sequence.

Results

Exemplar Protein Families

We focused on three families of protein domains that have been the subject of previous work using ML-APR. The first, the GK domain family, contains two structurally similar but functionally distinct groups: the guanylate kinase (gk) enzymes, which catalyze the transfer of a phosphoryl group from ATP to GMP, and the GK_{PID} domains, which mediate protein–protein interactions (Anderson et al. 2016). A biological function of GK_{PID}s is to bind a protein called Pins, an association that is crucial for orientation of the mitotic spindle in animal cells. All forms of life contain gk enzymes; a duplication of the gk gene and subsequent sequence divergence produced the GK_{PID}, which is found only in animals and closely related unicellular protists. A recent study (Anderson et al. 2016) reconstructed ancestral GK proteins and found that the last common ancestor of gk enzymes and GK_{PID}s—called Anc-gk$_{dup}$, which existed just before the gene duplication that produced the two separate proteins—was an effective guanylate kinase with no Pins-binding or spindle-orienting activity. Its daughter node on the tree—the ancestor of all GK_{PID}s—bound Pins with moderate affinity and could mediate spindle orientation in cultured cells in which endogenous GK_{PID} had been disabled. A single substitution (s36P) that occurred on the branch between these two ancestral proteins was sufficient to confer on Anc-gk$_{dup}$ both

Pins-binding and spindle orienting functions and to abolish the ancestral guanylate kinase enzyme activity.

The second family of protein domains we evaluated was the DNA binding domain (DBD) of steroid hormone receptors (SRs). SRs are ligand-activated transcription factors, the DBDs of which bind as dimers to palindromic response elements composed of inverted repeats of a six base half-site. There are two phylogenetic classes of SRs in vertebrates, which differ in their DNA-specificity at the two central bases of the half-site: estrogen receptors (ERs) bind preferentially to estrogen response elements (EREs), which are palindromes of AG<u>GT</u>CA; the other clade, which contains receptors for progestagens, androgens, and gluco- and mineralocorticoids (PR, AR, GR, and MR), binds to steroid response elements (SREs), the prototype of which is a palindrome of AG<u>AA</u>CA. A recent publication reconstructed the last common ancestor of the two clades (AncSR1DBD) and found that it bound and activated transcription specifically on EREs, with no activation and much lower affinity for SREs. SRE recognition emerged on the daughter branch leading to the ancestor of the PR, AR, GR, and MR (AncSR2DBD), which activated solely on SREs with no activation from and lower affinity for EREs. A set of 14 historical substitutions, when introduced into the AncSR1DBD, was sufficient to recapitulate the switch in specificity (McKeown et al. 2014; Anderson et al. 2015).

The third domain family is the ligand-binding domain (LBD) of the SRs. SR LBDs specifically bind steroid hormones and regulate transcription of target genes near the response element to which their DBD is anchored. The two phylogenetic classes of vertebrate SRs also differ in their ligand specificity. ERs are specifically activated by estrogens, which have an aromatized A-ring (where the four rings on the steroid backbone are denoted with the letters A-D). PR, AR, GR, and MR, in contrast, all bind steroids with a non-aromatized A-ring. A recent pair of publications reconstructed the LBD of AncSR1 and found that it had ER-like specificity for aromatized estrogens, whereas the LBD of AncSR2—like PR, AR, GR, and MR—specifically activated transcription in the presence of non-aromatized steroids but not estrogens (Eick et al. 2012; Harms et al. 2013). Two substitutions that occurred on the branch between AncSR1LBD and AncSR2LBD were sufficient to switch ligand preference, restoring estrogen-specific activation when the ancestral states were introduced into AncSR2LBD and recapitulating activation by non-aromatized steroids when the derived states were introduced into AncSR1LBD (Harms et al. 2013).

These studies all used the ML sequence as the basis for their published inferences about functions and mutational effects. The degree of uncertainty varied among the ancestral proteins studied, with Anc-gk$_{dup}$ having relatively high confidence (mean PP over sites = 0.94, 11% of sites ambiguously reconstructed), AncSR1DBD having medium confidence (mean PP = 0.87, 15% ambiguous sites), and AncSR1LBD having very high uncertainty (mean PP = 0.70, 26% ambiguous sites). Ambiguously reconstructed sites were defined as those at which more than one state had posterior probability >0.20

Table 1. Summary Statistics for Ancestral Reconstructions in this Study.

Ancestor	Ln Posterior Probability (PP)	Ln PP Units Worse Than ML	Average PP	# Expected Errors	# Sites Different From ML Ancestor
Anc-gk$_{dup}$					
ML	−14.6	–	0.94	12	0/187
AltALL	−27.8	13.2	0.91	18	20/187
Bayes 1	−25.4	10.8	0.92	15	7/187
Bayes 2	−28.1	13.5	0.92	15	5/187
Bayes 3	−29.9	15.3	0.91	17	11/187
Bayes 4	−34.9	20.4	0.91	17	13/187
Bayes 5	−36.7	22.1	0.91	17	10/187
AncSR1 DBD					
ML	−15.6	–	0.87	11	0/82
AltALL	−20.1	4.5	0.85	12	12/82
Bayes 1	−23.0	7.4	0.85	12	7/82
Bayes 2	−29.4	13.8	0.82	15	12/82
Bayes 3	−30.0	14.4	0.84	13	9/82
Bayes 4	−39.5	23.9	0.82	15	10/82
Bayes 5	−40.0	24.4	0.82	16	12/82
AncSR1 LBD					
ML	−123.0	–	0.69	76	0/249
AltALL	−153.6	30.6	0.64	88	65/249
Best 1	−176.6	53.6	0.64	89	57/249
Best 2	−177.3	54.3	0.65	88	55/249
Best 3	−177.5	54.5	0.64	90	57/249
Best 4	−177.8	54.8	0.65	87	55/249
Best 5	−177.5	54.5	0.65	87	59/249
Bayes 1	−228.9	105.9	0.60	99	75/249
Bayes 2	−233.2	110.2	0.61	97	77/249
Bayes 3	−229.5	106.5	0.61	96	76/249
Bayes 4	−236.3	113.3	0.59	101	86/249
Bayes 5	−227.1	104.1	0.61	96	76/249
AncSR2 LBD					
ML	−23.0	–	0.93	17	0/249
AltALL	−41.7	18.7	0.90	25	26/249

NOTE.—For each ancestral protein sequence, columns show the natural logarithm of the posterior probability (Ln PP), the difference in Ln PP compared to the maximum likelihood (ML) reconstruction, the mean PP over sites, the number of expected errors in the sequence given its PP, the number of sites that are different from the ML reconstruction, and the total number of residues in the protein.

(supplementary table S2, Supplementary Material online). Some but not all of the strategies we assess for characterizing robustness have been applied previously to these protein families (Eick et al. 2012; McKeown et al. 2014; Anderson et al. 2016); here we use both new and already published data to systematically evaluate these approaches.

ML-APR Functional Inferences are Qualitatively Robust to Sequence Uncertainty

We first evaluated the AltAll strategy by introducing all plausible alternate amino acids into the ML ancestor. The AltAll sequence contains the ML state at all unambiguously reconstructed sites and the plausible alternate state (the state with second-highest posterior probability, if >0.2) at all ambiguous sites. We expressed each AltAll protein, characterized its functions using the same methods previously used to characterize the ML ancestral protein, and compared the results.

For all three protein domains, the AltAll reconstructed sequences were very different from their corresponding ML ancestors, with differences at 12, 20, and 65 sites in the DBD, GK, and LBD ancestors, respectively. In the most extreme case (AncSR1LBD), the AltAll protein was different from the ML ancestor at almost 30% of residues (table 1). All AltAll sequences were far less likely to be precisely correct than the ML

ancestor, with total posterior probability lower than that of the ML ancestor by factors ranging from 10^{-2} to 10^{-13} (table 1). All three AltAll sequences contained more expected errors than their corresponding ML ancestors.

