THE UNIVERSITY OF CHICAGO

EPISTASIS, CONTINGENCY, AND EVOLVABILITY IN THE SEQUENCE SPACE OF ANCIENT PROTEINS

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN BIOCHEMISTRY AND MOLECULAR BIOPHYSICS

BY

TYLER NELSON STARR

CHICAGO, ILLINOIS

AUGUST 2018

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank my advisor, Joe Thornton, for his contributions to my growth as a scientist. Thank you, Joe, for creating a scientific atmosphere with the right blend of support and independence for me to develop intellectually during my graduate studies. Thank you for teaching me to think critically, write effectively, and conduct science rigorously. Thank you for encouraging me to think big, keep a broad perspective, and never lose sight of "what is the question?" Thank you for your support in attending meetings, completing applications, and the occasional celebratory drink. And thank you for encouraging and modeling a healthy work-life balance.

Thank you to my lab-mates, past and present, for broad scientific discussions, irreplaceable advice and feedback, and welcome social distractions. From trivia to journal club, from conferences to lab meetings, and from drinks out to lunch in – these interactions formed a core part of my graduate experience.

Thank you to my family – my parents, Matt and Gail, my siblings, Megan and Brian, my siblings-in-law, Joe and Charisse, my niece, Rylee, my aunts, uncles and cousins, and my newly inducted family-in-law. You have provided such a fantastic support network over the years. The family vacations, the phone and text conversations, the games, the holidays – though not academic endeavors, these experiences played a crucial role in my ability to successfully complete my graduate work. Your unwavering support, genuine interest, and deep presence in my daily life helped me in unexplainable ways. And a special thanks to my grandparents – including my grandmother, Marilyn Starr, who graduated with a bachelor's degree in Biology

from University of Chicago years ago – for your love and support, and for establishing a firm family value in education and intellectual achievement.

Thank you to my dog, Bishop, for forcing me to get out on the Lakefront Trail each day, rain or shine, wind or snow. Any novel idea that might be found in this work probably originated during one of these dog-walk musings, and your insatiable desire to snuggle helped me fall asleep on nights when my mind was reeling.

And last, thank you to my husband, Devon. After meeting during my first year of graduate school, our relationship has developed hand-in-hand with this graduate experience, culminating in our wedding two months ago. Your support during this process was invaluable. You listened to me when I was excited or needed to vent, you supported me (and joined me!) when I traveled for conferences, you freely discussed non-science topics when it was clear I needed something else, you forced me out into the city and into social situations, and you brought joy to every single day of the last five years. It may be cliché, but I really mean it when I say: it simply wouldn't have been the same without you.

Abstract

Many fundamental questions in the molecular evolution of proteins—the roles of contingency and determinism in evolutionary processes, the effect of epistatic substitutions in structuring available paths, and the characteristics of trajectories of functional innovation—depend on the distribution of functional proteins in sequence space and knowledge of how evolution proceeded across this space. However, empirical insight into the historical sequence spaces over which proteins evolved has only recently become accessible with the development of model systems in protein evolution and the advent of high-throughput deep mutational scanning approaches. In my thesis work, I combined two recently developed experimental tools: ancestral protein reconstruction – a phylogenetic technique for inferring the sequences of ancient proteins and experimentally charting their evolutionary history – and deep mutational scanning – an experimental strategy for functionally characterizing large libraries of protein variants. By combining ancestral protein reconstruction and deep mutational scanning for the first time, I explored the mechanistic basis for and evolutionary significance of epistasis, contingency, and evolvability in protein functional evolution. This work reveals how chance factors play a dominant role in the outcomes realized in evolution, how interactions between protein residues enhance the ability of evolution to reach protein sequences with novel functional properties, how the windows of mutational accessibility fluctuate over evolutionary time, and the genetic and biophysical features that give rise to these molecular evolutionary phenomena.

Chapter 1

Introduction

1.1 Sequence space and protein evolution

The diversity of modern protein sequences, folds, and functions was generated by a long and convoluted evolutionary process. Through cycles of mutation, drift, and selection, proteins have emerged with remarkable functional capabilities and beautiful three-dimensional structures. However, the canvas on which this process unfolds – the map between protein sequence, structure, and function (1) – remains elusive. Somewhere out there is a complete map, that annotates every possible polypeptide sequence with its physical and biochemical properties (2). Unfortunately, there are not enough atoms in the universe with which to store this information, and we, as biologists, will never be able to explore more than an infinitesimal fraction of this space. Yet this 'sequence space' and its functional annotation underlies much that we do as biologists – it contains the information that would allow us to predict the consequences of missense mutations for human health (3); it contains the information that would allow us to build new proteins not yet discovered in terrestrial evolution, for example, to combat pathogens (4) or catalyze novel chemistries (5); and it contains the information that would allow us to predict future (6) and rationalize historical (7) evolutionary trajectories.

Many sub-disciplines of biology traverse sequence space in their routine study. Much of modern biochemistry relies on the characterization of mutations to proteins of interest, taking short, pointed forays into sequence space. Protein design efforts jump into uncharted waters of sequence space, to identify protein folds and functions that evolution has not yet come close to exploring (8). And studies of protein engineering and evolution trace long pathways through

sequence space (7, 9). Although many disciplines work within its backdrop, until recently, we have lacked the methods to comprehensively map the sequence-function landscape in an efficient manner.

1.2 Deep mutational scanning

The advent of high-throughput DNA sequencing has opened up new experimental methods for charting protein sequence-function landscapes. These 'deep mutational scans' (10) characterize function within large protein variant libraries via high-throughput-sequencing-based readouts. Broadly, a deep mutational scan subjects a rationally designed library of protein mutants to a selection or screen in which the frequency of each genotype in the library changes in proportion to some function of interest; by using deep sequencing to characterize each genotype's frequency before and after the selection, a measure of the function of interest can be ascribed to each genotype in the library (11, 12). This method is therefore different from a more traditional library selection, because instead of just randomly selecting a subset of the most active library genotypes, it assigns scores to all genotypes in the library, enabling a comprehensive annotation of the sequence-function landscape within some limited mutational radius.

These technologies have allowed huge libraries of protein variants to be characterized in parallel, enabling the rapid and efficient annotation of segments of sequence space larger than could be done previously. By constructing libraries containing all single mutations across the length of a protein sequence, the site-specific amino acid preferences of a protein can be determined (13-17), revealing how protein architecture impacts site-specific mutational tolerance. By comparing single-mutant libraries across protein homologs, researchers have

investigated how amino acid preferences change over evolutionary time (18-20). By characterizing libraries containing all single and double mutants across the length of a protein, the prevalence and characteristics of epistasis in the local sequence neighborhood of proteins can be determined (21). And by assaying higher-order combinations of mutants, combinatorially complete subsets of sequence space can be fully annotated (22-26), enabling exploration into some of the deeper factors that contribute to the structure of sequence-function landscapes. Although these designs still only scratch the surface of the expansive space of possible sequences, they enable new insights into its properties and the factors that contribute to its structure.

1.3 Epistasis

Epistasis describes the non-additivity of mutational effects – for example, when a double mutant behaves differently than would be expected from characterization of either of its constituent single mutations alone. Epistasis is a central phenomenon influencing the structure of the protein sequence-function landscape (27) – if every mutation had the exact same functional effect across every possible protein background, what a simple world we would live in! Because of epistasis, evolutionary trajectories exhibit historical contingency, meaning that the mutations accessible to an evolving protein depend intimately on the mutations that occurred in its prior history (7). Because of epistasis, even proteins that fulfill the same function in all living organisms have not yet fully explored the constellation of genotypes that could possibly encode this function, indicating that the 'protein universe' is still expanding in its search through sequence space (28).

There remains disagreement, however, on the pervasiveness of epistasis in protein evolution: computational simulations (29, 30) and comparative sequence analysis (28, 31, 32) suggest that epistasis causes rampant turnover in site-specific mutational tolerance over long-term protein evolution, yet experimental studies have not always borne this result out (18, 33-35). Furthermore, there is not yet a consensus on the mechanisms by which epistatic interactions emerge, and how different types of epistasis might influence evolutionary processes. The literature on epistasis in protein evolution is complex and deserves a more detailed elaboration than can be done in this Introduction: Chapter 2 is devoted toward reviewing this literature in detail and identifying the central questions about epistasis in protein evolution that remain.

1.4 Chance and determinism

The degree to which the outcomes of evolution are deterministic or subject to chance is a longstanding question in evolutionary biology (36-43). If evolution is deterministic, then replicate evolutionary trajectories under the same conditions will reach the same outcome; however, if chance factors are at play, different outcomes emerge even under identical conditions. This question cannot typically be addressed by looking at natural evolution, which has only proceeded once – even when investigating convergent evolution, the starting points of each trajectory might subtly differ, and the selective conditions might not be identical. Instead, biologists have turned to laboratory evolution experiments to address this question (44): by evolving a protein or a population many times in parallel under the same selection pressure, the degree to which identical phenotypes and genotypes emerge reveals the conditions under which evolution behaves deterministically or idiosyncratically.

From these types of experimental evolution studies, some aspects of evolution do indeed seem deterministic: for example, selection will deterministically drive populations toward higher fitness – even if the underlying sequence changes that drive this adaptation might differ among populations (45). Recent studies that evolve a protein toward some defined function many times in parallel suggest that protein evolution can be deterministic at the level of function, but the specific genetic and biophysical changes that occur in each population can differ due to chance: either stochasticity (randomness), in which mutations simply happen to reach fixation in a given population due to chance (43), or contingency, in which early chance events dictate which steps are taken at later branch points (40, 41, 43).

The degree to which chance influences the outcomes of protein evolution ultimately depends on the density and connectivity between genotypes with some function in sequence space. If functional networks – the mutationally connected network of genotypes sharing some common function (1, 46) – are sparsely connected, then there are few paths that can be taken through sequence space under selection for some function, so replicate trajectories will tend to follow the same path toward the same functional optimum under identical conditions. On the other hand, if many routes through a densely-connected functional network are available, different outcomes will occur due to stochastic outcomes at each branch point along a trajectory, compounded by the dependence of later steps on previous ones.

Though the characteristics of functional networks have been pondered for years, only with the advent of deep mutational scanning can we comprehensively map their topology in defined regions of sequence space, thereby opening the possibility of determining how the distribution of functions in sequence space creates roles for chance versus determinism in protein evolution.

1.5 Evolvability

The capacity of proteins to evolve novel functions is a remarkable phenomenon: proteins can maintain their core structure and function while accumulating dozens or hundreds of substitutions (28), yet in other contexts, a single mutation can alter a protein's structure (47) or function (48). The accessibility of mutations that alter a protein's function is often referred to as its 'evolvability.' Previous work on the evolvability of protein functions has focused on the role of nonspecific stabilizing mutations in protein evolvability (49-51). In various systems, it has been observed that mutations that stabilize a protein's free energy of folding simultaneously enhance its evolvability, allowing it to tolerate previously-disallowed mutations that also endow it with new functional properties but are mildly destabilizing (49).

Other factors that contribute to protein evolvability, however, have remained underexplored. The above examples on the role of stabilizing mutations in protein evolvability relate to the phenomenon of 'nonspecific epistasis' (7), which we elaborate in Chapter 2. However, there is no rigorous understanding of how so-called 'specific epistasis' (also elaborated in Chapter 2) contributes toward the evolvability of protein functions.

1.6 Ancestral protein reconstruction

In this thesis, I will address the above phenomena in the context of ancestrally reconstructed proteins. Ancestral protein reconstruction is a powerful technique for framing studies of protein evolution (52, 53). This phylogenetic technique reconstructs representative sequences of ancestral proteins, which can be cloned and characterized through typical molecular biology and biochemical assays. This approach has proven effective at identifying the historical

substitutions that underlie shifts in protein functions and identifying the biophysical mechanisms by which they exert their effects.

To date, studies on ancestral proteins have focused on understanding the effects of only those substitutions that happened to occur along a particular evolutionary lineage. While fruitful, this approach places evolutionary 'blinders' on our view of protein evolution: just because some historical substitution altered protein function via some particular mechanism, that doesn't indicate that *all possible* routes toward the derived function needed to proceed in this very same way. A new synergy is now possible: by conducting deep mutational scans in ancestral proteins, we can illuminate the broader sequence-function landscape over which historical evolutionary trajectories unfolded. By conducting deep mutational scans across related ancestral protein backgrounds, we can ask how specific historical substitutions altered the availability and characteristics of trajectories of functional change. By conducting targeted mutational scans of the sites containing key function-switching substitutions, we can identify whether an evolutionary trajectory represented a unique, deterministic outcome, or whether it was simply one of many possible alternatives. And by decomposing the sequence-function landscape into its underlying genetic and structural determinants, we can address how the biophysics of a protein's structure give rise to the landscape properties that determine its evolution.

1.7 Research questions and approach

In this thesis, I combine mutational scanning approaches with ancestral protein reconstruction, to explore the prevalence and characteristics of epistasis among amino acid substitutions, how this epistasis impacts the structure of ancient protein sequence-function

landscapes, and how these landscapes create roles for chance factors in the outcomes of evolution.

In Chapter 2, I review the literature on epistasis in protein evolution to address several fundamental questions that are debated in the field: What is the prevalence of epistasis between pairs of random mutations in proteins? What is the prevalence of epistasis between substitutions that accumulate during long term protein evolution? And, what are the underlying mechanisms by which epistatic interactions emerge? By reviewing the literature, I describe an emerging picture of pervasive epistasis, in which the physical and biological effects of mutations change over the course of evolution in a lineage-specific fashion. I describe two broad classes of epistatic interactions, which arise from different biophysical mechanisms and have different effects on evolutionary processes.

In Chapter 3, in collaboration with Dan Bolon, Julia Flynn, and Parul Mishra at the University of Massachusetts Medical School, we perform a medium-throughput mutational scan of substitutions that occurred during 1 billion years of a model protein's evolution, revealing pervasive epistasis among the substitutions that occurred during this interval. These results unite experimental and computational approaches for determining the impact of epistasis on long-term protein evolution, and illustrate how epistasis continually opens and closes windows of mutational opportunity over evolutionary timescales to produce contingent and irreversible evolutionary histories.

In Chapter 4, I perform a deep mutational scanning approach to dissect the roles of determinism and chance in protein evolution. I take an evolutionary transition in protein function whose historical details are well understood, and survey a massive combinatorial library of alternative protein variants to identify and characterize alternative trajectories by which the same

functional transition that occurred during history could have unfolded. I find that the outcome of historical evolution was not unique in its function, biochemical mechanism, or evolutionary accessibility, identifying dominant roles for two forms of chance in the outcome of evolution.

In Chapter 5, I ask to what extent epistasis among mutations impacts the evolvability of new protein functions in sequence space. I analyze the deep mutational scanning dataset described in Chapter 4, to partition the determinants of this sequence-function landscape into its non-epistatic and epistatic components. I explore the biophysical basis for observed epistatic effects, and explore the role of epistasis in structuring the evolvability between two distinct functions in sequence space.

Taken together, this work identifies new characteristics, mechanisms, and evolutionary impacts of epistasis between amino acid mutations, and connects these behaviors to the global properties of the sequence-function landscape. The emerging picture of protein evolution is one that is dominated by the context-dependency of mutational effects, leading to dominant roles for chance factors in the details and outcomes of protein evolution.

Chapter 2

Epistasis in protein evolution

The work described in this chapter was published as: Tyler Starr and Joseph Thornton. "Epistasis in protein evolution." Protein Science 25:1204-1218 (2016).

2.1 Summary

The structure, function and evolution of proteins depend on physical and genetic interactions among amino acids. Recent studies have used new strategies to explore the prevalence, biochemical mechanisms, and evolutionary implications of these interactions – called epistasis – within proteins. Here we describe an emerging picture of pervasive epistasis in which the physical and biological effects of mutations change over the course of evolution in a lineage-specific fashion. Epistasis can restrict the trajectories available to an evolving protein or open new paths to sequences and functions that would otherwise have been inaccessible. We describe two broad classes of epistatic interactions, which arise from different physical mechanisms and have different effects on evolutionary processes. Specific epistasis – in which one mutation influences the phenotypic effect of few other mutations – is caused by direct and indirect physical interactions between mutations, which nonadditively change the protein's physical properties, such as conformation, stability, or affinity for ligands. In contrast, nonspecific epistasis describes mutations that modify the effect of many others; these typically behave additively with respect to the physical properties of a protein but exhibit epistasis because of a nonlinear relationship between the physical properties and their biological effects, such as function or fitness. Both types of interaction are rampant, but specific epistasis has stronger effects on the rate and outcomes of evolution, because it imposes stricter constraints and

modulates evolutionary potential more dramatically; it therefore makes evolution more contingent on low-probability historical events and leaves stronger marks on the sequence, structure, and function of protein families.

2.2 Introduction

A protein's biological functions emerge from its chemical and physical properties, which in turn are determined by the interactions between its amino acid residues in three-dimensional space. It is therefore not surprising that the functional effect of changing an amino acid often depends on the specific sequence of the protein into which the mutation is introduced. This dependency on genetic context has long been called epistasis by geneticists (54). Epistasis is invoked when the combined effect of two or more mutations deviates from that which would be predicted by adding their individual effects (Fig. 2.1).

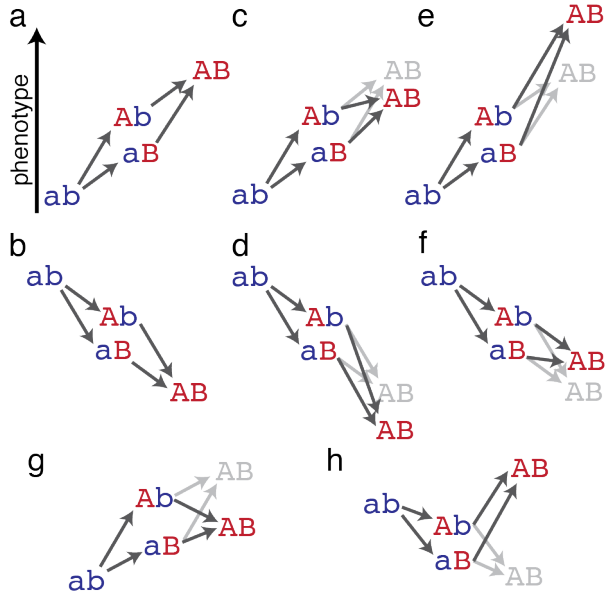


Figure 2.1. Patterns of epistasis between mutations. We use the terms positive and negative epistasis, as suggested by Phillips (54). **a,b**, Mutations a→A and b→B behave additively with respect to the measured phenotype (e.g. stability, fitness): the phenotypic effect of a state at one site is independent of the state of the other. **c,d**, The two mutations exhibit negative epistasis: the

(**Figure 2.1, continued**) double mutant AB has a lower phenotype than would be expected from the effect of A and B alone, regardless of the net direction of the mutational effect. **e,f**, The two mutations exhibit positive epistasis: the double mutant AB has a greater phenotype than would be expected from the effect of A and B alone, regardless of the net direction of the mutational effect. **g**, In contrast to the nonadditivity in *magnitude* of mutational effects in (**c**) through (**f**), the two mutations exhibit negative *sign* epistasis: the sign of the phenotypic effect of b→B changes with respect to the state of the other site. **h**, The two mutations exhibit positive reciprocal sign epistasis: the sign of the phenotypic effect of either mutation changes in the background of the other.

Although studies of epistasis have traditionally focused on genetic interactions between mutations at different loci (54), recent research has begun to address epistasis within proteins – its prevalence, biochemical mechanisms, and impacts on evolution. However, a consensus view of these subjects has not yet emerged. Some papers conclude that epistasis is "rampant" (33) or even the "primary factor" in protein evolution (31), whereas others claim that the frequency and magnitude of epistasis is "sufficiently low" such that it does not strongly affect the patterns of substitution in evolving proteins (55). There is also no clear picture of the mechanisms that cause epistasis: many papers have focused exclusively on epistasis mediated by effects on protein stability (29, 30, 34, 35, 56), although a few have addressed effects on protein conformation, ligand binding, and allostery (57-59).

These disagreements reflect, at least in part, the lack of a unified discussion of the parallels and contrasts now emerging from the diverse modes of analysis applied to epistasis and its effects on protein evolution. Here we attempt such a unified view, focusing on the following specific questions: How important a factor is epistasis in changing the effects of mutations during the course of evolutionary history? Does epistasis typically amplify or dampen the effect of individual mutations? Does most evolutionarily relevant epistasis reflect very specific interactions between mutations – for example, with only one potential "permissive" mutation that can open the path for another specific mutation – or are many-to-one, one-to-many, or

many-to-many interactions more common? What are the molecular mechanisms of interaction that produce each form of epistasis? And how do epistatic interactions of these various types influence the pathways and outcomes of long-term protein evolution?

2.3 Epistasis and protein sequence space

The concept of sequence space provides a useful metaphor for understanding the relationship between a protein's sequence, its physical or biological properties, and its evolution. Sequence space is a multidimensional representation of all possible protein genotypes, each connected to its neighbors by edges representing changes in a single residue (1). Assigning physical or biological properties to each genotype yields a "topological map" of the sequence space, just as a topological map of a geographic landscape assigns elevations to locations defined by their latitudinal and longitudinal coordinates. Epistasis makes the topology of sequence space "rugged" (27), in that the physical or biological effect of a mutation differs in sign or magnitude depending on the sequence background into which it is introduced; just as on a rugged geographical landscape, the change in elevation caused by a step in some direction varies dramatically depending on the starting point.

As proteins evolve, they follow trajectories through sequence space, so this topology also determines how mutation, drift, selection, and other forces can drive genetic and functional evolution. A typical trajectory in natural or directed protein evolution consists of iterative mutational steps between functional proteins, based on the idea that nonfunctional variants of biologically important proteins will usually reduce fitness and therefore be removed by natural selection (1, 60-62). In the absence of epistasis, any mutation that changes protein properties in a beneficial way can be fixed by natural selection, irrespective of the genetic background in which

it occurs; the result is a relatively large number of passable trajectories through sequence space to the functional optimum that combines all of the beneficial sequence states. When epistasis is present, however, a mutation may be beneficial in some backgrounds or deleterious (or neutral) in others; the number of passable trajectories becomes smaller, the fixation of any one mutation may be contingent on the prior occurrence of other specific mutations, and there may be multiple local optima, consisting of mutually conditional beneficial states, isolated from each other by trajectories of low fitness.

Epistasis can therefore create a strong path-dependency in trajectories of protein evolution (40, 41, 43, 63), because the mutations that are stochastically fixed may determine which functional optimum an evolving protein ultimately occupies; these optima may differ not only in primary sequence but also in interesting physical or biological properties. Epistasis can yield evolutionary "dead-ends" in sequence space, from which a potentially beneficial mutation is not immediately accessible; in such cases, a relaxation of selection or even selection for other protein properties is necessary before a trajectory is opened to a superior optimum (57, 64-70). Epistasis can also cause a mutation that confers or improves a function in one protein to have no effect or even be strongly deleterious in a related protein (33, 41, 71); as a result, attempts to leverage natural sequence variation or experimental observations to predict mutational effects or engineer proteins with desired properties often fails (72). These issues highlight why characterizing epistasis – including the breadth of its effect, its mechanistic underpinnings, and its evolutionary impact – is important for our basic understanding of protein biochemistry and evolution.

2.4 Prevalence and strength of epistasis

How prevalent is epistasis within proteins, how strongly does it modulate the effects of mutations, and to what extent does this context-dependence affect long-term evolution? Studies of these questions have used two primary approaches – deep mutational scanning of large numbers of mutations in individual proteins, and analyses of changes in mutational effects across long-term trajectories of protein evolution.

2.4.1 Epistasis in a protein's local sequence neighborhood

Using a recently developed technique called deep mutational scanning, a very large library of mutant versions of some protein of interest can be characterized en masse with respect to some physical or biological property. By analyzing many or all variants that differ by one or two amino acids from a starting protein, it is possible to comprehensively characterize pairwise epistatic interactions in that protein's local sequence neighborhood (10, 21, 73-75). In the absence of epistasis, one can predict the behavior of double mutants with perfect accuracy by adding the effects of their constituent single mutations (R^2 approaches 1 for the correlation between observed and predicted double mutant function). In contrast, in a completely epistatic landscape, the effect of a mutation is completely independent in every single background (R^2 approaches 0). Experiments reveal an intermediate prevalence of epistasis: the properties of single mutants predict double mutant behavior moderately well ($R^2 \sim 0.65$-0.75) (10, 73, 74). This result indicates that strong epistasis is not all-pervasive, but it also points to epistasis that is pervasive and weak or relatively rare and strong. In fact, it appears that both types of interactions are important: a comprehensive study of pairwise interactions in protein G domain 1 (GB1) found strong deviations from additivity (by a factor >2) in ~5% of all pairs of mutations, while

weak epistasis (<2-fold deviation) affected ~30% of pairs (21). Thus, small-effect epistasis is very common, and large-effect epistasis is less so, but still affects a substantial number of mutations.

Does epistasis tend to affect protein properties in one direction more than another? In "negative epistasis" a double mutant's measured phenotype has a smaller value than expected under additivity (e.g. Figs. 2.1c,d,g), whereas in "positive epistasis" the phenotype is greater than predicted (e.g. Figs. 2.1e,f,h) (54). In deep mutational scanning studies of ligand affinity and fitness effects, far more pairs exhibit negative than positive epistasis – with the former group outnumbering the latter by a factor of 3 to 20 (21, 74, 75). Most mutations have deleterious effects on these phenotypes, so negative epistasis in the majority of cases acts synergistically to make double mutants worse than either single mutant alone (Fig. 2.1d) (21, 74, 75). This kind of epistasis would cause weakly deleterious mutations to become progressively less evolutionarily accessible as modifying mutations accumulate.

Of particular importance for evolution is positive sign epistasis (e.g. Fig. 2.1h), in which a pair of deleterious or neutral mutations becomes beneficial when combined. Although far less prevalent than negative epistasis, positive sign epistasis still appears to be widespread. In GB1, most mutations that are deleterious have at least one or more interacting mutations elsewhere in the protein that make the first mutation's effects beneficial or neutral (21). Positive epistasis can open mutational trajectories to combinations of substitutions that would otherwise have been inaccessible. For example, in a high-throughput screen of a mutant protein library for variants that maintained the wild-type function, about 95% of the functional variants recovered would have been predicted to be non-functional from the effects of single mutations alone (22).

These deep mutational scanning studies provide important insights into how epistasis might affect the first stages of an evolutionary process that begins from present-day forms, initially closing many paths to beneficial combinations but sometimes opening new ones. But the strategy leaves untouched important questions about the effect of epistasis on long-term historical protein evolution. For example, mutational scans suggest that many mutations manifest sign epistasis in their interactions, but how frequently does the direction of a mutation's effect actually change during evolution? There is plenty of epistasis in the local sequence neighborhood of a protein, but does this epistasis actually matter in determining proteins' historical trajectories? Is the strength and pervasiveness of epistasis in the immediate neighborhood of extant proteins similar to that in the much larger tracts of sequence space traversed by proteins evolving over hundreds of millions of years? Answering these questions requires direct analysis of epistasis across long-term trajectories of protein evolution.

2.4.2 Epistasis in long-term protein evolution

One way to gain insight into epistasis in real protein evolution is to compare the effects of some mutation on physical or biological properties when it is introduced into different proteins related by evolutionary descent (homologs). Some studies have addressed this question experimentally, while others have used computational approaches to indirectly infer the prevalence and strength of epistasis during long-term evolution.

Experimental comparisons of mutational effects between homologs: Manipulative experiments on protein homologs point to both strong and pervasive effects of epistasis that cause the functional effects of mutations to differ between related proteins. One study tested the functional effect of 168 amino acid differences that separate orthologous enzymes that have

maintained the same function in two bacterial species (33). Each individual residue from one ortholog was introduced into the other: about one third of these "sequence swaps" severely decreased enzyme activity. This result indicates that permissive epistatic interactions made the residue tolerable in its native background, that restrictive epistatic mutations made it intolerable in the other, or both. A similar study examined all combinations of nine variable residues that differ between closely related orthologous proteins and statistically determined both the average effects of each residue on catalytic activity, as well as the variance of its effect across different combinations (76). The standard deviation of every mutation's effect was at least 45% of its average effect (up to 75% in the most extreme case), indicating significant epistasis among the nine sequence differences between the proteins.

These studies demonstrate widespread epistasis, but they do not trace the accumulation of epistatically interacting mutations over time. A recent study addressed this question using the recent evolution of influenza nucleoprotein (34). The mutational trajectory of the protein over the last 39 years was reconstructed. Each of the 39 substitutions that occurred during this trajectory was assessed for its effects on viral RNA transcription when introduced into the sequence context in which it occurred historically and into the sequence from an extinct strain that closely resembles an ancestral version of the protein. Every substitution was neutral in the background in which it occurred, but three were radically deleterious with respect to both function and fitness in the ancestral background, indicating relatively rare but extremely strong epistasis that allowed these mutations to be tolerated later.

The above examples illuminate the variability in mutational effect for states that were actually incorporated into diverging proteins during evolution. But what is the impact of epistasis on the effects of *all* mutations, including those that are never observed because they are

deleterious? A recent study compared site-specific mutational preferences between two influenza nucleoprotein orthologs by assessing the effect on viral fitness of all 19 possible single-amino acid replacement mutations at every site in the two proteins, whether or not they changed during evolution (55). The two proteins differ at only 6% of sites, but significant differences in site-specific amino acid preference were found at 3 to 15% percent of sites (depending on the statistical method used to evaluate differences). Thus, on average, each substitution during the evolution of these two closely related proteins modulated the amino acid preferences at one or two other sites.

Strong epistasis is also apparent in laboratory evolution studies. One study placed a protein under strong selective pressure to evolve a new activity and then reimposed selection for the original activity, a trajectory that involved 28 amino acid changes in all (41). The "ancestral" amino acid state at each of these 28 sites was then introduced singly into the "derived" protein, and the derived states were each introduced into the ancestral protein to test for context-dependence. Almost half of the substitutions were deleterious when swapped into the other background, pointing to widespread epistatic interactions among the sites and states that were substituted during the laboratory evolutionary process.

Comparative sequence analysis: Computational analyses of protein sequence data have investigated epistasis by seeking evidence that the effects of mutations differ among phylogenetic lineages. There are several major "signatures" of epistasis that have been detected in these kinds of studies (Fig. 2.2), which point to a strong and pervasive effect of epistasis on protein evolution.

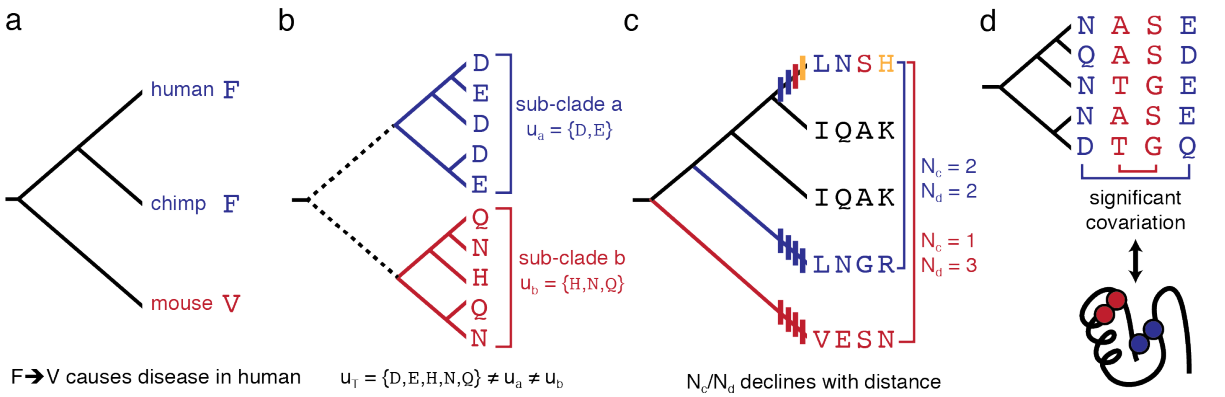


Figure 2.2. Evidence for epistasis in extant sequence data. a, Though a mutation from F to V at a given site is known to cause disease in humans, V is observed as the wild-type state in mouse. **b**, Amino acid usage for a site in a given sub-clade (u_a or u_b) only represents a fraction of the total amino acid usage observed for the site over its long-term evolution (u_T). **c**, Phylogenetic analysis is used to infer the directionality of amino acid substitutions (indicated by vertical bars; additional sequences used to polarize changes leading to bottom two sequences are not shown). The number of paired substitutions to the same amino acid state (convergent substitutions, N_c) decreases with increasing evolutionary distance, relative to the number of paired substitutions to different amino acid states (divergent substitutions, N_d). **d**, Pairs of sites that exhibit significant covariation correspond to protein structural contacts.

First, amino acid states that cause disease in one lineage frequently correspond to a wild-type state in the orthologous protein from other species (Fig. 2.2a) (77-83). These states do not cause disease in the lineages in which they have fixed, so other lineage-specific substitutions must have modulated their effects. Remarkably, of the sequence differences between orthologous proteins in humans and other vertebrates, about 1% of the states from other species are known to cause disease in humans (77, 79); when the incomplete nature of databases of pathogenic mutations is taken into account, it is estimated that up to 10% of differences between orthologs might correspond to epistatically modified pathogenic states.

Second, genome-scale alignments of orthologs with conserved functions from distant taxa point to extensive changes in the tolerability of specific mutations over the long term. For example, when clades of related species are examined, a typical sequence site samples only a small subset of the amino acid states that are sampled across its long-term evolution (Fig. 2.2b)

(31), pointing to lineage-specific epistatic constraints. Other studies found that the rate of convergent substitutions between orthologs declines as sequence divergence increases, as expected if the site-specific tolerability of each amino acid changes in a lineage-specific manner (Fig. 2.2c) (28, 84-86). By analyzing these data with an evolutionary model that incorporates epistatic interactions, one study found that an average amino acid substitution switches the tolerability of five other potential mutations, making deleterious mutations at other sites non-deleterious, or vice versa (32). Although the patterns – and particularly the quantitative estimates of the extent of epistasis – identified in these large-scale computational studies depend on assumptions and statistical models (87), their congruence with experimental studies of specific proteins suggests that epistasis is indeed likely to be pervasive during long-term protein evolution.

Finally, sequence signatures of covariation provide circumstantial evidence for pervasive epistasis in long-term evolution (Fig. 2.2d). Epistasis between residues causes sites within a protein to constrain each other's evolution, and thus the state present at one site in a sequence should provide information about the state at an interacting site. Such signatures of covariation in sequence alignments have been used to correctly predict protein folds from sequence and to produce novel sequences that fold in a desired conformation (88-93). They have also been used to predict and engineer protein-protein interactions (23, 94-96), to predict the functional effects of mutations (97, 98), and to understand other sequence-structure-function relations (99-101). Although the physical basis for the signal of covariation in these analyses has not been established, some of the strongest pairwise covariation terms have been experimentally validated as strong epistatic interactions (97, 98). Further, efforts to design and predict protein structure and function have been far more successful when mutual information among residues is included

in the analysis than when only site-specific amino acid frequency profiles are used, suggesting that covariation signatures have indeed captured distinct and biologically meaningful dependencies among sites (88, 97, 98). That such analyses can capture enough of the relevant details about protein architecture to do this kind of practical work suggests that epistasis is likely to be a strong determinant of protein structure and function.

In silico evolutionary simulations: Epistasis has a similarly pervasive role in computational simulations of neutral (29, 30) and adaptive (102) protein evolution. In these studies, some ancestral protein with a defined structure is allowed to evolve *in silico* under defined population genetic conditions. For example, a recent study simulated the evolution of *argT* for replicate evolutionary trajectories 30 substitutions long under purifying selection to maintain a stable fold, implemented by applying a stability prediction algorithm that uses the protein's known crystallographic structure and a very simple function that relates stability to fitness (30). Each substitution that occurred during a trajectory was then evaluated for its predicted effects on stability and fitness when introduced into every protein sequence that existed at some point during that trajectory. The vast majority of substitutions had different predicted effects on stability and fitness at the time they occurred during the trajectory than they would have if introduced at a different time. Specifically, most substitutions were neutral at the time of their fixation but would have been deleterious at earlier points in the trajectory, reflecting an important role for permissive mutations in the turnover of evolutionarily viable mutations. Epistasis also caused substantial irreversibility in the evolutionary process: once a substitution enabled by an earlier permissive substitution at some other sequence site occurs, then the ancestral state at the permissive site becomes deleterious, making reversion to that state unlikely.

Relationship to epistasis in the local sequence network: How does the epistasis observed in long-term evolution compare with the mixture of large- and small-effect epistasis observed in the local sequence networks of various proteins? Two kinds of analyses bear on this problem but do not clearly resolve it. Computational simulations suggest that both small- and large-effect epistasis have a pervasive influence on proteins' evolutionary trajectories (30); however, the models used in these simulations are highly simplified, so the real-world relevance of their quantitative conclusions is unclear (103). Experimental analyses of the effects of mutations introduced into homologous proteins have detected extensive large-effect epistasis (33, 34, 41, 76), demonstrating its relevance in real-world, historical evolutionary trajectories. The pervasive small-effect epistasis visible in local sequence networks, however, has not been generally observed in these kinds of experiments. It could be that small-effect epistasis does not meaningfully impact natural evolution, but we cannot rule out ascertainment bias as an alternative explanation; quantifying small deviations from additivity is considerably more difficult than establishing the significance of large deviations, and a general view that only large-effect epistasis is worth reporting may be at play, as well.

Taken together, analyses to date point to extensive large-effect epistasis in the local sequence neighborhoods of present day proteins and in the substitutions that become fixed during evolutionary trajectories. Small-effect epistasis is clearly present in sequence neighborhoods; the extent to which it affects and is incorporated into real-world evolutionary trajectories remains unresolved.

2.5 Specificity of epistasis and causal mechanisms

The above examples demonstrate that epistasis is pervasive and, in some cases, strong during the course of long-term evolution. A complete description of epistasis in protein evolution requires more detailed attention to the nature of the interactions, including their specificity, their effects on evolutionary trajectories, and the molecular mechanisms by which mutations interact.