We experimentally characterized the AltAll ancestors and found that all three had functions qualitatively similar to those of the ML ancestors, indicating general robustness to uncertainty. For Anc-gk$_{dup}$, both the AltAll and ML reconstructions were active guanylate kinases with kinetic parameters comparable to those of extant gk enyzmes, with no detectable Pins binding activity (fig. 1A and supplementary fig. S1, Supplementary Material online; Anderson et al., 2016). In the AncSR1DBD, the AltAll and ML reconstructions both activated luciferase reporter transcription robustly from ERE but not SRE (fig. 1B; McKeown et al., 2014). And, in the SR ligand binding domain, both versions of AncSR1LBD were highly sensitive to estradiol and insensitive to a non-aromatized steroid, whereas both versions of AncSR2 LBD were very sensitive to non-aromatized steroids and unresponsive to estrogens (fig. 1C; supplementary table S1 and fig. S2, Supplementary Material online). These experiments indicate that in all cases studied, the central inferences about ancestral proteins' functions are robust to statistical uncertainty about their precise sequence.
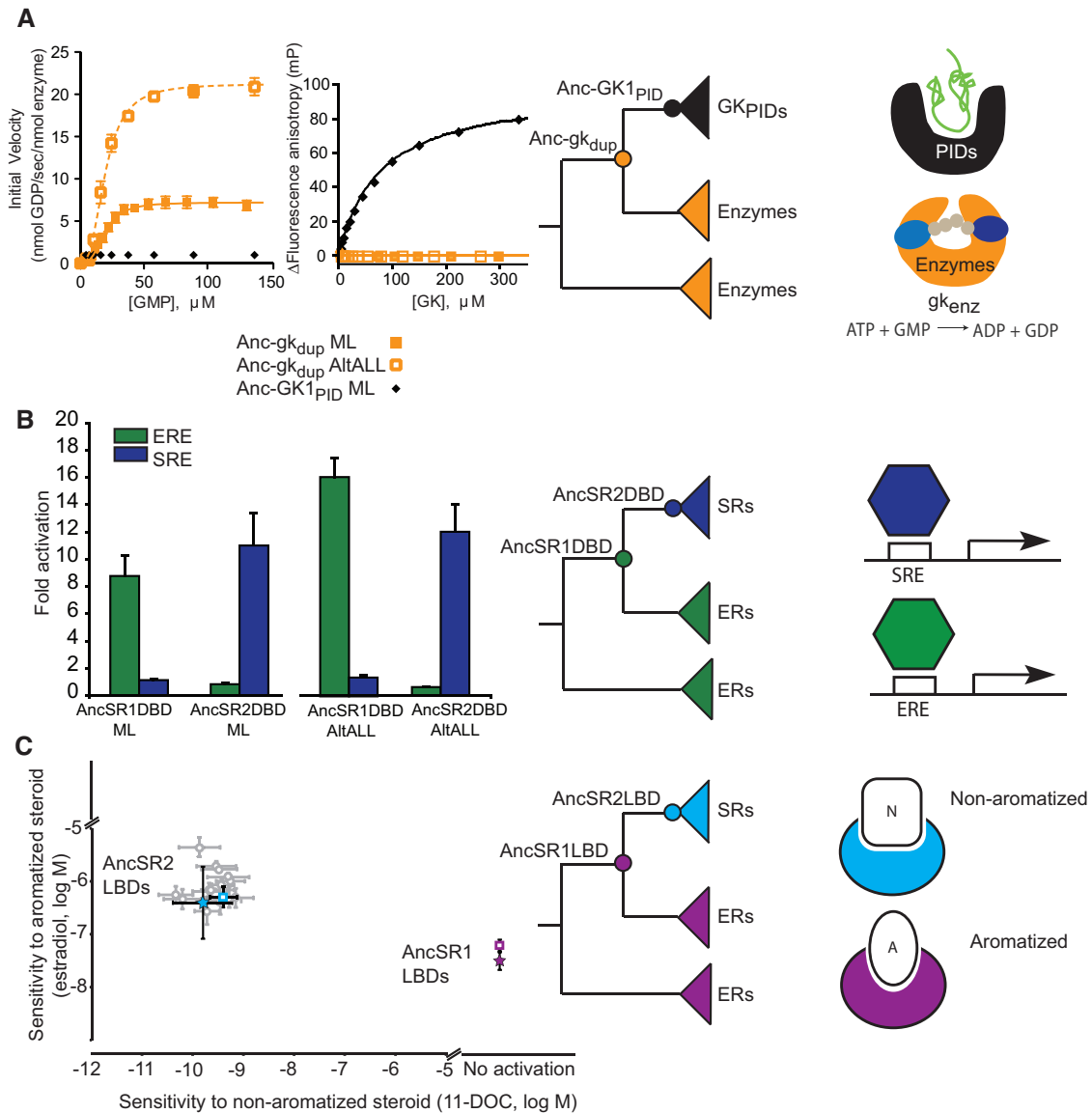
FIG. 1. Functional inferences based on ML-APR are robust to incorporation of uncertainty. For each protein domain family tested, the functional properties of the ML and AltAll ancestral reconstructions are shown (left panels), with a reduced phylogeny (center) and cartoon representation of the functions being assayed. (*A*) ML and AltAll versions of the ancestral guanylate kinase (Anc-gk$_{dup}$) both function as enzymes (orange), in contrast to the ancestor of GK protein interaction domains (AncGK1$_{PID}$, black). The left and right graphs show gunaylate kinase activity and peptide binding affinity, respectively. Error bars indicate SEM of three independent experiments. (*B*) Both the ML and AltAll AncSR1DBDs activate preferentially on the estrogen response element (ERE, green) rather than the steroid response element (SRE, blue) in a luciferase reporter assay. ML and AltAll versions of AncSR2 preferentially activate from SREs. Column height and error bars indicate mean and SEM from three experiments with three replicates each. (*C*) Both ML and AltAll AncSR1LBDs (purple square and star, respectively) are activated by estradiol (an aromatized steroid) but not 11-deoxycorticosterone (a non-aromatized steroid) in a luciferase reporter assay. In contrast, ML and AltAll AncSR2LBDs (blue square and star, respectively) are preferentially activated by the non-aromatized steroid. Single-substitution neighbors of ML AncSR2LBD containing a single plausible alternative state are indicated by open gray circles. Each point shows the concentration of hormones at which half-maximal activation is achieved (EC50); error bars show 95% confidence intervals. Some results in this figure were previously reported in Eick et al. (2012), McKeown et al. (2014), and Anderson et al. (2016).

For AncSR2LBD, we also experimentally characterized alternative reconstructions, each containing just one of the 26 plausible alternate states (Eick et al. 2012). These data provided useful context for interpreting the observed functions of the ML and Alt-All reconstructions. We found that these single-variant plausible ancestors occupied a relatively tight cloud of phenotypes defined by their sensitivity to non-aromatized steroids and insensitivity to estrogens; the ML and

AltAll ancestral forms were both contained within this cloud (fig. 1C). These data indicate directly that in this protein, there is little epistasis among plausible alternative states with respect to the form of ligand specificity studied.

Although the qualitative phenotypes were all consistent between AltAll and ML reconstructions, there were nontrivial quantitative differences in the measured functional parameters. Anc-gk$_{dup}$ AltAll displayed a turnover rate three times

faster than that of the ML reconstruction (fig.1A), but both are within the normal range for extant members of the family (supplementary fig. S1, Supplementary Material online). Both the ML and AltAll versions of AncSR1DBD activated transcription robustly from ERE and not at all from SRE, but the magnitude of activation on ERE ranged from 9-fold above control for ML to 16-fold for AltAll (fig. 1B). Similarly, the EC50s for activation of AncSR1LBD and AncSR2LBD varied by a factor of three between ML and AltAll reconstructions, a distinct but relatively small difference compared with the orders-of-magnitude preference each protein displays between classes of hormones (fig. 1C; supplementary table S1, Supplementary Material online).

We conclude that the functional inferences based on ML reconstructions for these proteins are robust to incorporation of large amounts of uncertainty. Qualitative measures of function—such as the presence/absence of a biochemical activity or the relative preference between distinct classes of binding partners—were identical between ML and AltAll versions of the protein, even when these sequences differed at a very large number of sites. Precise quantitative estimates of functional parameters, however, were less robust, with the true value presumably lying somewhere in or near a range defined by these observations.

Functional Effects of Historical Substitutions are Robust to Incorporating Uncertainty

We next evaluated the extent to which inferences about the functional effects of historical substitutions are robust to statistical uncertainty about the ancestral sequences into which they are introduced. If epistasis exists between the ambiguously reconstructed sites and the sites that putatively changed the protein's function, then inferences about the effects of substitution in the latter class of sites could depend strongly on the ancestral sequence chosen as the genetic background for the experiments. To address this issue, we assayed the functional effects of key historical substitutions on the AltAll ancestral reconstructions and compared them to their effects on the ML reconstructions.

We first examined the historical ser36Pro substitution in Anc-gk$_{dup}$. When introduced into the ML protein, this mutation was previously found to abolish the protein's enzyme activity and confer the acquisition of binding to the Pins protein (Anderson et al. 2016). We found that the mutation has the same qualitative effect when introduced into the AltAll version of the ancestral protein, abolishing catalytic activity and conferring Pins binding with similar affinity (fig. 2A).

We next examined the set of 14 large-effect historical mutations in AncSR1DBD. Introducing this group of substitutions into the ML reconstruction was previously shown to recapitulate the historical change in DNA recognition, abolishing luciferase activation from ERE and establishing activation on SRE (McKeown et al. 2014). We introduced them into the AltAll version of AncSR1DBD and found that they conferred a qualitatively and quantitatively similar shift in DNA specificity (fig. 2B).

Finally, we studied the effect on the ancestral SR ligand-binding domain of the derived amino acids Gln41 and Met75, which were previously shown to dramatically decrease sensitivity to estrogens and increase sensitivity to non-aromatized steroids when introduced into the ML reconstruction of AncSR1LBD. We found that introducing them into the AltAll version had a similar effect (fig. 2C). Conversely, reverting these residues to the ancestral states (glu41 and leu75) in both the ML and the AltAll reconstructions of AncSR2LBD restores the ancestral ligand preference, switching the protein's preferred ligands from non-aromatized steroids to estrogens (fig. 2D).