The biological properties of a protein are ultimately determined by its physical properties (such as stability, ligand affinity, or conformational dynamism), which in turn are determined by protein sequence (Fig. 2.3). Nonlinearity in either mapping – from protein sequence to physical property, or physical property to biological property – results in epistasis at the level of function or fitness. We can distinguish two broad classes of epistatic interaction – specific and nonspecific epistasis (7) – which refer to whether the epistasis involves mutations that modify the effects of few or many other potential mutations. The difference in specificity arises from a difference in the biophysical mechanisms that produces each class of epistasis. The two classes also differ, in turn, in the mapping the interaction affects – from sequence to physical property or from physical property to biological characteristic – and in their implications for evolutionary processes.

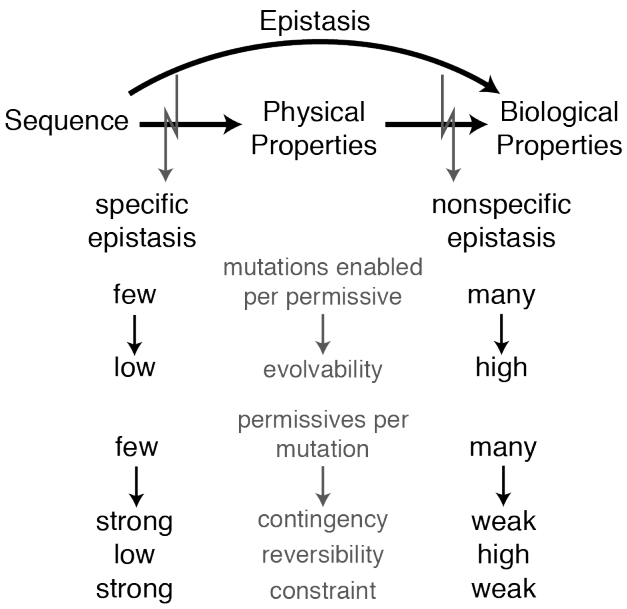


Figure 2.3. Mechanisms of epistasis and their evolutionary implications. Biological properties (e.g. function, fitness) depend on the physical properties of protein molecules (e.g. stability, solubility, affinity for ligand), which in turn depend on the peptide sequence. Evolutionarily relevant epistasis describes nonadditivity in the mapping from sequence to biological properties. Specific epistasis causes nonadditivity in the mapping from sequence to physical properties because of physical interactions between sites. Nonspecific epistasis arises from the intrinsic nonlinear relationship between various physical and biological properties. Specific permissive mutations enable fewer mutations than nonspecific permissive mutations, and therefore have a less dramatic impact on protein evolvability. Similarly, mutations that require a specific permissive mutation to be tolerated have fewer possible permissive mutations than a mutation that can be enabled through a nonspecific effect. This causes specific epistasis to underlie stronger historical contingency, lower reversibility, and stronger long-term evolutionary constraints.

2.5.1 Specific epistasis

In specific epistasis, a mutation modulates the effects of a small number of other mutations. Specific epistasis is typically mediated by physical interactions among residues; these may involve direct interactions between amino acid side chains (41, 43, 47), mutual interaction with other side chains (104) or ligands (68, 105, 106), or a dependence of one mutation on a structural change caused by another (41, 57, 107, 108). Because of these physical interactions,

two specifically epistatic mutations affect a physical property of the protein – such as stability, affinity, catalysis, or dynamic motions – in a nonadditive fashion (Fig. 2.4b).

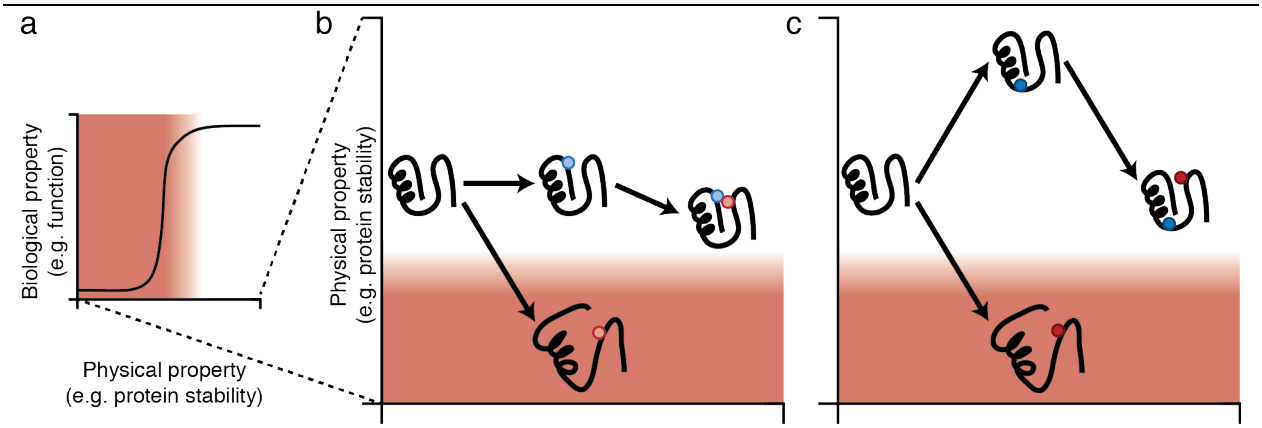


Figure 2.4. Examples of specific and nonspecific epistasis. a, The relationship between stability and function is often modeled by a sigmoidal function. This relationship is projected onto the *y*- and *z*-axes of the mutational reaction coordinates of (**b**) and (**c**) through a gradient from white (functional) to red (non-functional). **b**, A graphical example of specific epistasis. The red mutation in the parental background is destabilizing, resulting in a large functional defect. However, a blue mutation, which by itself does not alter stability, interacts with the red mutation to reduce its stability defect, and therefore, its functional defect. **c**, A graphical example of nonspecific epistasis. The red mutation in the parental background is destabilizing, resulting in a large functional defect. The blue mutation is stabilizing, which by itself has little impact on function. However, in this stability-buffered background, the red mutation (whose magnitude of destabilization is unchanged) can occur with no functional defect.

A comprehensive case study on specific epistasis has emerged from investigations on vertebrate steroid receptors using ancestral protein reconstruction and experimental analysis. This work traced the evolution of specificity in the glucocorticoid receptor for its ligand – the steroid hormone cortisol – from a promiscuous ancestral protein that was activated by both cortisol and a structurally related class of steroids called mineralocorticoids. Seven historical substitutions, when introduced into the ancestral protein, are sufficient to fully recapitulate the evolution of cortisol specificity. Introducing five of these substitutions, however, yields a completely nonfunctional receptor unless the other two "permissive" substitutions, which themselves have no effect on function, are in place first (57, 64). Structural analysis suggested a

direct conformational mechanism for specific epistasis: the function-switching substitutions dramatically shifted the position of a helix that lines the ligand pocket, destabilizing key elements of the active conformation but also allowing formation of a new cortisol-specific contact. The two permissive substitutions appeared to directly compensate, generating new physical interactions that stabilized the same structural elements destabilized by the function-switching mutations, thereby permitting the evolution of a receptor that could be activated only by cortisol.

To determine whether the permissive substitutions were truly specific, a follow-up experiment assessed how many other epistatically acting mutations might have been available that could have permitted the function-switching substitutions (38). A library of thousands of variants of the ancestral receptor was prepared, each of which contained the function-switching substitutions but neither of the permissive substitutions; this library was then screened for epistatic mutations that could rescue cortisol activation. In addition to the historical permissive substitutions, this screen uncovered three new compensatory mutations. However, none of these new compensatory mutations could have been permissive during evolution, because each dramatically compromised the ancestral receptor's function when introduced on its own. Thus, the epistatic interaction between the historical permissive and function-switching substitutions was extremely specific, with alternate permissive mutations in the neighborhood of the ancestral receptor being extremely rare. This genetic specificity arose directly from the biophysical basis of the interaction: both the permissive and compensatory mutations acted locally, restoring contacts to the same structural elements destabilized by the function-switching mutations, but only the historical permissive substitutions were also compatible with both the ancestral and derived ensembles of conformations that contribute to ligand-induced activation. Thus, the

direct, local relationship between permissive and function-switching substitutions caused them to interact nonlinearly in their effects on physical properties (ligand-activation and structure), giving rise to a very specific, few-to-few genetic interaction.

2.5.2 Nonspecific epistasis

In nonspecific epistasis, a mutation modulates the effects of a relatively large number of other mutations. Nonspecific epistasis occurs when two mutations interact nonadditively with respect to some biological property despite contributing additively at the level of physical protein properties. The epistasis arises because of a nonlinearity in the mapping from the physical property to the biological property (Fig. 2.4a), so a mutation with the same biophysical effect size has a different impact on function or fitness, depending on the current location of the parental protein on this property's landscape (Fig. 2.4c).

Nonspecific epistasis has been most thoroughly studied for mutations that independently affect the stability of the protein's native fold but exhibit epistasis in the protein's functionality or contribution to fitness (34, 49-51, 109-112). Nonspecific epistasis has also been observed for mutations that are additive with respect to other physical properties (folding, affinity, enzyme activity), but non-additive with respect to surface expression, transcriptional activity, or fitness, because of the nonlinear mapping between the two types of property (69, 70, 113).

Mutations that interact epistatically in this way manifest low specificity in their coupling to each other. Every mutation that affects a physical property that maps nonlinearly to biological function or fitness will epistatically interact with every other mutation that affects the same property. That is, a mutation that increases stability and is therefore permissive for another mutation that reduces stability should be permissive for any mutation that reduces stability by a

similar amount; further, its permissive effect could be replaced by that of any other mutation that has a similar positive effect on stability. Because the effects on the protein's physical property are independent of each other, the mechanisms that produce nonspecific epistasis typically involve no physical interaction, direct or indirect, between the relevant residues.

For example, in the evolutionary trajectory of influenza nucleoprotein discussed above, each of the three cases of epistasis involves a destabilizing mutation that depends on a counterbalancing stabilizing mutation to be tolerated (34). Importantly, one single stabilizing mutation can rescue any one of the individual destabilizing mutations, highlighting the nonspecific nature of this coupling. The stabilizing substitutions do not substantially alter the destabilizing effect of the interacting mutations; rather, they buffer the overall stability of the protein such that the double mutants are not substantially destabilized relative to the parent. Despite such additivity at the level of protein stability, these pairs of mutations exhibit strong epistasis at the level of protein function and viral fitness. This epistasis fits a model by which stability maps nonlinearly to function and fitness: in the simplest case, suppose that the biological function of a protein scales linearly with the quantity of folded protein in the cell (114). Two mutations that independently affect protein stability will epistatically affect the quantity of folded protein (and thus function), simply because of the sigmoidal shape of the Boltzmann distribution that relates changes in the energy of a conformational state to changes in the probability that the state is occupied. Since many proteins are only marginally stable, existing just slightly above the steep part of the curve that relates stability to fraction-folded, this particular type of nonspecific epistasis through stability appears to be a strong factor in protein evolution.

These two types of epistasis are not mutually exclusive. Mutations may interact nonadditively at the level of a physical property, and a nonlinear mapping from that property to function or fitness may further amplify (106) or buffer (114) the interaction. For example, mutations in ancestral transcription factors exhibited pervasive specific epistasis for DNA-binding affinity (106), and there is also a nonlinear relationship between DNA-binding affinity and transcriptional activation. Together, these two nonlinearities yield dramatic epistasis in the effects of mutations in the protein or DNA on transcriptional activation; as a result, mutations in the transcription factor can allow the DNA to tolerate affinity-reducing mutations, and subsequent mutations in the DNA can exert a permissive effect on the protein, opening up mutational pathways that lead to a new regulatory complex with entirely new specificity. In contrast, if two mutations interact to increase each other's effect on stability, but the protein is already far above the stability threshold, neither the individual mutations nor their combination will strongly affect the protein's function or fitness (114).

The distinction between two types of epistasis highlights the need for researchers to be transparent and cognizant of where epistasis comes from in their data. Some studies set a threshold to distinguish between functional and nonfunctional (or fit and unfit) genotypes and, in turn, to classify evolutionary pathways through sequence space as passable or not (22, 23, 106). Imposing this kind of nonlinearity will necessarily lead to apparent epistasis and determine a study's conclusions about the availability of mutational trajectories. If the threshold does not have a sound biological motivation – or if its role in determining a study's conclusions is not explored – spurious inferences about epistasis and evolution may result.

2.5.3 Specific positive versus nonspecific negative epistasis

What is the relative prevalence of specific and nonspecific epistasis? Mutational scanning studies have identified "hotspot" residues which interact epistatically with dozens or hundreds of other mutations, pointing to their nonspecific effect (21, 73-75). As expected, some of these hotspots contain stabilizing mutations that can permit many different destabilizing mutations to be tolerated (73), or destabilizing mutations that interact negatively with many other mildly destabilizing mutations (21, 74). This epistasis arises because the functions being assayed – fitness, or affinity for a binding partner – depend on the fraction of protein folded, which is nonlinearly related to stability.

Although nonspecific epistasis accounts for a large number of interactions, specific epistasis is also very important. In one mutational scanning study, the most densely connected hotspots – the most nonspecifically coupled mutations – accounted for less than 20 percent of all epistatic interactions between mutations (74). As expected, specific epistasis typically involves direct interactions that modulate the effects of two sequence changes on a protein's physical properties. For example, two individually destabilizing mutations to cysteine yield a stabilizing disulfide bond when combined (21).

Nonspecific mechanisms seem to be associated strongly with negative epistasis, while specific mechanisms are associated with positive interactions. In the mutational scan of the GB1 protein, nearly all of the strong, negatively epistatic pairs in GB1 involve combinations of two destabilizing mutations, and these pairs are distributed relatively uniformly in three-dimensional space on the protein structure (21). In contrast, most of the positively interacting pairs are in close structural proximity (Cβ distance <10Å), and many affect hydrogen bond networks – suggesting direct and specific physical interactions (21, 74). The association of nonspecific

interactions with negative epistasis makes sense, given that most mutations are destabilizing, and most proteins have only a small "stability reservoir" above the critical threshold, below which the proportion of folded protein drops off precipitously (Fig. 2.4a). A slightly destabilizing mutation – if it does not exhaust the stability reservoir – will therefore typically have only a weak effect on folding and function, but combining two such mutations may be strongly deleterious.

2.5.4 Specific and nonspecific epistasis in long-term evolution

Although both positive specific epistasis and negative nonspecific epistasis are prevalent in mutational scanning studies, this might not be true in long-term evolution. Deleterious combinations of mutations will usually be removed by purifying selection, so positive interactions – such as permissive epistasis – might be expected to dominate the record of long-term sequence evolution. Does specific epistasis dominate, as well? Case-studies of the historical evolution of individual proteins have uncovered both specific (38, 57) and nonspecific (34) mechanisms of permissive interactions, but such studies are insufficient to determine the general prevalence of the two classes of epistasis in protein evolution.

Several larger-scale studies do suggest that specific positive epistasis is pervasive. In the *in silico* evolution of *argT* discussed above, most substitutions were tolerated at the time of their fixation only because of prior permissive substitutions (30); that is, they were deleterious in at least some sequence backgrounds that existed earlier in the evolutionary trajectory. Further, these substitutions had smaller predicted stability effects when they occurred than they did at earlier times, pointing to a specific epistatic effect of the permissive mutations on the mapping from sequence to stability. The average epistatic effect with respect to stability was small (predicted

ΔΔΔG=0.5 kcal/mol), suggesting that – at least under the assumed conditions – moderate epistatic effects with respect to a protein's physical properties can have a meaningful impact on fitness and evolution.

Structural analyses of compensated mutations that are pathogenic in humans but fixed in other species support a similar conclusion. Most pathogenic mutations are predicted to be destabilizing when introduced into the human structure *in silico*. However, when they are introduced into structures in which the nearby residues have the amino acids found in the species in which they have fixed, they are predicted to be less destabilizing, again by about half a kcal/mol (80). This points to a general role for specific epistasis in this mode of permissive evolution, but it does not rule out an additional role for nonspecific, structurally distal epistatic modifiers.

Is the widespread specific epistasis for stability suggested by these *in silico* predictions observed when epistasis is experimentally characterized in real proteins? Two studies have experimentally measured the effects on folding stability of a handful of mutations when introduced into divergent orthologs, which differ at up to 28 and 43 percent of sites (35, 56). Both studies found that it is rare for mutations that are destabilizing in one ortholog to become stabilizing in the other, or vice versa. But the magnitude of many mutations' effects on stability does change notably, with the correlation of a residue's stability effects in the two orthologs degrading as sequence distance increases, reaching $R^2 = 0.8$ among the most divergent orthologs. One study observed that the stability effects of a mutation when introduced into different orthologs frequently differ by more than 0.5 kcal/mol (35), consistent with computational predictions.

Taken together, these studies point to a pervasive effect of specific, positive epistasis in evolution that is retained in protein sequences over long periods of time. They do not rule out an additional role for nonspecific epistasis. Although combinations of amino acids that negatively interact are typically removed from the sequence record by purifying selection, this does not mean they are not important during historical evolution, because determining the paths that evolution does not take is as important in evolutionary outcomes as shaping those it may pass through.

2.6 Evolutionary implications of epistasis

We have elaborated a body of research that suggests frequent epistasis between and among substitutions that fix along evolutionary trajectories. How does this epistasis impact the evolutionary process? Furthermore, we have examined two broad classes of mechanism by which these mutations interact nonadditively. How do these two types of epistasis differentially affect the evolutionary properties of molecules?

2.6.1 Evolvability and robustness

A protein's evolvability is determined by the accessibility of mutations that confer new functions; its robustness is determined by the set of available neutral mutations. Positive epistasis – permissive substitutions of either the specific or nonspecific type – increases both evolvability and robustness, because it opens some mutational trajectories that would otherwise have involved deleterious steps. A nonspecific permissive mutation, however, has the potential to open many more evolutionary pathways than specific epistatic mutations do. For example, when stabilizing permissive mutations are introduced into a cytochrome P450 enzyme, it can tolerate a

wider range of mutations that confer new functions but are moderately destabilizing (49). Similarly, in the antibiotic resistance gene TEM-1 β-lactamase, a "global suppressor" mutation M182T stabilizes the protein, relieving the otherwise deleterious effect of many other mutations that reduce protein stability, including many that enhance the protein's activity on new antibiotics (51, 109, 111). Permissive mutations that globally buffer other physical properties should promote similar increases in evolvability and robustness, though this remains to be demonstrated.

In contrast, a permissive substitution of specific effect can influence the evolutionary potential of, at most, the subset of residues with which it is physically coupled. Specific permissive substitutions are thus inherently limited in the range of mutational trajectories they can enable. Assessing cases in which specific epistatic interactions has narrow effects on evolvability and robustness is more challenging, because it requires a robust, negative result to demonstrate a few-to-few relationship between mutations. Mutational scanning studies have met this challenge to some extent, identifying interactions that increase evolvability and robustness at specifically coupled positions, without as global an impact on evolvability as nonspecifically permissive mutations (21, 73, 74). Further case studies will be required to assess the generality of this result and assess the effects of nonspecific and specific epistasis on these properties during historical evolution.

2.6.2 Historical contingency

When positive epistasis is highly specific, then the outcomes of evolution will be contingent on low-probability chance events, because selection for the new function cannot drive acquisition of the permissive substitutions that are a prerequisite for the function-switching

changes, and the chance that they will fix by mutation and drift alone is very small (38). In such cases, evolutionary processes exhibit radical stochasticity in their outcomes: parallel populations evolving under the same dynamics will reach different endpoints in response to some selective pressure, because the permissive substitutions that happen to fix will generally be different in each population, and each set of permissive substitutions in turn will open trajectories to distinct functional optima (40, 43).

In contrast, evolutionary contingency should be much weaker when nonspecific permissive epistasis is at play. In such cases, a large number of possible permissive mutations could permit any particular function-switching mutation; across parallel evolving populations, the probability is reasonably high that one of these permissive mutations would occur eventually, opening paths to similar or identical outcomes. Indeed, in the evolution of drug resistance in TEM-1 β-lactamase (115) and influenza neuraminidase (116) or immune escape in influenza nucleoprotein (34), many different mutations were discovered that permit a particular function-enhancing mutation through nonspecific buffering of properties such as folding, stability, and expression. Thus, nonspecific epistasis appears to be associated with much less stochasticity in the outcomes of evolutionary trajectories.

2.6.3 Reversibility

The reversibility of evolution has long been a topic of interest to evolutionary biologists, because irreversibility implies that the accessibility of some genetic or phenotypic state – the ancestral one – depends on the moment in the genetic history of the organism when it occurs. Specific and nonspecific epistasis appear to differentially affect the reversibility of evolution.

Specific epistasis contributes to evolutionary irreversibility. In several cases, some time after a substitution that affects function takes place, restrictive epistatic substitutions have occurred, making the ancestral state at the first site deleterious. In each case, specific steric clashes are involved: the restrictive amino acid is compatible with the derived state at the first site but not the ancestral state, either because of a direct clash between side chains or because of conformational changes that produce conflicts at other sites (41, 107, 117). The physical interaction between these residues affects in a nonadditive fashion the protein's propensity to fold into its functional conformations, making the fitness effect of a reversal strongly deleterious – and thus very unlikely – once the restrictive substitution has occurred.

In contrast, nonspecific epistatic substitutions appear to be reversible over long-term evolution. Several studies have shown that destabilizing substitutions that were initially permitted by a stabilizing substitution can revert, even after relatively long evolutionary intervals (34, 35, 56). In contrast to the specific examples above, the relative stability of the ancestral state remains unchanged, and reversion is freely accessible in the subsequent evolution of the protein.

2.6.4 Long-term evolutionary constraints

Specific and nonspecific epistasis also imprint themselves differently in the constraints that leave marks on modern-day sequences. The rate of evolution at a site reflects the strength of selective constraint that acts on that sequence position. A mutation can tighten or relax the selective constraints at a site it interacts with, slowing or accelerating its rate and changing the set of amino acid states that it tolerates. These dynamics lead to signatures of amino acid covariation in extant sequence data, which reflect the extent and nature of epistatic constraints among sites.

Nonspecific epistasis – irrespective of its prevalence and strength – should leave only sparse signals of covariation in the evolutionary record. Consider some destabilizing substitution that was initially permitted by a prior stabilizing substitution, leading to a temporary epistatic dependence: any subsequent stabilizing substitution at the same or another site could relieve the constraint that this coupling creates. The association between the two originally dependent states would then break down (56).

In contrast, specific epistasis permanently changes the effects of interacting mutations, altering the native preference of sites for particular amino acid states. These types of interactions should thus generate strong, precise signals of covariation, as restraints of co-occurrence are not easily reduced with subsequent evolutionary change. The fact that signals of covariation in sequence alignments are strongly related to the co-localization of amino acid pairs in protein's three-dimensional structures supports the notion that specific epistasis underlies most retained signals of epistasis in the evolutionary record. Indeed, across a large number of protein families, the majority of the sites with the strongest signal of covariation are in direct structural contact (90, 91).

2.7 Conclusions and future directions

The studies we have discussed paint an emerging picture of pervasive epistasis among the sites and states that substitute during protein evolution. Mutational scanning indicates that both specific and nonspecific coupling between residues contribute strongly to nonadditivity in mutational effects at any moment in time. Over long-term evolution, permissive substitutions – either specific or nonspecific – play a particularly critical role in opening evolutionary

trajectories. The class of specific epistatic interaction, however, appears to most profoundly affect the long-term outcomes of evolution.

We emphasize that this picture is emerging, not complete. Many more case studies, particularly of historical evolution – over various time scales using proteins with different functions and architectures – are required to understand the full range of biophysical mechanisms and evolutionary implications of epistatic interactions within proteins. Several specific questions and approaches seem particularly ripe for study.

First, combining high-throughput mutational scanning techniques with ancestral protein reconstruction provides insight unavailable to either technique alone. Library-based explorations of sequence space around some extant protein tells us how epistasis shapes the pathways that evolution *could* follow from the present, but it does not tell us anything about the history that produced that protein. Conversely, mechanistic dissection of reconstructed ancestral proteins and their trajectories can tell us how epistasis shaped the pathway that evolution *did* follow. By characterizing the sequence space around ancestral proteins, we can begin to address key questions about evolutionary processes: How did epistasis shape the sequence space of an evolving protein? How many pathways could have been followed to the same or similar outcomes? How many different outcomes could have been achieved under a given selection pressure, and how did epistasis influence their accessibility? To what extent were mutational trajectories transiently opened and closed by permissive and restrictive substitutions? What are the roles and particular mechanisms of specific and nonspecific epistasis that mediated these effects? To date, only one study has sought to characterize an ancestral sequence space – the study of specific epistasis in the steroid receptors (38). The extent to which the strong evolutionary contingency observed in that case pertains to the evolution of other proteins – and

what biophysical and genetic factors contribute to the ensuing evolutionary dynamics – remains to be determined.

Second, higher-order interactions are likely important in evolution. Most of the studies of epistasis in protein evolution discussed here focus on pairwise epistasis (the interaction between mutations at two sites). However, there is no reason why the impact of epistasis exposed for the interactions between two sites cannot extend to higher-order combinatorial interactions (e.g. the joint effect of two mutations varies with the identity of a third (106), etc.). A detailed understanding of higher-order epistasis requires the characterization of numbers of variants inaccessible to traditional experimental techniques (118). However, the technological innovations underlying high-throughput mutational scanning techniques (10, 119), coupled with quantitative formalisms for higher-order epistasis (118, 120) are expanding explorations into how the importance and mechanisms of epistasis inferred at the pairwise level extends to higher-order combinations of sites.

Third, epistasis between interacting molecules is a ripe matter for evolutionary biochemical analysis. Proteins interact with other proteins, nucleic acids, and small molecules. Epistasis between mutations in molecules that interact with each other should also have important evolutionary ramifications, but it is unclear how the relative prevalence and evolutionary implications compare to those associated with intramolecular epistasis. Recent experimental dissections (106, 121), high-throughput mutational scans (122), and systems-level approaches (123) have begun to address this question, revealing molecular mechanisms and evolutionary implications of epistasis that are unique to interactions between molecules. Further work in this area has the potential to broaden our understanding of the impact of epistasis on protein biochemistry and evolution.

Finally, we would like to highlight a conceptual issue that frames our understanding of epistasis and protein evolution. Epistasis is frequently discussed as a "constraint" in molecular evolution; this view may reflect the role of epistasis in constraining the outcomes of protein engineering efforts, in confounding genetic predictions from single-site data, or in structuring sequence space to produce local optima. But epistasis is not only a brake on evolution: dissecting the dense network of genetic interactions in multidimensional sequence space reveals how epistasis can also make possible the evolution of new genotypes, functions, and phenotypes. Permissive mutations can relieve constraints that would otherwise make potentially beneficial mutations inaccessible (21, 22), allowing proteins to evolve new functions in a very small number of mutational steps. Thus, more functional diversity may exist within the local sequence landscape of any given protein than is typically appreciated, and epistasis may allow proteins to travel along connected paths among these functionally distinct regions of sequence space.

Because of the size of sequence space and the ways that epistasis structures it, even some of the most ancient proteins have not yet fully explored the boundaries of their networks of neutral divergence (28). As we develop a more complete picture of these sequence spaces, the ways in which epistasis structures their topologies, and how proteins traverse them during evolution, our capacity to understand present-day proteins, their histories, and their possible futures should become deeper and more precise.

Chapter 3

Pervasive contingency and entrenchment in a billion years of Hsp90 evolution

The work described in this chapter was published as: Tyler Starr, Julia Flynn*, Parul Mishra*, Daniel Bolon, and Joseph Thornton. "Pervasive contingency and entrenchment in a billion years of Hsp90 evolution." Proceedings of the National Academy of Sciences, USA 115:4453-4458 (2018). *co-first authors*

3.1 Summary

Interactions among mutations within a protein have the potential to make molecular evolution contingent and irreversible, but the extent to which epistasis actually shaped historical evolutionary trajectories is unclear. To address this question, we experimentally measured how the fitness effects of historical sequence substitutions changed during the billion-year evolutionary history of the heat shock protein 90 (Hsp90) ATPase domain beginning from a deep eukaryotic ancestor to modern *Saccharomyces cerevisiae*. We found a pervasive influence of epistasis. Of 98 derived amino acid states that evolved along this lineage, about half compromise fitness when introduced into the reconstructed ancestral Hsp90; further, the vast majority of ancestral states reduce fitness when introduced into the extant *S. cerevisiae* Hsp90. Overall, more than 75% of historical substitutions were contingent on permissive substitutions that rendered the derived state non-deleterious, became entrenched by subsequent restrictive substitutions that made the ancestral state deleterious, or both. This epistasis was primarily caused by specific interactions among sites rather than a general effect on the protein's tolerance to mutation. Our results show that epistasis continually opened and closed windows of mutational opportunity

over evolutionary timescales, producing histories and biological states that reflect the transient internal constraints imposed by the protein's fleeting sequence states.

3.2 Introduction

Epistatic interactions can, in principle, affect the sequence changes that accumulate during evolution. A deleterious mutation's expected fate is to be purged by purifying selection, but it can be fixed if a permissive substitution renders it neutral or beneficial (30, 34, 57). Conversely, a neutral mutation – which by definition is initially reversible to the ancestral state without fitness cost – may become entrenched by a subsequent restrictive substitution that renders the ancestral state deleterious (29, 30, 107); reversal of the entrenched mutation would then be unlikely unless the restrictive substitution were itself reversed or another permissive substitution occurred.

The extent to which epistasis-induced contingency and entrenchment actually affected protein sequence evolution remains unclear, however, because there is no consensus on the prevalence, effect size, or mechanisms of epistasis among historical substitutions. Deep mutational scans have revealed frequent epistasis among the many possible mutations within proteins (21, 22, 74, 75, 124), but how these interactions affect the substitutions that actually occurred during historical evolution is not known. Historical case studies have shown that particular substitutions were contingent (42, 57, 69, 70) or became entrenched during evolution (107), but whether these are examples of a general phenomenon is unknown. Computational approaches suggest pervasive contingency and entrenchment among substitutions (28-31, 79, 83, 86), but some of these analyses rely on models of uncertain adequacy (56, 87, 125), and their claims have not been experimentally validated. Swapping sequence states among extant

orthologs reveals frequent epistasis among substitutions (33), but this "horizontal" approach, unpolarized with respect to time, leaves unresolved whether permissive or restrictive interactions are at play (126). Some experimental studies have systematically examined epistasis among substitutions in an historical context, but most have measured effects on protein function (33, 34) or stability (35, 56), leaving unexamined the prevalence of epistasis with respect to fitness – the phenotype that directly affects evolutionary fate. Others have focused on fitness but used methods that cannot detect effects of relatively small magnitude, which could be both widespread and consequential for evolutionary processes (18, 34).

We directly evaluated the roles of contingency and entrenchment on historical sequence evolution by precisely quantifying changes over time in the fitness effects of all substitutions that accumulated during the long-term evolution of heat shock protein 90 (Hsp90) from a deep eukaryotic ancestor to *S. cerevisiae*. Hsp90 is an essential molecular chaperone that facilitates folding and regulation of substrate proteins through an ATP-dependent cycle of conformational changes, modulated by co-chaperone proteins. Orthologs from other fungi, animals, and protists can complement Hsp90 deletion in *S. cerevisiae* (127, 128), indicating that the protein's essential molecular function is conserved over large evolutionary distances. To quantify the context-dependence of historical sequence changes in Hsp90, we used a sensitive deep sequencing-based bulk fitness assay (119) to characterize protein libraries in which each ancestral amino acid is reintroduced into an extant Hsp90 and each derived state is introduced into a reconstructed ancestral Hsp90. We focused our experiments on the N-terminal domain (NTD) of Hsp90, which mediates ATP-dependent conformational changes.

3.3 Results

3.3.1 The historical trajectory of Hsp90 sequence evolution

We inferred the maximum likelihood phylogeny of Hsp90 protein sequences from 261 species of Amorphea (the clade comprising Fungi, Metazoa, Amoebozoa, and related lineages (129)), rooted using green algae and plants as an outgroup (Fig. 3.1a, Appendix 1 Fig. A1.1). We reconstructed ancestral NTD sequences at all nodes along the trajectory from the common ancestor of Amorphea (ancAmoHsp90) to extant *S. cerevisiae* (ScHsp90) and identified substitutions as differences between the most probable reconstructions at successive nodes.

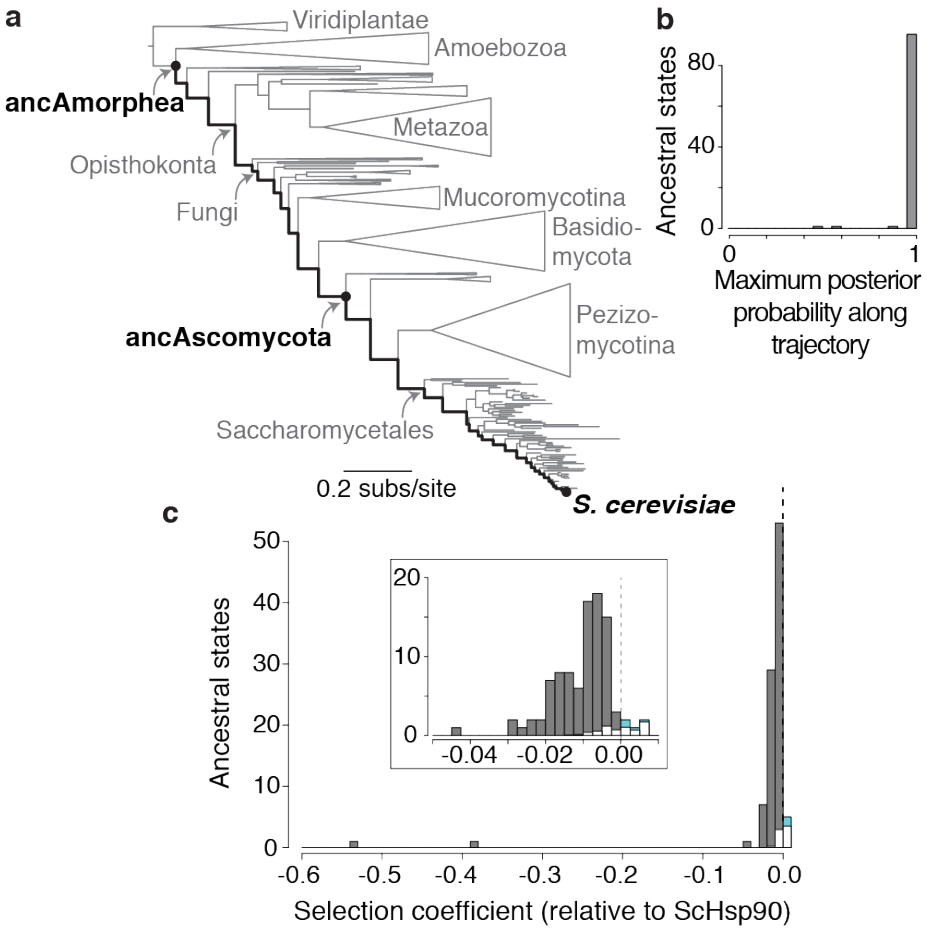


Figure 3.1. Ancestral states are deleterious in the yeast Hsp90 NTD. a, Maximum likelihood phylogeny of Hsp90 protein sequences from Amorphea. The evolutionary trajectory studied, from the last common ancestor of Amorphea to modern *S. cerevisiae*, is indicated by a dark

(**Figure 3.1, continued**) black line. Major taxonomic groups are labeled in gray. Ancestral and extant genotypes characterized in this study are in black. Complete phylogeny with taxon names is in Appendix 1 Fig. A1.1. **b**, Statistical confidence in ancestral amino acid states. For each of the 98 inferred ancestral states in the NTD, the highest posterior probability of the state at any internal node along the trajectory is shown. **c**, Distribution of selection coefficients of individual ancestral states when introduced into ScHsp90, measured as the logarithm of relative fitness compared to ScHsp90 in a deep-sequencing based bulk competition assay. Dashed line indicates neutrality. Inset, close view of the region near $s = 0$. In each histogram bin, colors show the proportion of ancestral states with selection coefficients in that range that are estimated to be neutral (white), deleterious (gray), or beneficial (blue) using a mixture model that takes account of experimental error in measuring fitness (see Appendix 1 Fig. A1.4).

Along this entire trajectory, substitutions occurred at 72 of the 221 sites in the NTD; because of multiple substitutions, 98 unique ancestral amino acid states existed at these sites at some point in the past and have since been replaced by the ScHsp90 state. The vast majority of these 98 ancestral states are reconstructed with high confidence (posterior probability >0.95) in one or more ancestors along the trajectory (Fig. 3.1b), and every ancestral sequence has a mean posterior probability across sites of >0.95 (Appendix 1 Fig. A1.2a-c).

3.3.2 Entrenchment and irreversibility

To measure the fitness effects of ancestral amino acids when they are re-introduced into an extant Hsp90, we created a library of ScHsp90 NTD variants, each of which contains one of the 98 ancestral states. We determined the per-generation selection coefficient (s) of each mutation to an ancestral state relative to ScHsp90 via bulk competition monitored by deep sequencing, a technique with highly reproducible results (Appendix 1 Fig. A1.3). Our assay system reduces Hsp90 expression to ~1% of the endogenous level (130), which magnifies the fitness consequences of Hsp90 mutations, enabling us to detect effects of small magnitude.