We again observed some quantitative differences—typically by less than a factor of three—between the effects of the mutations in the ML and AltAll backgrounds, but the qualitative effects were unchanged (fig. 2). Taken together, these results indicate that the effects of historical substitutions are robust to simultaneous incorporation of all plausible alternate states, even when large degrees of uncertainty are present in the ML reconstruction.

Bayesian Sampling from the Posterior Probability Distribution

The AltAll strategy incorporates all non-ML states that are plausible, defined as having a posterior probability at some defined cutoff, and it excludes any that do not meet this criterion. A Bayesian sampling approach, in contrast, will typically produce ancestral sequences that exclude most plausible alternate states (because their posterior probabilities are by definition <0.50) in favor of the ML state, but it will include some implausible states with very low probabilities and exclude some ML states that have very high posterior probabilities.

We characterized the utility of the Bayesian sampling method and the robustness of inferred ancestral functions to uncertainty incorporated using this approach, by applying it to the three protein domain families. For each ancestral protein, we computationally generated at least one million Bayesian reconstructions by sampling an amino acid from the posterior probability distribution at each site in the protein. The variation in the amount of uncertainty among the three proteins led to ensembles of Bayesian sequences with different characteristics (fig. 3). For the Anc-gk$_{dup}$ and AncSR1DBDs, most of the generated sequences differed from the ML sequence at fewer than 20 sites and were up to 40 log-units less likely. In the far more uncertain AncSR1LBD, the Bayesian sequences differed from the ML sequence by 50–100 residues and 60–170 ln-likelihood units. The Bayesian sequences typically had lower posterior probabilities than the AltAll reconstructions but differed from the ML sequences at a similar number of sites.

For experimental analysis of each domain family, we randomly chose five sequences from the generated ensemble of Bayesian sequences (fig. 3 and supplementary table S2, Supplementary Material online). In all cases, the sampled Bayesian ancestors were many log units less likely to be correct and contained more expected errors than the ML reconstructions. Compared with the AltAll sequences, all but one
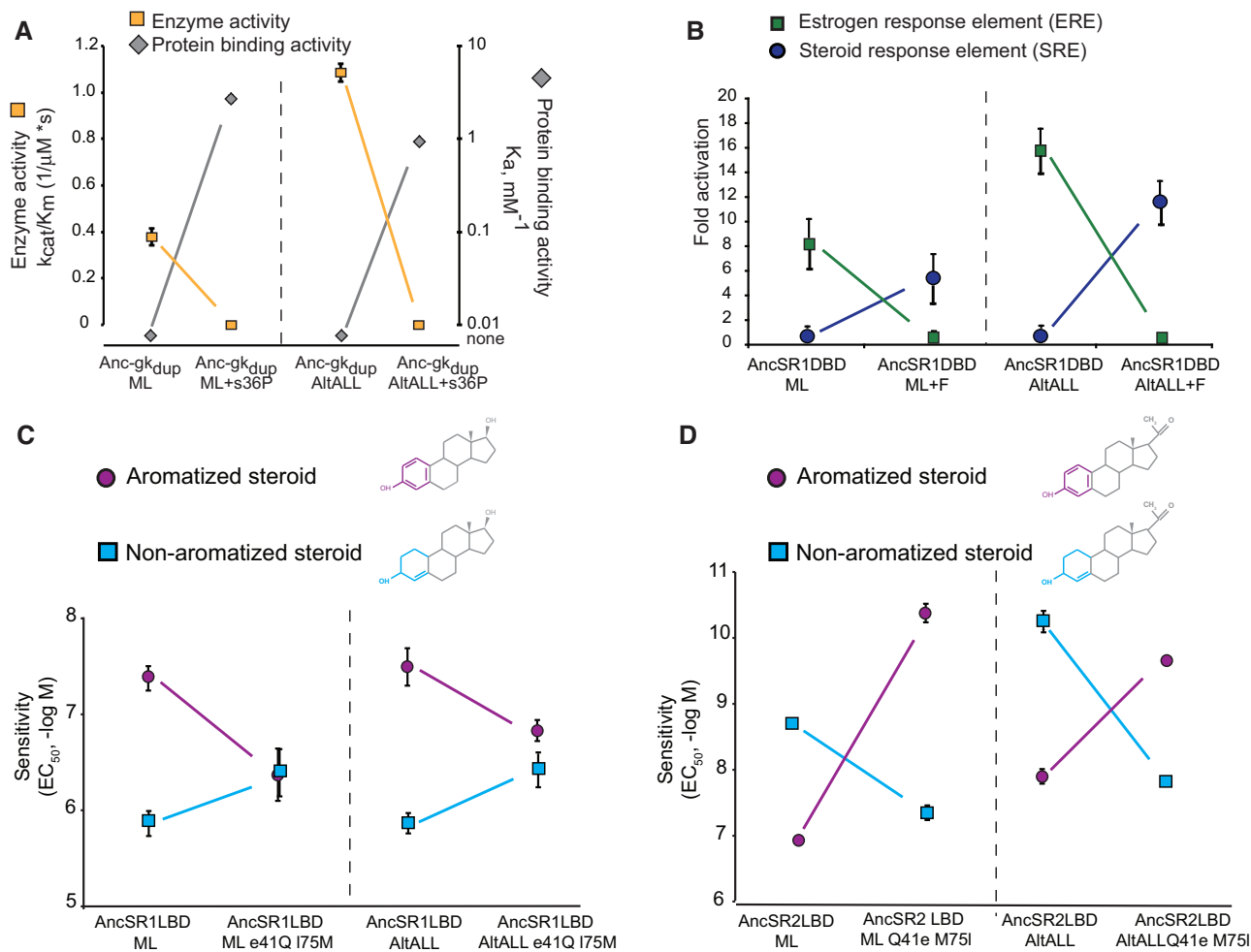
Fig. 2. Functional effects of historical substitutions are robust to incorporation of uncertainty. (*A*) The s36P substitution (upper case, derived; lower case, ancestral) changes the function of Anc-gk$_{dup}$ from a nucleotide kinase (orange squares, plotted using the y-axis at left) to a protein binding domain (grey diamonds, plotted using the right y-axis), in both the ML and AltAll backgrounds. Error bars indicate SEM from three independent experiments. Colored lines show the effect of the mutation on each function. (*B*) Function switching mutations (+F) switch the response element specificity of AncSR1DBD from ERE (green squares) to SRE (purple circles) in both the ML or AltAll ancestral backgrounds. (*C*) Introducing two major-effect historical substitutions into either AncSR1LBD ML or AltAll reconstructions decreases preference for an aromatized steroid (magenta with Lewis structure shown), and increases sensitivity to a non-aromatized steroid (blue). Points and error bars shown mean and SEM of EC50s from three independent dose-response experiments. (*D*) Reversal of the two key historical substitutions to their ancestral states (upper case, derived; lower case, ancestral) switches the preference of the ML and AltAll AncSR2LBDs from non-aromatized (blue) to aromatized (pink) steroids. Some data in this figure were previously reported in Anderson et al. 2016, McKeown et al. 2014, Eick et al. 2012, and Harms et al. 2013.

of the 15 sampled Bayesian ancestors had lower posterior probability (by factors ranging from 10^{-4} to 10^{-49}), and most contained more expected errors (table 1). The various reconstructions covered wide regions of sequence space, differing from each other by up to 104 residues in the case of AncSR1LBD (table 2 and fig. 3).

We synthesized and experimentally characterized the sampled sequences and found that functional inferences were generally robust to incorporating uncertainty using the Bayesian approach. For Anc-gk$_{dup}$, all five Bayesian sequences were, like the ML and AltAll proteins, effective guanylate kinases with similar catalytic performance (fig. 4A). For AncSR1DBD, all five Bayesian reconstructions of AncSR1DBD activated from ERE but not from SRE, just as both the ML and AltAll sequences did, and the degree of activation they elicited was similar in all cases (fig. 4B).