We found that the vast majority of reversions to ancestral states in ScHsp90 are deleterious (Fig. 3.1c). Using a mixture model to account for experimental noise in our fitness measurements, we estimate that 92% of all reversions reduce the fitness of ScHsp90 (95% CI 83–99%; Fig. 3.1c, Appendix 1 Fig. A1.4). Three other statistical methods that differ in their assumptions yielded estimates that between 54% and 95% of reversions are deleterious (Appendix 1 Fig. A1.5a). Two reversions cause very strong fitness defects (s = -0.38 and -0.54), but the typical reversion is only mildly deleterious (median s = -0.010, $P = 1.2 \times 10^{-16}$, Wilcoxon rank sum test; Fig. 3.1c). This conclusion is robust to excluding ancestral states that are reconstructed with any statistical ambiguity ($P = 4.5 \times 10^{-14}$). The magnitude of each mutation's negative effect on fitness correlates with indicators of site-specific evolutionary, structural, and functional constraint, corroborating the view that they are authentically deleterious (Appendix 1 Fig. A1.6).

These results do not imply that reversions can never happen—12 sites did undergo substitution and reversion at some point along the lineage from ancAmoHsp90 to ScHsp90. Rather, our observations indicate that at the current moment in time, most ancestral states are selectively inaccessible, irrespective of whether they were available at some moment in the past or might become so in the future (131).

3.3.3 Intramolecular versus intermolecular epistasis

Reversions to ancestral states might be deleterious because the derived states were entrenched by subsequent substitutions within Hsp90 (intramolecular epistasis) (29, 30); alternatively, they might be incompatible with derived states at other loci in the *S. cerevisiae* genome (intermolecular epistasis), or the derived states might unconditionally increase fitness.

Entrenchment because of intramolecular epistasis predicts that introducing into ScHsp90 sets of deleterious ancestral states that existed together at ancestral nodes should not reduce fitness as drastically as would be predicted from the individual mutations' effects. To test this possibility, we reconstructed complete NTDs from two ancestral Hsp90s on the phylogenetic trajectory (Fig. 3.2a, Appendix 1 Fig. A1.2) and assayed their relative fitness in *S. cerevisiae* as chimeras with ScHsp90's other domains. This design provides a lower-bound estimate of the extent of intramolecular epistasis, because it does not eliminate interactions between substitutions in the NTD and those in other domains of Hsp90.

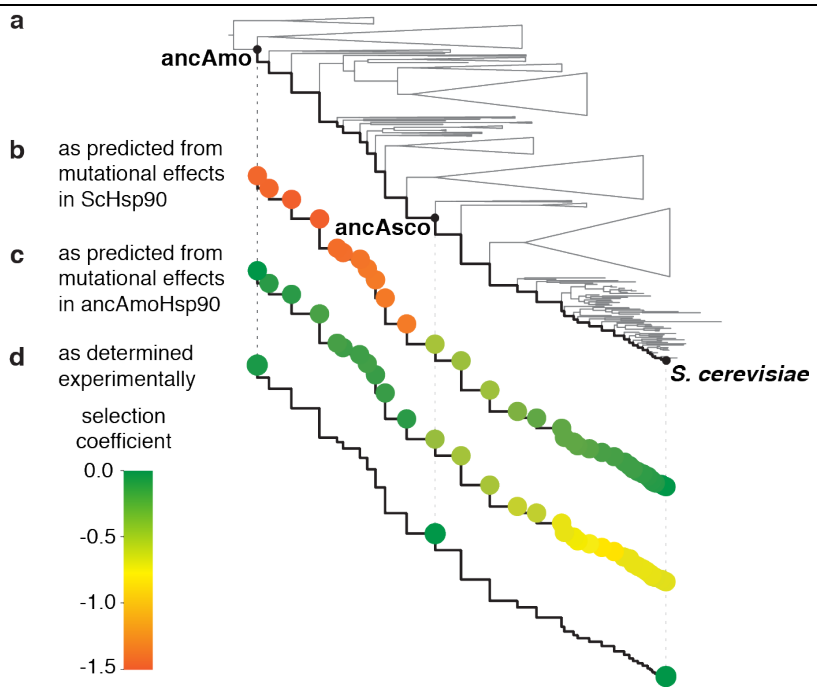


Figure 3.2. Fitness effects of historical substitutions are modified by intramolecular epistasis. For each node along the trajetory from ancAmoHsp90 to ScHsp90 (black line), the predicted or actual selection coefficient of the entire NTD genotype is represented from green (s = 0) to orange (s = -1.5). **a**, The Hsp90 phylogeny, represented as in Fig. 3.1a. **b**, The predicted selection coefficient of each ancestral sequence relative to ScHsp90 was calculated as the sum of the selection coefficients of each ancestral state present in that ancestor when measured individually in ScHsp90. **c**, The predicted selection coefficient of each sequence relative to ancAmoHsp90 was calculated as the sum of the selection coefficients of each derived state present at that node when measured individually in ancAmoHsp90. **d**, Experimentally determined selection coefficients for ancAmoHsp90 and ancAscoHsp90 relative to ScHsp90. For selection coefficients of each genotype, see Appendix 1 Fig. A1.7.

We found that intramolecular epistasis is the predominant cause of entrenchment. The first reconstruction, ancAscoHsp90, from the ancestor of Ascomycota fungi (estimated age ~450 million years (132)), differs from ScHsp90 at 42 NTD sites. If the fitness effects of these ancestral states when combined were the same as when introduced individually, they would confer an expected fitness of 0.65 (95% CI 0.61–0.69; Fig. 3.2b). When introduced together, however, the actual fitness is 0.99 (Fig. 3.2d, Appendix 1 Fig. A1.7a), indicating that the current fitness deficit of ancestral states is caused primarily by deleterious epistatic interactions within the NTD.

The older ancestor, ancAmoHsp90 (estimated age ~1 billion years (133)), differs from ScHsp90 at 60 NTD sites. When combined, these differences would confer an expected fitness of 0.23 (95% CI 0.21–0.26; Fig. 3.2b), but the actual fitness of the NTD is 0.43 (Appendix 1 Fig. A1.2d, Fig. A1.7a), again indicating strong epistasis within the NTD. We hypothesized that the remaining fitness deficit caused by the ancAmoHsp90 NTD could be attributed to intramolecular epistasis between the NTD and substitutions in the other Hsp90 domains. We identified a candidate substitution in the protein's middle domain that physically interacts with NTD residues to form the ATP-binding site (134); reverting this substitution to the ancAmorphea state (L378i) together with the ancAmorphea NTD increases fitness to 0.96 (Fig. 3.2d, Appendix 1 Fig. A1.7a).

These findings indicate that virtually all the context-dependent deleterious effects of ancestral states are caused by intramolecular interactions within the NTD and with one other site in the Hsp90 protein. Derived states that emerged along the Hsp90 trajectory have been entrenched by subsequent substitutions within the same protein, which closed the direct path back to the ancestral amino acid without causing major changes in function or fitness (30).

3.3.4 Contingency and permissive substitutions

We next determined whether the derived states that evolved during the protein's history were contingent on prior permissive substitutions. We constructed a library of variants of the ancAmoHsp90 NTD, the deepest ancestor of the trajectory, each of which contains one of the 98 forward mutations to a derived state. We cloned this NTD library into yeast as a chimera with ScHsp90's other domains (with site 378 in its ancestral state) and used our deep sequencing-based bulk fitness assay to measure the selection coefficient of each mutation relative to ancAmoHsp90 (Appendix 1 Fig. A1.3d,e).

We found that about half of mutations to derived states were selectively unfavorable (Fig. 3.3a). After accounting for experimental noise using a mixture model, we estimate that 48% of derived states reduce ancAmoHsp90 fitness (95% CI 29–74%), and 32% are neutral (95% CI 0–57%; Appendix 1 Fig. A1.8); three other statistical approaches gave similar results (Appendix 1 Fig. A1.5a). Twenty percent of the derived states are beneficial in our assay (95% CI 9–42%), which could be because they are unconditionally advantageous or because of epistatic interactions with other loci in *S. cerevisiae* or other regions of ScHsp90. Two derived states had very strong fitness defects, but the typical derived state is weakly deleterious (median s = -0.005, $P = 5.8 \times 10^{-4}$, Wilcoxon rank sum test; Fig. 3.3a).

As with the reversions to ancestral states, the effects of individual derived states, as measured in the ancestral background, predict fitness consequences far greater than observed when the derived states are combined in the Hsp90 genotypes that existed historically along the phylogeny (Fig. 3.2c,d, Appendix 1 Fig. A1.7b). Thus, many derived states would have been deleterious if they had occurred in the ancestral background, but they became accessible following subsequent permissive substitutions that occurred within Hsp90. Taken together, the

data from the ancestral and derived libraries indicate that 77% of the amino acid states that occurred along this evolutionary trajectory were contingent on prior permissive substitutions, entrenched by subsequent restrictive substitutions, or both (Fig. 3.3b).

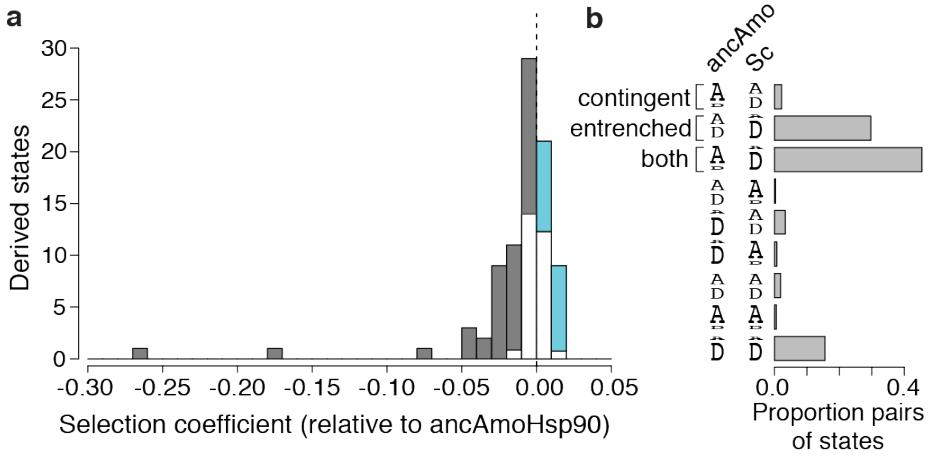


Figure 3.3. Widespread contingency and entrenchment. a, Distribution of measured selection coefficients of derived NTD states when introduced singly into ancAmoHsp90. Dashed line indicates neutrality. In each histogram bin, colors show the proportion of derived states with selection coefficients in that range that are estimated to be neutral (white), deleterious (gray), or beneficial (blue) using a mixture model that takes account of experimental error in measuring fitness (see Appendix 1 Fig. A1.8). **b**, The fraction of pairs of ancestral and derived states that are inferred to be contingent, entrenched or both. Pairs of ancestral and derived states at each site can be classified by the relative fitness of the two states when measured in ancAmoHsp90 or in ScHsp90: ancestral state more fit (A larger than D), derived state more fit (D larger than A), or fitnesses indistinguishable (A and D same size). The fraction of pairs in each category was estimated as the product of the posterior probabilities that each pair of sites is in the relevant selection category (ancestral state with fitness greater than, less than, or indistinguishable from the derived state) in the ScHsp90 and the ancAmoHsp90 backgrounds.

3.3.5 Specificity of epistatic interactions

Epistatic effects on fitness can emerge from specific genetic interactions between substitutions that directly modify each other's effect on some molecular property, or from nonspecific interactions between substitutions that are additive with respect to bulk molecular properties (e.g. stability (34, 114)) if those properties nonlinearly affect fitness (7, 135, 136).

To explore which type of epistasis predominates in the long-term evolution of Hsp90, we first investigated the two strongest cases of entrenchment, the strongly deleterious reversions

V23f and E7a (upper-case letters indicate the ScHsp90 state and lower-case the ancestral state). We sought candidate restrictive substitutions for each of these large-effect reversions by examining patterns of phylogenetic co-occurrence. Substitution f23V occurred not only along the trajectory from ancAmoHsp90 to ScHsp90 but also in parallel on another fungal lineage; in both cases, candidate epistatic substitution i378L co-occurred on the same branch (Appendix 1 Fig. A1.9a,b). As predicted if i378L entrenched f23V, we found that introducing the ancestral state i378 in ScHsp90 relieves the deleterious effect of the ancestral state f23 (Fig. 3.4a). These two residues directly interact in the protein's tertiary structure to position a key residue in the ATPase active site (Appendix 1 Fig. A1.9c,d), explaining their specific epistatic interaction.

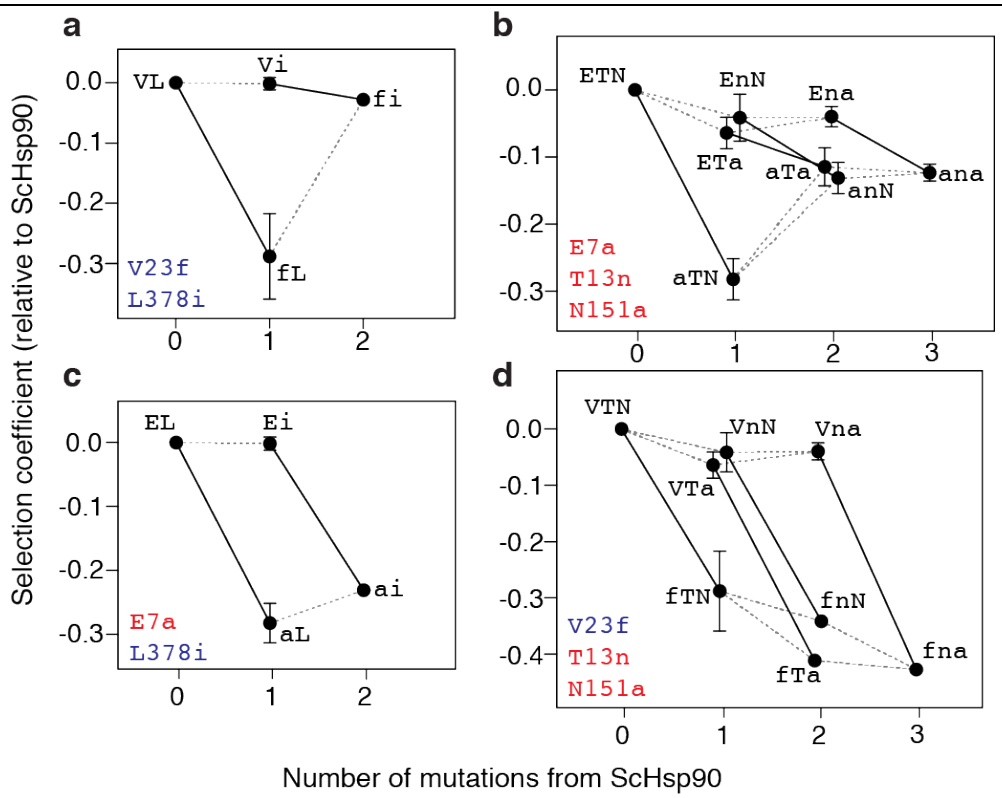


Figure 3.4. Epistatic interactions are specific. Large-effect deleterious reversions and restrictive substitutions that contributed to their irreversibility. For each single, double, or triple mutant in ScHsp90, the selection coefficient relative to ScHsp90 is shown, as assessed in monoculture growth assays. Lines connect genotypes that differ by a single mutation; solid lines indicate the effect of the large-effect reversions in each background. Error bars, SEM for 2 to 4 replicates (see Methods; absence of error bar indicates one replicate). Data points are labeled by

(**Figure 3.4, continued**) amino acid states: lower case, ancestral state; upper case, derived state. Mutations tested in each cycle are in the bottom-left corner; those in the same color interact specifically with each other. **a**, Deleterious reversion V23f is ameliorated by L378i. **b**, Deleterious reversion E7a is partially ameliorated by N151a or T13n. **c**, L378i does not ameliorate E7a. **d**, N151a and T13n do not ameliorate V23f.

In the case of E7a—the other reversion strongly deleterious in ScHsp90—the ancestral state was reacquired in a closely related fungal lineage. We reasoned that the substitutions that entrenched a7E on the lineage leading to ScHsp90 must have themselves reverted or been further modified on the fungal branch in which reversal E7a occurred. We identified two candidates (n13T and a151N) that met these criteria (Appendix 1 Fig. A1.10a,b,c). As predicted, experimentally introducing the ancestral states n13 or a151 into ScHsp90 relieves much of the fitness defect caused by the ancestral state a7, indicating that substitutions n13T and a151N entrenched a7E (Fig. 3.4b). These three sites are on interacting secondary structural elements that are conformationally rearranged when Hsp90 converts between ADP- and ATP-bound states (Appendix 1 Fig. A1.10d,e).

To test whether these modifiers specifically restrict particular substitutions or are general epistatic modifiers, we asked whether the restrictive substitutions that entrenched one substitution also modify the effects of the other (34). As predicted if the interactions among these sets of substitutions are specific, introducing L378i does not ameliorate the fitness defect caused by E7a, and introducing T13n or N151a does not ameliorate the fitness defect caused by V23f (Fig. 3.4c,d). These data indicate that specific biochemical mechanisms underlie the restrictive interactions for these large-effect examples of epistatic entrenchment.

Finally, we investigated whether the epistatic interactions among the set of small-effect substitutions in this trajectory are also specific or the nonspecific result of a threshold-like relationship between fitness and some bulk property such as stability (34, 114). If epistasis is

mediated by a nonspecific threshold relationship, mutations that decrease fitness in one background will never be beneficial in another, although they can be neutral if buffered by the threshold (Fig. 3.5a) (18, 34, 56). In contrast, specific epistatic interactions can switch the sign of a mutation's selection coefficient in different sequence contexts (Fig. 3.5b) (135). As predicted under specific epistasis, we found that for most differences between ancAmoHsp90 and ScHsp90 (65%), the ancestral state confers increased fitness relative to the derived state in the ancestral background but decreases it in the extant background (Fig. 3.5c). The selection coefficients of mutations are negatively correlated between backgrounds (P=0.009), indicating that the substitutions that became most entrenched in the present also required the strongest permissive effect in the past. This pattern is expected if the structural constraints that determine the selective cost of having a suboptimal state at some site are conserved over time, but the specific states preferred depend on the residues present at other sites.

Taken together, these findings indicate that most epistasis during the long-term evolution of Hsp90 involved specific one-to-one (or few-to-few) interactions among sites, not general effects on the protein's tolerance to mutation.

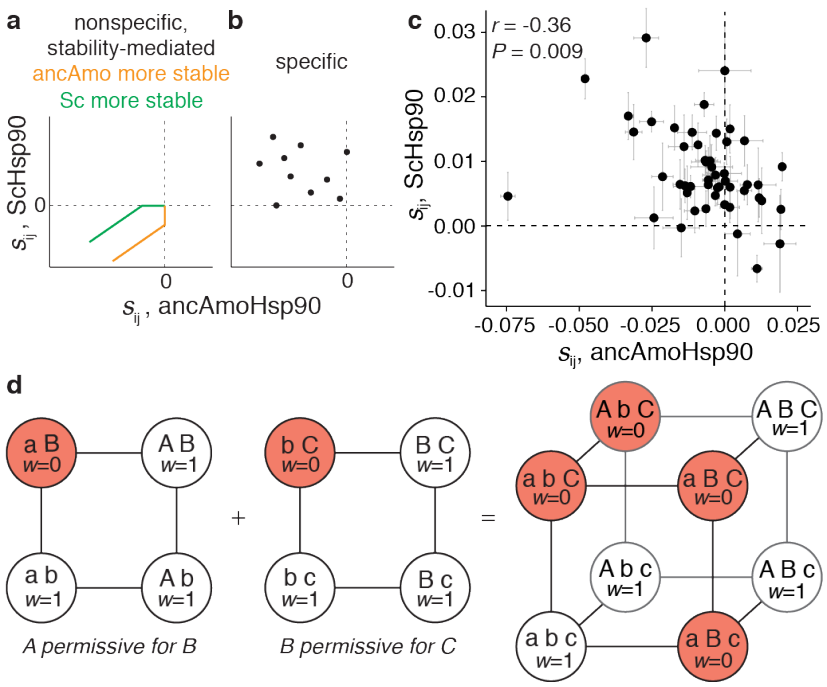


Figure 3.5. A daisy-chain model of epistasis. a,b, Expected relationship under two models of epistasis between selection coefficients of ancestral-to-derived mutations (s_{ij}) when introduced into ancestral (*x*-axis) or derived (*y*-axis) backgrounds. **a**, Nonspecific epistasis: if genetic interactions are the nonspecific result of a threshold-like, buffering relationship between stability (or another bulk property) and fitness (34, 114), then the effects of strongly deleterious mutations will be positively correlated between the two backgrounds, but weakly deleterious mutations in the less stable background may be neutral in the more stable background (yellow, ancAmoHsp90 more stable; green, ScHsp90 more stable). **b**, Specific epistasis: if interactions reflect specific couplings between sites, then mutations from ancestral to derived states can be deleterious in the ancestral background but beneficial in the derived background (upper left quadrant). **c**, Measured selection coefficients for ancestral-derived state pairs that differ between ancAmoHsp90 and ScHsp90. Dashed lines, $s = 0$. Error bars, SEM from two replicate bulk competition measurements. *r*, Pearson correlation coefficient and associated *P* value. Full distribution showing strongly deleterious outliers in the ScHsp90 or ancAmoHsp90 data is shown in Appendix 1 Fig. A1.10f. **d**, Daisy-chain model of specific epistatic interactions. Each square shows the mutant cycle for a pair of substitutions (A and B or B and C; lower-case, ancestral state; upper-case, derived), one of which is permissive for the other. Each circle is a genotype colored by its fitness (*w*): white, neutral; red, deleterious. Edges are single-site amino acid changes. The cube shows the combined mutant cycle for all three substitutions. Permissive substitutions become entrenched when the mutation that was contingent upon it occurs. Substitutions in the middle of the daisy-chain, which require a permissive mutation and are permissive for a subsequent mutation, are both contingent and entrenched.

3.4 Discussion

3.4.1 Relation to prior work

We observed widespread and specific epistasis over the course of a billion years of Hsp90 evolution, during which the protein's function, physical architecture, and fitness were conserved. The fraction of historical substitutions that were contingent on permissive substitutions, entrenched by restrictive substitutions, or both—about 80%—is considerably higher than suggested by previous experimental work (33-35) and some computational analyses (79), rivaling the highest estimates from computational studies (28, 30, 31). One explanation for the more widespread epistatic interactions in our study may be our method's capacity to detect much smaller growth deficits than have been discernable in previous experimental studies.

Another difference from previous research is that we primarily observed specific epistasis, whereas several studies have found a dominant role for nonspecific stability-mediated epistasis, particularly during the short-term evolution of viruses (34, 56, 69). This disparity could be attributable to a difference in selective regime or in time scale: the epistatic constraints caused by specific interactions are expected to be maintained over far longer periods of time than those caused by nonspecific interactions, which are easily replaced by other substitutions because of the many-to-many relationship between permissive and permitted amino acid states (34, 56, 135). The prevalence and type of epistasis may also vary because of differences in proteins' physical architectures. Additional case studies will be necessary to evaluate the causal role of these and other factors in determining the nature of epistatic interactions during evolution.

3.4.2 Limitations

Our strategy has some known limitations, but none are likely to change our major conclusions. For example, we assayed the effect of long-past substitutions in the context of extant yeast cells. Our experiments, however, indicate that there is only very weak epistasis for fitness between historical substitutions within Hsp90 and those at other loci, because the reconstructed ancestral Hsp90 chimeras cause a fitness deficit of only 0.01 to 0.04 when introduced into *S. cerevisiae* cells—much smaller than the sum of intramolecular incompatibilities revealed by introducing the ancestral states individually. Our finding of widespread contingency and entrenchment is therefore not an artifact of incompatibilities between ancestral Hsp90 states and the genotype of present-day *S. cerevisiae* at other loci.

A second potential limitation is that the ancestral states we tested were reconstructed phylogenetically, not known empirically. But the vast majority of states were inferred with high statistical confidence, because the Hsp90 NTD is well conserved and we used a densely-sampled alignment. The ambiguity that was present primarily concerned the specific ancestral node at which an inferred ancestral state was present, not whether or not it was ancestral somewhere along the trajectory, which is the key inference for our purposes. Further, even when all states with any degree of statistical uncertainty in the ancestral reconstruction were excluded from the analysis, the remaining data strongly supported our conclusions concerning contingency and entrenchment.

Finally, we measured fitness under a particular set of experimental conditions. Our assay system reduces Hsp90 expression to ~1% of the endogenous level (130). Based on previous work quantifying the relationship between Hsp90 function, expression, and growth rate (130), we estimate that the average selection coefficient of -0.01 we observed among contingent or

entrenched substitutions corresponds to a fitness deficit of approximately $s = -5 \times 10^{-6}$ under native-like expression levels. Mutations with selection coefficients in this range would likely be subject to purifying selection in large microbial populations (137-139). Our assay also tests fitness under log-phase growth conditions in rich media. A more heterogeneous or demanding environment would likely increase the magnitude of selective effects of Hsp90 mutations, because stress should amplify the fitness consequences of mutations in the proteostasis machinery.

3.4.3 Implications

Our observation that contingency and entrenchment affected the majority of historical substitutions suggests a daisy-chain model by which genetic interactions structured long-term Hsp90 evolution (Fig. 3.5d). A permissive mutation becomes entrenched and irreversible once a substitution contingent upon it occurs; if the contingent substitution subsequently permits a third substitution, it too becomes entrenched (28, 30).

Most of the substitutions along the trajectory from ancAmoHsp90 to ScHsp90 were both contingent and entrenched, suggesting that they occupy an internal position in this daisy chain. Each of these changes closed reverse paths at some sites and opened forward paths at others, which—if taken—would then entrench the previous step. Evolving this way over long periods of time, proteins come to appear exquisitely well-adapted to the conditions of their existence, with most present states superior to past ones. The conditions that make today's states so fit, however, include—or are even dominated by—the transient internal organization of the protein itself.

3.5 Methods

Phylogenetic analysis and ancestral reconstruction. We obtained Hsp90 protein sequences from the Amorphea clade (129) from NCBI, the JGI Fungal Program, the Broad Institute Multicellularity Project, the literature (140), and Iñaki Ruiz-Trello. Each protein was used as a query in a BLASTp search against the human proteome to identify and retain Hsp90A orthologs. We used CD-HIT (141) to filter proteins with high sequence similarity. We removed sequences with >67% missing characters and highly diverged, unalignable sequences. Remaining sequences were aligned with Clustal Omega (142). Lineage-specific insertions were removed, as were unalignable linker regions (ScHsp90 sites 1-3, 225-237, 686-701). We added six Hsp90A sequences from Viridiplantae as an outgroup, resulting in a final alignment of 267 protein sequences and 680 sites.

We inferred the maximum likelihood (ML) phylogeny given our alignment and the LG model (143) with gamma-distributed among-site rate variation (4 categories) and ML estimates of amino acid frequencies, which was the best-fit model as judged by AIC. The phylogeny was inferred using RAxML version 8.1.17 (144). The ML phylogeny reproduces accepted relationships between major taxonomic lineages (129, 145-149). Most probable ancestral sequences were reconstructed on the maximum likelihood phylogeny using the AAML module of PAML version 4.4 (150) given the alignment, ML phylogeny, and LG+Γ model. The trajectory of sequence change was enumerated from the amino acid sequence differences between successive ancestral nodes on the lineage from the common ancestor of Amoebozoa + Opisthokonta (ancAmorphea) to *S. cerevisiae* Hsp82 (ScHsp90, Uniprot P02829).

Coding sequences for the most probable ancestral amino acid sequences of the Hsp90 N-terminal domain (NTD) from ancAmorphea (ancAmoHsp90) and the common ancestor of

Ascomycota yeast (ancAscoHsp90) were synthesized by IDT. These sequences were cloned as chimeras with the ScHsp90 middle and C-terminal domains and intervening linkers via Gibson Assembly. AncAmoHsp90 also carries an additional reversion to the ancAmorphea state at site 378 in the middle domain (Appendix 1 Fig. A1.2d), which is part of a loop that extends down and interacts with ATP and the NTD (134, 151).

Generating mutant libraries. ScHsp90 and ancAmoHsp90 gene constructs were expressed from the p414ADHΔTer plasmid (130). The ScHsp90 library consists of variants of the ScHsp90 NTD, each containing one mutation to an ancestral amino acid state. The ancAmoHsp90 library consists of variants of the ancAmoHsp90 NTD, each containing one mutation to a derived state. Two sets of PCR primers were designed for each mutation, to amplify Hsp90 NTD fragments N-terminal and C-terminal to the mutation of interest; primers introduce the mutation of interest and generate a 25-bp overlap between fragments, as well as 20-bp overlaps between each fragment and the destination vector for gene re-assembly. PCR was conducted with Pfu Turbo polymerase (Agilent) for 15 amplification cycles. The resulting PCR fragments were stitched together with a 10-cycle assembly PCR, pooled, and combined via Gibson Assembly (NEB) with a linearized p414ADHΔTer Hsp90 destination vector excised of the NTD.

Barcode labeling of library genotypes. Following construction of the plasmid libraries, each variant in the library was tagged with a unique barcode to simplify sequencing steps during bulk competition (152). A pool of DNA constructs containing a randomized 18 base-pair barcode sequence (N18) and Illumina sequencing primer annealing regions (IDT) was cloned 200 nucleotides downstream from the hsp90 stop codon via restriction digestion, ligation, and transformation into chemically-competent *E. coli*. Cultures with different amounts of the

transformation reaction were grown overnight and the colony forming units in each culture were assessed by plating a small fraction. We isolated DNA from the transformation that contained approximately 10-20 fold more colony-forming units than mutants, with the goal that each mutant would be represented by 10-20 unique barcodes.

To associate barcodes with Hsp90 mutant alleles, we conducted paired end sequencing of each library using primers that read the N18 barcode in the first read and the Hsp90 NTD in the other. To generate short DNA fragments from the plasmid library that would be efficiently sequenced, we excised the gene region between the NTD and the N18 barcode via restriction digest, followed by blunt ending with T4 DNA polymerase (NEB) and plasmid ligation at a low concentration (3 ng/μL) to favor circularization over bi-molecular ligations. The resulting DNA was re-linearized by restriction digest, and Illumina adapter sequences were added via an 11-cycle PCR. The resulting PCR products were sequenced using an Illumina MiSeq instrument with asymmetric reads of 50 bases and 250 bases for Read1 and Read2 respectively. After filtering low quality reads (Phred scores < 10), the data were organized by barcode sequence. For each barcode that was read more than 3 times, we generated a consensus sequence of the N-domain indicating the mutation that it contained.

Bulk growth competitions. For bulk fitness assessments, we transformed *S. cerevisiae* with the ScHsp90 library along with wildtype ScHsp90 and a no-insert control; we also transformed *S. cerevisiae* with the ancAmoHsp90 library along with wildtype ScHsp90, wildtype ancAmoHsp90, and a no-insert control. Concentrations of plasmids were adjusted to yield a 2:6:1 molar ratio of wildtype: no-insert control: average variant in the library. Plasmid libraries and corresponding controls were transformed into the DBY288 Hsp90 shutoff strain (16, 153), resulting in ~150,000 unique yeast transformants representing 50-fold sampling for the average

barcode. Following recovery, transformed cells were washed 5 times in SRGal-W (synthetic 1% raffinose and 1% galactose lacking tryptophan) media to remove extracellular DNA, and then transferred to plasmid selection media SRGal-W and grown at 30°C for 48 hours with repeated dilution to maintain the cells in log phase of growth. To select for function of the plasmid-borne Hsp90 allele, cells were shifted to shutoff conditions by centrifugation, washing and re-suspension in 200 mL SD-W (synthetic 2% dextrose lacking tryptophan) media and ampicillin (50μg/mL), and grown at 30°C 225 rpm. Following a 16-hour growth period required to shut off expression of the wildtype chromosomal Hsp90, we collected samples of ~10^8 cells at 8 or more time points over the course of 48 (ScHsp90 library) or 31 (ancAmoHsp90 library) hours and stored them at -80°C. Cultures were maintained in log phase by regular dilution with fresh media, maintaining a population size of 10^9 or greater throughout the bulk competition. Bulk competitions of each library were conducted in duplicate from independent transformations.

DNA preparation and sequencing. We collected plasmid DNA from each bulk competition time point as previously reported (154). Purified plasmid was linearized with AscI. Barcodes were amplified by 18 cycles of PCR using Phusion polymerase and primers that add Illumina adapter sequences, as well as an 8-bp identifier used to distinguish among libraries and time points. Identifiers were designed so that each differed by more than two bases from all others to avoid misattributions due to sequencing errors. PCR products were purified two times over silica columns (Zymo research), and quantified using the KAPA SYBR FAST qPCR Master Mix (Kapa Biosystems) on a Bio-Rad CFX machine. Samples were pooled and sequenced on an Illumina NextSeq (ancAmoHsp90 library) or HiSeq 2000 (ScHsp90 library) instrument in single-end 100 bp mode.

Analysis of bulk competition sequencing data. Illumina sequence reads were filtered for Phred scores >20, strict matching of the sequence of the intervening bases to the template, and strict matching of the N18 barcode and experimental identifier to those that were expected in the given library. Reads that passed these filters were parsed based on the identifier sequence. For each identifier, the data was condensed by generating a count of each unique N18 read. The unique N18 count file was then used to identify the frequency of each mutant using the variant-barcode association table. For each variant in the library, the counts of each associated barcode were summed to generate a cumulative count for that mutant.

Determination of selection coefficient. The ratio of the frequency of each variant in the library relative to wildtype (ancAmoHsp90 or ScHsp90) was determined at each time point, and the slope of the logarithm of this ratio versus time (in number of generations) was determined as the raw per-generation selection coefficient (s) (155):

$$s = \mathrm{d}/\mathrm{dt}\,[\,\ln(n_m / n_{wt})\,]$$

where n_m and n_{wt} are the number of sequence reads of mutant and wildtype, respectively, and time is measured in number of wildtype generations. No-insert plasmid selection coefficients were determined from the first three time points because their counts drop rapidly over time. Mutants with selection coefficients within three standard deviations of the mean of no-insert variants were considered null-like and also analyzed based on the first three time points. For all other variants, selection coefficients were determined from all time points. Final selection coefficients for each variant were scaled in relative fitness space ($w = e^s$) such that the Hsp90 null allele, which is lethal, has a relative fitness of 0 ($s = -\infty$). This definition of relative fitness, unlike that which defines $w = 1 - s$, has the advantage of making selection coefficients additive

and reversible (the selection coefficient of mutation from state *i* to *j* is the opposite of the selection coefficient of that from *j* to *i*) (156).

Generation of individual mutants and monoculture analysis of yeast growth. To measure the relative fitness of ancAscoHsp90, mutations missed in the bulk libraries, and genotypes in mutant cycles that we sought to test in combination for epistatic interactions, we assayed growth rate in monoculture and related this to fitness, which assumes the relative rate of growth of two genotypes is the same in isolation as in direct competition (155). The growth rate of individually cloned mutants was estimated over 30 hours of growth with periodic dilution to maintain log-phase growth, as per Jiang et al. (130). Growth rates were determined as the slope of the linear model relating the log-transformed dilution-corrected cell density to time. The growth rate was converted to an estimate of the selection coefficient by taking the difference in growth rate (Malthusian parameter) between mutant and wildtype and multiplying this by the wildtype generation time (155), then rescaling selection coefficients in relative fitness space such that a null mutant analyzed in parallel has relative fitness 0 ($s = -\infty$).

Individual mutants of ancAmoHsp90 and ScHsp90 were generated in the p414ADHΔTer background by Quikchange site-directed mutagenesis, confirmed by Sanger sequencing. Mutations that were generated and assayed in ancAmoHsp90 (with number of replicate measurements in parentheses) include: S49A (n=1), T137I (n=1), V147I (n=1), I158V (n=1), R160L (n=1), G164N (n=1), E165P (n=1), L167I (n=1), K172I (n=1), L193I (n=1), and V194I (n=1). Mutations generated and assayed in ScHsp90 include: T5S (n=3), E7A (n=4), T13N (n=3), V23F (n=2), N151A (n=3), L378I (n=2), double mutants E7A/T13N (n=3), E7A/N151A (n=3), T13N/N151A (n=3), V23F/T13N (n=1), V23F/N151A (n=1), E7A/L378I (n=1), V23F/L378I (n=1) and triple mutants E7A/T13N/N151A (n=2) and V23F/T13N/N151A (n=1).

Robustness of results to statistical uncertainty and technical variables. The conclusion that the typical ancestral state is deleterious in ScHsp90 is robust to the exclusion of 20 ancestral states that have posterior probability < 1.0 at all ancestral nodes along the trajectory ($P = 4.5 \times 10^{-14}$, Wilcoxon rank sum test with continuity correction). The mutation to one ancestral state was missed in the bulk competition: its selection coefficient was inferred separately via monoculture, and including it in the analysis still leads to the conclusion that the typical ancestral state is deleterious ($P = 7.8 \times 10^{-17}$, Wilcoxon rank sum test with continuity correction).

The conclusion that the average derived state is deleterious in ancAmoHsp90 is retained when we include only the 32 mutations for which the ancAmoHsp90 state is inferred with a posterior probability of 1.0 and the derived state is inferred with posterior probability 1.0 in at least one node along the trajectory ($P = 1.1 \times 10^{-4}$, Wilcoxon rank sum test with continuity correction). The conclusion is also robust if we include selection coefficients as determined separately via monoculture for mutations to 11 derived states that were missed in the bulk competition ($P = 5.4 \times 10^{-4}$, Wilcoxon rank sum test with continuity correction).

We assessed relative fitness for six genotypes (ScHsp90+E7a, ScHsp90+V23f, ScHsp90+N151a, ScHsp90+T13n, ancAmoHsp90, and ancAmoHsp90+i378L) both by monoculture and by bulk competition. These two measures are well correlated (Pearson $R^2 = 0.95$), although the magnitude of a fitness effect is smaller when measured by monoculture growth assays (Appendix 1 Fig. A1.3f), perhaps because of differences in experimental conditions for bulk versus monoculture growth, such as the type of growth vessel and culture volume (and consequential aeration). The only conclusion involving a comparison between these two kinds of measurements is that ancAscoHsp90 (measured via monoculture) is more fit than

would be predicted from the sum of selection coefficients of its component states (measured via bulk competition) (Fig. 3.2, Appendix 1 Fig. A1.7). We therefore used the observed linear relationship between the two types of fitness assays to transform ancAscoHsp90's fitness as measured by monoculture (0.991); the expected fitness of ancAscoHsp90 in a bulk competition is 0.986, still much larger than the predicted fitness of 0.65 in the absence of epistasis.