For the AncSR1LBD, however, the five Bayesian ancestors were apparently nonfunctional, failing to activate transcription in the presence of any ligand tested in either major class; ligand-independent constitutive activity was also lacking (fig. 4C and supplementary fig. S2, Supplementary Material online). It is very unlikely that the true ancestral protein lacked transcriptional activity, because descendant proteins from a wide variety of species have been tested, and virtually all of them function as transcriptional activators; for the Bayesian sequences to represent the true ancestral function, a very large number of independent gains of transcriptional activity in various SR and outgroup lineages would be required, an unlikely and non-parsimonious scenario. We therefore conclude that the Bayesian versions of AncSR1LBD are likely to be artifactually non-functional, presumably because of the
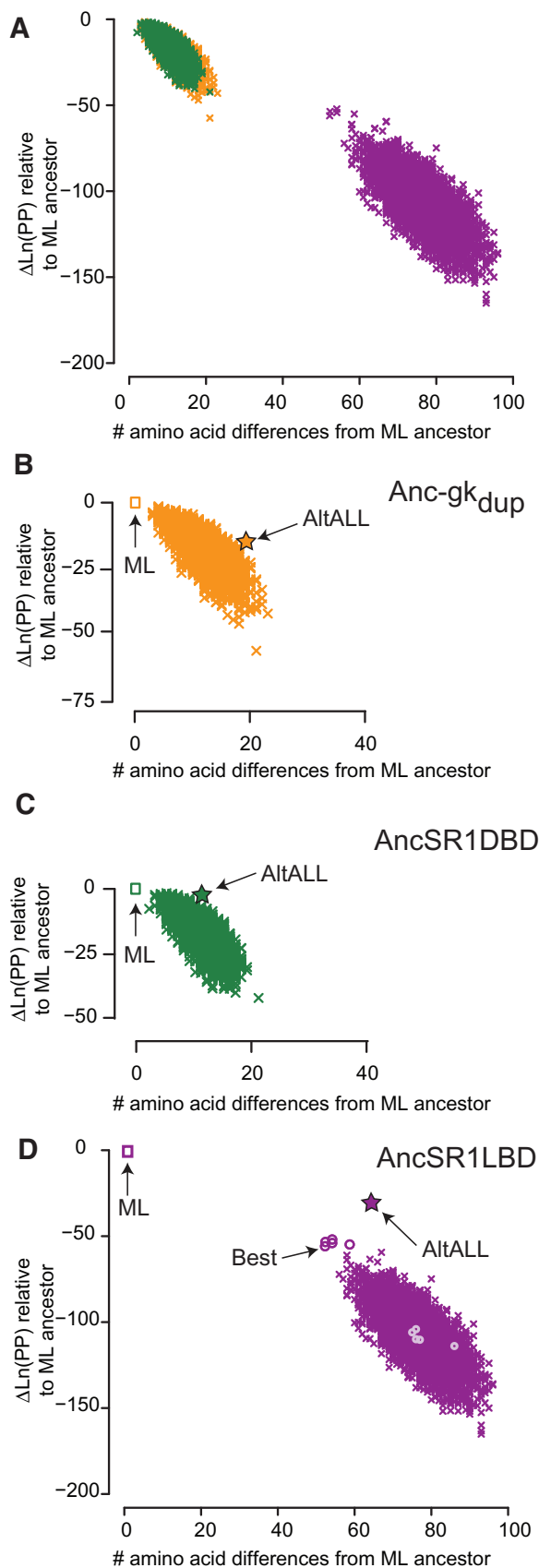
FIG. 3. Probability distributions of ancestral sequences generated by Bayesian sampling from the posterior probability distribution of ancestral amino acid states at each site. Each point represents one protein sequence, plotted by the number of amino acid differences and the

greater uncertainty of this ancestral protein and the resulting incorporation of more errors into Bayesian sequences. The AncSR1LBD Bayesian samples are ~100 ln-units less likely than the ML ancestor, differ from it by 75–86 amino acids, and contain 20–25 more expected errors—far more uncertain by every measure than is the case for the other protein domain families studied (fig. 3A and table 1).

Causes of Error in Bayesian-Sampled Ancestors

In contrast to the Bayesian reconstructions of AncSR1LBD, the AltAll version was transcriptionally active and had specificity similar to that of the ML ancestor, despite differing from it at 65 sites—almost as many differences as the Bayesian versions had. We hypothesized that the non-functionality of the Bayesian reconstructions of AncSR1LBD might be due to sampling of some very low-probability states—which have a high probability of being incorrect—and the consequent failure to include states reconstructed with very high probability, which are almost certainly correct and are typically strongly conserved among extant proteins.

To gain further insight into this possibility, we specifically evaluated additional sequences generated from the AncSR1LBD posterior probability distribution that contained fewer low-probability states. To accomplish this, we chose from the ensemble the five sequences with the highest total posterior probability (Best1-5; fig. 3D; reported in Eick et al. 2012). These Best sampled sequences covered a wide range of space, differing substantially from the ML ancestor (by 55–59 residues), from the AltAll version (74–79 differences), and from each other (63–82 differences). The Best sequences contained between 87 and 90 expected errors, about the same as AltAll and slightly fewer than those of the Bayesian ancestors (tables 1 and 2). However, the Best sampled ancestors contained only about half as many very low-probability states (PP < 0.1) than the Bayesian samples, and AltAll contained zero. The Bayesian samples also contained seven times more sites at which very high-probability states (PP ≥ 0.9) were excluded than did the Best sampled ancestors, and the AltAll and ML sequences by definition contained zero (fig. 5).

When experimentally characterized (Eick et al. 2012), all five Best sampled ancestors were estrogen-specific transcriptional activators, activating in response to estradiol at concentrations ranging from 10 to 1000 nM (fig. 4D; see Fig. S5B in Eick et al. 2012) but displaying no response to non-aromatized hormones (supplementary fig. S2, Supplementary Material online; see Fig. S5B in Eick et al., 2012). Thus, whereas all the Bayesian reconstructions were nonfunctional, all seven non-Bayesian

Fig. 3 Continued

difference in posterior probability relative to the ML reconstruction. (A) Ensemble of Bayesian sampled reconstructions of Anc-gk$_{dup}$ (yellow), AncSR1DBD (green), and AncSR1LBD (pink). (B) Distribution of 1 million Anc-gk$_{dup}$ sequences. (C) Distribution of 1 million AncSR1DBD sequences. (D) Distribution of 10 million AncSR1LBD sequences. ML (open square), AltAll (star), five Bayesian sampled sequences chosen at random for experimental characterization (light colored circles), and five sampled sequences with the highest lnPP (Best, open circles) are indicated.

Table 2. Number of Amino Acid Differences among Alternative Ancestral Reconstructions for Three Protein Domains.

Anc-gk$_{dup}$	ML	AltALL	Bayes 1	2	3	4	5
ML	*						
AltALL	20	*					
Bayes 1	7	21	*				
Bayes 2	5	23	11	*			
Bayes 3	11	21	14	13	*		
Bayes 4	13	14	18	17	20	*	
Bayes 5	10	19	17	15	19	17	*

AncSR1DBD	ML	AltALL	Bayes 1	2	3	4	5
ML	*						
AltALL	12	*					
Bayes 1	7	11	*				
Bayes 2	12	15	12	*			
Bayes 3	9	11	11	14	*		
Bayes 4	10	14	12	10	10	*	
Bayes 5	12	15	14	20	15	16	*

AncSR1LBD	ML	AltALL	Bayes 1	2	3	4	5	Best 1	2	3	4	5
ML	*											
AltALL	65	*										
Bayes 1	75	89	*									
Bayes 2	77	94	101	*								
Bayes 3	76	89	93	101	*							
Bayes 4	86	99	104	100	97	*						
Bayes 5	76	83	98	94	101	97	*					
Best 1	57	74	96	83	86	89	89	*				
Best 2	55	79	96	90	90	85	84	79	*			
Best 3	57	74	94	79	87	100	86	77	82	*		
Best 4	55	78	91	91	89	89	87	71	73	76	*	
Best 5	59	75	89	88	86	93	78	63	79	76	79	*

Note.—Anc-gk$_{dup}$, AncSR1DBD, and AncSR1LBD domains are 187, 82, and 249 amino acids long, respectively.

reconstructions, spanning a very large region of sequence space, have similar functional phenotypes, indicating that the inference of ancestral function is highly robust to incorporation of large amounts of statistical uncertainty but not, in this case, when the Bayesian sampling method is used. Notably, however, the Bayesian samples did not suggest a different ancestral function but instead failed to function at all.

The biochemical properties of the low-probability states incorporated into the Bayesian ancestors reinforce the view that they are a likely cause of the non-functionality of these sequences. Compared with the Best sampled ancestors, the Bayesian sequences incorporated ∼3 times more low-probability states at buried sites in the core of the protein, which are generally subject to greater constraints for proper folding than surface sites (fig. 5B). Further, the Bayesian sequence contained more non-conservative amino acid differences from the ML ancestor than did the Best Sampled or AltAll sequences (fig. 5D and supplementary table S3, Supplementary Material online). Thus, the Bayesian reconstructions of AncSR1LBD were more likely to incorporate

erroneous amino acids that disrupt protein structure and function.

Discussion

Robustness of Functional Inferences

Our findings have both practical and conceptual implications. Practically, they suggest that qualitative functional inferences— the presence/absence of a molecular function or relative preference/specificity among ligands—about reconstructed ancestral proteins are often robust to stochastic uncertainty about the precise ancestral sequences, even when a very large amount of uncertainty is present. We found that all the functions we examined are qualitatively unchanged across plausible sequence reconstructions in all three protein domain families, even when very large numbers of alternate plausible amino acids are incorporated into the protein. However, precise quantitative estimates of function—such as EC50, dissociation constant, or enzyme kinetic parameters—did vary among alternative reconstructions. This variation was relatively small—a factor of two or three, in most cases—and was far less than the differences between paralogs or ancestral proteins with biologically distinct functions. Thus, inferences about the precise quantitative parameters of molecular function should be made with caution, and experimental characterization of the range of plausible values for such parameters is particularly important. Our results are broadly consistent with a variety of case-studies of other proteins, which have found that qualitative inferences about the functions of ancestral proteins are generally robust to uncertainty about the ancestral sequence (Thomson et al. 2005; Ortlund et al. 2007; Finnigan et al. 2012; Boucher et al. 2014; Howard et al. 2014). Similarly, a recent study generated a small library of variant proteins containing shuffled combinations of possible ancestral states (not weighted by their posterior probabilities) and found a general pattern of qualitative robustness among the variants, although differences were again observed in quantitative measures of function (Bar-Rogovsky et al. 2015).