Expected versus observed fitness. To identify epistasis between candidate interacting sites (e.g. Fig. 3.4a-d) or among the broader set of substitutions (e.g. Fig. 3.2), we compared the observed fitness of genotypes with multiple mutations to that expected in the absence of epistasis. In the absence of epistatic interactions, selection coefficients combine additively (156). We therefore calculated the expected selection coefficient of a genotype as the sum of selection coefficients of its component mutations as measured independently in a reference background (ancAmoHsp90 or ScHsp90). The standard error of a predicted fitness given the sum of selection coefficients was calculated as the square root of the sum of squared standard errors of the individual selection coefficient estimates, as determined from the duplicate bulk competition measurements. Epistasis was implicated if the observed fitness of a genotype differed from that predicted from the sum of its corresponding single-mutant selection coefficients.

Estimating the fraction of deleterious mutations. We sought to determine the fraction of mutations in each dataset that are deleterious using a modeling approach that incorporates measurement error and which does not require individual mutations to be classified as deleterious, neutral, or beneficial. We used the mixtools package (157) in R to estimate mixture models of underlying Gaussian distributions that best fit the empirical distributions of mutant selection coefficients in each library. First, we fit a single Gaussian distribution to the measured selection coefficients of replicate wildtype sequences that were present in the library but

represented by independent barcodes. We then required one of the Gaussian distributions in each mixture model to have a mean and standard deviation fixed to that of the corresponding wildtype measurements (ScHsp90 or ancAmoHsp90), with a freely estimated mixture proportion. The other Gaussian components in each mixture model had a freely fit mean, standard deviation, and mixture proportion. Mixture models were fit to all non-outlier selection coefficients, because the presence of strongly deleterious selection coefficients ($s < -0.04$), which are unambiguously deleterious, interfered with model convergence. We assessed mixture models with a variable number of mixture components ($k = 2$ to 6 for the ancAmoHsp90 library and 2 to 5 for the ScHsp90 library, because the 6-component model would not converge), and obtained the maximum likelihood estimate of each component's mean, standard deviation, and mixture proportion via an expectation-maximization algorithm as implemented in mixtools. We compared the models built for each k using AIC. For ScHsp90, the 3-component mixture model was favored by AIC (Appendix 1 Fig. A1.4a). For ancAmoHsp90, the 2-component and 5-component mixture models had virtually indistinguishable AIC (Appendix 1 Fig. A1.8a), but the 2-component mixture model had a visually suboptimal fit (Appendix 1 Fig. A1.8c,d) and attributed a larger proportion of mutations as being deleterious (0.53 versus 0.48 for the 5-component mixture model), so we selected the more conservative and visually superior 5-component mixture model.

The mixture component derived from the wildtype sampling distribution was taken to represent neutral mutations in the library, and the other mixture components were taken to reflect non-neutral mutations. We estimated the fraction of neutral mutations in each distribution as the mixture proportion of the neutral mixture component. We then estimated the proportion of deleterious (or beneficial) mutations as the cumulative density of all non-neutral components that

was below (or above) zero. These proportions were re-weighted after including the strongly deleterious states that were excluded from the model fitting step, as described above.

Uncertainty in the estimated fraction of mutations that are deleterious or beneficial was determined via a bootstrapping procedure. For each of 10,000 bootstrap replicates, measured selection coefficients from the bulk competition were resampled with replacement. Mixture models with fixed k were fit to each bootstrap sample, and the estimated fractions of mutations that are deleterious, neutral, or beneficial were determined as above.

For the representations in Figs. 3.1c and 3.3, we sought to assign a probability of being deleterious, neutral, or beneficial to each individual mutation. For each mutation, we calculated the posterior probability that it is neutral (PP_{neut}) as the probability density of the neutral mixture component at the observed selection coefficient measured for the mutant (s_{obs}), divided by the sum of the probability density of the neutral and non-neutral components at s_{obs}. The posterior probability that a mutation is non-neutral ($PP_{non\text{-}neut}$) is 1-PP_{neut}. The posterior probability that a mutation is beneficial (PP_{ben}) was calculated as $PP_{ben} = PP_{non\text{-}neut}$ if $s_{obs} > 0$, and zero if $s_{obs} < 0$; the posterior probability that a mutation is deleterious (PP_{del}) was calculated as $PP_{del} = PP_{non\text{-}neut}$ if $s_{obs} < 0$, and zero if $s_{obs} > 0$. For variants with $s < -0.04$ that were excluded from the model fitting step, $PP_{del} = 1$. To generate the representations in Figures 3.1c and 3.3a, PP_{del}, PP_{ben}, and PP_{neut} were each summed for the for the set of mutations that fall within each histogram bin.

To estimate the probability that a pair of states exhibit contingency and/or entrenchment (Fig. 3.3b), we calculated the joint posterior probability as the product of the probabilities that each pair of sites is in the relevant selection category (ancestral state with fitness greater than, less than, or indistinguishable from the derived state) in the ScHsp90 and the ancAmoHsp90 backgrounds. For sites that substituted from the ancAmoHsp90 state i to the ScHsp90 state j

($i \rightarrow j$, n = 35), i is the ancestral state and j the derived state for measurements in both backgrounds. For sites that substituted from the ancAmoHsp90 state i to an intermediate state j before substituting back to i in ScHsp90 ($i \rightarrow j \rightarrow i$, n=12), then i is the ancestral state and j derived in ancAmoHsp90 assay, and j is the ancestral state and i derived in ScHsp90. For sites that substituted from the ancAmoHsp90 state i to an intermediate state j that was further modified to k in ScHsp90 ($i \rightarrow j \rightarrow k$, n=25), two comparisons were made: in the first, i was ancestral and k was derived for measurements in both backgrounds, while in the second comparison, i was ancestral and j derived in ancAmoHsp90, and j ancestral and k derived in ScHsp90.

In addition to the mixture model approach presented above, we report three independent methods for estimating the fraction of mutations in each dataset that are deleterious (or beneficial) (Appendix 1 Fig. A1.5a). First, we estimated the fraction of mutations in each distribution that are deleterious (or beneficial) as the fraction of observed selection coefficients (s_{obs}) that are less (or greater) than zero. This counting approach assumes that, at some magnitude, all mutations have a true $s > 0$ or $s < 0$, in contrast to the mixture model approach that designates some mutations as neutral. This method would be unbiased if experimental errors are random and if the number of truly beneficial and truly deleterious mutations is equal. In our data, experimental errors are unbiased with respect to s_{obs} (Appendix 1 Fig. A1.5b,c), but there appear to be more deleterious than beneficial mutations. As a result, measurement error is likely to cause the number of mutations with true $s < 0$ and $s_{obs} > 0$ to exceed the number with true $s > 0$ and $s_{obs} < 0$; this approach is therefore expected to underestimate the fraction of mutations with true $s < 0$.

Second, we used an empirical Bayes approach. For each mutation, we compute the posterior probability that it is non-neutral by comparing the likelihoods of two hypotheses: the

null hypothesis, that a variant is neutral and therefore $s \sim N(0, SEM_{wt})$, and the alternative hypothesis, that a variant is non-neutral and therefore $s \sim N(s_{obs}, SEM_{mut})$. We estimated SEM_{wt} by calculating the sample standard deviation of repeated wildtype fitness measurements (Appendix 1 Figs. A1.4b, A1.8b), divided by $\sqrt{2}$, because mutant s_{obs} are calculated from duplicate measurements. We estimated SEM_{mut} over all duplicate bulk fitness measurements, which makes the assumption that all variants have the same experimental error (Appendix 1 Fig. A1.5b,c). SEM_{mut} was calculated as:

$$SEM_{mut} = \sqrt{\frac{\sum_{i=1}^{N}(s_i - \overline{s_i})^2}{N}}$$

where s_i is a measured selection coefficient of a mutant in a single replicate, $\overline{s_i}$ is the corresponding mean selection coefficient for that mutant as calculated from both replicates, and N is the total number of observations from both replicates. The estimated values for SEM_{wt} and SEM_{mut} were similar (0.0040 and 0.0041, respectively). The posterior probability that a variant is non-neutral is calculated from the relative likelihoods of the two hypotheses, with a uniform prior on the two hypotheses:

$$P(\text{non-neutral}) = \frac{P(s_{obs}|s \sim N(s_{obs}, SEM_{mut})}{P(s_{obs}|s \sim N(s_{obs}, SEM_{mut}) + P(s_{obs}|s \sim N(0, SEM_{wt})}$$

If a variant has $s_{obs} > 0$, then P(non-neutral) corresponds to a probability that a mutation is beneficial; if a variant has $s_{obs} < 0$, then P(non-neutral) corresponds to a probability that a mutant is deleterious. To estimate the proportion of mutations that are in each fitness category, we summed the probabilities for each category across all mutants.

Last, we constructed a 95% confidence interval (CI) for each mutation given its mean selection coefficient and the estimated SEM_{mut} described above. We then counted the fraction of mutations that are below (or above) zero and whose 95% CI excludes zero. This yields a

conservative estimate for our parameter of interest, the total fraction of mutations that are deleterious (or beneficial), as it is designed to indicate whether any particular mutation is deleterious (or beneficial), not to estimate the proportion (which does not depend on unequivocally classifying any one individual mutation as neutral or not).

Data and code availability. Processed sequencing data and scripts to reproduce all analyses are available at github.com/JoeThorntonLab/Hsp90_contingency-entrenchment. Raw sequencing data from each bulk competition have been deposited in the NCBI Sequence Read Archive under accession SRP126524.

Chapter 4

Alternative evolutionary histories in the sequence space of an ancient protein

The work described in this chapter was published as: Tyler Starr, Lora Picton, and Joseph Thornton. "Alternative evolutionary histories in the sequence space of an ancient protein." Nature 549:409-413 (2017).

4.1 Summary

To understand why molecular evolution turned out as it did, we must characterize not only the genetic path that evolution followed across the space of possible molecular sequences but also the many alternative trajectories that could have been but taken but were not. A large-scale comparison of real and possible histories would help to establish whether the outcome of evolution represents a unique or optimal state driven by natural selection or the contingent product of historical chance events (158); it would also reveal how the underlying distribution of functions across sequence space shaped historical evolution (1, 61). Here we combine ancestral protein reconstruction (126) with deep mutational scanning (10, 22-24, 119, 124) to characterize alternate histories in the sequence space around an ancient transcription factor, which evolved a novel biological function through well-characterized genetic and biochemical mechanisms (70, 106). We found hundreds of alternative protein sequences, distributed in clusters across sequence space, that use diverse biochemical mechanisms to perform the derived function at least as well as the historical outcome. These alternatives all require prior permissive substitutions that do not enhance the derived function, but not all require the same permissive changes that occurred during history. We found that if evolution had begun from a different starting point within the network of sequences encoding the ancestral function, outcomes with different genetic and

biochemical forms would almost certainly have resulted; this contingency arises from the distribution of functional variants in sequence space and epistasis between amino acids. Our results illuminate the topology of the vast space of possibilities from which history sampled one path, highlighting how the trajectory and outcome of evolution depend on a serial chain of compounding chance events.

4.2 Introduction

We applied deep mutational scanning to the DNA-binding domain of a reconstructed ancestral steroid hormone receptor, whose historical trajectory of functional, genetic, and biochemical evolution is well understood. Steroid receptors are a family of paralogous metazoan transcription factors that mediate the classic biological responses to sex and adrenal steroids by binding to specific DNA sequences and regulating the expression of target genes. Steroid receptors fall into two clades that differ in their DNA-binding specificity (Fig. 4.1a): estrogen receptors specifically bind an inverted palindrome of the sequence AGGTCA (estrogen response element, ERE)(159), and receptors for androgens, progestogens, and corticosteroids prefer inverted palindromes of AGAACA (steroid response element, SRE)(160). Although steroid receptors tolerate some degeneracy in their response elements, these sequences represent the high-affinity consensus binding sites for each class (159, 160) and have therefore been the focus of extensive biochemical characterization (161-164).

In previous work, the ancestral protein from which all steroid receptors descend (AncSR1) was reconstructed and shown experimentally to specifically bind ERE (70, 106). After duplication of AncSR1, one of the daughter proteins diverged in function to yield the subsequent ancestral protein AncSR2, which strongly prefers SRE. When three substitutions from this

historical interval are reintroduced into AncSR1, they radically shift relative affinity for ERE and SRE, an effect that is robust to statistical uncertainty about the ancestral sequence (165). These key substitutions are located on the protein's recognition helix (RH), which directly contacts the response element's major groove (161-163). Although they shift specificity, the RH substitutions yield a protein with affinity too low to activate transcription. Another eleven substitutions (11P) from this evolutionary interval – located outside the RH – were permissive, increasing affinity for both DNA targets, thereby allowing the protein to tolerate the function-switching RH substitutions (70).

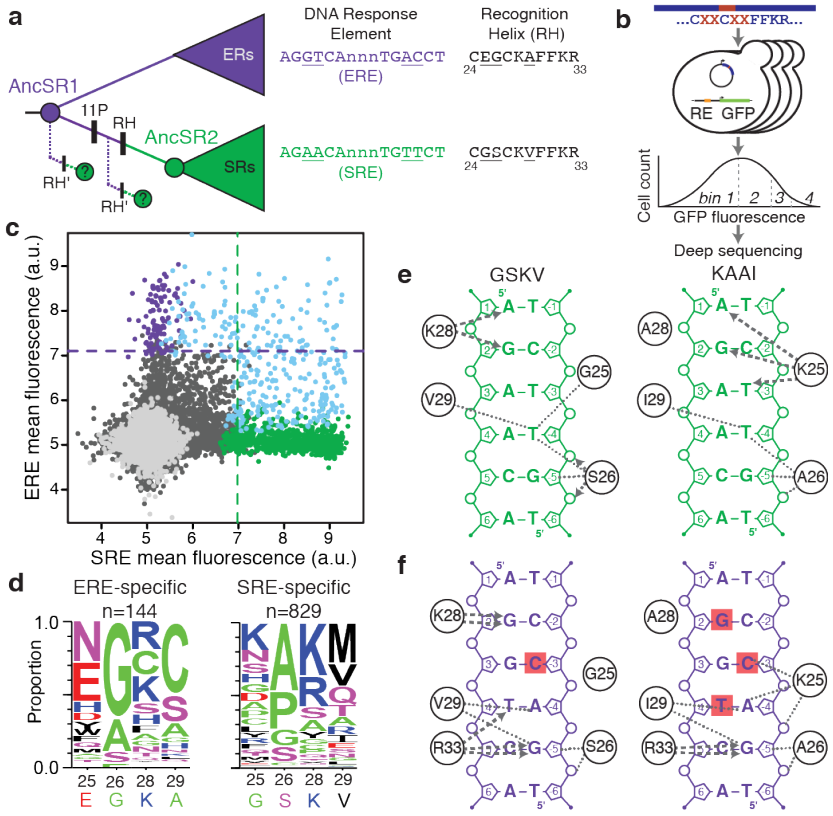


Figure 4.1. Diverse sequences and mechanisms can yield the derived DNA specificity. a, The historical transition in DNA-binding specificity in steroid receptors occurred through 3 changes in the recognition helix (RH), which required permissive substitutions (11P). Here we ask whether alternate recognition helix mutations (RH′) existed that recapitulate the derived function, either before or after the historical permissive substitutions occurred. Preferred DNA response elements for each clade are shown, along with the amino acid sequence of the protein's

(**Figure 4.1, continued**) RH (residue numbers 24-33); underlines indicate historically variable nucleotide and amino acid states. ERs, estrogen receptors; APGMRs, receptors for androgens, progestogens, glucocorticoids, and mineralocorticoids. Reconstructed ancestral proteins are labeled and colored by their RE preference. Dotted lines indicate possible alternate trajectories. **b**, A yeast-based FACS-seq assay for steroid receptor DNA-binding. A combinatorial library of 160,000 RH variants (top; Xs denote variable sites) was cloned into yeast reporters containing ERE or SRE upstream of yEGFP. The activity of each RH variant on each response element was determined by coupling FACS to deep sequencing. **c**, GFP activation on ERE and SRE by each RH variant in the AncSR1+11P background. Purple dots, ERE-specific variants; green, SRE-specific; blue, promiscuous; dark gray, non-functional; light gray, stop-codon-containing variants. Purple line, activity of AncSR1:EGKA on ERE; green line, activity of AncSR1+11P:GSKV on SRE. a.u., arbitrary fluorescence units. **d**, Logos showing the frequency of amino acid states at each variable RH position among ERE-specific and SRE-specific variants; n, number of variants in each class. States are colored by biochemical category (red, acidic; blue, basic; magenta, polar uncharged; black, large nonpolar; green, small nonpolar). Ancestral states for AncSR1 and AncSR2 are indicated along with corresponding residue numbers. **e**,**f**, Diverse biochemical mechanisms for recognition of SRE (**e**) or ERE (**f**) by the historical derived RH (GSKV) and an alternative SRE-specific variant (KAAI). Contacts in FoldX-generated structural models are shown between RH residues (circles) and DNA bases (letters), backbone phosphates (small circles) and sugars (pentagons, numbered by position in the DNA motif; dashed numbers refer to the complementary strand). Hydrogen bonds are shown as dashed arrows from donor to acceptor; dotted lines, non-bonded contacts. Red squares, bases that form hydrogen bonds in the EGKA-ERE structure that are unsatisfied in complex with an SRE-specific RH. Only DNA contacts that vary among the analyzed structures (Appendix 2 Figure A2.4) are shown.

4.3 Results

4.3.1 Deep mutational scanning of an ancient evolutionary transition

To characterize alternative outcomes and pathways by which SRE specificity could have evolved (Fig. 4.1a), we focused on the RH, the only portion of the protein that directly contacts the nucleotides that vary between ERE and SRE. We prepared a library that contains all 160,000 combinations of all 20 amino acids at four key sequence sites in the RH – the three residues that historically shifted DNA specificity, plus a physically adjacent lysine that varies among the broader nuclear receptor superfamily. The library was constructed in AncSR1+11P, the genetic background that enabled the historical RH substitutions to alter DNA specificity. We engineered

yeast reporter strains in which the consensus ERE or SRE sequence drives expression of a fluorescent GFP reporter; GFP activation in this system directly relates to DNA affinity (Appendix 2 Fig. A2.1a) (70, 106, 164). We transformed the RH library into each reporter and used FACS coupled to deep sequencing (FACS-seq) to quantify the ability of each variant in the library to bind ERE or SRE (Fig. 4.1b, Appendix 2 Figs. A2.1-3, Table A2.1). We validated this method by directly measuring GFP activation of many randomly chosen variants and comparing these values to their FACS-seq activities (Appendix 2 Fig. A2.1e). We classified genotypes as ERE-specific, SRE-specific, promiscuous, or inactive; the results of all downstream analyses were robust to the specific classification criteria (Appendix 2 Table A2.2).

4.3.2 The historical outcome is not unique in its function

We found 828 new RH variants that are SRE-specific, binding SRE as well or better than the historical outcome and displaying no activity on ERE (Fig. 4.1c). The historical outcome of evolution was therefore not unique in encoding specificity for SRE over ERE. These alternative SRE-specific genotypes are not subtle variations on the historical genotype; rather, they employ amino acid states with diverse biochemical characteristics (Fig. 4.1d), and they discriminate between SRE and ERE using different sets of physical contacts (Figs. 4.1e, f, Appendix 2 Fig. A2.4). For example, the historical outcome (sequence GSKV) binds SRE in part via polar contacts from residue Lys28 to nucleotides A1 and G2, but the alternative RH genotype KAAI makes no polar contacts using residue 28, instead making hydrogen bonds from Lys25 to A1, G2, and the opposite-strand nucleotide T−3 (Fig. 4.1e). It also exhibits novel mechanisms of ERE-exclusion: whereas GSKV leaves the hydrogen bonding potential of C−3 unsatisfied, KAAI also leaves G2 and T4 unpaired, because Ala28 – unlike Lys28 of GSKV – cannot bond to G2,

and Ile29 interferes with a hydrogen bond to T4 made by the conserved Arg33 residue (Fig. 4.1f, Appendix 2 Fig. A2.4c).

4.3.3 The historical outcome is not unique in its accessibility

Although the historical outcome is not unique in its function or biochemical mechanism of SRE specificity, it might have been uniquely accessible from the ancestral RH. To investigate the distribution of functions across sequence space, we constructed a force-directed graph of functional RH variants (Fig. 4.2a). Each node represents a functional RH genotype, and edges connect RH genotypes separated by one nonsynonymous nucleotide mutation. Although the vast majority of RH variants are nonfunctional, virtually all of the 1351 functional variants are part of a single connected network that can be traversed by single-nucleotide mutations (steps) without visiting nonfunctional genotypes (1). ERE-specific, SRE-specific, and promiscuous variants are interspersed throughout the network, resulting in a very large number of potential evolutionary paths. The network contains several clusters of highly interconnected variants that share distinguishing amino acid state patterns, with epistasis and the structure of the genetic code magnifying separation among clusters.

The ancestral and derived RHs (EGKA and GSKV, respectively) are connected by a path of just three steps, whereas the most distant proteins in the network are 13 steps apart. From the ancestral starting point, GSKV is not a uniquely accessible outcome: 64 other SRE-specific RHs are accessible in three or fewer steps without passing through nonfunctional intermediates. Some of these alternative outcomes can be reached in just one or two steps, and these too exhibit biochemically diverse amino acid states (Appendix 2 Fig. A2.5a). Similar conclusions emerge under a variety of evolutionary scenarios. For example, if selection against too-tight or too-weak

binding allows access only to genotypes with SRE affinity in a narrow range indistinguishable from the historical outcome, there are still hundreds of alternative variants, many of which are easily accessible from the historical starting point (Appendix 2 Table A2.2, column E). Even when trajectories are allowed only if SRE affinity increases at every step – as would occur under positive selection for that function – there are numerous alternative SRE-specific genotypes with a nontrivial probability of evolving from the ancestral RH, and all of these are more likely than the historical outcome (Appendix 2 Fig. A2.5a,b,c). Taken together, these data indicate that the historical trajectory was not the only path, or even the shortest, from the ancestral RH to a derived protein that is SRE specific.

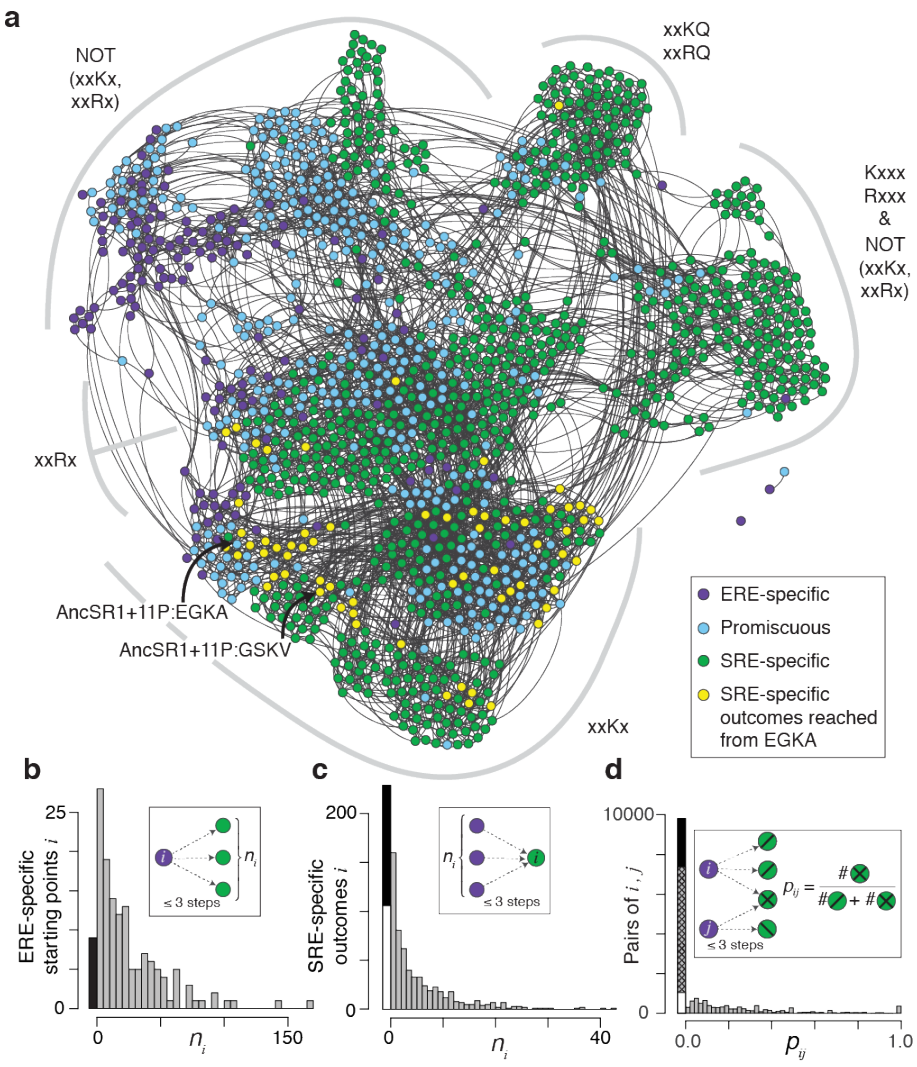


Figure 4.2. Evolvability of SRE specificity in an ancestral sequence space. a, Functional topology of the RH sequence space in AncSR1+11P is shown as a force-directed graph. Nodes are functional RH variants, colored by specificity class; SRE-specific variants accessible from EGKA in three or fewer mutational steps are yellow. Edges connect variants separated by one nonsynonymous nucleotide mutation. Densely connected sets of nodes cluster close together. Large clusters (grey arcs) are labeled by their defining genetic features; x's denote variable sites within a cluster. Historical ancestral and derived RH genotypes are indicated. **b**, Distribution of ERE-specific starting points by number of SRE-specific outcomes reached in a trajectory no longer than the historical 3-step path. Black bar, ERE-specific variants that reach zero SRE-specific outcomes because of epistasis (all possible paths of ≤ 3 steps are blocked by nonfunctional intermediates). **c**, Distribution of SRE-specific outcomes by number of ERE-specific starting points that reach it by a trajectory of length ≤ 3 steps. Black bar, outcomes reached from zero starting points because of epistasis; white, outcomes reached from zero starting points because of the diameter of the functional network of genotypes (all starting points would require >3 nonsynonymous mutations irrespective of the functionality of intermediate genotypes). **d**, Distribution of pairs of ERE-specific starting points by the fraction of SRE-

(**Figure 4.2, continued**) specific outcomes they reach in ≤ 3 steps that are shared. Black bar, pairs of starting points with zero shared outcomes because of epistasis; white, pairs with zero shared outcomes because of the diameter of the functional network of genotypes; grey hatched, pairs with zero shared outcomes because of the distribution of functions across sequence space (genotypes could have been reached in ≤ 3 nonsynonymous mutations from both starting points but none are SRE-specific).

4.3.4 The historical starting point is not unique in its evolvability

The ancestral RH was just one of many possible starting points within a large network of mutually accessible ERE-specific genotypes. To determine whether the evolution of SRE specificity depended on the starting point within this network, we identified trajectories to SRE specificity from all ERE-specific RHs. All but 2 ERE-specific starting points can access SRE specificity without passing through nonfunctional intermediates (Fig. 4.2a), and more than 90% can do so by paths no longer than the historical trajectory (Fig. 4.2b). This observation indicates that evolution of the derived specificity per se was not strongly dependent on the starting point. Whether any particular SRE-specific genotype would evolve, however, could be contingent on where in the network of ERE-specific variants an evolutionary trajectory begins. For each SRE-specific RH, we asked how many ERE-specific starting points could access it by a path no longer than the historical three-step trajectory (Fig. 4.2c). About one-third of possible outcomes are not easily reached from any possible starting point – some because the large diameter of the functional network means that the minimum genetic distance to the closest ERE-specific variant is more than three nonsynonymous mutations, and some because epistasis requires trajectories longer than the minimum genetic distance (22, 24). Of the remaining SRE-specific variants, most (including the historical outcome GSKV) are readily accessible from just one or a few starting points, and even the most accessed outcome is easily reached from less than one-third of all possible starting points. As a result, most pairs of ERE-specific starting points lead to entirely

non-overlapping sets of SRE-specific outcomes (Fig. 4.2d), which contain genetically and biochemically distinct sets of amino acids (Appendix 2 Fig. A2.6a). The evidence for dependence on starting point persists when path lengths much longer than the historical trajectory are considered (Appendix 2 Fig. A2.6b,c,d) and when alternate evolutionary models are applied (Appendix 2 Table A2.2). Taken together, these data indicate that the derived specificity for SRE could have evolved in many ways from AncSR1+11P, but the underlying genetic and biochemical form depended strongly on the starting RH genotype.

4.3.5 Historical permissive substitutions are broadly permissive

We next asked how the historical permissive substitutions affected the accessibility of the derived specificity and its dependence on starting point. We constructed and characterized the same four-site combinatorial RH library, this time in the AncSR1 background without 11P (Fig. 4.1a, Fig. 4.3a, Appendix 2 Fig. A2.1). Removing 11P causes a large reduction in the number of functional variants (Fig. 4.3b) and a striking reduction in the size and connectivity of the functional network (Fig. 4.3c). Unlike the RH network in AncSR1+11P, many functional variants in AncSR1 are isolated and therefore cannot be reached from most other genotypes without passing through nonfunctional intermediates. Nevertheless, most functional RHs – including the ancestral RH (EGKA) – remain interconnected in the primary sub-network, within which numerous SRE-specific RHs are accessible. This indicates that although the historically derived RH genotype GSKV requires the historical permissive substitutions, other RH genotypes with the derived specificity could have evolved in the absence of 11P. The set of trajectories from AncSR1 to genotypes encoding SRE-specific variants, however, are more complex than when 11P are present: in the absence of 11P, the shortest path from the ancestral RH to the

closest SRE-specific genotype is 5 steps long, all paths require permissive RH steps that do not enhance SRE activity, and all paths pass through promiscuous intermediate genotypes (Appendix 2 Fig. A2.7a,b). Thus, in the absence of the historical permissive substitutions, other permissive mutations would have been required for SRE specificity to evolve from the ancestral genotype.

Comparison of the two networks shows that the 11P substitutions enhanced the accessibility of SRE specificity not only from the ancestral genotype but from all potential ERE-specific starting points. Whereas virtually all starting points in the AncSR1+11P network could access at least one SRE-specific node without passing through nonfunctional intermediates, more than one-fourth of ERE-specific variants in the network without 11P have no connected path to the derived specificity, and those that can access SRE specificity require longer paths (Fig. 4.3d). Removing the historical permissive substitutions also increases the proportion of ERE-specific starting points that require a permissive step prior to acquiring activity on SRE (Fig. 4.3e). And, unlike the AncSR1+11P network, every path from the ancestral to the derived specificity in the AncSR1 space must pass through a promiscuous intermediate (Fig. 4.3e).

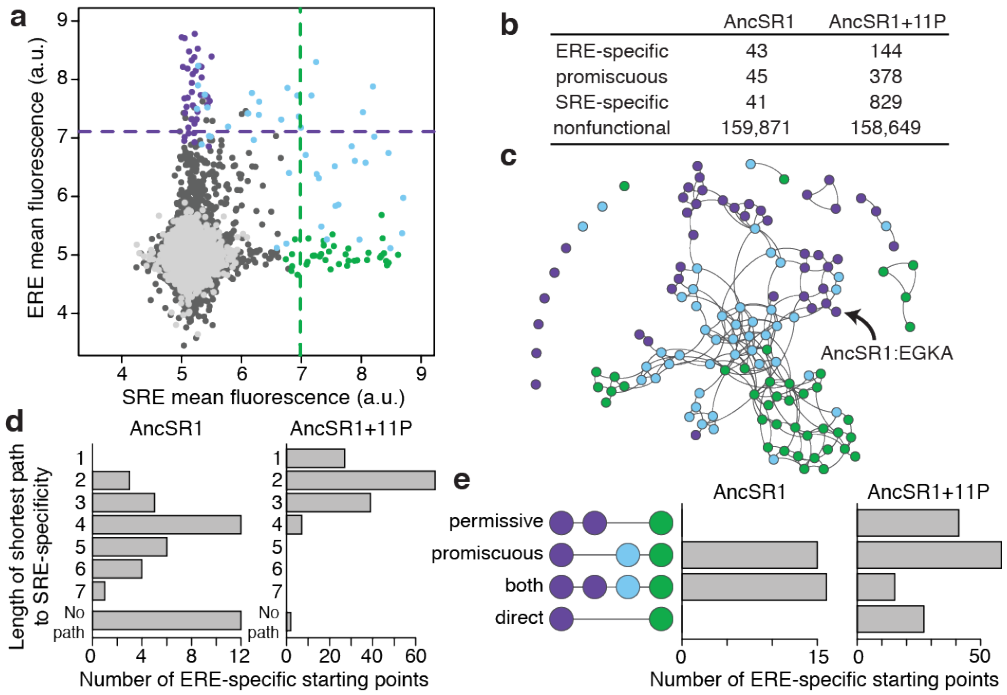


Figure 4.3. Historical permissive substitutions broadly enhanced evolvability of SRE specificity. a, Each RH variant's GFP activation on ERE and SRE in the AncSR1 background; a.u., arbitrary fluorescence units. Figure details as in Fig. 4.1c. **b**, Number of RH variants in each functional class in the AncSR1 and AncSR1+11P backgrounds. **c**, Functional topology of the RH sequence space in AncSR1 is shown as a force-directed graph, represented as in Fig. 4.2a. **d**, Distribution of ERE-specific starting points by the length of the shortest possible path to an SRE-specific variant in the AncSR1 (left) and AncSR1+11P (right) functional networks. 11P reduce the shortest RH path length ($P < 10^{-12}$, Wilcoxon rank sum test with continuity correction). **e**, For all ERE-specific starting points in AncSR1 (left) and AncSR1+11P (right), the shortest path(s) to SRE specificity classified by characteristics of the trajectory: permissive (one or more ERE-specific intermediates), promiscuous (one or more promiscuous intermediates), both, direct (one-step path with neither permissive steps nor promiscuous intermediates), or no path (all paths require nonfunctional intermediates). If a starting point has multiple equally short paths, it contributes to each category proportionally. Distributions differ between the networks ($P < 10^{-7}$, Chi-squared test with simulated p-value).

Finally, we investigated the mechanism by which the historical permissive substitutions changed the topology of the RH sequence space and enhanced the potential for evolution across it. 11P were broadly permissive, increasing the number of SRE-specific genotypes in the network by a factor of 20 (Fig. 4.3b). Previous work has suggested that increases in protein stability often mediate this kind of generalized permissive effect (7, 34, 69, 135), but 11P has

been shown not to increase the stability of AncSR1 (70). It was previously proposed that 11P permitted the historical RH substitutions, which shift DNA preference but reduce affinity to a level below that required for activation, by nonspecifically increasing affinity for both kinds of response element (70). This hypothesis, which could also explain the broadly permissive effect of 11P on many RH genotypes, makes four testable predictions. First, those RH variants that do not require 11P to encode a functional SRE-specific DNA-binding domain should have greater affinity for SRE and higher mean fluorescence in FACS-seq than those that require 11P, whether or not 11P are present; we compared the predicted affinity and mean fluorescence of all 11P-independent and 11P-dependent SRE-specific variants and found that this prediction holds true (Fig. 4.4a, Appendix 2 Fig. A2.8a-d). Second, if 11P nonspecifically increase affinity, they should not change the genetic determinants of binding within the RH; as predicted, the amino acids that are most enriched among SRE-specific variants do not change, but the magnitude of preference becomes more evenly distributed among tolerated states (Fig. 4.4b, Appendix 2 Fig. A2.8e). Third, if 11P are nonspecific enhancers of affinity, they should not change the biochemical mechanisms by which the RH confers specificity, a prediction we tested by identifying the site-specific biochemical properties in the RH that determine specificity for ERE and SRE (Appendix 2 Fig. A2.8f): we found that the determinants of SRE specificity are not dramatically altered by 11P (Fig. 4.4c). Fourth, if 11P nonspecifically enhance affinity by all RHs, they should add new functional genotypes across sequence space; we found that the set of variants permitted by 11P are not localized to some region of the network but instead surround the sparser set of variants that functioned independently of 11P (Fig. 4.4d,e). As a result, 11P's nonspecific effect on affinity enhanced the connectivity of the ancestral sequence network, increasing the number and reducing the length and complexity of paths from ERE to SRE.