The impacts of large-effect historical mutations on qualitative measures of function also appear to be generally robust to uncertainty about the ancestral sequences. This result suggests that the genetic causes for the evolution of ancient protein functions can be identified with some confidence using APR. The consistency of these substitutions' effects across reconstructions indicates that epistasis is limited between large-effect historical mutations and plausible alternative sequence states in the reconstruction.

Our study was not comprehensive. We specifically examined the robustness of inferences about protein function to sequence uncertainty—particularly functions that changed among paralogs over evolutionary time. There could be other biochemical properties, such as stability or protein dynamics, that are less robust to uncertainty. In cases in which uncertainty has been assessed, however, estimates of ancestral protein stability have generally proven to be fairly robust to uncertainty about the ancestral sequence (see refs. Akanuma et al. 2013, 2015; Wheeler et al. 2016).
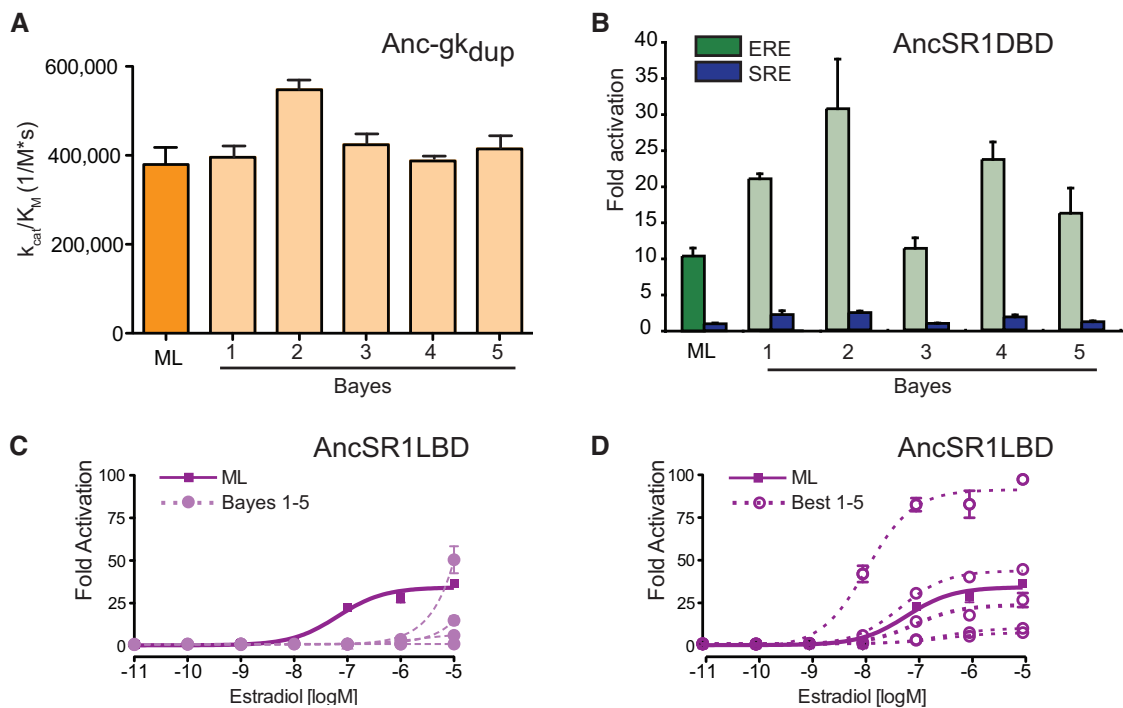
FIG. 4. Functions of ancestral sequences generated by Bayesian sampling. (*A*) Five randomly sampled Bayesian versions of Anc-gk$_{dup}$ had similar catalytic performance (k_{cat}/K_M) as the ML ancestors. Column height and error bars indicate mean and SEM of three independent experiments. (*B*) All five random Bayes AncSR1DBD ancestors activated from ERE (green) but not SRE (blue), like the ML ancestor. Column height and error bars indicate mean and SEM from three experiments with three replicates each. (*C*) The randomly sampled AncSR1LBD Bayes ancestors were not activated by physiological and greater concentrations ($>1\ \mu M$) of estradiol, unlike the ML ancestor. Error bars indicate SEM of three technical replicates. (*D*) The five "Best" AncSR1LBD Bayes ancestors were all activated by estradiol (previously reported in Eick et al. 2012). Error bars indicate SEM of three technical replicates. Some data in this figure were previously published in Anderson et al. 2016, McKeown et al. 2014, and Eick et al. 2012.

We examined only one source of uncertainty in ancestral proteins: statistical ambiguity about ancestral sequences given the data, phylogeny, and evolutionary model. Previous theoretical and simulation studies showed that incorporating stochastic uncertainty about the underlying phylogeny does not strongly affect ancestral sequence reconstruction (Hanson-Smith et al. 2010). Consistent with this finding, several empirical case studies have found that assuming different plausible phylogenies has only weak effects on inferences of ancestral protein functions (Gaucher et al. 2003; Akanuma et al. 2015; Clifton and Jackson 2016; Steindel et al. 2016) Systematic error in the tree, however, could still affect ancestral sequences and their functions (Groussin et al. 2015). Further, comprehensive work has not been conducted on the effects of the assumed model on ancestral reconstructions; however, research to date suggests that functional inferences are generally robust to sequence uncertainty associated with using different models and methods (Ugalde et al. 2004; Thomson et al. 2005; Chang et al. 2007; Devamani et al. 2016; Steindel et al. 2016). Further research will be required to thoroughly assess the effect of these sources of uncertainty and potential error on functional inferences about ancestral proteins.

Mutation, Epistasis, and the Robustness of Ancestral Functions

The extraordinary functional robustness of the reconstructed ancestral sequences we studied may seem surprising, given the high degree of statistical uncertainty and the fact that some of the alternate sequences differed from each other and from the ML ancestors at scores of amino acid sites—up to a third of the entire length of the protein. Most random amino acid replacements in proteins are deleterious (Guo et al. 2004), so how can ancestral reconstructions be so robust to sequence change?

Plausible alternate states are not random substitutions; rather, they are drawn from the much smaller set of states that are found in present-day members of the family, particularly those descending on branches near the ancestral node of interest. Unambiguous reconstructions occur when the ancestral state is conserved in large numbers of extant sequences descending from the node of interest and its close outgroups, reflecting strong functional constraints on that site. In contrast, ambiguous reconstructions occur when information about the ancestral state is lost in most descendant taxa, due to multiple exchanges between amino acid states on branches near the ancestral node of interest, a situation that leads to relatively small differences in the likelihoods of these states at the node. This kind of evolutionary dynamic occurs when functional constraints discriminate weakly or not at all between these states, which occurs only when they are all compatible with the protein's functions. Thus, uncertainty about the ancestral state tends to occur when multiple states are compatible with functional constraints, and the protein's function in turn is robust to which of these states is incorporated. The distribution of uncertainty and
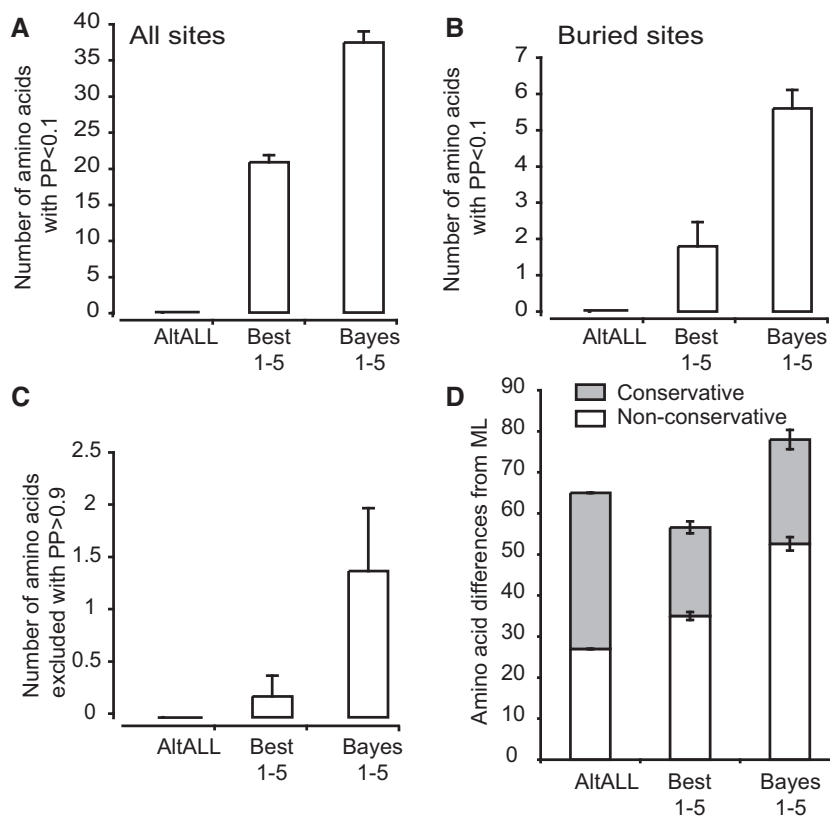
Fig. 5. Sequence characteristics of functional and non-functional AncSR1LBD reconstructions. (A,B) Non-functional Bayesian ancestors incorporate more low-probability states than functional ancestors. Columns show the average number of residues per protein in each category that differ from AncSR1LBD ML and have PP < 0.1, for all residues (A), and buried residues (B). (C) Non-functional Bayesian ancestors exclude more high-probability states than functional ancestors, as indicated by the average number of amino acid residues with PP > 0.9 that were excluded per protein in each category. (D) Types of amino acid differences between AncSR1LBD ML and alternative reconstructions. Amino acids were characterized as hydrophobic aliphatic (I, V, L), hydrophobic aromatic (F, Y, W), hydrophobic other (A, G, M), polar negative (E, D), polar positive (K, R), or polar neutral (Q, N, C, S, T), or other (P, H). Changes within a class were classified as conservative, those between were designated non-conservative.

plausible alternative reconstructions around a protein's three-dimensional structure may also provide useful information for efforts to engineer or design proteins with useful properties while maintaining core functions.

Epistasis is rampant among potential replacements in proteins and appears to be a widespread phenomenon among the substitutions that occurred during evolution (Starr and Thornton 2016; Storz 2016). In contrast, our observation that key historical substitutions have very similar effects on function when introduced into the ML or AltAll versions of a protein suggests that there are few strong epistatic interactions between these substitutions and the ambiguously reconstructed states. Similarly, our observation that introducing large numbers of alternative residues simultaneously into an ML protein does not strongly change its function suggests that there are few strong epistatic interactions among the plausible states at ambiguously reconstructed sites. How can a methodology that infers ancestral states on a site-by-site basis—with no attention to potential interactions or covariation among sites—not be undermined by epistasis?

It appears that strong epistatic interactions, to the extent that they functionally constrain the protein, are built into the phylogenetic structure of conservation and variability in the data with respect to ancestral reconstructions. That is, if some function depends on a specific combination of states across sites, those states tend to be consistently inherited together on the phylogeny and therefore to be reconstructed with little ambiguity. As a result, the states that make up compatible combinations will be generally conserved within clades and will be reconstructed with high confidence in the ancestors of those clades. That is, functional dependence across sites is reflected in the coupling of states at those sites on the tree—in the phylogenetic signal within the data—so this coupling tends to be reconstructed in the ancestors even when the probabilistic model itself does not explicitly incorporate it into the calculation of likelihoods. Supporting this view, previous studies of mutational trajectories in specific proteins have identified very strong epistatic interactions among historical substitutions, but those interacting sites were reconstructed in ancestors across the tree without ambiguity because changes in each site tended to lock in the derived state in the descendant clade (Ortlund et al. 2007; Lynch et al. 2011; Smith et al. 2013; McKeown et al. 2014).

Methods for Characterizing Robustness

Although we examined three protein domain families with different kinds of functions, structural architectures, and degrees of uncertainty, some other ancestral proteins or families might not be so robust to ambiguity about the precise ancestral sequence. It therefore remains essential for studies using APR to experimentally characterize the robustness of their functional conclusions to uncertainty about ancestral sequence reconstruction.

Our findings provide some guidance concerning the best way to accomplish this task. The AltAll strategy offers several advantages for characterizing uncertainty. First, it is conservative, simultaneously incorporating an alternative plausible residue at every ambiguously reconstructed site in the protein. The AltAll sequence therefore occupies a position at or near the far edge of the cloud of plausible reconstruction sequences, which contains all possible combinations of the most probable and second-most probable plausible residues. Depending on the number of ambiguously reconstructed sites, this cloud may contain a very large number of individual sequences—from 4096 in the case of AncSR1 DBD to 3.7×10^{19} for AncSR1 LBD. The true ancestral sequence is almost certainly closer to the ML reconstruction than the AltAll sequence is, making AltAll a conservative test of robustness. When the ML and AltAll sequences are functionally similar, it is likely that all or most combinations of the residues they contain will also share similar functions; our experiments on variants of AncSR2LBD-ML containing single plausible alternative states are consistent with this expectation.

A second advantage is that the AltAll strategy incorporates potential epistasis among ambiguously reconstructed residues by testing all plausible reconstructed amino acids together. Third, the strategy is highly efficient, allowing a strong test of robustness with a single experiment, even when a reconstructed protein contains many ambiguously reconstructed sites. Some caution is still required, however: although the AltAll sequence is a conservative test, it is not a comprehensive one, because it tests a single alternative sequence rather than all—or even many—possible combinations of plausible amino acids. In some cases, it may be useful to also test the effects of specific alternate residues on function, such as cases where ambiguously reconstructed sites are in positions of obvious structural importance, or where the possibility of specific epistatic couplings is of interest.

Bayesian sampling has the advantage of testing multiple ancestral sequences, and Bayesian inference is a generally appealing statistical framework for hypothesis evaluation and parameter estimation. Attempting to characterize the functions of reconstructed ancestral proteins by sampling a limited number of sequences from the posterior distribution, however, has some apparent disadvantages. When ambiguity about the reconstruction is low or moderate, we found that this strategy, like the AltAll approach, yielded ancestral proteins with functions similar to the ML reconstruction. When ambiguity was high, however, the Bayesian samples were strongly skewed towards reduced or abolished function, and this is unlikely to accurately reflect the state of the true ancestor. We propose three possible causes of this apparent bias.

First, unlike the AltAll reconstructions, the Bayesian-sampled sequences incorporated very low-probability residues, which by definition are more likely to be erroneous than high-probability states. Further, compared with the alternative residues in AltAll proteins, the low-probability states found in Bayesian sequences are more likely to occur at sites in the protein's core and to change biochemical properties. Such errors are more likely to affect the protein's structure, stability, and function and may therefore have caused the Bayesian-sampled ancestors to be artifactually non-functional.

Second, the use of an incorrect evolutionary model may contribute to the susceptibility of Bayesian ancestors to a loss of function. Model violation—particularly when a state has higher equilibrium frequency in the model than it does at a site in a real protein—might inflate the posterior probability of a deleterious residue that is observed rarely or never in extant sequences; if such states are associated with posterior probabilities of a few percent or even less, it is likely that one or more will be sampled somewhere along a protein's length. In contrast, the AltAll proteins incorporate only states with probability greater than the plausibility cutoff (0.20 in our work), which is far more likely to occur when the state is present in multiple sequences descending from branches near the node of interest rather than due to model violation alone.

Third, sampling error may tend to cause a bias towards non-functionality in small samples of Bayesian ancestral proteins. In principle, it should be possible to estimate the function of the unknown ancestral sequence by sampling from the distribution of possible ancestral sequences, characterizing the function of each sampled protein, and taking the average of those observations, each weighted by the posterior probability of its sequence. In practice, however, obtaining a representative sample may be impossible. When uncertainty is high, the number of sequences with very low probabilities and compromised functions is immense. Random sampling will capture these sequences, but may often miss the rarer, functional sequences with higher posterior probabilities. This occurred for AncSR1-LBD: there are a large number of high-probability functional sequences in the universe of possible ancestors, including the ML sequence, the AltAll sequence, the five proteins with the highest probabilities in the sampled ensemble and many of the virtually innumerable proteins between these sequences. None of these, however, was included in the random sample we picked for experimental characterization, which consisted entirely of poorly functioning proteins with posterior probabilities many orders of magnitude lower than the functional sequences. If even one functional sequence with higher posterior probability had been sampled, it would have dominated the weighted estimate of the ancestral protein's function. But because these sequences are so rare relative to the vast number of nonfunctional sequences, they can be stochastically excluded from small samples, leading to a bias towards non-functionality. When uncertainty is

high, the sample size required to ensure a reasonably accurate representation of function across the posterior distribution of sequences is likely to be experimentally intractable.

Distinguishing among these possible explanations will require further research. Whatever the causes of the susceptibility of Bayesian samples of ancestral sequences towards non-functionality when uncertainty is high, the approach is best used to evaluate the posterior probabilities of specific hypotheses about function, and the evidence about function provided by any specific sequence should be weighted by the sequence's posterior probability. Bayesian sampling should not be used to evaluate robustness per se, where one or even several observations of a nonfunctional or poorly functioning sequence is taken as evidence against the hypothesis that the ancestral protein functioned.

Vast Networks of Functional Proteins in Sequence Space

Finally, our results have implications for our understanding of the distribution of protein functions in sequence space. The proteins we studied allowed up to 90 amino acids—more than a third of all residues in the sequence—to be substituted without changing the function. This result indicates that the network of protein sequences that shares the same function is truly vast. This does not imply that all sequences so distant from the ML reconstruction have similar functions, of course, because so many mutations are deleterious, so the probability that any protein would tolerate 90 random replacements is virtually zero. Rather, our observations indicate that the network of functionally similar sequences extends, presumably sparsely, over an extraordinarily large region of sequence space, and that the substitutions that occurred during evolution constitute a set of very long bridges—possibly quite narrow ones—through this space. Ancestral sequence reconstructions therefore appear to provide useful information about the distant addresses of presumably rare proteins with similar functions.