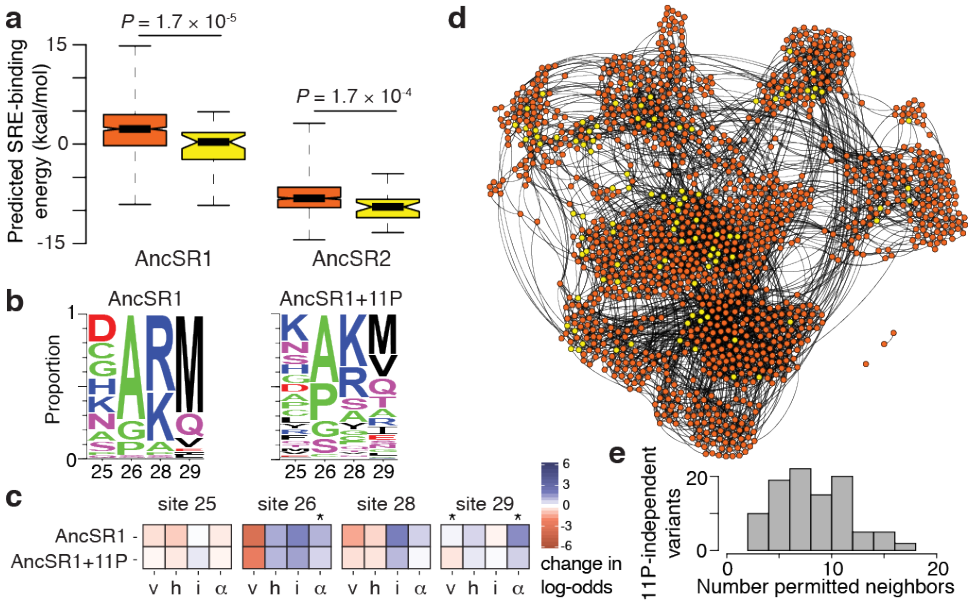


Figure 4.4. The effect of historical permissive substitutions is mediated by nonspecific increases in affinity. a, Predicted SRE-binding affinity of SRE-specific RH variants that are functional in the absence of 11P (yellow, n=41) or that require 11P to be functional (orange, n=790), modeled by FoldX in the AncSR1 (left) and AncSR2 (right) crystal structures. In each category, the median (bar), approximate 95% confidence interval (notch), interquartile range (colored box), and range (whiskers) are shown. The *P*-value for the difference in medians is shown (Wilcoxon rank sum test with continuity correction). **b**, Logos showing the frequency of amino acid states at each variable RH position among SRE-specific variants in the AncSR1 (left) and AncSR1+11P (right) backgrounds. States are colored by biochemical category as in Fig. 4.1d. **c**, Biochemical determinants of SRE specificity in the AncSR1 (top) and AncSR1+11P (bottom) backgrounds. A multiple logistic regression model predicts the probability that a variant is SRE-specific from the biochemical properties of its amino acid state at each of the four variable RH sites: v, volume; h, hydrophobicity; i, isoelectric point; α, α-helix propensity. Colored boxes show the best-fit coefficients of this model as the change in log-odds of being SRE-specific per unit change in each property. Asterisks indicate site-specific determinants that differ significantly between the AncSR1 and AncSR1+11P background (*Z*-test, $P < 0.05$). **d**, The AncSR1+11P RH functional network, indicating the location of variants that are functional in the absence of 11P (yellow) and those that require 11P to be functional (orange). **e**, 11P permit immediate neighbors of 11P-independent variants. For each RH genotype that was functional in the absence of 11P, the number of single-mutant neighbors that became functional when 11P was introduced.

4.4 Discussion

Our results shed light on the roles of determinism and chance in protein evolution (7, 43, 61, 158). The primary deterministic force is natural selection, which drives the evolution of

forms that optimize fitness. Chance appears in two non-exclusive ways: as historical contingency – when the accessibility of some outcome depends on prior events that cannot be driven by selection for that outcome – and as stochasticity – when there are numerous possible paths to genotypes of similar function, and which one is realized is random (Appendix 2 Fig. A2.7c). Previous work has shown that historical function-switching substitutions in some proteins were contingent on prior permissive mutations (38, 42, 57, 69, 70), but the overall roles of chance and determinism in the evolution of a new function can be understood only by characterizing other ways that the function could have evolved. Our results point to a major role for stochasticity and contingency in the many possible histories by which SRE specificity could have evolved from AncSR1. Hundreds of genotypes encoding SRE specificity were accessible from AncSR1, but selection for that function alone could not have deterministically driven evolution down any of those paths, because all were contingent on permissive mutations – either the historical 11P substitutions or the alternative permissive mutations we discovered within the RH. Which particular permissive mutations happened to occur determined which SRE-specific genotypes then became accessible. Further, given some permissive set of first steps, paths to numerous SRE-specific genotypes typically become available. Thus, evolution of any particular SRE-specific outcome – including the one that evolved during history – would be contingent on the initial stochastic acquisition of some set of permissive mutations, followed by the subsequent stochastic realization of one of many possible ways to encode the derived function. These serial stochastic choices result in compounding contingency, magnifying the role of chance in evolution.

Some aspects of biological history cannot be reconstructed, but our conclusions are likely to be robust to major forms of uncertainty. For example, the probability of any evolutionary

trajectory across sequence space depends on both the quantitative relationship between molecular function and organismal fitness and on population size, but neither of these is known. Nevertheless, we found that contingency and stochasticity were important not only under scenarios that maximize their effects – such as when evolution proceeds primarily by drift and purifying selection, avoiding nonfunctional genotypes – but also under those that favor determinism, as when selection drives continuous enhancement of the derived function or allows affinity within only a narrow range. Second, sequence space is so vast that we could comprehensively explore only a limited portion, studying variability at a relatively small number of key sites and evaluating the presence or absence of all 11 historical permissive substitutions as a group. But contingency and stochasticity are likely to remain important when larger regions of sequence space are considered. If – as seems likely – these unexplored regions contain additional trajectories to SRE-specific outcomes, then the role of stochasticity in the "choice" among these many options would be even more important. The contingency on starting point that arises from the broad distribution of SRE-specific genotypes across sequence space would persist even if new potential outcomes were discovered, and it would be magnified if those outcomes were even more distant than those we characterized. Finally, the dependence on permissive mutations that we observed would be eliminated only if there is a mutation at some other site that could confer SRE activity on AncSR1 in a single step; this seems implausible, because all other residues in the protein are distant from the variable DNA bases and therefore cannot confer SRE-binding without causing a major conformational change that would somehow bring a pre-existing surface compatible with SRE into contact with the response element.

Despite the abundance of accessible SRE-specific genotypes in the sequence space near the ancestral and derived RHs (Appendix 2 Fig. A2.5d,e), the RH genotype that historically

evolved is conserved among present-day descendants. We cannot rule out the possibility that some unknown property made this sequence selectively superior to the hundreds of other genotypes that are at least as effective at recognizing SRE and excluding ERE. But its conservation could also have been caused by factors that accumulated after its stochastic realization. For example, a substitution can become epistatically entrenched by subsequent restrictive substitutions at other sites in the protein (30, 107). A transcription factor's sequence may also become pleiotropically entrenched by subsequent mutations in the ensemble of response elements it binds (166). If one of the many alternative SRE-specific outcomes had instead evolved from the ancestral protein by chance, it too could have been subsequently locked in, yielding conservation and the illusion that it evolved deterministically. The singularity of the present seems to rationalize the past. History leaves no trace of the many roads it did not take, or of the possibility that evolution turned out as it did for no good reason at all.

4.5 Methods

Construction and validation of a yeast assay for steroid receptor DNA-binding domain function. All work was performed in *S. cerevisiae* strain K20 (CEN.PK 102-5B, *URA3*$^-$, *HIS3*$^-$, *LEU*$^-$) (167). We constructed yeast reporter strains containing yEGFP under the control of a minimal *CYC1* promoter with two upstream ERE or SRE palindromes, integrated into the *ADE2* locus (167). Colony PCR and Sanger sequencing confirmed correct integration of the ERE_2-*yEGFP* or SRE_2-*yEGFP* reporter. An additional 20 µg/mL adenine hemisulfate was added to all media to ameliorate *ADE2* disruption.

The yeast expression plasmid pTNS33 contains the AncSR1 DNA-binding domain (DBD, GenBank AJC02122.1) (70) with an N-terminal SV40 nuclear localization sequence and

Gal4 Activation Domain (AD) connected by a 9-residue linker (IQQGGSGGS). Expression of the AD-DBD fusion protein is controlled by the galactose-inducible *GAL1* promoter, in the background of the pRS413 plasmid (168) containing a *HIS* selection marker. We assembled pTNS33 by yeast homologous recombination using the LiAc/ssDNA/PEG method (169), selecting for growth on SC-His plates with 2% Dextrose (+D). We confirmed correct plasmid assembly via Sanger sequencing.

To validate the *ERE_2-yEGFP* and *SRE_2-yEGFP* reporters, a selection of previously assayed DBDs spanning a range of DNA-binding affinities (70, 106) were cloned into the pTNS33 background and transformed into each yeast reporter strain. Individual colonies were inoculated in 3mL SC-His with 2% raffinose (+R), and incubated for 16 hours at 30 ºC 225 rpm in an orbital shaker incubator. Cells were back-diluted to 0.25 OD_{600} in SC-His with 2% galactose (+G) to induce DBD expression and grown for an additional 24 hours. Cells were pelleted and suspended to 1 OD_{600} in 1x TBS. We analyzed 10,000 cells of each genotype by flow cytometry on a BD LSR-Fortessa 4-15, with 488 nm excitation and 530 nm emission. We used gates drawn empirically on FSC/SSC and FSC-H/FSC-A scatter plots (e.g. Appendix 2 Fig. A2.2) to isolate a homogeneous cell population, from which we determined the mean per-cell green fluorescence. The relationship between mean GFP activation and previously measured binding affinities was fit to a segmented-linear relationship in R with the 'segmented' package (170).

Library generation. AncSR1 and AncSR1+11P RH libraries were constructed by synthesizing pools of oligonucleotides containing degenerate NNK codons at four variable sites in the recognition helix and inserting these into coding sequences for the previously reconstructed AncSR1 DBD or the AncSR1+11P DBD, which contains the 11 previously

identified historical permissive mutations (70). These libraries encode all combinations of all 20 amino acids at the three RH sites that changed during the historical evolution of SRE specificity (sites 25, 26, and 29) and at the adjacent position (site 28), which physically interacts with the substituted residues (70) and varies among the broader nuclear receptor superfamily (171). Each RH library contains 1,048,576 genetic variants, encoding 160,000 full-length proteins and 34,481 stop-codon-containing variants. To construct the libraries, 53-nt single-stranded DNA oligonucleotides were synthesized (DNA2.0, Newark, California), containing variable RH sites and invariant flanking sequence identical to the respective plasmid sequences. Oligonucleotide pools were converted to dsDNA by primer extension with Klenow polymerase and purified on a Qiagen MinElute column. Yeast expression plasmids containing AncSR1 or AncSR1+11P were modified by site-directed mutagenesis to introduce EcoRI and NcoI sites, which were cut to excise the native RH and linearize the vector to receive the oligonucleotide pool. Plasmid libraries were assembled via Gibson Assembly, incubating 0.56 pmol gel-purified linear vector, 8.4 pmol oligonucleotide pool, and 120 µL 2× GA Master Mix (NEB) at 50 ºC for 1 hr. Assembled libraries were purified over DNA Clean & Concentrator columns (Zymo) and transformed into electrocompetent NEB5α *E. coli* cells with a 2.5 kV electroporation pulse in 0.2 mm gap cuvettes. Aliquots of cells were serially diluted and plated on LB+carbenicillin to estimate transformation efficiencies. Remaining cells were grown overnight, and plasmids were harvested using the GenElute Midiprep plasmid purification kit. For both the AncSR1 and AncSR1+11P RH libraries, we obtained at least 20 times more transformants than the effective size of the library (Appendix 2 Table A2.1).

Each RH library (AncSR1 and AncSR1+11P) was independently transformed twice into each yeast reporter strain (ERE and SRE) for replicate FACS-seq analyses. We followed a yeast

electroporation protocol (172), scaled up for 10 times the number of cells and a total of 120 μg of library plasmid in 600 μL H_2O. An aliquot of cells was serially diluted and plated on SC-His+D to estimate transformation yield, which averaged 1.25×10^7 cfus across the 8 transformations (Appendix 2 Table A2.1). The remaining cells were grown to saturation in 500 mL SC-His+D. Consistent with previous observations (173), we observed that seven out of eight colonies post-transformation were multiple-vector transformants. We performed an additional passage, at which point multiple-vector clones were detected at less than one in eight colonies. A total of five passages occur prior to quantification (see below), so multiple vector transformants are expected to occur at a frequency no greater than 0.007 in the library. Furthermore, if our conclusion that there are many functional RH variants were caused by false positives due to co-transformation of nonfunctional genotypes with functional ones, this would result in stop-codon-containing variants being classified as functional, but this was never observed. Passaged yeast library aliquots of 3×10^9 cells were flash frozen in liquid nitrogen and stored at -80 ºC as 25% glycerol stocks.

Library induction and FACS. Yeast library aliquots were thawed on ice, added to 500 mL SC-His+D, and grown for 12 hours at 30 ºC 225 rpm. Cells were diluted to 0.25 OD_{600} in 500 mL SC-His+R, and grown for an additional 12 hrs at 30 ºC 225 rpm. Cells were then diluted to 0.25 OD_{600} in 200 mL SC-His+G to induce DBD expression, and grown for 24 hrs at 30 ºC 225 rpm. Induced cells were spun at 3,000 *g* for 5 min, suspended to 3×10^7 cells/mL in 1× TBS, passed through a 40 μm nylon cell strainer, and stored on ice for sorting. Alongside each library induction, we induced isogenic controls expressing known DBDs according to the same protocol but at 3 mL volumes.

Each library was sorted into 4 bins on a BD FACSAria II. Initial gates were drawn to isolate homogenous cells and exclude doublets, using SSC/FSC and FSC-H/FSC-A scatterplots (Appendix 2 Fig. A2.2). We assigned sort gate boundaries to the AncSR1+11P/SRE library to correspond to the observed mean fluorescence of a stop-codon-containing variant, of AncSR1+11P:GSKV, and AncSR1+11P:GGKA, the variant with the highest previously known activation; these gates yielded four bins that captured 45%, 45%, 9.5%, and 0.5% of the library population, respectively. Gates for other libraries were assigned to yield the same bin sizes. To calibrate the arbitrary-unit fluorescence scales of sorting experiments conducted on different days, we transformed fluorescence values by a linear model fit to the relationship between mean fluorescence of reference isogenic cultures induced and analyzed in parallel to each library sorting experiment. Cells were sorted into SC-His+D with 34 μg/mL chloramphenicol to prevent bacterial contamination and stored on ice until ~10^8 cells were sorted. An aliquot of cells sorted into each bin was serially diluted and plated to estimate cfu recovery (Appendix 2 Table A2.1). Remaining cells were suspended to an estimated 200,000 cells/mL in SC-His+D+chloramphenicol, and grown for 16 hours at 30 ºC 225 rpm. Plasmids were extracted from each outgrowth according to the protocol of Fowler et al. (174), which was scaled up 16-fold for bins 1 and 2, 8-fold for bin 3, and 3-fold for bin 4 to avoid bottlenecks. Extracted plasmids were estimated to be present at a concentration of 2×10^6 plasmids/μL by comparing bacterial transformation efficiencies of yeast-extracted plasmid to pUC19 and bacterial-purified plasmid standards.

Sequencing and processing. We used PCR to amplify the variable RH region from post-sort plasmid aliquots; primers appended in-line barcodes (175) to identify the experiment and sort bin, along with binding sites for sequencing primers and Illumina flow cell adapter

sequences. Barcodes were of different lengths to stagger reads across clusters and were assigned to bins to optimize the distribution of base calls at each position during the initial rounds of sequencing. Multiple barcodes were used for bins 1 and 2, which contained the majority of cells. For each bin-barcode combination, PCR was conducted in 8 replicate 50-µL aliquots, with 10 µL of plasmid template, 10 µL 5x HF buffer, 1 µL 10 mM dNTPs, 2.5 µL 10 µM forward and reverse primer, and 0.5 µL Phusion polymerase per reaction. PCRs were assembled on ice, transferred to a thermocycler block preheated to 98 ºC, and subjected to 20 PCR cycles with 60 ºC annealing. PCRs were gel-purified, quantified via BioAnalyzer and qPCR, and then pooled for sequencing according to the relative numbers of cells acquired in each bin. Single-end 50bp reads spanning the barcode and RH sequence were acquired on an Illumina HiSeq2500.

We discarded sequence reads with an average Phred score <30 and sequences that did not perfectly match the barcode and invariant portion of the template. Reads were demultiplexed by barcode and further processed using tools from the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). RH variants with inconsistent read numbers between barcodes in the same bin were considered uncharacterized for that entire experiment. This procedure yielded filtered read counts in each sort bin greater than the number of cells sorted into that bin (Appendix 2 Table A2.1). To estimate the number of cells of a genotype that were sorted into a bin, we divided the number of sequence reads of a genotype in a bin by the average number of reads per cell in that bin.

Estimating mean fluorescence and standard error. We estimated the mean fluorescence of each variant in the library from the distribution of its reads across fluorescence sort bins using a maximum likelihood approach (176). We first assessed the fit of various distributions to the observed per-cell fluorescence of a series of isogenic cultures of different RH

genotypes analyzed in isolation via flow cytometry, and found the logistic distribution to have the best fit by AIC (Appendix 2 Fig. A2.1b,c). We then used the 'fitdistrplus' (177) package in R to find the maximum likelihood mean fluorescence for each library variant given its distribution of cell counts across sort bins, the fluorescence boundaries of those bins, and the logistic distribution; this approach explicitly takes into account the fact that the fluorescence of a cell within a sort bin is not precisely measured and has been shown to be an unbiased approach for estimating underlying activities in FACS-seq analyses (176). Estimates of mean fluorescence from the FACS-seq library characterization were compared between independent replicates (Appendix 2 Fig. A2.1d). Interval-censored per-cell observations from the two independent replicates were then pooled, and the maximum likelihood mean fluorescence for each variant estimated from this pooled data. These final estimates were compared to fluorescence observed directly for isogenic cultures of randomly selected clones from each library, which were isolated post-sort, genotyped, re-induced in isogenic cultures and analyzed via flow cytometry according to the protocol above (Appendix 2 Fig. A2.1e).

We estimated the standard error of mean fluorescence (SEM) for genotypes based on their depth of coverage (number of cells sampled) in two ways. First, we estimated SEMs from stop-codon-containing variants in each library by binning them according to their depth of coverage and calculating the standard deviation of the sampling distribution of estimated mean fluorescence for variants in each bin. Second, we leveraged variability in the mean fluorescence estimates from the two replicate FACS-seq experiments for each library: using coding variants for which the number of cells sampled between replicates is within 20% of each other, we calculated the difference between the estimate of mean fluorescence from the pooled data and the estimates from each of the two replicates, binned variants by their average depth of coverage for

the two replicates, and calculated the standard deviation of the distribution of differences for each bin. Every variant in the library was then assigned the SEM for the appropriate coverage depth bin. These two approaches yielded a similar relationship between SEM and sampling depth, but the second approach estimated higher SEMs at higher coverage depths (Appendix 2 Fig. A2.1g); to be conservative, we therefore used the second approach for further analyses.

Classifying strength of activation on each response element. We used mean fluorescence estimates to classify the strength with which each library variant binds to ERE and SRE using nonparametric comparisons to distributions of reference genotypes. A variant was classified as active on a response element if its mean fluorescence was significantly greater than that of stop-codon-containing variants contained in the library: for each variant, the *P*-value for the null hypothesis that a variant is inactive was calculated as the proportion of stop-codon-containing variants of similar sampling depth with greater mean fluorescence than that of the variant of interest; variants were labeled "active" if the null hypothesis could be rejected at a 5% false discovery rate (using the Benjamini-Hochberg procedure) or "inactive" if the null hypothesis could not be rejected.

Each active variant was then subclassified as a weak or strong activator by comparing its mean fluorescence to that of the relevant ancestral genotypes (AncSR1:EGKA on ERE, or AncSR1+11P:GSKV on SRE). Specifically, for each active variant we performed a test of noninferiority within an equivalence margin of 20% of the range between the average mean fluorescence of stop-codon-containing variants and the mean fluorescence of the ancestral reference. This test compares the mean fluorescence of a variant of interest to the fluorescence of cells with the relevant ancestral genotype, shifted to 80% of the range between the mean of stop-codon-containing variants and the ancestral reference. To determine whether a variant's

fluorescence is greater than this shifted ancestral reference, we generated 10,000 bootstrap replicates from the shifted distribution of ancestral cellular fluorescence, with replicate size of similar sampling depth to the variant of interest; the mean fluorescence of each bootstrap replicate was calculated using the FACS gates and maximum likelihood procedure described above. The *P*-value for the null hypothesis that a variant is a weak activator was calculated as the proportion of bootstrap replicates with fluorescence greater than that of the variant of interest; variants were classified as "strong" if the null hypothesis could be rejected at a 5% false discovery rate (using the Benjamini-Hochberg procedure) or "weak" if the null hypothesis could not be rejected. AncSR1:EGKA was represented by relatively few cells in the ERE library, resulting in an artificially low mean fluorescence determined by FACS-seq and a "weak" classification, so it was manually classified as a strong activator on ERE by definition. For library classifications, we determined the reference activity of AncSR1:EGKA on ERE from an isogenic culture analyzed in parallel to library sorts. Using the lower FACS-seq mean fluorescence measurement as the reference activity for this genotype does not alter our conclusions (Appendix 2 Table A2.2, column A).

Extrapolation to missing genotypes. Classification of variants that are rare in the library may not be reliable. We examined how agreement in classification between FACS-seq replicates is affected by sampling depth, and we found that the probability that a variant is classified as positive in one replicate if it is classified as positive in the other depends on sampling depth below 15 cells (Appendix 2 Fig. A2.1f). We therefore considered variants with 15 or fewer cells to be experimentally undetermined, accounting for 2.0% to 8.8% of all variants across the four DBD/response element combinations (Appendix 2 Table A2.1). To predict the classification of these variants, we used a continuation ratio ordinal logistic regression model that predicts the

probability that a variant is strong, weak, or inactive from its genotype, trained on the empirical classification of all the determined genotypes in the library. We modeled amino acid states as potentially contributing first-order main effects (20 states × 4 positions = 80 parameters) and pairwise epistatic effects (${}_4C_2 \times 20^2$ = 2,400 parameters). We fit these models to the observed classifications in each library using a coordinate-descent fitting algorithm with L_1 penalization, as implemented in the 'glmnetcr' package (178) in R. We used 10-fold cross validation to determine the quality of model predictions and to select the penalization parameter λ. We set $\lambda = 10^{-5}$ to obtain a high true positive rate without compromising the positive predictive value (Appendix 2 Fig. A2.3).

Classifying response element specificity. The specificity of each variant was determined from its functional classification on ERE and SRE. ERE-specific variants are strong on ERE and inactive on SRE; SRE-specific variants are strong on SRE and inactive on ERE; promiscuous variants are strong on one response element and strong or weak on the other; and nonfunctional variants are not strong on either response element. The false positive rate was very low, with no stop-codon-containing variants classified as functional. AncSR1+11P:EGKA is classified as promiscuous, because it has very strong ERE activity and SRE activity that is very weak but statistically distinguishable from background, consistent with previous observations (70).

A small number of RH variants were unexpectedly inferred to be functional in AncSR1 but nonfunctional in AncSR1+11P (Appendix 2 Fig. A2.8a-c). To validate this observation, we re-cloned the three SRE-specific variants with the largest reduction in fluorescence when 11P were included (CARV, HARV, HPRM) and assessed their SRE activation in the AncSR1 and AncSR1+11P backgrounds in isogenic cultures via flow cytometry; for comparison, we also

validated a putatively 11P-independent genotype (KASM) and two 11P-dependent variants (SPKM, YGKQ), alongside GSKV for reference. Inductions were conducted in triplicate, each from an independent transformant. Classifications of the three comparison genotypes were all confirmed; however, the three genotypes that were putatively restricted by 11P showed no reduction in fluorescence in this assay, indicating that they were falsely classified as nonfunctional in the AncSR1+11P FACS-seq assay (Appendix 2 Fig. A2.8c). Notably, the predictive logistic regression correctly predicts that these three variants are strong SRE-binders in the AncSR1+11P background. These three variants manifested strong growth defects in the AncSR1+11P background, even in the ERE strain in which they do not activate GFP expression.

Robustness of results to classification method. We tested the robustness of our conclusions to alternative methods for classifying variants as functional. These include: (A) using the internal library AncSR1:EGKA mean fluorescence estimated by FACS-seq as the reference level of AncSR1 activation on ERE; (B) increasing the margin of equivalence to 50% of the activity difference between ancestral and stop-codon-containing variants; (C) classifying any active variant (weak or strong) as functional; (D) using the 80% mark of the range from stop-codon-containing to ancestral variants as a hard threshold rather than a null hypothesis for statistical testing; (E) defining functional variants as between 80% and 120% of the ancestral activity, so that extremely strong binders are classified as nonfunctional; (F) using predicted classifications for all variants, with experimental classifications used only to train predictive models; (G) using no predicted classifications, and labeling all undetermined genotypes as nonfunctional; (H) using for each variant the strongest functional class as predicted or determined by experiment; (I) using the experimental classification for a variant only if it was identical between replicates and predicting all others; (J) and using the per-variant estimates of

the standard error of mean fluorescence based on coverage depth to calculate a *P*-value that a variant is inactive or weakly active given a normal distribution, rejecting each null hypothesis at a 5% FDR as above. When appropriate, ordinal logistic regression models were re-trained to predict missing genotypes under each scheme. These alterations made no qualitative differences to our conclusions (Appendix 2 Table A2.2).

Network construction and trajectories through sequence space. Network representations of functional RH variants in the AncSR1 and AncSR1+11P backgrounds were constructed using the R package 'rgexf' and the network visualization program Gephi (179). Nodes representing RH variants were connected by edges if any genetic encoding of their protein-coding sequences could be interconverted with a single nucleotide mutation given the standard genetic code. The network was represented as a force-directed graph, which clusters nodes in two-dimensional space based on connectivity: nodes tend to repel each other, but each edge between connected nodes provides an attractive force; in the "equilibrium" layout, sets of densely interconnected nodes tend to cluster to the exclusion of less connected nodes. Force-directed graph layouts were constructed with the ForceAtlas2 method in LinLog mode, Gravity 1.0 and Scaling 0.8 (AncSR1) or 0.125 (AncSR1+11P).

We used the 'igraph' package in R to characterize the set of paths between functional nodes. A step was defined as a nonsynonymous nucleotide mutation between two functional variants; synonymous mutations within a single node were not considered as contributing to trajectory length. The graph was directed, so that trajectories can proceed from ERE to SRE specificity directly or via a promiscuous intermediate; nonfunctional intermediates[2] and functional reversions were not allowed, but "neutral" steps within a functional class were allowed. Epistasis was inferred when the shortest path between two nodes was longer than the

minimum genetic distance between genotypes (22, 24); epistasis may arise because the state at one site specifically modulates the functional effect of some state at another site or because of nonlinearity in the genotype-phenotype map (180), such as the threshold we used to classify variants as functional.

The distribution of shortest path length to SRE specificity from ERE-specific starting points in the AncSR1 and AncSR1+11P networks was compared via a Wilcoxon rank sum test with continuity correction, as observations were not normally distributed. The number of ERE-specific starting points in each network that require permissive steps and/or promiscuous intermediates on their shortest path to SRE specificity was compared via a Chi-squared test. Only one category (AncSR1 network, no path) had an expected value less than 5; the Chi-squared test remains significant when excluding this class from the comparison ($P < 10^{-7}$).

To compare genotypic states among outcomes reached from different ERE-specific starting points, we calculated the frequency distribution of amino acid states at each sequence site for the set of outcomes reached from each starting point; we then calculated the Jensen-Shannon (J-S) distance between these distributions for pairs of starting points. To capture a true amino acid state distribution across outcomes, we only considered ERE-specific starting points that access at least 15 outcomes (the median across all ERE-specific starting points). We compared these observed J-S distances to a null expectation of J-S distances in the absence of structure in sequence space, in which we randomly sampled two sets of variants from all possible SRE-specific outcomes according to the same sample sizes used in each real comparison, and calculated the J-S distance between these randomly sampled distributions.

We also considered a regime in which SRE-binding affinity is under strong selection, such that SRE-binding affinity is required to increase with each step; such a scenario has a strong

potential to make evolution deterministically favor a single outcome. In this scheme, a step from one genotype to a neighbor was allowed only if the lower bound of the 90% confidence interval of the neighbor's mean SRE fluorescence, estimated from its mean and SEM, was greater than the upper bound of the confidence interval of the starting genotype. We then calculated the probability of each accessible trajectory using two previously described models (24): in the equal fixation model, any step that enhances SRE affinity from a particular node is equally likely to occur; in the correlated fixation model, the probability that an SRE-affinity-enhancing step occurs is directly proportional to the degree to which it increases SRE mean fluorescence, relative to the other SRE enhancing steps available from the given node.

Structural modeling and predictions of RE-binding affinity. We used FoldX (181) to predict the affinity to SRE of all RH variants that were 11P-dependent (SRE-specific in the AncSR1+11P background and nonfunctional in AncSR1), or 11P-independent (SRE-specific in AncSR1) (Appendix 2 Fig. A2.8a). For structure-based affinity prediction, we used the crystal structures of AncSR1/ERE (PDB 4OLN) and AncSR2/SRE (PDB 4OOR) as starting points, with crystallographic waters and non-zinc ions removed. We removed chains E, F, K and L from the 4OOR structure. We used the RepairPDB function to optimize both structures according to the FoldX force field, and we used the BuildModel function to mutate the AncSR1/ERE structure to AncSR1:GSKV/SRE. The BuildModel function was then used to model each SRE-specific RH variant in complex with SRE on each of the AncSR1 and AncSR2 structures, and the AnalyzeComplex function was used to predict the total DNA-binding energy of each protein variant with SRE. The predicted binding energies of 11P-dependent and 11P-independent variants were compared using a nonparametric Wilcoxon rank sum test with continuity correction, as data were not normally distributed. This test was conducted independently for

energies predicted using the AncSR1 and AncSR2 structures. To compare these same groups as directly estimated in FACS-seq, a Wilcoxon rank sum test with continuity correction was used, as data were not normally distributed.

To characterize the diversity in biochemical mechanisms of SRE specificity, we analyzed FoldX models of the 10 most active SRE-specific variants that were identified in our AncSR1+11P FACS-seq experiment. We modeled binding to SRE using the AncSR2/SRE structure as described above and binding to ERE using the crystal structure of AncSR2:EGKA/ERE (4OND), with water and non-zinc ions removed and optimized using the RepairPDB function. To illustrate protein-DNA contacts made in each structural model, we used NUCPLOT (182) to identify all hydrogen bonds with distance ≤3.35Å between non-hydrogen atoms and non-bonded packing contacts ≤3.90Å. Summary figures display the union of contacts made by a residue in either of the half sites of the response element palindrome; we only illustrate residues whose contacts vary among the analyzed structures.

To ensure structural inferences converge, we built each SRE- and ERE-bound FoldX model a second time. We observed convergence in all polar contacts (and absence thereof in ERE structures) illustrated in Fig. 4.1 and Appendix 2 Fig. A2.4. Only several non-bonded contacts were not replicated: I29/T–4 in KAAI/SRE; Q29/A4 and Q29/T–4 in YGKQ/SRE; M29/T–3 in KSAM/SRE; and K25/G2 and K25/T–3 in KASM/SRE. To determine whether electrostatic clashes in ERE-bound structures could be satisfied by bridged water molecules (183), models were built again using the BuildModel function with predicted waters. In some cases (GGRT, YGKQ, DSKM, CGRV), but not all (GSKV, KAAI, PAKE, KSAM, DPKQ, SAKE, KASM), polar groups on ERE that were not satisfied by direct interaction with protein side chains are predicted to be satisfied by water bridges between protein and DNA.

Biochemical determinants of RE-binding specificity. Logos illustrating the frequency with which each amino acid state is found at each position among variants of a functional class were constructed using WebLogo (184). Since our sequence space is combinatorially complete (all 160,000 genotypes are classified, either by FACS-seq or via prediction), the logo plots do not need to be normalized by background input frequencies. To evaluate similarity of the frequency profiles between classes of variants, the frequency of each amino acid state in a class was centered logratio-transformed, the appropriate transformation before computing correlations among compositional data; a pseudocount of one was added to the number of observations of each amino acid to allow log-transformation of states observed zero times. The Spearman rank correlation coefficient was computed for the correlation between functional classes.

To identify the biochemical properties of amino acids that contribute to DNA specificity, we developed a multiple logistic regression model that describes the probability that an RH variant specifically binds a response element as a function of the biochemical properties of the amino acid states at each of its four variable RH positions. The model includes four properties (hydrophobicity, volume, isoelectric point, and α-helix propensity), with the values for each amino acid's properties from ref. (185), which we then centered and standardized; the effect of each property at each site on the probability of being a specific binder is reflected in a model coefficient, which represent the model's free parameters. We used R to find the values of these coefficients that best fit the observed classifications for each DBD/RE combination. Differences in the contribution of a property to specificity were identified if its associated coefficients in two models differed by a *Z*-test ($P<0.05$ with no correction for multiple testing).

Data and code availability. Raw sequencing data were deposited to the NCBI SRA under BioProject number PRJNA362734. Processed data and scripts to reproduce analyses are available at github.com/JoeThorntonLab/nature-2017_RH-scanning.

Chapter 5

Epistasis and evolvability in a protein sequence-function landscape

5.1 Summary

Though epistasis between amino acid mutations is prevalent and influential in case studies of protein functional evolution, we lack an understanding of how epistasis' contribution to the structure of protein sequence-function landscapes impacts the evolution of new protein functions on a global scale. Here, we directly probe the pattern of epistasis in a combinatorial sequence-function landscape and explore its impact on the evolvability of protein functions. We computationally decompose a dual-function sequence-function landscape into its genetic determinants, both main-effect and epistatic. We explore the molecular basis for epistatic interactions, and explore how the genetic determinants differ between the two distinct protein functions. We find that epistatic interactions are key in enabling single mutations to cause a shift in protein function, allowing new functions to be quickly gained along mutational trajectories. This indicates that epistasis has a constructive role in promoting the evolvability of new functions from the global perspective of the sequence-function landscape.

5.2 Introduction

Epistasis – the non-additivity of mutational effects – is an important factor in the molecular evolution of proteins (7, 135). Whether epistasis contributes positively or negatively to the functional evolution of proteins remains an open question. Though it is often a matter of perspective, epistasis is often conceived as 'constraining' evolutionary trajectories (34, 186, 187), and epistasis often hinders our ability to engineer new proteins with desirable properties

(40, 62, 63, 188). At the same time, a constructive role for epistasis has also been recognized in molecular evolution. For example, deep mutational scans have revealed pervasive positive epistasis (21, 26, 189), which could enable evolutionary trajectories to distant regions of sequence space with new properties, and theoretical considerations and simulation have suggested that epistatic interactions might underlie the fundamental ability of protein sequences to spontaneously fold into their native structures (190).

The distribution of functions in sequence space has long been known to be influenced by epistasis: early theoretical models on protein sequence-function landscapes considered epistasis as a central determinant of the landscape structure (e.g. (191)). In related work, the presence of extended functional networks – mutationally connected networks of sequences that share some common function – has been theoretically (192-194) and experimentally (195-197) shown to enhance the evolvability of new functions, by enabling proteins to drift to regions of the functional network that are mutationally adjacent to novel molecular functions. However, the precise role of epistasis in determining the mutational adjacency of distinct functions in sequence space has not been thoroughly explored.

New methods for the analysis and description of epistasis (118, 136, 198, 199) coupled with new high-throughput methods for characterizing large mutant libraries (11) enable us to interrogate the role of epistasis in the structure of protein sequence-function landscapes and how this structure impacts protein functional evolution. Epistasis has traditionally been conceived from a 'reference-based' perspective: the non-additivity of mutational effects is often judged based on their individual and joint effects as measured when introduced into some specific reference background. More recently, a global approach to describing and characterizing epistasis has emerged (198). In this global approach, epistasis is not defined as the non-additivity

of mutations when added to some particular reference background, but instead, it is defined as the non-additivity of the effects of mutations when averaged across a global ensemble of backgrounds. This background-averaged epistasis can be mathematically related to reference-based epistasis (198), but it represents a more parsimonious and informative way to decompose a protein sequence-function landscape into its underlying genetic determinants, particularly when considered from the global perspective of sequence space (25, 118, 200).

Here, we employ a global epistasis analysis to describe epistasis in an empirical sequence-function landscape and consider how this epistasis contributes to the evolvability of protein functions in sequence space. We focused on a previously collected combinatorial mutational scanning dataset of four key residues at the protein-DNA interface of ancestral steroid receptor transcription factors (201). Steroid receptors are a clade of paralogous metazoan transcription factors, responsible for translating steroid hormone signaling molecules into changes in cellular physiology via their sequence-specific DNA-binding function. The ancestral steroid receptor bound specifically to a DNA motif called the estrogen response element (ERE); after a gene duplication, one lineage of steroid receptors lost this ERE-specificity, and evolved specificity for a DNA motif called the steroid response element (SRE) (70). The genetic and biophysical mechanism of this historical transition in DNA-binding specificity was previously described (70, 106). More recently, we interrogated the key specificity-determining residues with a four-site combinatorial mutational scanning approach (201), in which we characterized the ability of 160,000 protein variants to bind to the ERE, SRE, or both. Here, we computationally decompose this sequence-function landscape into its underlying genetic determinants, to reveal how epistasis contributes to the juxtaposition of these two functions in sequence space. We find

that epistasis plays a constructive role, enhancing the ease with this evolutionary transition can occur on a global scale.

5.3 Results & Discussion

5.3.1 A global model of the sequence-function landscape

To identify the genetic determinants of ERE- and SRE-binding, we built regression models that decompose the ERE- and SRE-binding sequence-function landscapes into the first-order main effects of individual states and second-order pairwise epistatic effects of pairs of states across the four variable sites. To perform the regression given complex and unknown nonlinearities between the observed high-throughput FACS-seq measurements and the scale to which genetic terms combine additively (136, 199), we used a proportional-odds ordinal logistic regression model. This generalized linear model expresses a discretized response variable (in our case, discretized measurements of ERE- or SRE-activation) as a function of a series of predictor variables (in our case, categorical variables representing genetic states or pairs of states). The coefficients corresponding to the predictors present in each genotype sum together to produce a continuous latent phenotype, which maps to one of n ordered classes based on the location of this phenotype relative to the n-1 inferred cut points. The proportional-odds assumption posits that if we collapsed our data into successive binary classifications and inferred a series of n-1 logistic regression models, the regression coefficients for the predictor variables for each of these binary logistic regressions would be consistent.

As described previously (201), we discretized our FACS-seq mean fluorescence measurements for ERE- and SRE-binding into null, weak, and strong binding categories. We verified that the proportional-odds assumption holds by plotting the log-odds of each main-effect

genotypic state across the null/weak and weak/strong boundaries (Fig. 5.1). We restricted the number of amino acid states that we modeled at each position to those that are found at greater than 0.01 frequency among weak and strong binders for either or both DNA motifs, reducing the number of states to 20 at position 1, 5 at position 2, 15 at position 3, and 16 at position 4 (24,000 genotypes). Separately for the ERE- and SRE-binding data, we then fit a proportional-odds regression model, with the first-order main-effect and second-order epistatic terms as predictor variables:

$$\text{class} \sim A_1 + A_2 + A_3 + A_4 + A_{12} + A_{13} + A_{14} + A_{23} + A_{24} + A_{34}$$

We used a no-intercept, one-hot encoding scheme for the categorical predictor variables, which makes this a background-averaged representation of the genetic determinants in this sequence-function landscape. We inferred the model using L_1 (LASSO) penalization, which shrinks coefficients to exactly zero to encourage sparsity in the fit and prevent overfitting. We judged model performance and selected the LASSO penalization parameter λ via 10-fold cross-validation (Fig. 5.2a-d). Though this proportional-odds implementation was too computationally demanding to fit a model that also incorporates third-order epistatic effects, a different but related ordinal logistic regression strategy (continuation ratio ordinal logistic regression) was also fit with third-order effects, which performed considerably worse in cross-validation than the second-order model and exhibited evidence of over-fitting, presumably because of the large number of third-order terms (Fig. 5.2e-f).

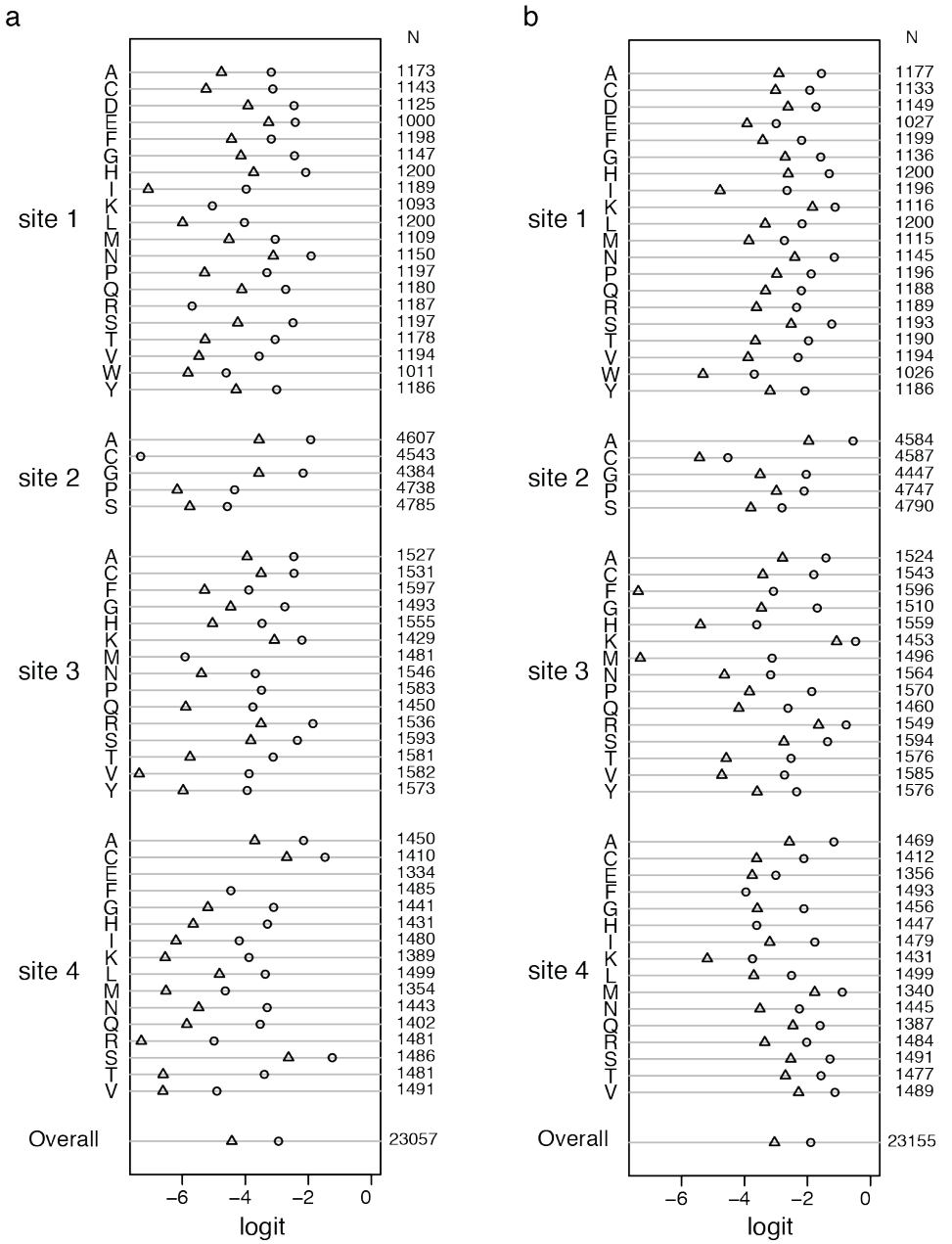


Figure 5.1. A graphical method for evaluating the proportional-odds assumption. For ERE-binding (**a**) and SRE-binding (**b**) models, the log-odds of each amino state across the null:weak+strong (circle) and null+weak:strong (triangle) binary classifications. The column marked with "N" gives the number of observations of each amino acid state in the experimental dataset. If the proportional-odds assumption is perfectly met, the difference in log-odds between the two binary classifications will be the same for each amino acid state (the distance between the triangle and circle symbols will be the same in each row). This pattern holds decently in the data; the amino acid states whose log-odds differ the most dramatically between these two classifications tend to be those with lower log-odds, meaning their log-odds are estimated from a lower observation count of that genotype among strong or weak+strong binders, which will produce higher-variance estimates of the log-odds.

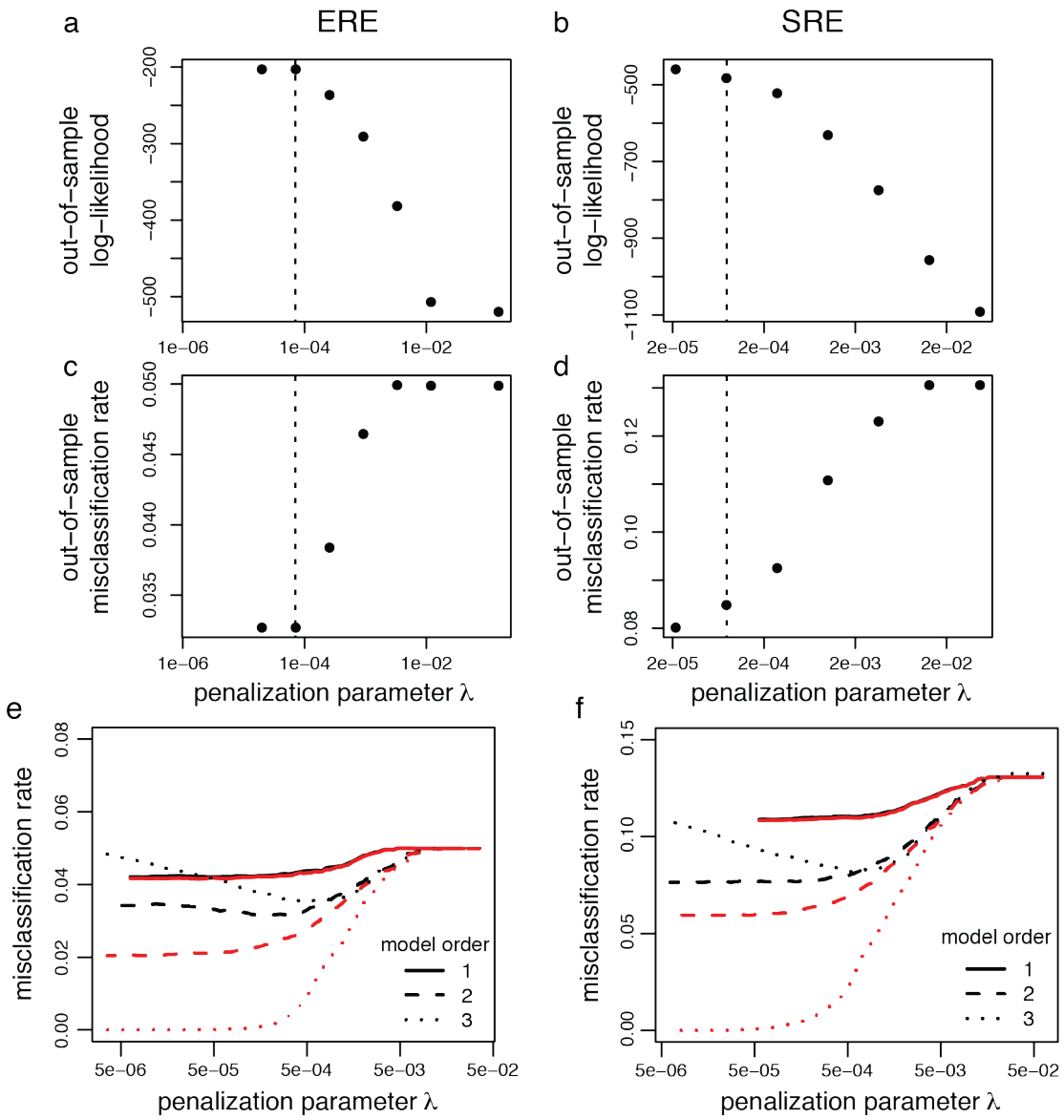


Figure 5.2. Ordinal logistic regression model performance. a-d, Cross-validation of proportional-odds regression models incorporating main-effect and pairwise epistatic terms. For ERE (**a**,**c**) and SRE (**b**,**d**) models, out-of-sample log-likelihood (**a**,**b**) and misclassification rates (**c**,**d**) were determined across a range of lambda penalization values via 10-fold cross-validation. The chosen lambda values are indicated by the vertical dashed line. **e**,**f**, Continuation ratio ordinal logistic regression models indicate that third-order epistasis terms result in over-fitting. For ERE (**e**) and SRE (**f**) models across a range of lambda values, out-of-sample misclassification rates were determined via 100-fold cross validation (black lines) for nested models incorporating first-order main effects only (model order 1), pairwise epistasis terms (model order 2), or third-order epistasis terms (model order 3). The third-order model performs worse in cross-validation than the second-order model. Red lines show the in-sample (self-trained and self-tested) misclassification rates; the large difference between the black and red lines for the third-order model illustrates over-fitting.

Under the hood, the proportional-odds model builds a linear model relating the genetic determinants to some latent phenotype, which by definition is whatever phenotype the genetic states contribute to linearly and additively. Though the proportional-odds model is built purely on the discretized null/weak/strong classification (without having ever been trained on the underlying continuous mean fluorescence measurements), we see that the latent phenotype maps closely to the experimental mean fluorescence (Fig. 5.3a,b): it reproduces the known censoring that we had previously observed in the relationship between mean fluorescence and $\log(K_{a,mac})$ at low affinities (see Appendix 2 Fig. A2.1a), and it also reveals a previously-unknown saturation of our mean fluorescence metric at high values (at least for SRE-binding). While we cannot know what this latent phenotype truly represents, we do see that it maps well to $\log(K_{a,mac})$ (Fig. 5.3a,b), and $\Delta G_{binding}$ is a physical quantity that genotypic terms should contribute to additively and linearly. While we cannot say that our terms are exactly analogous to $\Delta\Delta G_{binding}$ terms, we will refer to our model parameters as "effects on binding" due to this relationship.

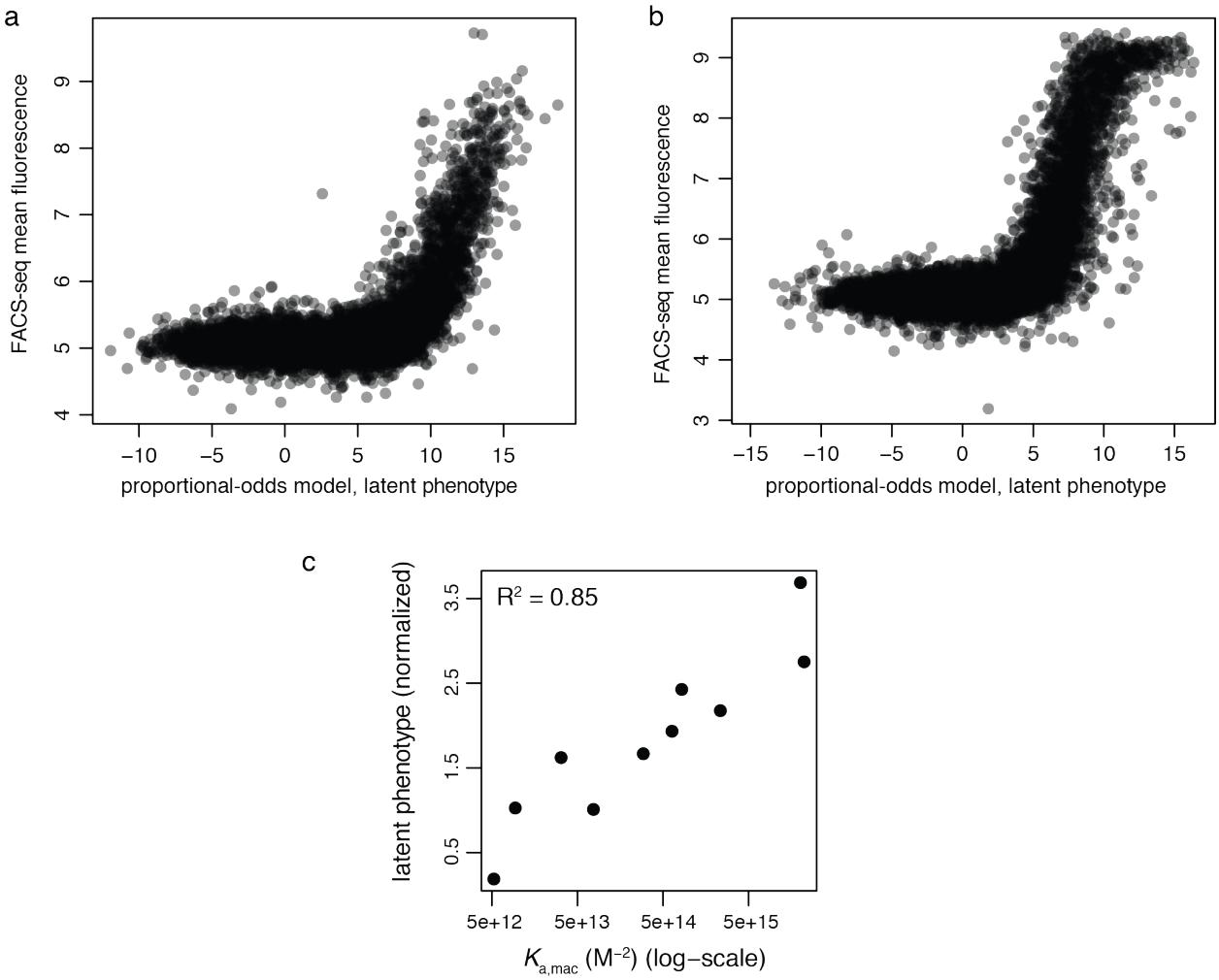


Figure 5.3. The latent phenotype of the proportional-odds model relates to binding affinity. **a**,**b**, For ERE (**a**) and SRE (**b**) models, the relationship between the proportional-odds model latent phenotype and the experimental FACS-seq mean fluorescence estimate for each genotype. **c**, For a handful of genotypes whose affinity for ERE or SRE was previously measured *in vitro* (106), the relationship between the proportional-odds model latent phenotype and log($K_{a,mac}$). R^2, Pearson correlation between latent phenotype and log($K_{a,mac}$) To illustrate the ERE- and SRE-binding latent phenotypes on an equivalent scale, the latent phenotypes for each model were normalized (divided by the standard deviation of the latent phenotypes for all genotypes on ERE or SRE).

5.3.2 The genetic determinants of ERE- and SRE-binding

The above approach estimates each state's background-averaged main effect on binding, and each pair of states' background-averaged epistatic contribution. For each genotype, its latent phenotype is calculated by summing the ten coefficients (four main-effect, six epistatic) that

describe its resident amino acid states at the variable positions. A positive term indicates that a state or pair of states, on average, improves binding to the respective DNA motif, while a negative term indicates that the state or pair of states, on average, worsens binding to the respective DNA motif.

The heat maps in Fig. 5.4 illustrate the model coefficients for each of the main-effect (margins of the matrix) and pairwise epistatic (sub-matrices) terms, for ERE-binding (upper-right triangle) and SRE-binding (lower-left triangle). To guide the interpretation of these coefficients, sequence logos show the fraction of strong ERE-binders (right) or SRE-binders (left) in the library that contain each amino acid state at each of the variable positions.

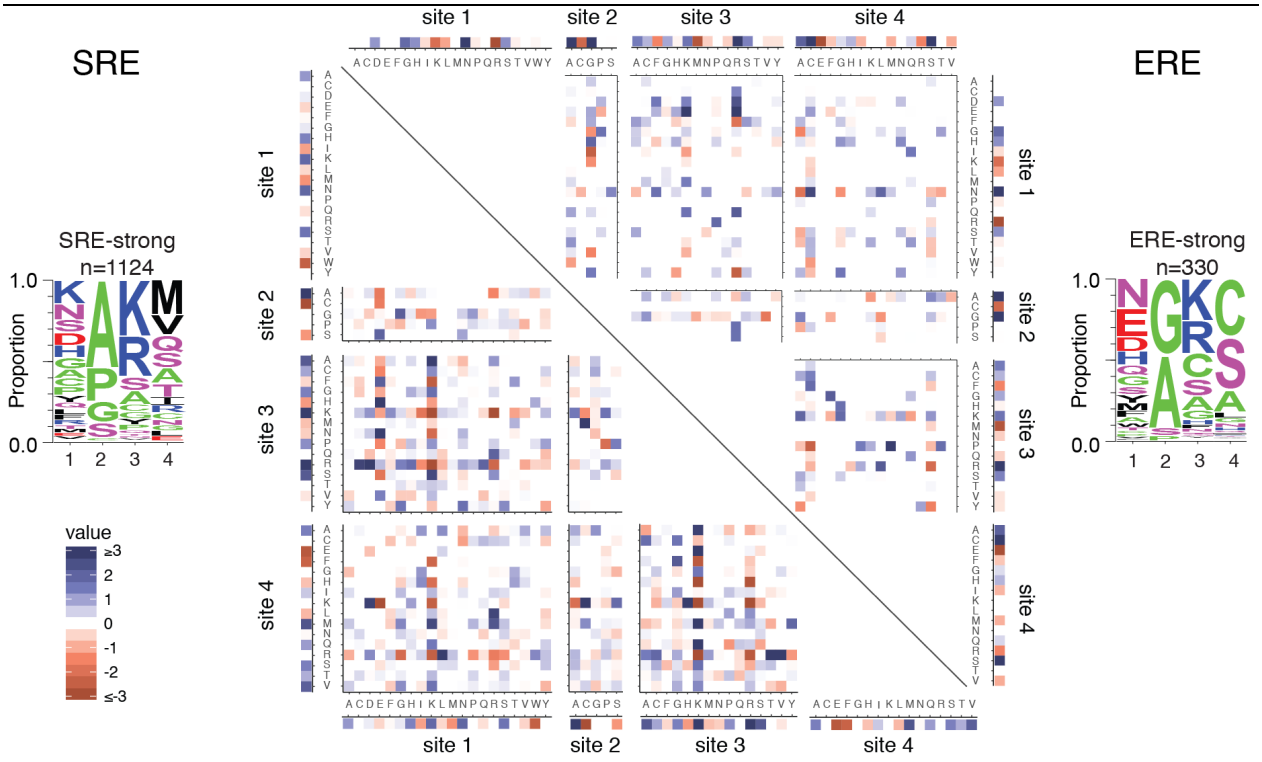


Figure 5.4. The genetic determinants of ERE- and SRE-binding. For the ERE (upper-right) and SRE (lower-left) models, a heat map representation of the model coefficients inferred from the second-order proportional-odds model. Main-effect terms at each of the four variable sites are illustrated on the matrix margins; pairwise epistasis terms between each pair of sites are illustrated in each of the sub-matrices. Heatmap scale is given in the lower-left corner. For interpretation, logo plots show the proportion of genotypes in the library that are strong binders to ERE (right) or SRE (left) that contain each amino acid state at each variable position.

Asparagine (N) at position 1 is the most prevalent state among strong binders to ERE, and the regression model ascribes to N1 a strong, positive main-effect coefficient, with no strongly negative interaction terms, indicating that this state's prevalence comes from a more-or-less universally positive contribution to ERE-binding. In contrast, glutamate (E) is the second most prevalent state at position 1 among strong ERE-binders, yet its main effect coefficient is shrunk to 0. This indicates that the relatively large number of strong ERE-binders that use E1 must have a specific amino acid state (or states) at another position (or positions), reflecting epistatic coupling of E1 with other residues. Indeed, two of the strongest positive epistatic interactions of any residues are E1 with lysine (K) or arginine (R) at position 3. This strong context-dependency for the beneficial effect of E1 has a straightforward biophysical basis: E1 makes a key polar contact to the C–3 DNA base in the crystal structure of the wildtype sequence (EGKA) bound to ERE (Fig. 5.5a); E1 also makes a salt bridge interaction with K3, which makes additional polar contacts to DNA. Presumably, without the presence of a positively charged residue at position 3, E1's negative charge is not entirely satisfied at the protein-DNA interface, which would dramatically weaken ERE-binding affinity.

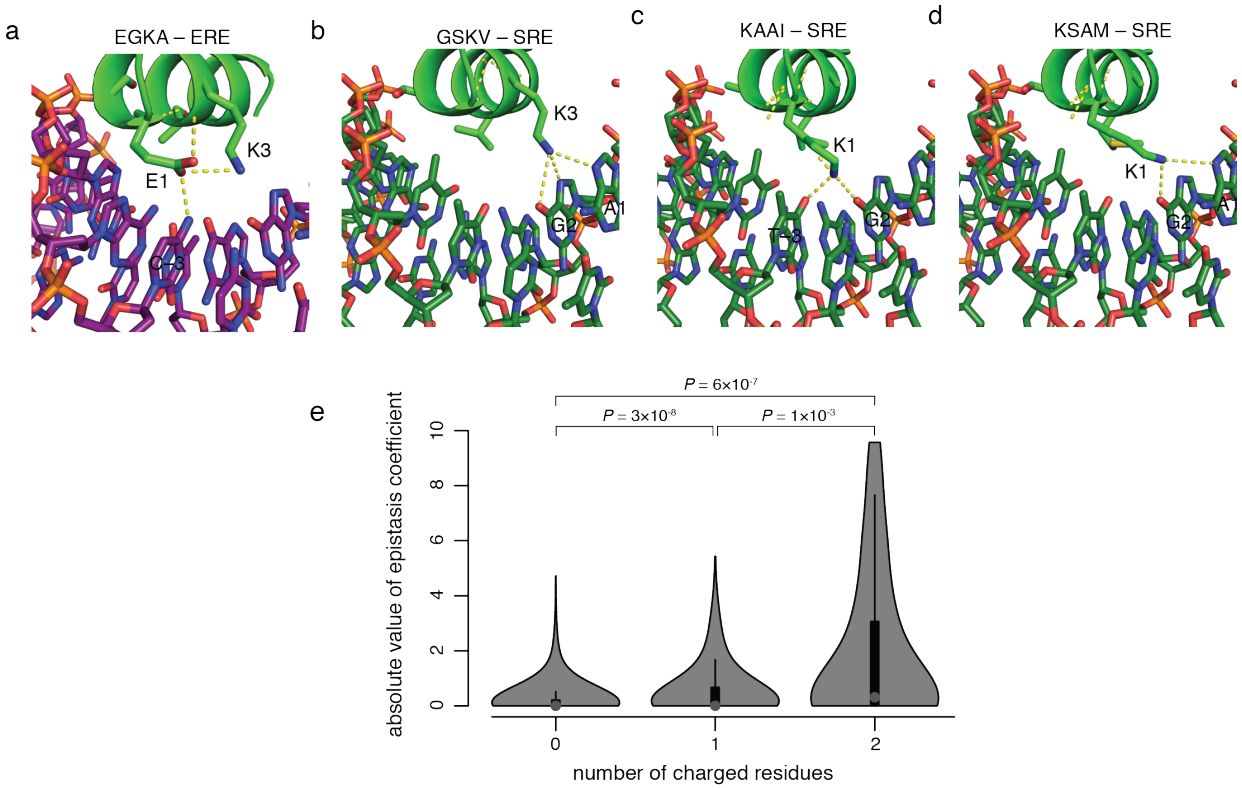


Figure 5.5. Biophysical basis for large epistatic interactions. a-d, Crystal structures (**a**) and molecular models (**b-d**) illustrating biophysical basis for large effect epistatic interactions. RH sequence and DNA motif are labeled above each structure. Yellow dashed lines, polar contacts. Important protein residues and DNA bases are labeled. Crystal structure of EGKA-ERE from PDB ID 4OLN (70); molecular models built using FoldX, as described in (201). **e**, Epistatic interactions involving one or two charged residues are enriched for larger magnitudes. Violin plots for the magnitude of epistatic coefficients, versus the number of charged residues (D, E, K, or R) involved in the interaction. The gray violin shows the overall density of the distribution; dark gray dot, median; thick black bar, interquartile range. *P*-values, Wilcoxon rank-sum test with continuity correction.

Some of the strongest negative epistatic interactions in this heat map also have intuitive biophysical underpinnings. Lysine (K), the most prevalent state at position 1 among strong binders to SRE, exhibits two of the strongest negative epistatic coefficients together with lysine (K) or arginine (R) at position 3. In the structure of the wildtype sequence (GSKV) bound to SRE, K3 makes polar contacts to the A1 and G2 DNA bases (Fig. 5.5b). In structural models of alternative SRE-specific genotypes (KAAI and KSAM) bound to SRE, K1 makes polar contacts

to some of the same DNA moieties as K3 (Fig. 5.5c,d). Presumably, if both K1 and K3 were present in the same genotype, one of the two residues would be unable to make its preferred polar DNA contacts due to steric or electrostatic clashes, leaving an unsatisfied charged residue at the protein-DNA interface, severely compromising SRE-binding affinity. Like the examples described above, many of the most extreme epistasis coefficients involve charged residues (D, E, K, and R; Fig. 5.5e), indicating that charged amino acids in particular exhibit strong context-dependency in their effects on DNA-binding.

These observations emphasize the importance of the background-averaged encoding of genetic determinants, compared to the reference-based approach (198): from the perspective of the wildtype ERE-binding genotype (EGKA), E1 has a positive main effect contribution to binding, because single mutations to E1 decrease binding (70, 106, 163); however, this is only because this reference background also has K3, which has a strong, positive epistatic interaction with E1. Considered globally across the broader array of backgrounds, E1 is not considered a universally beneficial state for ERE-binding, but rather is only beneficial when introduced in conjunction with a positively charged residue at site 3. Similarly, from the perspective of the wildtype SRE-binding genotype (GSKV), K1 is a deleterious state, because this reference background also has K3, with which K1 interacts negatively. It is only by considering the background-averaged effects of each of these states that we can observe that K1 is, on average, a beneficial state for SRE-binding, and that E1 is not a universally beneficial state for ERE-binding.

5.3.3 Partitioning the determinants of ERE- and SRE-binding

Next, we sought to determine how important different sets of model terms are for explaining variation among genotypes in ERE- and SRE-binding. We predicted the latent phenotype of all 24,000 genotypes from models truncated to include only a subset of model terms, and compared this phenotype to the 'true' latent phenotype of each genotype predicted from the full-order regression model. For example, to determine the proportion of variance in the DNA-binding latent phenotype that is explained by first-order main effect terms, we set all pairwise epistasis terms to zero, predicted the latent phenotype from the main effect terms only, and determined R^2 between the full-order latent phenotype and this truncated main-effect-only phenotype. This reveals that, across the 24,000 modeled genotypes, main effect terms explain 91.6% of the variance in ERE-binding and 75.0% of the variance in SRE-binding (Fig. 5.6a). The same approach reveals that pairwise epistasis terms account for 16.7% of the variance in ERE-binding and 22.5% of the variance in SRE-binding (Fig. 5.6a; these values need not sum to 1 if there is correlation among the predictor values). Therefore, main effect terms explain a substantially larger fraction of variation in ERE- and SRE-binding than epistasis terms across the entire landscape, though epistasis makes a sizeable contribution to variation in the latent phenotypes.

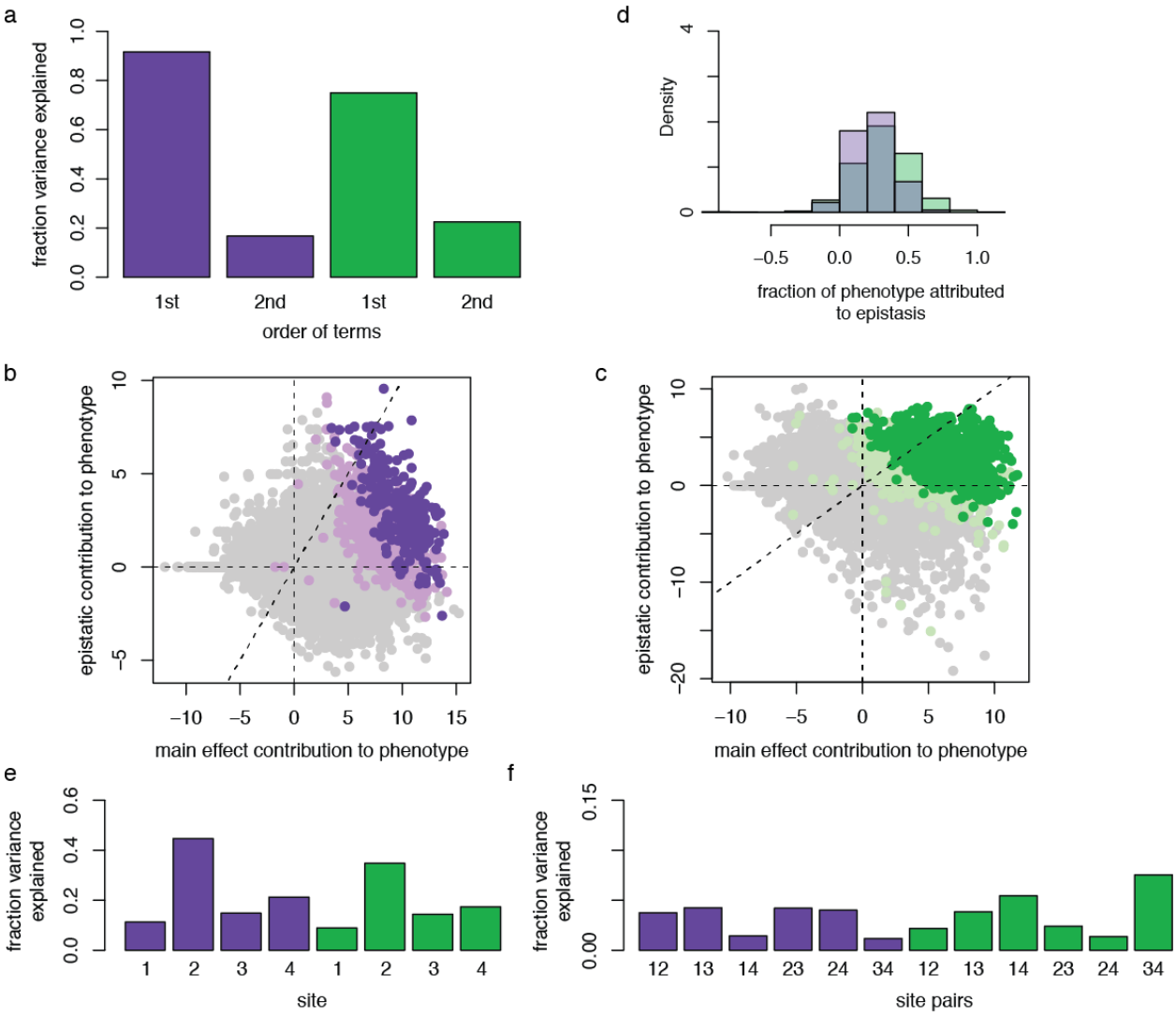


Figure 5.6. Partitioning the genetic determinants of ERE- and SRE-binding. a, The coefficient of determination for truncated models including only 1st order main-effect terms or only 2nd order epistatic terms, for ERE-binding (purple) or SRE-binding (green). **b**,**c**, For each genotype, the sum of its component main-effect (*x*-axis) and epistatic (*y*-axis) terms, for ERE-binding (**b**) or SRE-binding (**c**). For each scatterplot, dark purple or green dots are experimentally determined "strong" binders, light purple or green dots are experimentally determined "weak" binders, and gray dots were inactive in the experimental dataset. Diagonal dashed line is the 1:1 line; the density of dark purple and green dots to the right of this diagonal indicates that main effect terms provide a greater contribution, on average, to the phenotype of strong binders. **d**, For "strong" ERE-binders (purple) and SRE-binders (green), the fraction of their phenotype that comes from epistatis coefficients. **e**, The coefficient of determination for truncated models including only the main effect terms at the indicated sites, for ERE-binding (purple) or SRE-binding (green). **f**, The coefficient of determination for truncated models including only the epistatic terms at the indicated site-pairs, for ERE-binding (purple) or SRE-binding (green).

Similarly, for both ERE- (Fig. 5.6b) and SRE-binding (Fig. 5.6c), we observe that genotypes experimentally classified as strong binders typically depend more heavily on positive contributions from the first-order effect of their genetic states than they do on the marginal epistatic effects of their particular combinations of states. Nonetheless, epistasis still makes a sizeable contribution to the strong phenotype of many binders, with an average contribution of 23.7% and 31.1% to strong ERE- and SRE-binders, respectively (Fig. 5.6d). Many genotypes are strong binders for ERE or SRE despite a near-zero or negative epistatic contribution; on the other extreme, several genotypes (LGVR, LPTR, LPVR) are strong SRE-binders despite a negative contribution of their main-effect contributions: L1, V3, T3, and R4 have zero or weakly negative main effects, but R4 has very strong positive interactions with L1, V3, and T3 (Fig. 5.4), indicating that the contribution of these states to SRE-binding is highly context-dependent.

Next, we asked how main-effect and epistatic terms at each site or pair of sites explain the variance in strength of ERE- and SRE-binding. Through the same approach as described above, we predicted phenotypes including terms only from a site or pair of sites, and computed the coefficient of determination between the full-order and truncated phenotypes. The pattern of explanatory power of main-effect terms at each of the four sites is similar for ERE- and SRE-binding (Fig. 5.6e): site 2, which has the most stringent state-preferences (see logos in Fig. 5.4), explains the most variation in both ERE- and SRE-binding phenotypes, followed by sites 4, 3, and 1. This suggests that the overarching importance of each site for DNA-binding is conserved between these two functions. However, ERE- and SRE-binding differ in the architecture of their epistatic determinants (Fig. 5.6f): for example, the three site-pairs involving site 2 are among the most explanatory epistatic site-pairs for ERE-binding, but the three least explanatory site-pairs for SRE-binding. Conversely, epistasis with site 4 is important for accounting for variation in

SRE-binding, but not so for ERE-binding. Epistasis between sites 1 and 3 is moderately important for explaining both ERE- and SRE-binding phenotypes – the only site-pair whose explanatory power is consistent between the two DNA motifs. Consistent with this analysis, we find that the magnitudes of main effect coefficients are more closely correlated between the ERE and SRE models than are the magnitudes of epistasis coefficients (Fig. 5.7). Overall, though ERE- and SRE-binding are both strongly determined by the main effects of residues at positions 2 and 4, this site-4-dependency for ERE-binding is relatively context independent but site 2's influence is context dependent, whereas for SRE-binding the inverse is true: the explanatory power of site 4 is much improved by considering epistatic context, while site 2's explanatory power is relatively independent of the other residues in the genotype.

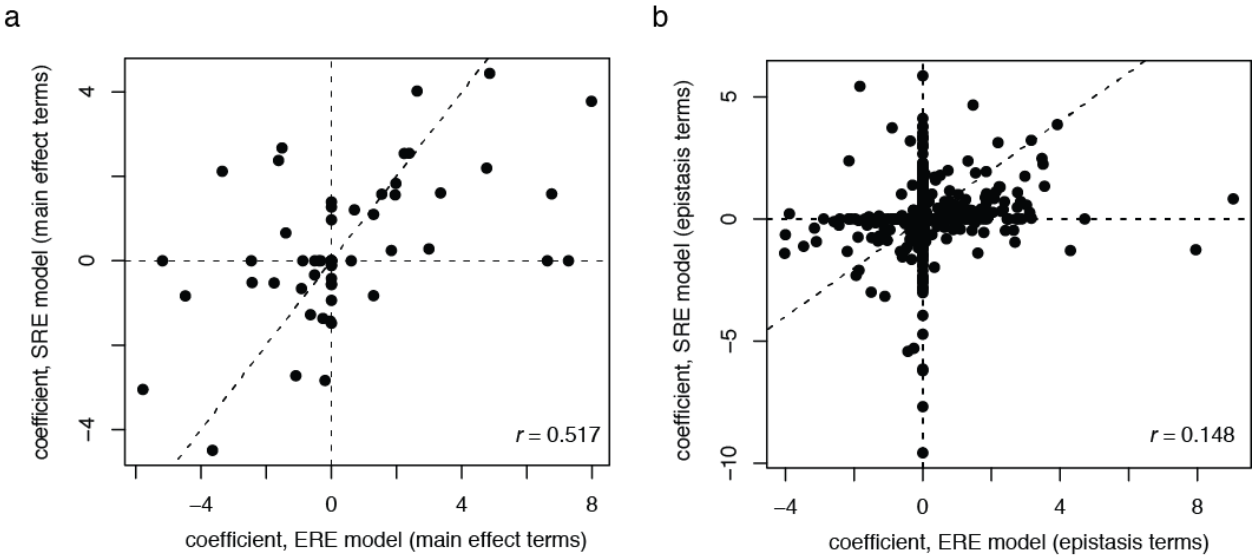


Figure 5.7. Epistasis terms differ more than main effect terms between ERE- and SRE-binding models. For main-effect (**a**) or epistatic (**b**) terms, the magnitude of each coefficient in the SRE model versus its magnitude in the ERE model. *r*, Pearson correlation coefficient. Diagonal dashed line, 1:1 line.