Methods

Ancestral Sequence Reconstruction

Posterior probability distributions at each site were inferred using the same data, evolutionary model, and phylogeny as in the original published reports (Eick et al. 2012; McKeown et al. 2014; Anderson et al. 2016). Specifically, Anc-gk$_{dup}$, AncGK1$_{PID}$, AncSR1DBD, AncSR1LBD, and AncSR2LBD were inferred from protein alignments of 224, 213, 184, and 213 sequences, respectively, using the best-fitting evolutionary model chosen using Prottest (Abascal et al. 2005), the ML phylogeny as described in each paper, and the maximum likelihood method of Yang et al. (1995) as implemented in Lazarus software. Further details are provided in references (Eick et al. 2012; McKeown et al. 2014; Anderson et al. 2016).

Each ancestral sequence is a string consisting of one amino acid state at each site in the protein. The ML sequence comprises the state with the maximum posterior probability at each site. The AltAll sequence comprises the state with the second-highest posterior probability if that state has PP > 0.20 and the maximum likelihood state otherwise.

Each Bayesian sequence was generated by sampling at each site one state from the posterior probability distribution at that site. For each protein, an ensemble of at least 1 million sequences was generated, from which 5 were chosen at random for experimental characterization. For AncSR1LBD, the five sequences in the ensemble with the highest posterior probability were also assessed. Sequences for all ancestral proteins have been deposited with GenBank (supplementary table S4, Supplementary Material online).

The posterior probability of a sequence is the product over sites of the posterior probability of the state at each site. The expected number of errors in a reconstructed sequence is the sum over sites of one minus the posterior probability at each site.

GK Domain Functional Characterization

Coding sequences for ancestors were generated and synthesized (GenScript) after codon optimization for expression in *E. coli* and then inserted into standard plasmid vectors for *E. coli* expression. Plasmids were transformed into BL21(DE3) competent *E. coli* that were grown on Luria broth (LB) plates under ampicillin (100 μg/mL) selection overnight. Fifty-milliliter starter cultures were generated and allowed to grow until turbid. Four liters of LB were inoculated with starter culture, allowed to grow to an OD600 of 0.7, and then induced by addition of 200 μM IPTG overnight at 20 °C. Protein purification was carried out using sequential NiNTA affinity, anion exchange, and size-exclusion chromatographies. All proteins eluted as predicted monomers from the size-exclusion column, and their purities were assessed to be >95% by Coomassie staining of a SDS-PAGE gel. Proteins were concentrated using Vivaspin concentrators (Sigma-Aldrich), flash-frozen in liquid nitrogen, and stored at −80 °C in buffer (20 mM Tris, pH 7.5, 150 mM NaCl, 1 mM DTT).

To assess the catalytic function of ancestral guanylate kinases, we used a coupled-enzyme assay as described previously (Agarwal et al. 1978). Briefly, phosphoryl-transfer is detected by oxidation of NADH by lactate dehydrogenase, which results in a decrease in absorbance at 340 nm. Guanylate kinase enzyme was assayed at 50–200 nM in assay buffer (100 mM Tris, pH 7.5, 100 mM KCl, 10 mM MgCl$_2$, 1.5 mM sodium phosphoenol pyruvate, 300 mM NADH, 4 mM ATP, pyruvate kinase, and lactate dehydrogenase). Initial GMP concentrations ranged from 500 to 1 mM. The reaction was initiated by addition of GMP and brief mixing, followed by absorption measurement using a Tecan Safire plate reader. Reactions were carried out at 30 °C and measured at 15 s intervals for 30 cycles. Data were analyzed and plotted using GraphPad Prism. Reaction rates are plotted as initial rate of ADP production.

Fluorescence anisotropy binding assays were performed on a Tecan Sapphire 96-well plate reader equipped with automated polarizers at room temperature. A FITC-labeled ∼20 amino acid peptide (GenScript) of the *D. melanogaster* Pins-Linker peptide (0.25 μM) was titrated with each binding partner in phospho-buffered saline solution with 1 mM DTT. Titration curves were fitted with a classical one-site binding model.

Steroid Receptor DBD Functional Characterization

DNA sequences encoding the AncSR1 ML, AltAll, and Bayes DBDs were codon-optimized for expression in mammalian cells, synthesized by GenScript, cloned into the mammalian expression vector pCMV-AD (Stratagene), and fused in-frame with the NF-κB activation domain. We used the reporter plasmid containing the estrogen response element, pGL3-4(EREc38) described by Tyulmenkov et al. (2000) for functional assays. This reporter contains four copies of the estrogen receptor recognition sequence upstream of a luciferase reporter gene. The SRE reporter plasmid was made by introducing AGAACA half sites into this sequence and having this promoter sequence synthesized by Blue Heron Biotechnology, followed by cloning of the promoter into the pGL3-4(EREc38) plasmid.

Activation of the reporter constructs was assessed by luciferase reporter assays in CV-1 cells as described in detail in McKeown et al. (2014). Mutants were generated using site-directed mutagenesis (QuikChange Lightning, Stratagene), and all clones were verified by sequencing (Genewiz, Inc.).

Steroid Receptor LBD Functional Characterization

DNA sequences encoding 11 alternate AncSR1LBDs were synthesized by GenScript after codon optimization for expression in Chinese hamster ovary cells (CHO-K1). The AltAll version of AncSR2LBD (26 amino acid differences from the ML ancestor) was also synthesized by Genscript. All LBD sequences were cloned in-frame into the pSG5-Gal4 DBD vector with the human GR hinge and verified by Sanger sequencing analysis. Sensitivity of the alternative LBDs to various hormones was determined by luciferase reporter assays in CHO-K1 cells as described previously (Eick et al. 2012; Harms et al. 2013). Dose–response relationships were estimated using nonlinear regression in Prism4 software (GraphPad Software, Inc.); fold increases in activation were calculated relative to the vehicle-only (ethanol) control. Sensitivity to hormone is reported as the EC50 value—the hormone concentration at which 50% activation of the receptor was observed. All assays were repeated at least in triplicate. Mutagenesis was performed using Stratagene's QuickChange Lightning site-directed mutagenesis kit and clones were verified by Sanger sequencing (Genewiz, Inc.).

Sequence Analysis

Amino acids were classified (Betts and Russell 2003) based on their biochemical properties: hydrophobic aliphatic (I, V, L), hydrophobic aromatic (F, Y, W), hydrophobic other (A, G, M), polar negative (E, D), polar positive (K, R), polar neutral (Q, N, C, S, T), or other (P, H). Substitutions among sites within a class were classified as conservative, whereas those between classes were designated non-conservative. Non-radical non-conservative changes were considered to be the following: hydrophobic aliphatic to/from hydrophobic other; hydrophobic aromatic to/from His; polar neutral to/from polar negative; polar neutral to/from polar positive; polar positive to/from His. All other types of changes were considered radical. Residues were also classified as buried or exposed based on their solvent exposure in the crystal structure of the human ERα LBD (PDB 1ERE), as assessed using ASAView (Ahmad et al. 2004). Residues with a relative exposed area of $\leq$10% were considered buried, whereas the remaining residues were classified as exposed (Le and Gascuel 2010). Results are summarized in supplementary table S3, Supplementary Material online.

Supplementary Material

Supplementary figures S1–S2 and tables S1–S4 are available at *Molecular Biology and Evolution* online.

Acknowledgments

References

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

Agarwal KC, Miech RP, Parks REJ. 1978. Guanylate kinases from human erythrocytes, hog brain, and rat liver. *Methods Enzymol.* 51:483–490.

Ahmad S, Gromiha M, Fawareh H, Sarai A. 2004. ASAView: database and tool for solvent accessibility representation in proteins. *BMC Bioinformatics* 5:51.

Akanuma S, Nakajima Y, Yokobori S, Kimura M, Nemoto N, Mase T, Miyazono K, Tanokura M, Yamagishi A. 2013. Experimental evidence for the thermophilicity of ancestral life. *Proc Natl Acad Sci U S A.* 110:11067–11072.

Akanuma S, Yokobori S, Nakajima Y, Bessho M, Yamagishi A. 2015. Robustness of predictions of extremely thermally stable proteins in ancient organisms. *Evolution* 69:2954–2962.

Anderson DP, Whitney DS, Hanson-Smith V, Woznica A, Campodonico-Burnett W, Volkman BF, King N, Thornton JW, Prehoda KE. 2016. Evolution of an ancient protein function involved in organized multicellularity in animals. *Elife* 5:e10147.

Anderson DW, McKeown AN, Thornton JW. 2015. Intermolecular epistasis shaped the function and evolution of an ancient transcription factor and its DNA binding sites. *Elife* 4:e07864.