5.3.4 Epistatic and main-effect terms synergize to cause single-step transitions in specificity

Last, we sought to understand how epistasis shapes the fundamental evolvability between ERE- and SRE-specificity in sequence space. We previously observed the presence of many single-step transitions between these two DNA-binding specificities, in which just a single mutation is sufficient to simultaneously decrease ERE-binding from 'strong' to 'null,' and increase SRE-binding from 'null' to 'strong.' Though two-step transitions involving a promiscuous intermediate can also be important or even necessary for protein functional evolution (23), the presence of such single-step 'discrete' transitions is a remarkable feature of this sequence-function landscape and is more straightforward to dissect.

A single mutation can modify up to four genetic terms (one main-effect term, and three epistatic terms). Fig. 5.8 illustrates the change in terms for ERE- and SRE-binding caused by the mutation underlying each of these single-step transitions. Overall, we observed 13 unique mutations between 16 ERE-specific and 22 SRE-specific genotypes that cause 23 different discrete switches from ERE- to SRE-specificity. Across all 23 transitions, many different main-effect and epistatic terms at different sites or site-pairs are leveraged to alter specificity, indicating that it is not just a handful of terms that are altered by every single-step transition.

In each of the 23 transitions, there is not a single term whose change is sufficient to simultaneously decrease ERE-binding and increase SRE-binding the minimum amount to recapitulate the switch in specificity. For example, the main effect of some mutations decreases ERE-binding sufficiently (e.g. N1K, C4R), but these same mutations do not sufficiently increase SRE-binding; on the other hand, some mutations increase SRE-binding with just the main effect (e.g. M1K, L4M), but do not decrease ERE-binding. Similarly, no individual epistasis term is

sufficient to cause the functional transition singlehandedly. Instead, each mutation causes its functional effect through the synergy between two or more of the four possible terms that are changed by the mutation. In this way, epistasis between amino acids appears to be integral to the adjacency of genotypes with distinct molecular functions within the sequence-function landscape. By opening up the degrees of freedom with which a mutation can alter protein functions, epistasis allows a combination of genetic determinants to be unleashed with just a single mutation, thereby causing a discrete switch in protein function.

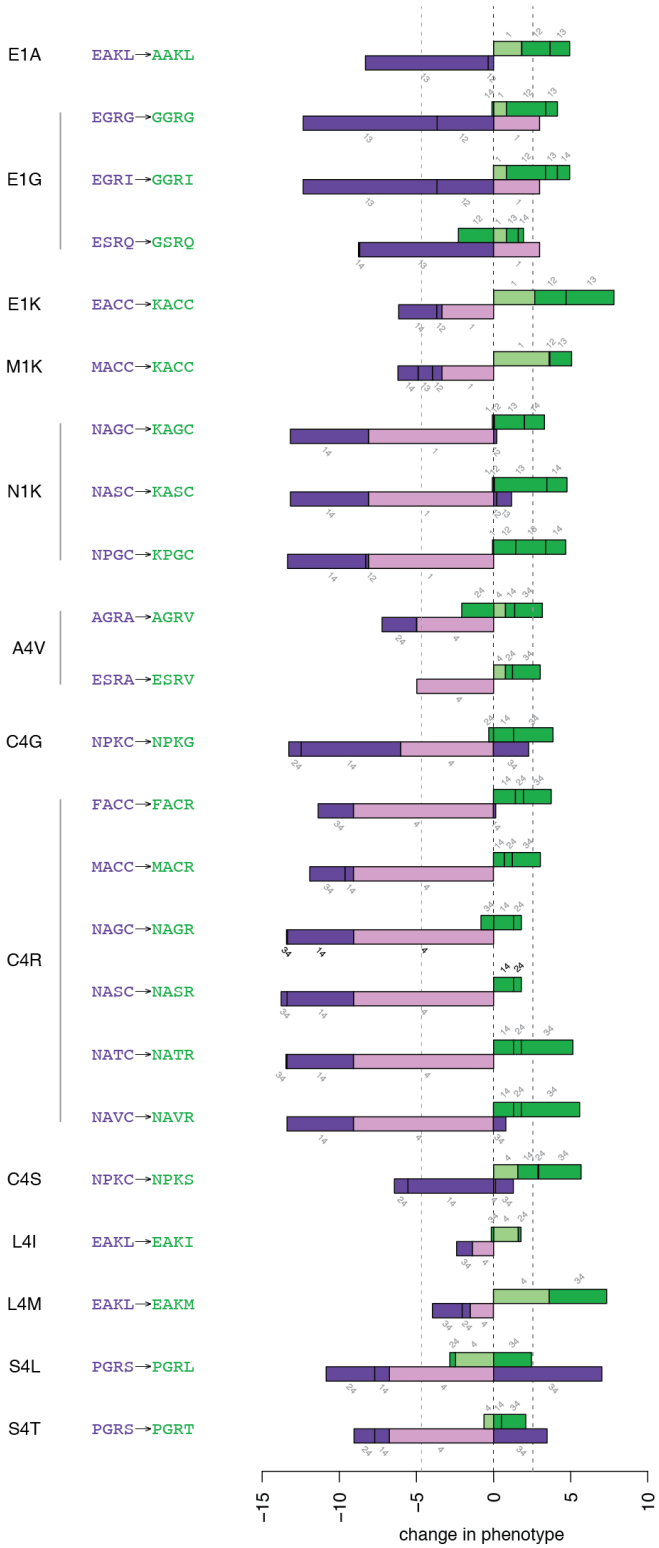


Figure 5.8. Epistasis and main effect terms synergize in single-mutant switches in specificity. For each of the 23 transitions in which a single mutation causes a discrete switch from ERE- to SRE-specificity, the change in the coefficients underlying this transition are shown. The genetic mutation is shown along the left-hand side, with the ERE-specific starting

(**Figure 5.8, continued**) sequence in purple and the SRE-specific outcome in green. The width of a bar shows the magnitude of effect caused by the change in the relevant coefficients for ERE-binding (purple) or SRE-binding (green). The change attributed to main-effect coefficients is shown in a pale shade, with the changes in epistatic coefficients shown in a dark shade. For each change, the specific site (main effect) or site-pairs (epistasis) of coefficients that changed are annotated above (SRE) or below (ERE) the bar chart. Gray dashed lines show the minimum decrease in ERE-binding required to change from 'strong' to 'null' (left) and the minimum increase in SRE-binding required to change from 'null' to 'strong' (right) from the proportional odds model; some experimentally determined transitions are not predicted by the model to be specificity switches, and so the sum of coefficients may not surpass this threshold from the model.

5.4 Conclusions and Future Directions

By analyzing the steroid receptor deep mutational scanning data via a global epistasis model, we reveal the full suite of genetic determinants that structure the sequence-function landscape of ERE- and SRE-specificity within these four specificity-determining sites. Though the main effect of genetic states accounts for a large fraction of variation in the DNA-binding phenotypes, we find that epistasis makes key contributions to the sequence-function landscape, particularly among those genotypes that are strong binders to either DNA motif. We find that the architecture of epistatic constraints differs more dramatically for the two functions than main effects, suggesting epistasis might underlie the origins of DNA-binding specificity within this system. Finally, we identify a necessary role for epistatic interactions in the adjacency of distinct molecular functions in sequence space, suggesting that epistasis might underlie the evolvability of novel protein functions on a global scale.

The idea that expansive functional networks shaped by epistasis enhance evolvability of novel functions has received much theoretical and experimental support (192-197). With just the two functions considered here, we see evidence that epistasis contributes to the evolvability of novel functions; if a broader diversity of DNA-binding specificities were considered, we believe

the role of epistasis would become even more pronounced. Other sequence-function landscapes, such as computational models of gene regulatory networks (202) and experimental maps of transcription factor binding sites (197, 203), have been elaborated with numerous functional annotations. Though not protein-based, the principle we suggest here could be tested in these systems, to see if the consideration of many additional function amplifies epistasis' constructive influence on evolvability.

An important future direction will be to perform structural analyses to understand the biophysical basis for our observations. First, what are the biophysical characteristics of the ERE and SRE DNA motifs that cause a difference in the architecture of epistatic determinants between these two functions? One possibility is that these two DNA motifs present crucial chemical moieties at different positions in the DNA major groove, altering the architecture of interactions among recognition helix residues necessary to satisfy the binding requirements of the motifs. Indeed, previous work suggests that the ancestral ERE-binder and derived SRE-binder exhibit stringent requirements for different base positions in the DNA sequence motif (106); structural analysis of a broader diversity of ERE- and SRE-binders should reveal if this pattern holds true among the broader diversity of genotypes that bind each element.

Second, structural analysis can be used to reveal why particular amino acid states are especially sensitive to context. At the most extreme, how are genotypes like LPVR strong SRE-binders despite a negative contribution of the states' main effects? Most strong SRE-binders uncovered in our scan use a positively charged residue at positions 1 or 3; does moving this positively charged residue to position 4 satisfy the same polar contacts, and if so, why is it so dependent on hydrophobic residues at positions 1 and 3?

Third, we suggest that epistasis underlies the ability of single mutations to cause discrete switches in DNA-binding specificity: why do these single mutations have amplified effects in some backgrounds but not others? What types of biophysical interactions potentiate these effects? Determining the structures of proteins with and without a particular mutation in different contexts – both epistatically potentiating backgrounds, and main-effects-dominated backgrounds – could reveal the biophysical features that underlie this effect. Coupled with our genetic observations, the structural analyses outlined here would connect our understanding of this global sequence-function landscape to its underlying biophysical determinants, satisfying a central goal of evolutionary biochemistry.

5.5 Methods

Preparing the data. We used the experimental classifications of null, weak, and strong as described in Starr et al. (201). To reduce model size and complexity, we only considered amino acid states that are found at a frequency of >0.01 among weak+strong binders for either DNA motif, reducing the dataset from 160,000 to 24,000 genotypes. Following this previous work, we considered 'undetermined' any genotype that was observed with 15 or fewer cells in the FACS-seq data (3.9% of ERE-binding and 3.5% of SRE-binding measurements in our reduced dataset).

Evaluating the proportional-odds assumption. The log-odds of an amino acid state for each of the null:weak+strong and null+weak:strong binary classifications was calculated from the experimental dataset by determining the frequency p of an amino acid state among weak+strong or strong binders, respectively, and calculating the log-odds:

$$\log\left(\frac{p}{1-p}\right)$$

The appropriateness of the proportional-odds assumption is judged by whether the difference in log-odds for these two classifications is the same across amino acid states (indicating that each amino acid state has a consistent effect on the underlying latent phenotype across the range of this latent phenotype); we see visually that this assumption generally holds (Fig. 5.1). The coefficients whose difference in log-odds differ most tend to be those with lower log-odds, indicating that the log-odds is calculated from a smaller number of observations of that state in the strong or weak+strong binders, which will produce higher-variance estimates of the log-odds.

Fitting the proportional-odds model. We used the 'ordinalNet' (204) package in R to fit proportional-odds ordinal logistic regressions with LASSO penalization, and to perform cross-validation. Separately for ERE and SRE, we modeled the 3-class null/weak/strong observations as a function of the categorical amino acid state at each position, including main-effect first order terms, and pairwise epistasis interaction terms between amino acids at all pairs of sites. Categorical variables were encoded in a one-hot scheme with no intercept. An example of the contrasts matrix for site 2 (five amino acid states) is shown below:

			model term			
		AA1A	AA1C	AA1G	AA1P	AA1S
	A	1	0	0	0	0
	C	0	1	0	0	0
amino acid state	G	0	0	1	0	0
	P	0	0	0	1	0
	S	0	0	0	0	1

In this scheme, no state is arbitrarily selected as a reference, and coefficients reflect a global, background-averaged term. This encoding is appropriate since we have combinatorially complete sampling, and fitting the model with LASSO penalization allows this encoding to have a unique solution.

To judge model fit and select the penalization parameter λ, we performed 10-fold cross validation. For each fold, we trained a model on 90% of the data, and tested it on the remaining 10%, evaluating the log-likelihood of the test set and calculating the misclassification rate (the fraction of null/weak/strong classifications that differ between the experimental classification and the model prediction).

Chapter 6

Conclusion

> Among all the occurrences possible in the universe the *a priori* probability of any particular one of them verges upon zero. Yet the universe exists; particular events must take place in it, the probability of which (before the event) was infinitesimal ... Destiny is written concurrently with the event, not prior to it ... The universe was not pregnant with life nor the biosphere with man. Our number came up in the Monte Carlo game. Is it any wonder if, like the person who has just made a million at the casino, we feel strange and a little unreal?
>
> –Jacques Monod, *Chance and Necessity* (158)

By combining mutational scanning approaches with ancestral protein reconstruction, this work provides the first map of sequence-function landscapes over which proteins evolved during history. The structure of this landscape – determined by epistasis between amino acid mutations, and the large number of ways of encoding a functional molecule – indicates important roles for chance factors in the outcomes of molecular evolution. As the quotation from Jacques Monod above suggests, the presence of a unique reality belies the role that chance played in its formation; by revisiting historical molecules and testing the effects of mutations other than that which occurred in evolution, we can begin to understand this role.

This approach is not without its limits (many of which are discussed within each individual chapter). Knowledge of the complete set of functional constraints that determined an

ancient protein's trajectory across sequence space is all-but-inaccessible to us. As such, it is possible that other functional constraints that we did not consider might make evolution more deterministic than we propose.

For Hsp90, this issue is unlikely to negate our conclusion that epistasis between substitutions creates pervasive contingency and entrenchment. We revealed a pervasive but subtle effect of intramolecular epistasis in the effects of historical substitutions on Hsp90's essential molecular function – the undefined set of functional constraints on Hsp90 that determine yeast growth under mild laboratory conditions. If other molecular functions are affected by these substitutions that do not result in observable fitness defects under the laboratory conditions that we tested, then even more substitutions might manifest an evolutionarily relevant defect, producing an even stronger signal of intramolecular epistasis. This premise could be tested by quantifying the fitness effects of our two libraries of historical substitutions under different environmental conditions, and in different yeast strains or species. Under different conditions, I would expect some mutations that were deemed neutral under our conditions to exhibit growth defects.

For the evolution of steroid receptor DNA-binding specificity, it is not as easy to test the robustness of our conclusions on contingency and stochasticity to the consideration of alternative functions. Even if we were to test the ability of alternative RH combinations to bind the spectrum of non-consensus, genomically distributed SRE motifs in a modern cell, we cannot be sure whether any or all of these SREs were present in the ancient organism in which AncSR2 evolved. Inspired by theoretical consideration (166), our model is that a subset of possible SREs emerged or co-evolved with the origin of AncSR2; following the evolution of AncSR2's new DNA-binding specificity, additional SREs across the genome emerged or drifted to the margins

of the protein's DNA-binding capability, giving rise to the diversity of SRE sequences present in modern-day organisms and entrenching the derived steroid receptor DNA-binding domain, slowing its evolution. Though this model might not be testable in this particular system, other more-recently diverged transcription factors in amenable model systems such as yeast might be experimentally tractable. Nonetheless, the observation that contingency and stochasticity play dominant roles in the outcomes of replicate experimental evolution trajectories – which happen over observable timeframes – are consistent with our observations about ancient protein evolution (41, 43).

At a broader level, the sheer size of sequence-function landscapes emphasizes that we will never be able to experimentally probe the depths of its entirety. Comprehensive combinatorial scans are currently limited in scale, more or less, to the four-site library that we considered in the steroid receptors, and because of epistasis, low-order deep mutational scans do not reliably predict the structure of sequence-function landscapes at higher-order mutational distances (200). Though the structure of the protein sequence-function landscape may never be exhaustively determined, the work described here illustrates how well-crafted evolutionary questions can be approached in new ways because of this technology, lending new insight into the molecular evolution of proteins.

Appendix 1

Supplementary figures for Chapter 3: Pervasive contingency and entrenchment in a billion years of Hsp90 evolution

---

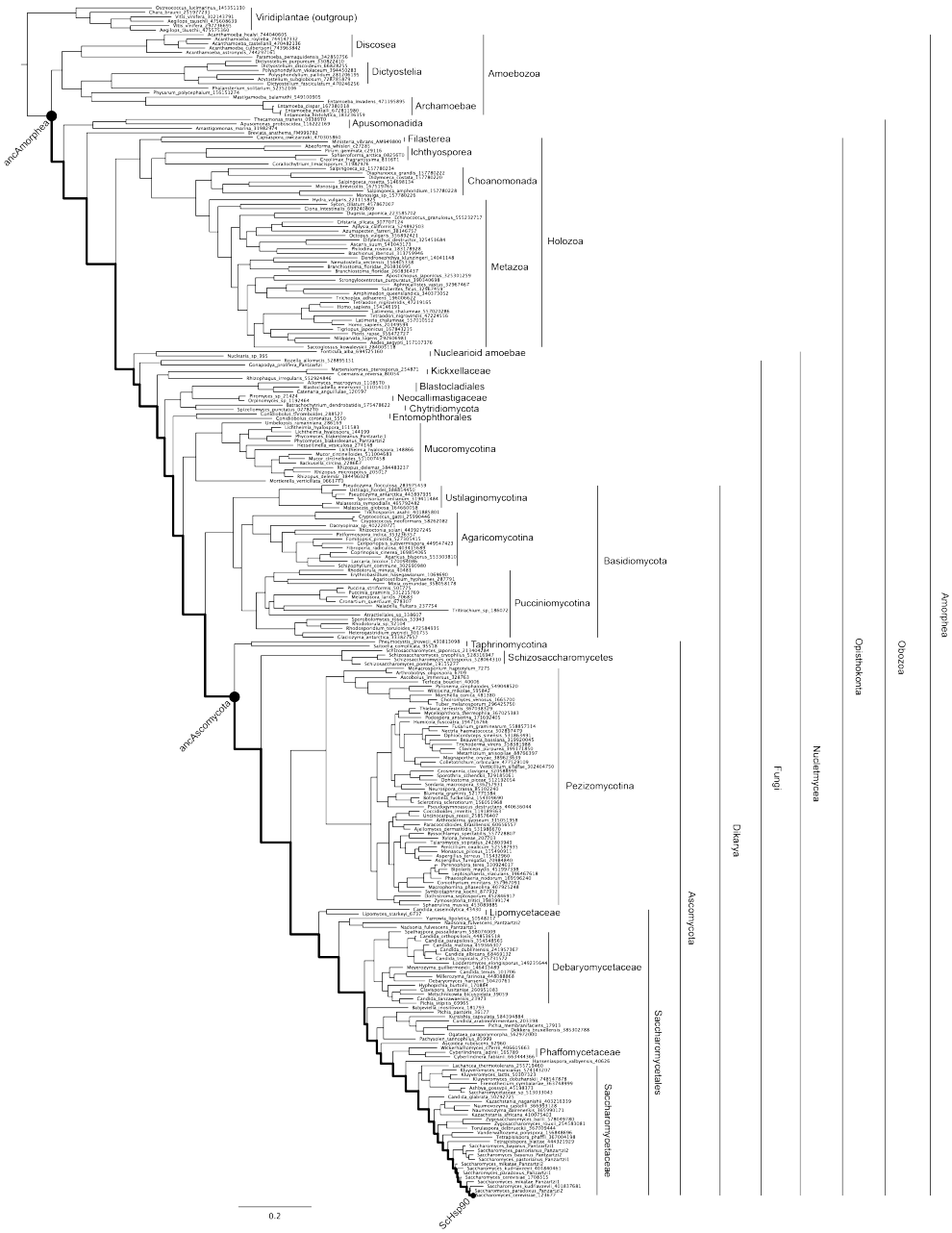


Figure A1.1. Hsp90 phylogeny. The maximum likelihood phylogeny of 267 Hsp90 protein sequences, with major taxonomic groups labeled. Taxon names indicate genus, species, and an accession number or sequence identifier. Nodes characterized in this study are shown as black dots; the trajectory studied is shown as a thick black line.

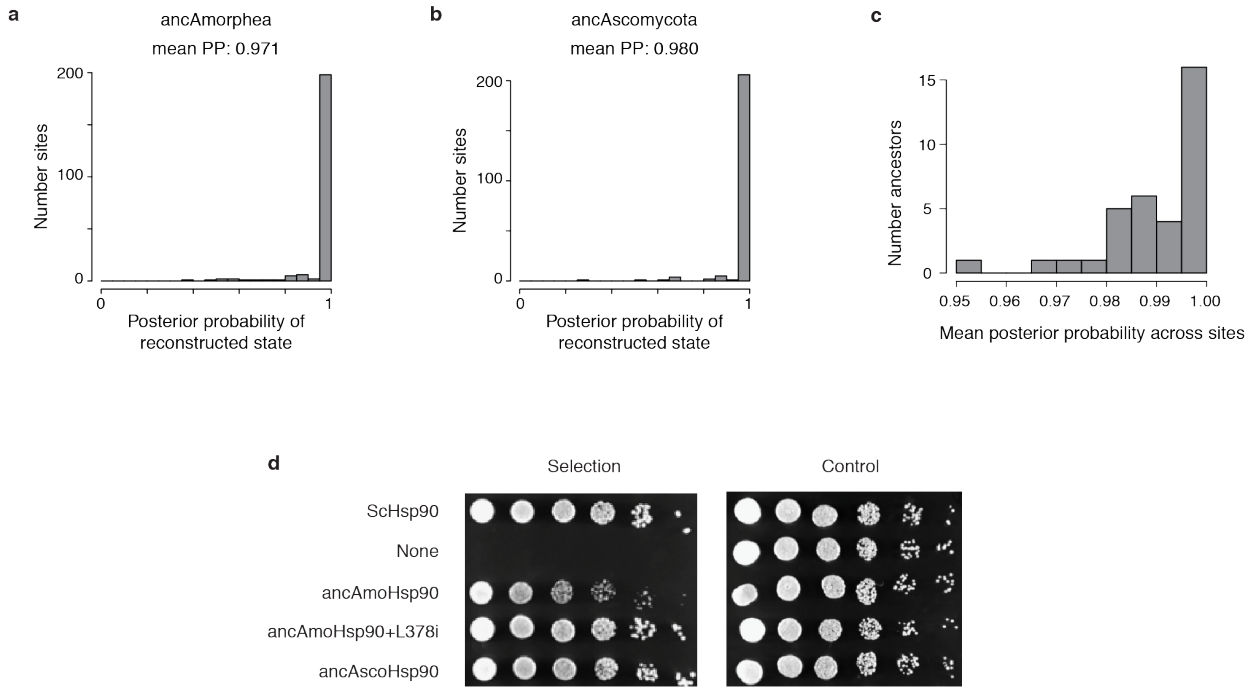


Figure A1.2. Ancestral Hsp90 sequences have high statistical support and complement yeast growth. a,**b**, For the ancestral NTD sequences reconstructed in this study, the distribution of posterior probability of ancestral states across NTD sites is shown as a histogram. The mean posterior probability of the most probable state across sites (mean PP) is shown for each ancestor. **c**, The distribution of mean PP for all reconstructed ancestral sequences along the trajectory from ancAmoHsp90 to ScHsp90. **d**, Growth of *S. cerevisiae* Hsp90 shutoff strains complemented with ancestral Hsp90 NTD variants. Spots from left to right are 5-fold serial dilutions. Control plates represent conditions in which the native ScHsp90 allele is expressed. Under selection conditions, the native ScHsp90 allele is turned off, and growth can only persist when a complementary Hsp90 allele is provided. The ancAmoHsp90 NTD expressed as a chimera with the Sc middle and C-terminal domains exhibits a slight growth defect; this is rescued by adding an additional reversion to the ancAmoHsp90 state in the middle domain (L378i), which occurs on a middle domain loop that extends down and interacts directly with the N-terminal domain and contributes to the NTD ATP-binding pocket. We subsequently refer to ancAmoHsp90+L378i as ancAmoHsp90.

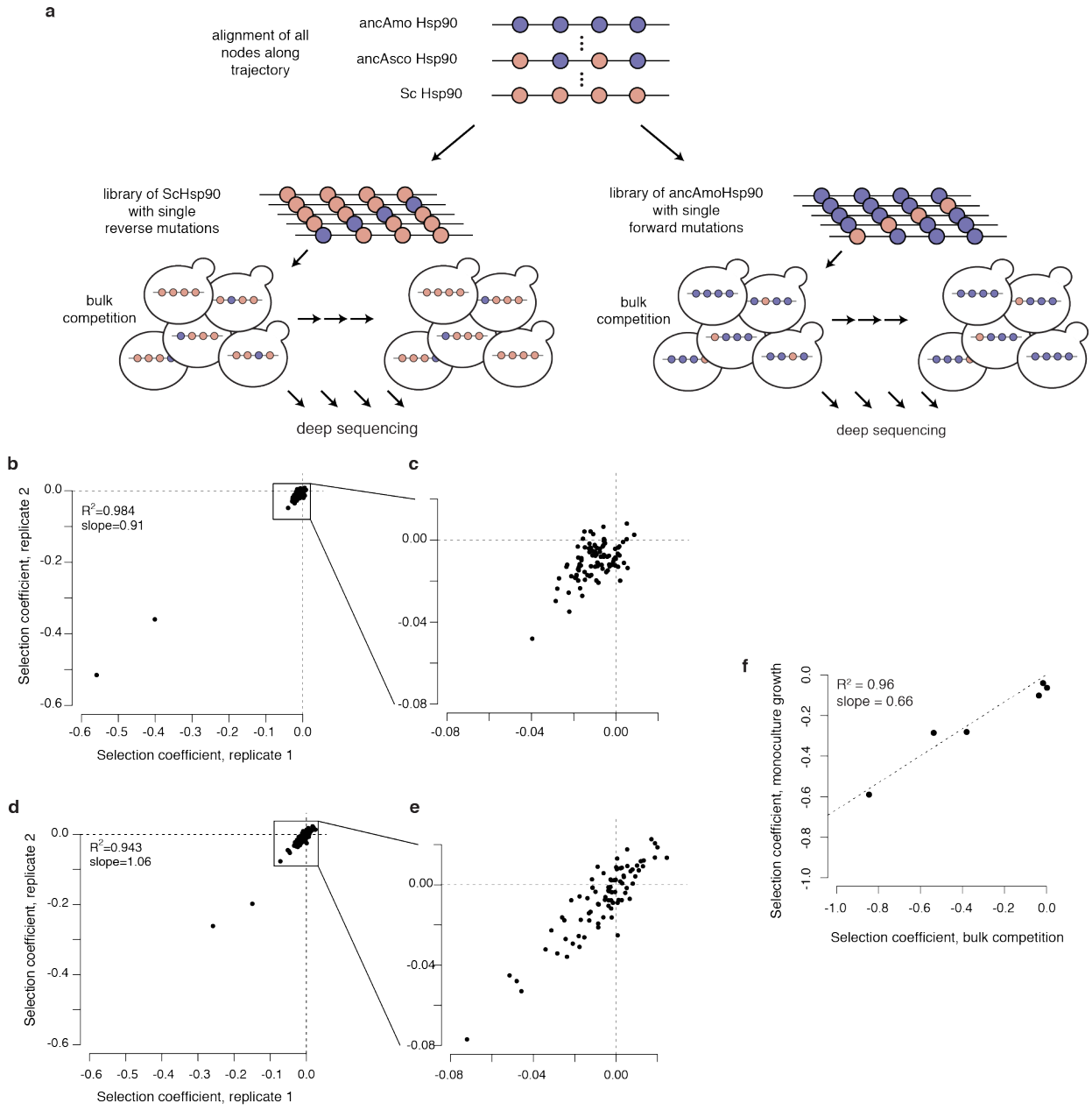


Figure A1.3. Experimental scheme and reproducibility. a, Experimental scheme for testing the fitness effects of individual mutations to ancestral states in ScHsp90 (left) or individual mutations to derived states in ancAmoHsp90 (right). An alignment of all ancestors along the focal trajectory was constructed to identify the trajectory of Hsp90 NTD sequence change from ancAmoHsp90 to ScHsp90. In the ScHsp90 and ancAmoHsp90 backgrounds, libraries were constructed consisting of the wildtype sequence and all individual mutations to ancestral or derived states. These libraries were transformed into yeast, which grew through a bulk competition. The frequency of each genotype at each time point was determined by deep sequencing, allowing us to calculate a selection coefficient for each mutation relative to the respective wildtype sequence. **b**, Reproducibility in selection coefficient estimates for replicate

(**Figure A1.3, continued**) bulk competitions of the ScHsp90 library. R^2, Pearson coefficient of determination. **c**, For visual clarity, zoomed in representation of the boxed region in (**b**). **d**, Reproducibility in selection coefficient estimates for replicate bulk competitions of the ancAmoHsp90 library. R^2, Pearson coefficient of determination. **e**, For visual clarity, zoomed in representation of the boxed region in (**d**). **f**, Correlation in fitness as measured via bulk competition or monoculture growth assay. R^2, Pearson coefficient of determination. The line was forced to go through (0, 0); when freely fit, the intercept term was not significantly different from zero.

a

Number components	log-likelihood	Number parameters	AIC	Estimated proportion deleterious	Estimated proportion beneficial
2	334.4	3	-662.7	0.921	0.079
3	341.0	6	-669.9	0.918	0.013
4	341.8	9	-665.6	0.918	0.016
5	343.2	12	-662.3	0.951	0.049

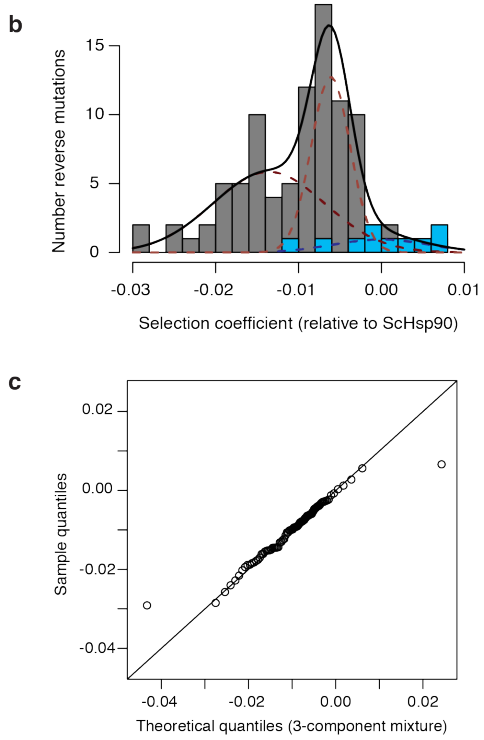


Figure A1.4. Estimating the proportion of mutations to ancestral states that are deleterious with a mixture model. a, Observed selection coefficients of reversions were fit to mixture models containing a variable number of Gaussian distributions; in each case, one neutral mixture component is fixed to have the mean and standard deviation of the sampling distribution of replicate wildtype ScHsp90 sequences present in the library, the mixture proportion of which is a free parameter; each additional non-neutral mixture component has a free mean, standard deviation, and mixture proportion. The empirical data were best fit by a 3-component mixture model, as assessed by AIC. Proportion deleterious (or beneficial) was estimated by summing the cumulative density below (or above) zero of the non-neutral components. **b**, The best-fit mixture model. Gray bars, observed distribution of selection coefficients of ancestral reversions; blue bars, distribution of observed selection coefficients of wildtype ScHsp90 sequences present in the library. Black line, best-fit mixture model; red dashed lines, individual non-neutral mixture components; blue dashed line, neutral (wildtype) mixture component. The area under the curve for each mixture component corresponds to the proportion it contributes to the overall mixture model. **c**, Quantile-quantile plot showing the quality of fit of the 3-component mixture model (*x*-axis) to the empirical distribution of selection coefficients of ancestral reversions (*y*-axis). The mixture model assigns more extreme selection coefficients to the tails than is observed in the empirical distribution, but provides a reasonable fit along the bulk of the distribution.

a

Method	Proportion deleterious, Sc	Proportion neutral, Sc	Proportion beneficial, Sc	Proportion deleterious, ancAmo	Proportion neutral, ancAmo	Proportion beneficial, ancAmo
Mixture model	0.92	0.07	0.01	0.48	0.32	0.20
Count	0.95	0.00	0.05	0.66	0.00	0.34
Empirical Bayes	0.81	0.16	0.03	0.55	0.19	0.27
Confidence Interval	0.54	0.45	0.01	0.48	0.26	0.26

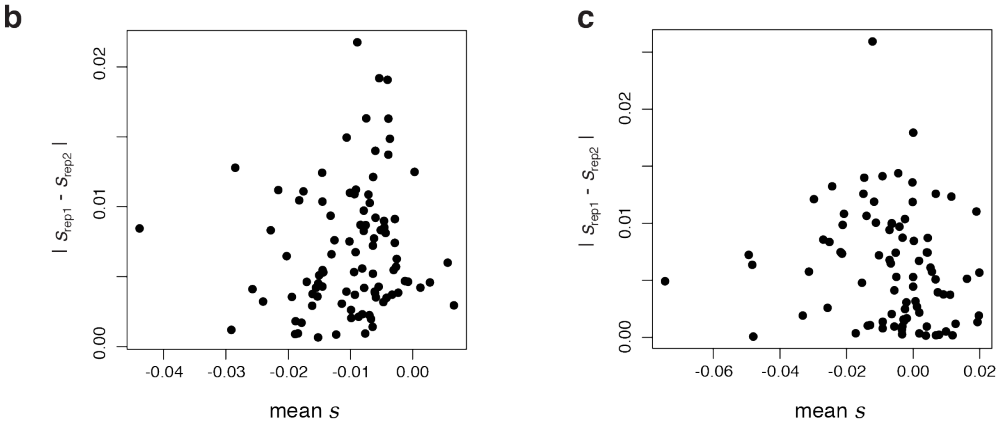
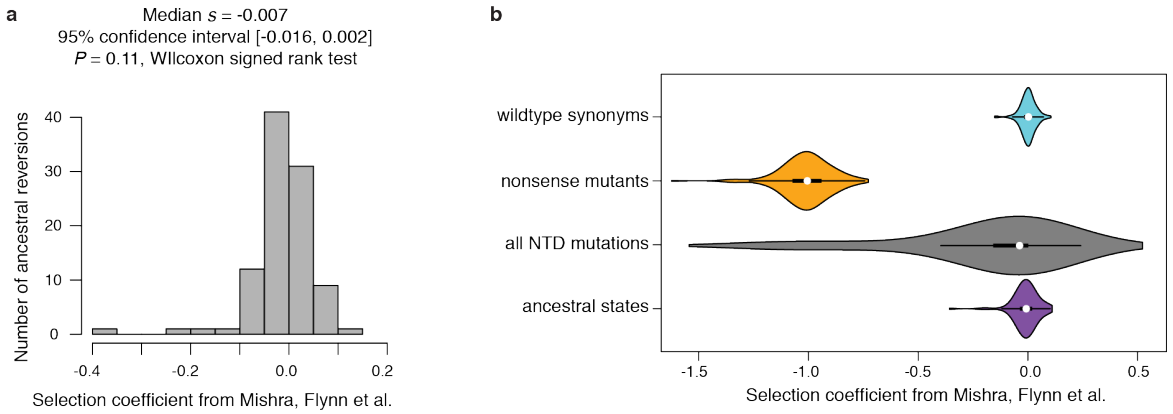


Figure A1.5. Alternate approaches for estimating the proportion of mutations that are deleterious. a, The estimated proportion of mutations that are deleterious, neutral, or beneficial in each background, as determined by each of four statistical methods. See SI Methods for descriptions of each method. **b,c**, Experimental errors are unbiased with respect to the observed selection coefficient. For the ScHsp90 (**b**) and ancAmoHsp90 (**c**) backgrounds, the absolute difference in *s* as determined in each replicate is shown versus their mean. In each background, there is no significant linear relationship between experimental error and s_{obs} ($P = 0.27$ and 0.24, respectively, Pearson's correlation).



c

Properties of sites	Spearman's rank correlation, $s > -0.1$		Spearman's rank correlation, all mutations	
	rho	*P*-value	rho	*P*-value
Average effect of all mutations at position (Mishra, Flynn et al.)	0.267	0.009	0.308	0.002
Relative evolutionary rate	0.294	0.004	0.326	0.001
Relative solvent accessibility (PDB 2CG9)	0.331	0.002	0.313	0.003
Distance to ATP gamma-phosphate	0.380	0.0003	0.395	0.0001
Properties of mutations				
BLOSUM62 score	0.048	0.65	0.077	0.45
Difference in hydrophobicity	-0.209	0.04	-0.178	0.08
Absolute difference in hydrophobicity	-0.121	0.24	-0.136	0.18
Difference in side chain volume	0.134	0.19	0.129	0.21
Absolute difference in side chain volume	-0.042	0.68	-0.072	0.48
Evolutionary distance from ScHsp90 to most recent ancestor with ancestral state	0.020	0.84	0.005	0.96

Figure A1.6. Ancestral states are deleterious in yeast Hsp90. a, The signature of deleterious ancestral states is present in the independent but lower-resolution dataset of Mishra, Flynn et al. (16). For each mutation to an ancestral state, the selection coefficient as determined by Mishra, Flynn et al. is shown. The median selection coefficient is -0.007, close to that estimated in the current study; however, this median selection coefficient is not significantly different than zero ($P = 0.11$). Because Mishra, Flynn et al. tested a much larger panel of mutations (all single mutations across the entire NTD), experimental variability of estimated selection coefficients was much larger, possibly explaining the lack of significance of this result in this dataset. **b**, Violin plots show the distribution of mutant effects in the dataset of Mishra, Flynn et al. (16). Ancestral states are less detrimental than the average random mutation in the NTD ($P = 3.5\times10^{-9}$, Wilcoxon rank sum test with continuity correction). **c**, Reversions exhibit properties typical of genuinely deleterious mutations. For various properties of sites at which we measured the fitness of ancestral variants (top) or properties of the specific amino acids mutated (bottom), we asked whether there was a significant correlation between the property and the selection coefficients of mutations via Spearman's rank correlation. Ancestral states tend to be more deleterious at positions that are less robust to any mutation, evolve more slowly, are less solvent accessible, and are closer to the gamma-phosphate of bound ATP. These properties are not completely independent; for example, there is a significant positive correlation between relative solvent accessibility and distance to ATP gamma-phosphate. Biochemical properties particular to the amino acid states in each mutation are generally not significantly correlated with the selective effect. Furthermore, we see no evidence for older states being more entrenched, as has been observed by others (29, 30, 79).