Bar-Rogovsky H, Hugenmatter A, Tawfik DS. 2013. The evolutionary origins of detoxifying enzymes: the mammalian serum paraoxonases (PONs) relate to bacterial homoserine lactonases. *J Biol Chem.* 288:23914–23927.

Bar-Rogovsky H, Stern A, Penn O, Kobl I, Pupko T, Tawfik DS. 2015. Assessing the prediction fidelity of ancestral reconstruction by a library approach. *Protein Eng Des Sel.* 28:507–518.

Betts MJ, Russell RB. 2003. Amino acid properties and consequences of substitutions. In: Barnes MR, Gray IC, editors. Bioinformatics for Geneticists. New Jersey: John Wiley & Sons, Ltd.

Bickelmann C, Morrow JM, Du J, Schott RK, van Hazel I, Lim S, Müller J, Chang BS. 2015. The molecular origin and evolution of dim-light vision in mammals. *Evolution* 69:2995–3003.

Boucher JI, Jacobowitz JR, Beckett BC, Classen S, Theobald DL. 2014. An atomic-resolution view of neofunctionalization in the evolution of apicomplexan lactate dehydrogenases. *Elife* 3.

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

Bridgham JT, Keay J, Ortlund EA, Thornton JW. 2014. Vestigialization of an allosteric switch: genetic and structural mechanisms for the evolution of constitutive activity in a steroid hormone receptor. *PLoS Genet.* 10:e1004058.

Carrigan MA, Uryasev O, Frye CB, Eckman BL, Myers CR, Hurley TD, Benner SA. 2015. Hominids adapted to metabolize ethanol long before human-directed fermentation. Proc Natl Acad Sci U S A. 112:458–463.

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

Chandrasekharan UM, Sanker S, Glynias MJ, Karnik SS, Husain A. 1996. Angiotensin II-forming activity in a reconstructed ancestral chymase. Science 271:502–505.

Chang BS, Jonsson K, Kazmi MA, Donoghue MJ, Sakmar TP. 2002. Recreating a functional ancestral archosaur visual pigment. Mol Biol Evol. 19:1483–1489.

Chang BSW, Matz MV, Field SF, Müller J, van Hazel I. 2007. Dealing with model uncertainty in reconstructing ancestral proteins in the laboratory: examples from ancestral visual pigments and GFP-like proteins. In: Liberles D, editor. Ancestral Sequence Reconstruction. Oxford: Oxford University Press.

Clifton BE, Jackson CJ. 2016. Ancestral protein reconstruction yields insights into adaptive evolution of binding specificity in solute-binding proteins. Cell Chem Biol. 23:236–245.

Devamani T, Rauwerdink AM, Lunzer M, Jones BJ, Mooney JL, Tan MA, Zhang ZJ, Xu JH, Dean AM, Kazlauskas RJ. 2016. Catalytic promiscuity of ancestral esterases and hydroxynitrile lyases. J Am Chem Soc. 138:1046–1056.

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

Finnigan GC, Hanson-Smith V, Stevens TH, Thornton JW. 2012. Evolution of increased complexity in a molecular machine. Nature 481:360–364.

Gaucher EA, Govindarajan S, Ganesh OK. 2008. Palaeotemperature trend for Precambrian life inferred from resurrected proteins. Nature 451:704–707.

Gaucher EA, Thomson JM, Burgan MF, Benner SA. 2003. Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins. Nature 425:285–288.

Goldschmidt V, Ciuffi A, Ortiz M, Brawand D, Munoz M, Kaessmann H, Telenti A. 2008. Antiretroviral activity of ancestral TRIM5alpha. J Virol. 82:2089–2096.

Groussin M, Hobbs JK, Szöllősi GJ, Gribaldo S, Arcus VL, Gouy M. 2015. Toward more accurate ancestral protein genotype–phenotype reconstructions with the use of species tree-aware gene trees. Mol Biol Evol. 32:13–22.

Guo HH, Choe J, Loeb LA. 2004. Protein tolerance to random amino acid change. Proc Natl Acad Sci U S A. 101:9205–9210.

Hanson-Smith V, Kolaczkowski B, Thornton JW. 2010. Robustness of ancestral sequence reconstruction to phylogenetic uncertainty. Mol Biol Evol. 27:1988–1999.

Harms MJ, Eick GN, Goswami D, Colucci JK, Griffin PR, Ortlund EA, Thornton JW. 2013. Biophysical mechanisms for large-effect mutations in the evolution of steroid hormone receptors. Proc Natl Acad Sci U S A. 110:11475–11480.

Hart KM, Harms MJ, Schmidt BH, Elya C, Thornton JW, Marqusee S. 2014. Thermodynamic system drift in protein evolution. PLoS Biol. 12:e1001994.

Hobbs JK, Shepherd C, Saul DJ, Demetras NJ, Haaning S, Monk CR, Daniel RM, Arcus VL. 2012. On the origin and evolution of thermophily: reconstruction of functional precambrian enzymes from ancestors of Bacillus. Mol Biol Evol. 29:825–835.

Howard CJ, Hanson-Smith V, Kennedy KJ, Miller CJ, Lou HJ, Johnson AD, Turk BE, Holt LJ. 2014. Ancestral resurrection reveals evolutionary mechanisms of kinase plasticity. eLife 3.

Jermann TM, Opitz JG, Stackhouse J, Benner SA. 1995. Reconstructing the evolutionary history of the artiodactyl ribonuclease superfamily. Nature 374:57–59.

Kaiser SM, Malik HS, Emerman M. 2007. Restriction of an extinct retrovirus by the human TRIM5alpha antiviral protein. Science 316:1756–1758.

Le SQ, Gascuel O. 2010. Accounting for solvent accessibility and secondary structure in protein phylogenetics is clearly beneficial. Syst Biol. 59:277–287.

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

McKeown AN, Bridgham JT, Anderson DW, Murphy MN, Ortlund EA, Thornton JW. 2014. Evolution of DNA specificity in a transcription factor family produced a new gene regulatory module. Cell 159:58–68.

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

Pollock DD, Chang BSW. 2007. Dealing with uncertainty in ancestral sequence reconstruction: sampling from the posterior distribution. Oxford: Oxford University Press.

Pupko T, Pe'er I, Shamir R, Graur D. 2000. A fast algorithm for joint reconstruction of ancestral amino acid sequences. Mol Biol Evol. 17:890–896.

Risso VA, Gavira JA, Gaucher EA, Sanchez-Ruiz JM. 2014. Phenotypic comparisons of consensus variants versus laboratory resurrections of Precambrian proteins. Proteins 82:887–896.

Risso VA, Gavira JA, Mejia-Carmona DF, Gaucher EA, Sanchez-Ruiz JM. 2013. Hyperstability and substrate promiscuity in laboratory resurrections of Precambrian beta-lactamases. J Am Chem Soc. 135:2899–2902.

Smith SD, Wang S, Rausher MD. 2013. Functional evolution of an anthocyanin pathway enzyme during a flower color transition. Mol Biol Evol. 30:602–612.

Starr TN, Thornton JW. 2016. Epistasis in protein evolution. Protein Sci. 25:1204–1208.

Steindel PA, Chen EH, Wirth JD, Theobald DL. 2016. Gradual neofunctionalization in the convergent evolution of trichomonad lactate and malate dehydrogenases. Protein Sci. 25:1319–1331.

Storz JF. 2016. Causes of molecular convergence and parallelism in protein evolution. Nat Rev Genet. 17:239–250.

Thomson JM, Gaucher EA, Burgan MF, De Kee DW, Li T, Aris JP, Benner SA. 2005. Resurrecting ancestral alcohol dehydrogenases from yeast. Nat Genet. 37:630–635.

Thornton JW, Need E, Crews D. 2003. Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling. Science 301:1714–1717.

Tyulmenkov VV, Jernigan SC, Klinge CM. 2000. Comparison of transcriptional synergy of estrogen receptors alpha and beta from multiple tandem estrogen response elements. Mol Cell Endocrinol. 165:151–161.

Ugalde JA, Chang BS, Matz MV. 2004. Evolution of coral pigments recreated. Science 305:1433.

Wheeler LC, Lim SA, Marqusee S, Harms MJ. 2016. The thermostability and specificity of ancient proteins. Curr Opin Struct Biol. 38:37–43.

Williams PD, Pollock DD, Blackburne BP, Goldstein RA. 2006. Assessing the accuracy of ancestral protein reconstruction methods. PLoS Comput Biol. 2:e69.

Williams SG, Harms MJ, Hall KB. 2013. Resurrection of an urbilaterian U1A/U2B"/SNF protein. J Mol Biol. 425:3846–3862.

Wilson C, Agafonov RV, Hoemberger M, et al. 2015. Kinase dynamics. Using ancient protein kinases to unravel a modern cancer drug's mechanism. Science 347:882–886.

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

Yokoyama S, Radlwimmer FB. 2001. The molecular genetics and evolution of red and green color vision in vertebrates. Genetics 158:1697–1710.

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

Zhang J, Rosenberg HF. 2002. Complementary advantageous substitutions in the evolution of an antiviral RNase of higher primates. Proc Natl Acad Sci U S A. 99:5486–5491.