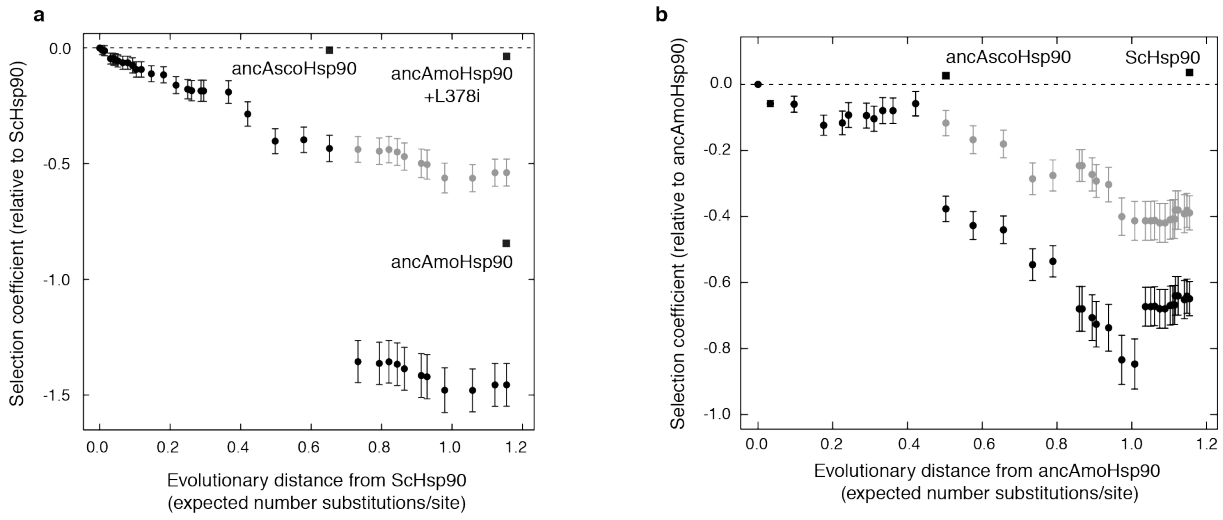


Figure A1.7. Fitness effects of historical substitutions are modified by intramolecular epistasis. Each black circle represents an ancestral protein along the trajectory from ancAmoHsp90 to ScHsp90. Position along the *x*-axis shows the evolutionary distance that separates it from ScHsp90 (**a**) or ancAmoHsp90 (**b**); *y*-axis position shows the predicted selection coefficient assuming no epistasis relative to ScHsp90 (**a**) or ancAmoHsp90 (**b**). Predicted selection coefficients were calculated as the sum of individual selection coefficients for all sequence differences present in its sequence as measured in ScHsp90 (**a**) or ancAmoHsp90 (**b**). Error bars show the 95% confidence interval for the predicted value, calculated by propagating the standard errors of individual site-specific selection coefficient measurements. Light gray dots show the same data, but excluding the effects of the two strongly deleterious outliers in each library. Labeled squares indicate experimentally determined selection coefficients for complete genotypes: ancAscoHsp90, ancestral Ascomycota (fitness determined via monoculture growth); ancAmoHsp90, ancestral Amorphea (fitness determined via bulk competition); ancAmoHsp90+L378i, ancAmoHsp90 with a candidate epistatic substitution in the Middle Domain also reverted to its ancAmorphea state (fitness determined via bulk competition). Dashed line, $s = 0$.

a

Number components	log-likelihood	Number parameters	AIC	Estimated proportion deleterious	Estimated proportion beneficial
2	243.9	3	-481.7	0.532	0.250
3	245.5	6	-479.0	0.587	0.330
4	246.2	9	-474.4	0.412	0.132
5	252.7	12	-481.5	0.482	0.197
6	252.7	15	-475.5	0.482	0.197

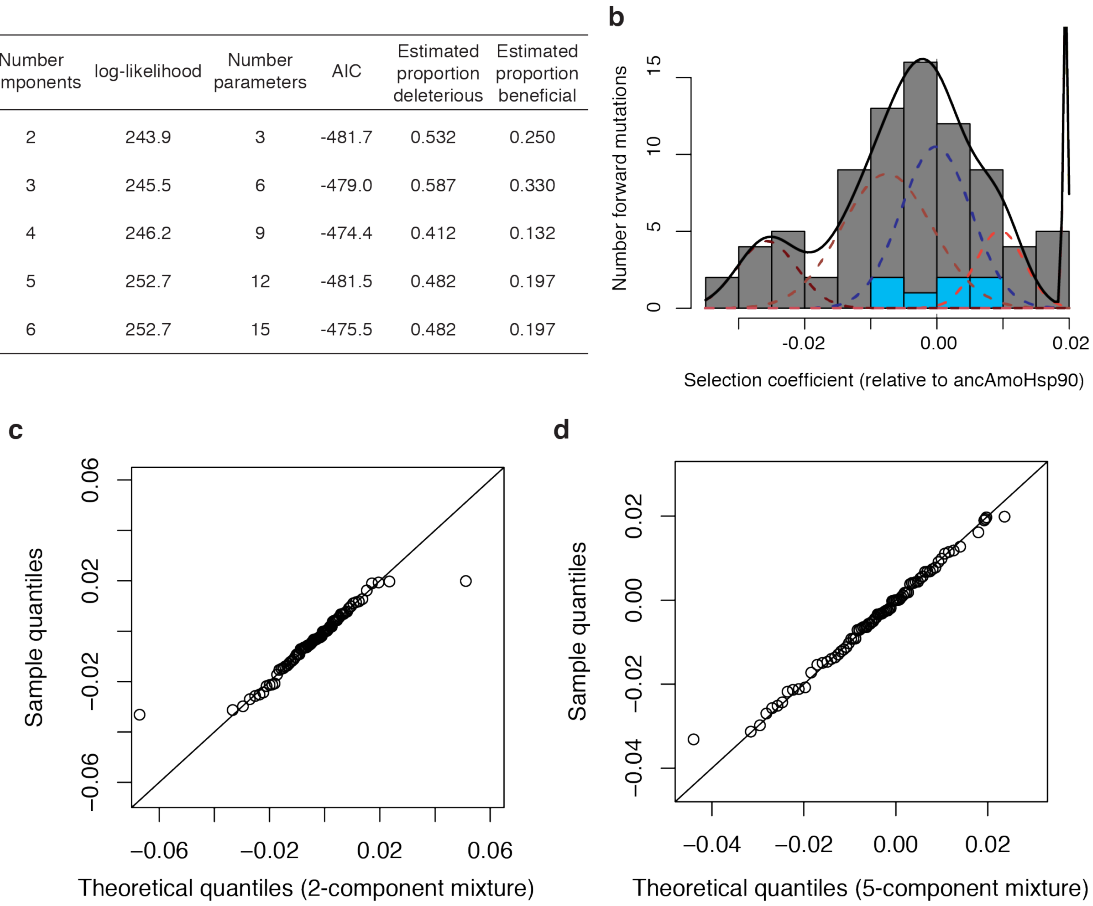


Figure A1.8. Estimating the proportion of mutations to derived states that are deleterious with a mixture model. a, The distribution of selection coefficients of mutations to derived states was fit by mixture models containing a variable number of Gaussian distributions; in each case, one neutral mixture component is fixed to have the mean and standard deviation of the sampling distribution of replicate wildtype ancAmoHsp90 alleles in the library, the mixture proportion of which is a free parameter; each additional non-neutral mixture component has a free mean, standard deviation, and mixture proportion. The empirical data were best fit by a 2-component mixture model, as judged by AIC, with a 5-component mixture being almost equally well fit; the 5-component mixture resulted in a more conservative estimate of the proportion of mutations that were deleterious than the 2-component mixture, and so was chosen despite the AIC difference of 0.2. Proportion deleterious (or beneficial) was estimated by summing the cumulative density below (or above) zero of the non-neutral components. **b**, The fit of the 5-component mixture model. Gray bars, distribution of selection coefficients of mutations to derived states; blue bars, distribution of selection coefficients of independent ancAmoHsp90 alleles present in the library. Black line, five-component mixture model; red dashed lines, individual non-neutral mixture components; blue dashed line, neutral (wildtype) mixture component. The area under the curve for each mixture component corresponds to the proportion it contributes to the overall mixture model. **c**,**d**, Quantile-quantile plot showing the quality of fit of the 2-component (**c**), or 5-component (**d**), mixture models (*x*-axis) to the empirical distribution of selection coefficients of mutations to derived states (*y*-axis).

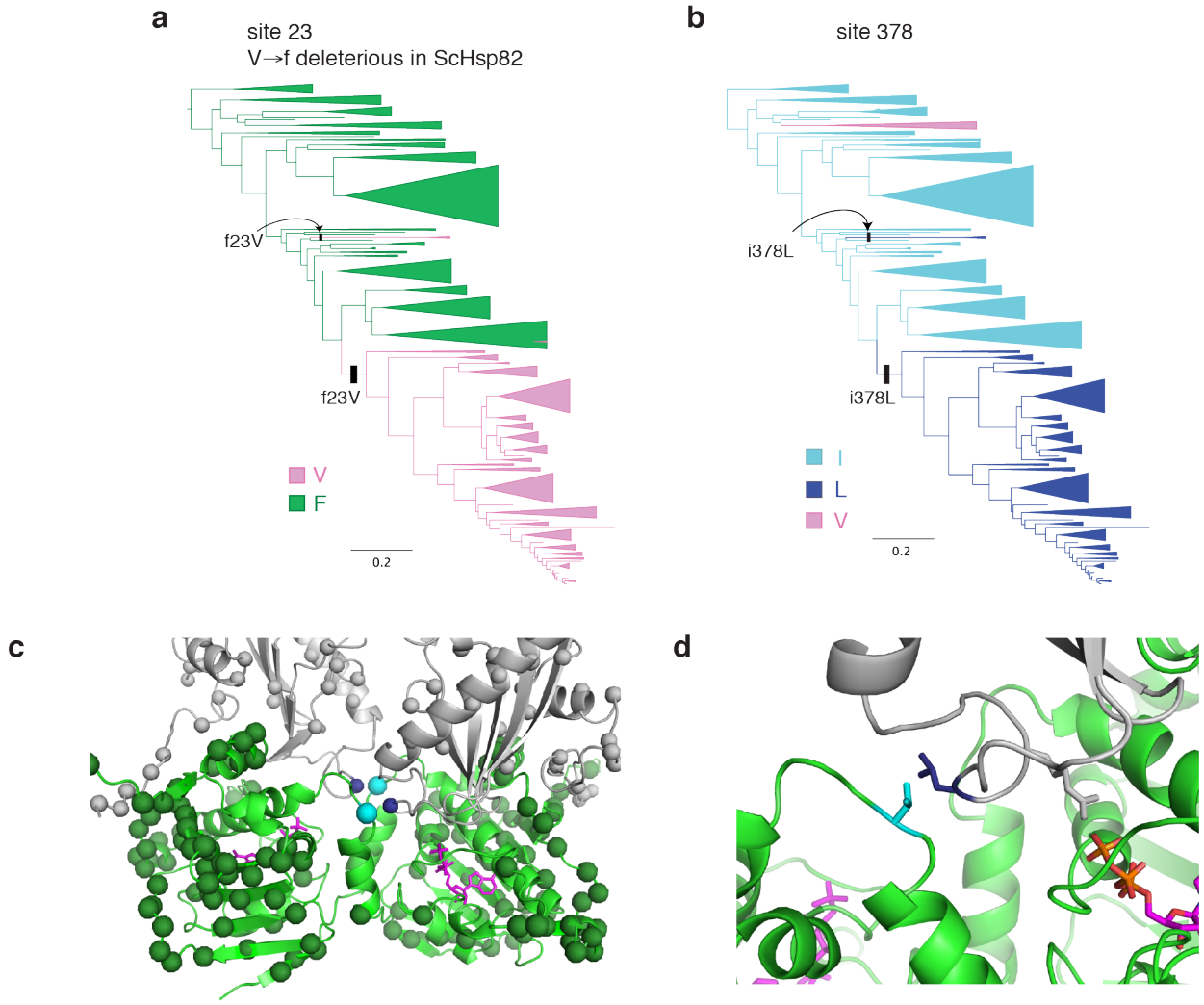


Figure A1.9. The deleterious V23f reversion is ameliorated by L378i. a,**b** Character state patterns at sites 23 (**a**) and 378 (**b**). On the lineage to ScHsp90, f23V co-occurred with i378L before the common ancestor of Ascomycota. The same two substitutions also co-occur on an independent lineage on this phylogeny (Kickxellaceae fungi), and in the distantly related Rhodophyta red algae (GenBank ADB45333.1 and RefSeq XP_005715129.1). **c**, The locations of sites 23 and 378 on the ATP-bound Hsp90 dimer structure (PDB 2CG9). Cyan spheres, site 23; dark blue; site 378; dark green, other variable NTD sites; gray, other variable middle and C-terminal domain sites. Magenta sticks, ATP. **d**, Zoomed view of sites 23 and 378. These side chains are in direct structural contact, and may be important for the positioning of the middle domain loop that bears R380 (gray sticks), which forms a salt bridge with the ATP gamma-phosphate and is critical for ATP binding and hydrolysis (134, 151).

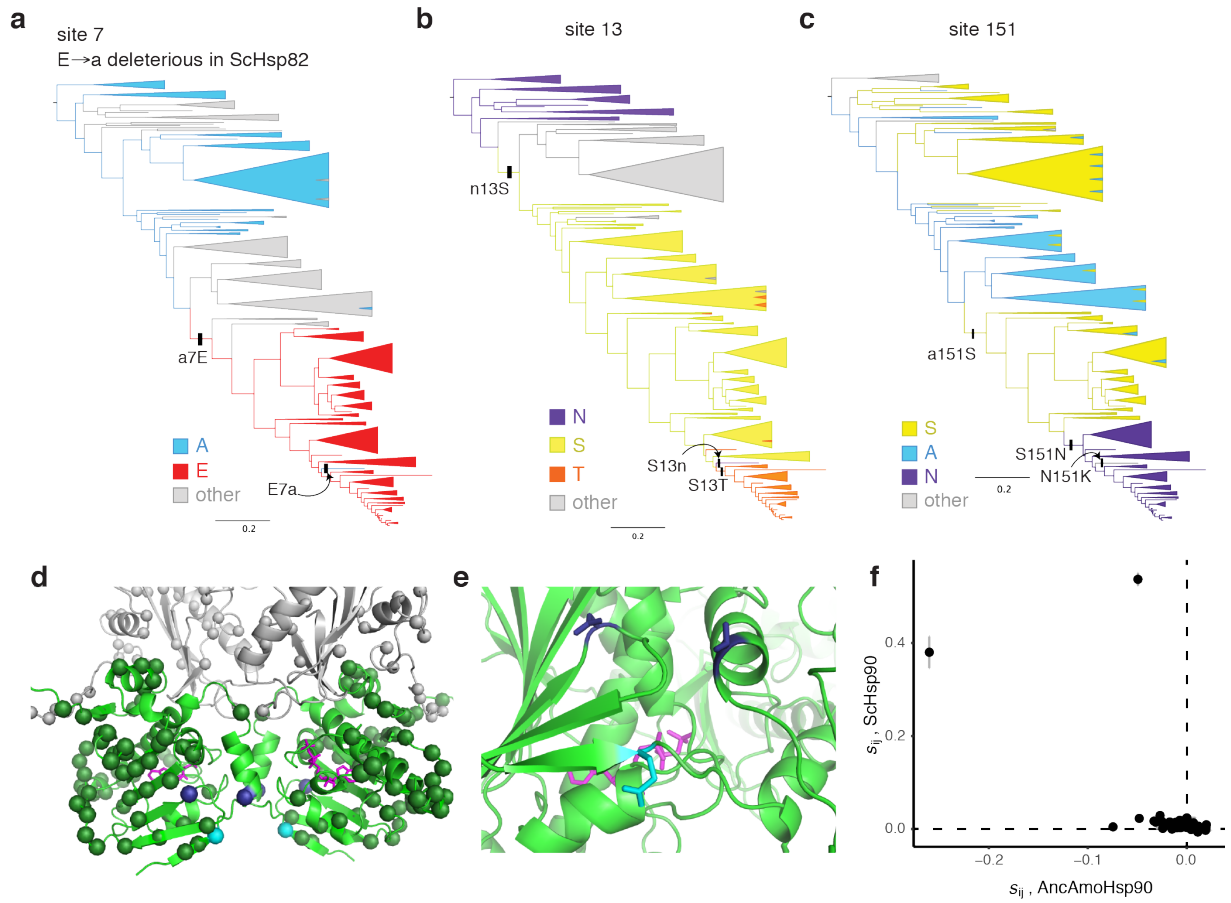
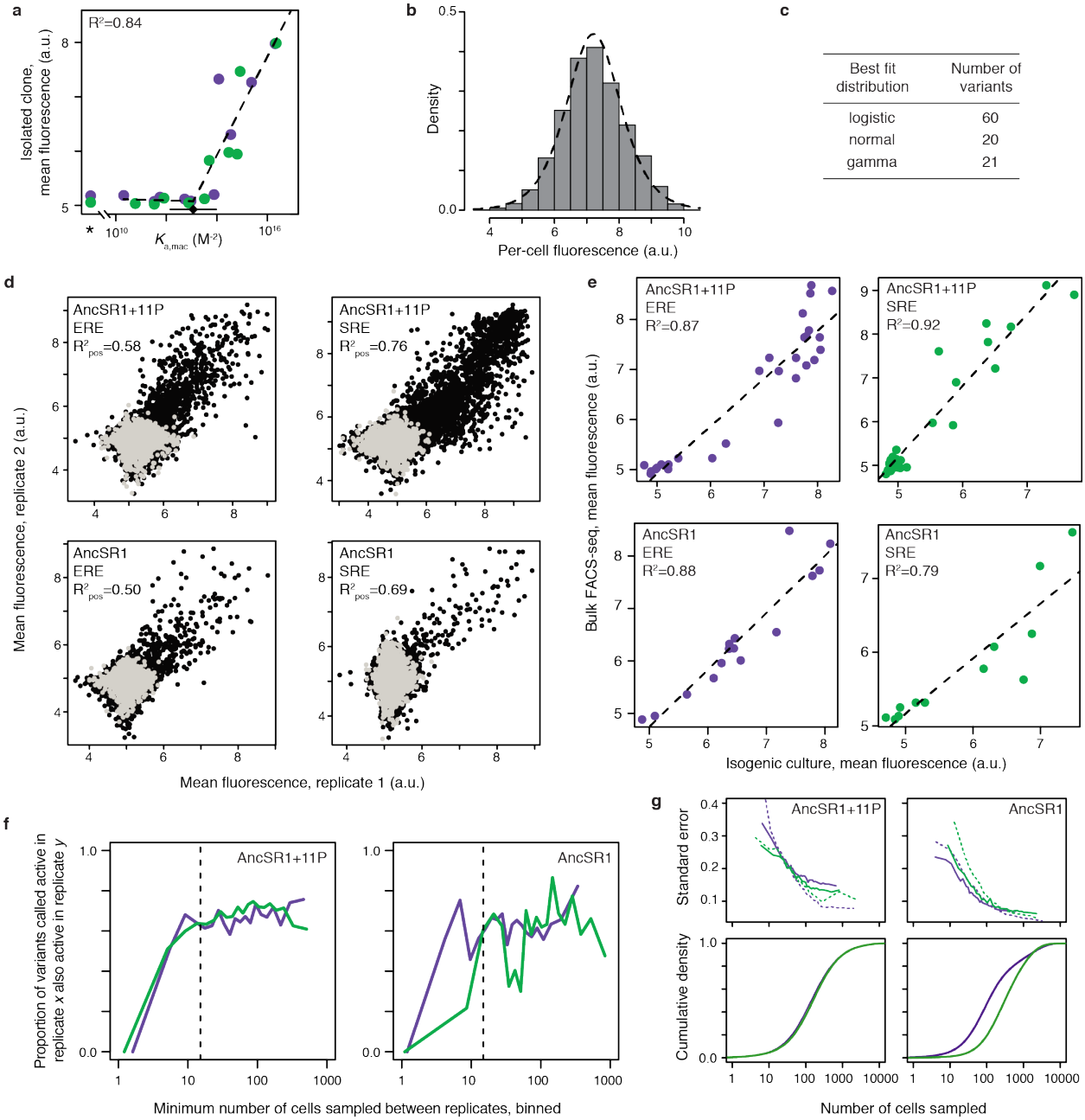


Figure A1.10. The deleterious E7a reversion is partially ameliorated by N151a or T13n. **a**,**b**,**c**, Character state patterns at sites 7 (**a**), 13 (**b**) and 151 (**c**). On the trajectory to ScHsp90, a7E occurred before the common ancestor of Ascomycota, then later reverted in the lineage leading to *Ascoidea rubescens* (arrow)*;* on this latter lineage, site 13 also reverted to the ancestral state asparagine, and site 151 substituted to a third state lysine. **d**, The locations of sites 7, 13, and 151 on the ATP-bound Hsp90 structure (2CG9), represented as in Fig. S9c. Cyan spheres, site 7; dark blue, sites 13 and 151. **e**, Zoomed in view of sites 7, 13, and 151. These side chains are not in direct physical contact; however, site 7 is on a beta strand that undergoes extensive conformational movement when Hsp90 converts between ADP- and ATP-bound states. **f**, The same plot as Fig. 5c is shown, including the two strongly outliers V23f and E7a. See Fig. 5c legend for details.

Appendix 2

Supplementary figures and tables for Chapter 4: Alternative evolutionary histories in the sequence space of an ancient protein

---



Best fit distribution	Number of variants
logistic	60
normal	20
gamma	21

Figure A2.1. Design and validation of a yeast FACS-seq assay for steroid receptor DNA-binding function. a, GFP activation in ERE (purple) and SRE (green) yeast reporters correlates with previously measured protein-DNA binding[11,12]. Asterisk, stop-codon-containing variant.

(**Figure A2.1, continued**) Dashed line, best fit segmented-linear relationship between GFP activation and $\log_{10}(K_{a,mac})$. **b**, Histogram of the per-cell green fluorescence for AncSR1 on ERE measured via flow cytometry, fit to a logistic distribution (dashed line). **c**, Distributions that provide the best fit to flow cytometry data for isogenic cultures of 101 DBD variants, using Akaike Information Criterion (AIC). **d**, Comparisons of mean fluorescence estimates between FACS-seq replicates of each protein/response element combination. Black points, coding RH variants; light gray, stop-codon-containing variants. R^2_{pos}, squared Pearson correlation coefficient for variants with mean fluorescence significantly higher than stop-codon-containing variants in either or both replicates. **e**, Comparisons between mean fluorescence as determined in FACS-seq and via flow cytometry analysis of isogenic cultures for a random selection of clones from each library. Dashed line, best-fit linear regression. **f**, Robustness of classification to sampling depth. Variants were binned according to the minimum number of cells with which they were sampled in either replicate. Below 15 cells sampled (dashed line), the probability that a variant called active in one replicate was also called active in the other is dependent on sampling depth; to minimize errors due to sampling depth, we eliminated as "undetermined" all variants with less than 15 cells sampled after pooling replicates. **g**, Standard error of mean fluorescence estimates (SEM) in each library as a function of sampling depth. Top plots show for each background, the relationship between SEM and sampling depth for ERE (purple) and SRE (green) libraries, as estimated from the sampling distribution of stop-codon-containing variants (dotted lines) or variability in mean fluorescence estimates between replicates (solid lines). Bottom panels show the cumulative fraction of coding variants in each library that have a certain number of cells sampled in the pooled data.

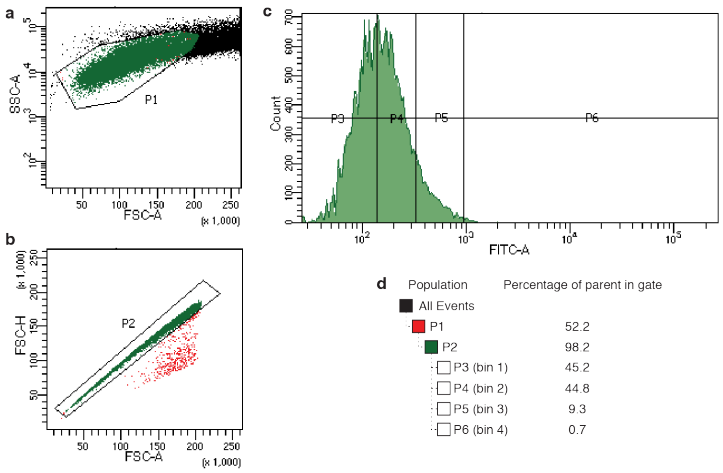
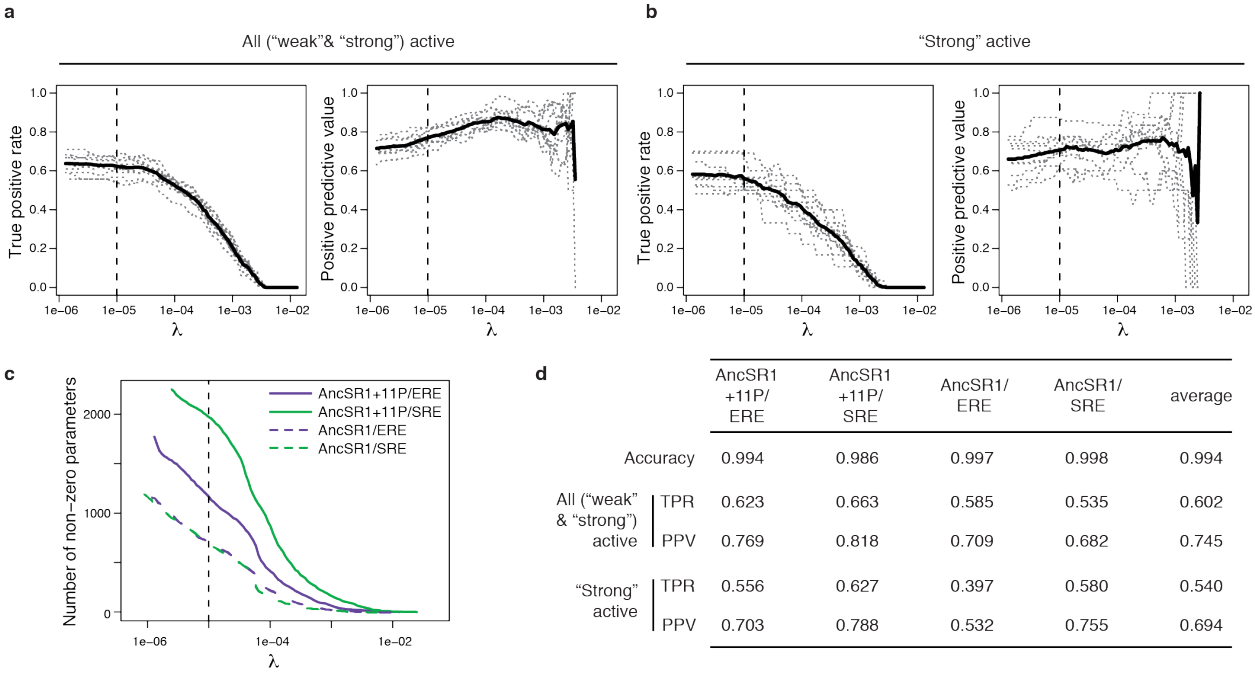


Figure A2.2. Representative FACS gates for library sorting. a, A scatterplot of side-angle scattering (SSC-A) and forward-angle scattering (FSC-A) selects for a homogenous cell population (P1). **b**, A scatterplot of the height of the per-cell forward scatter peak (FSC-H) compared to the integrated area of this peak (FSC-A) excludes events where multiple cells pass through the detector simultaneously (P2). **c**, Final sort bins (P3 – P6) are drawn on the distribution of green fluorescence (FITC-A). **d**, Table showing the hierarchical parentage of sort gates and the percentage of events that fall in each bin.



		AncSR1 +11P/ ERE	AncSR1 +11P/ SRE	AncSR1/ ERE	AncSR1/ SRE	average
Accuracy		0.994	0.986	0.997	0.998	0.994
All ("weak" & "strong") active	TPR	0.623	0.663	0.585	0.535	0.602
	PPV	0.769	0.818	0.709	0.682	0.745
"Strong" active	TPR	0.556	0.627	0.397	0.580	0.540
	PPV	0.703	0.788	0.532	0.755	0.694

Figure A2.3. Models to predict the function of missing genotypes. For each protein/response element combination, a continuation ratio ordinal logistic regression model was constructed to predict the functional class of a variant as a function of its four RH amino acid states, including possible first order main effects and second order pairwise epistatic effects. 10-fold cross-validation was used to select the penalization parameter λ and evaluate performance. **a**,**b**, True positive rate (left, TPR, the proportion of experimental positives that are predicted positive) and positive predictive value (right, PPV, the proportion of predicted positives that are experimentally positive) are shown as a function of λ for AncSR1+11P on ERE. Classifications were evaluated for (**a**) all active (weak and strong) versus inactive variants and (**b**) strong active versus weak active and inactive variants. Gray dotted lines, cross-validation replicates; solid line, mean. Dashed line shows the chosen value of $\lambda = 10^{-5}$; as λ continues to decrease beyond $\lambda = 10^{-5}$, TPR plateaus but PPV continues to decline. **c**, The number of non-zero parameters included in each model as a function of λ. Dashed line, $\lambda = 10^{-5}$. **d**, Summary of performance metrics from 10-fold cross-validation for each model with $\lambda = 10^{-5}$. Accuracy is the proportion of predicted classifications (strong, weak, and inactive) that match their experimentally determined classes.

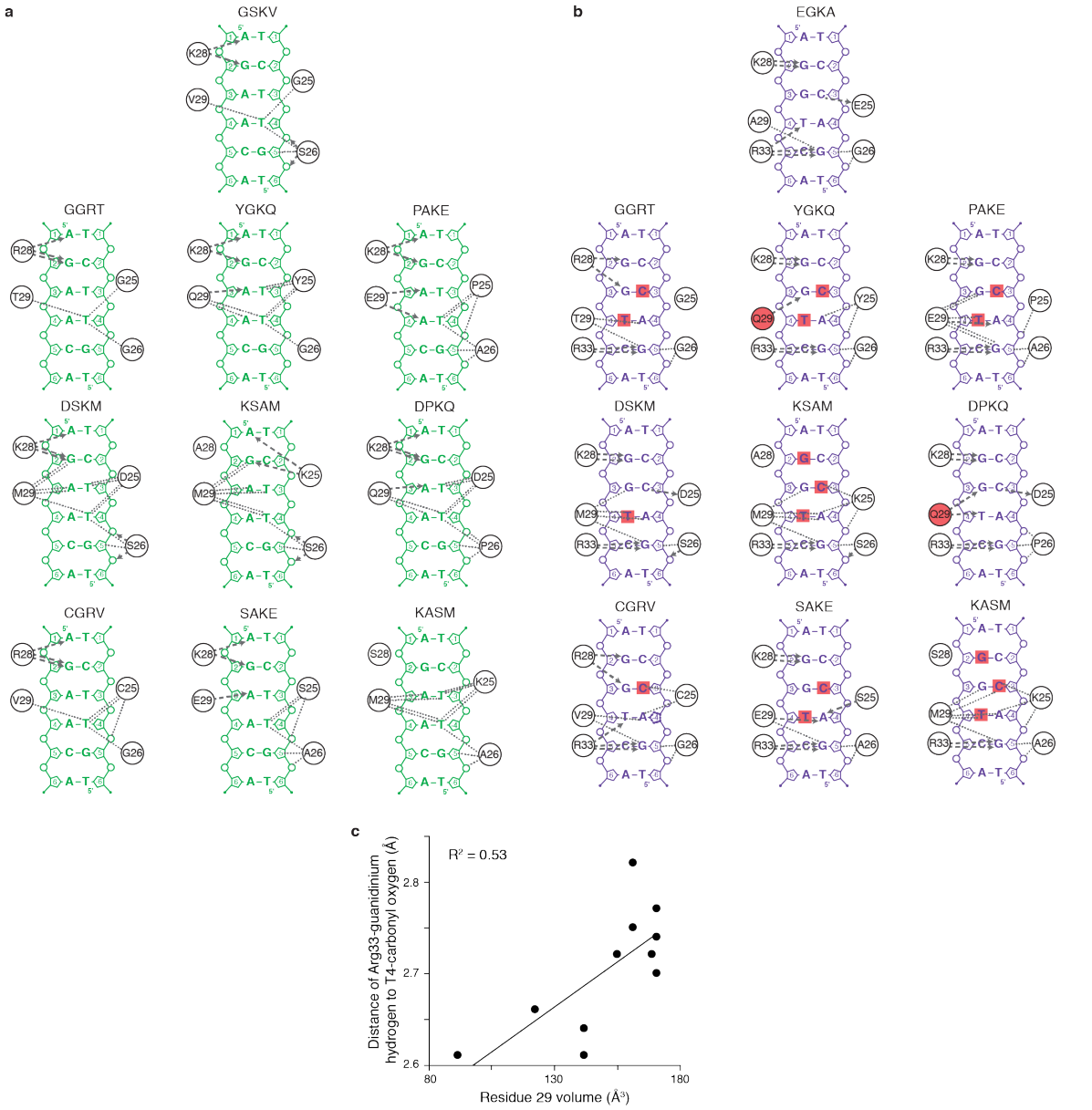
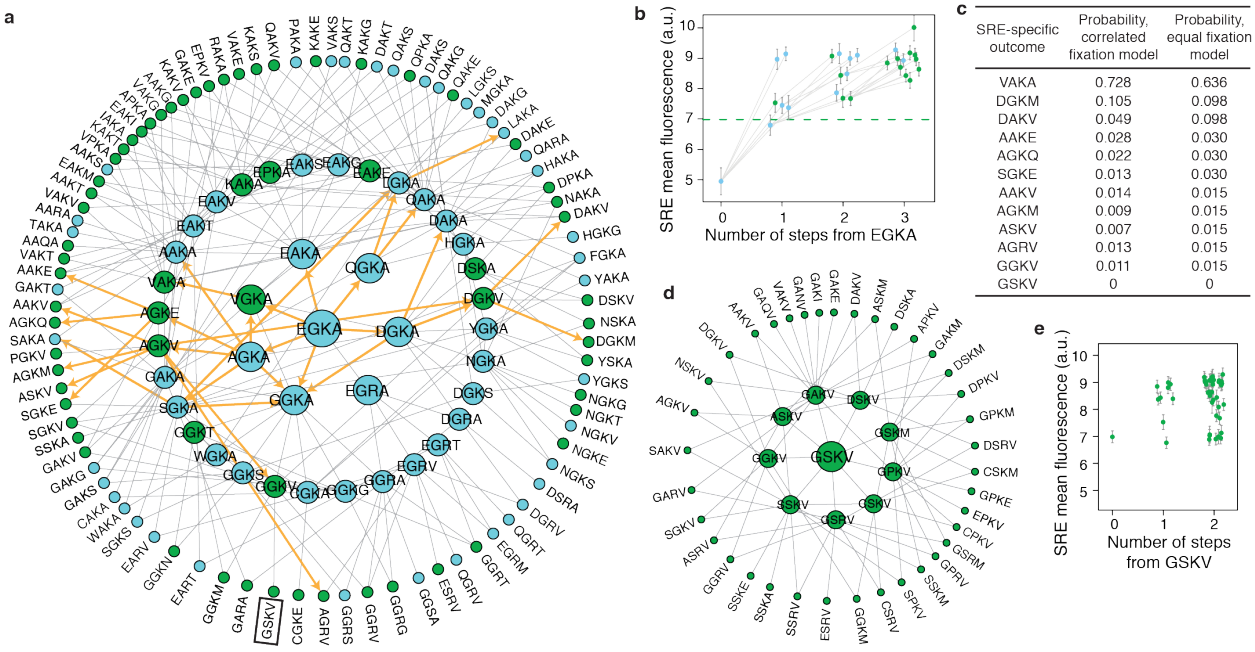


Figure A2.4. Biophysical diversity in DNA recognition. a,**b**, Diverse mechanisms for recognition of SRE (**a**) or ERE (**b**) by the historical RH genotypes (GSKV and EGKA) and alternative SRE-specific variants. Contacts from FoldX-generated structural models are shown between RH residues (circles) and DNA bases (letters), backbone phosphates (small circles) and sugars (pentagons, numbered by position in the DNA motif; dashed numbers refer to the complementary strand). Hydrogen bonds are shown as dashed arrows from donor to acceptor; dotted lines, non-bonded contacts. Red squares, bases that form hydrogen bonds in the EGKA-ERE structure that are unsatisfied in complex with an SRE-specific RH; red circles, side chains with polar groups that are not satisfied in complex with ERE. Only DNA contacts that vary among the analyzed structures are shown. **c**, Large side chains at position 29 correlate with the loss of a conserved R33 hydrogen bond to ERE. For ERE-bound structural models, the distance of the Arg33 guanidinium hydrogen to the ERE T4 carbonyl oxygen was measured and compared to the atomic volume of the residue at position 29 in that variant.



SRE-specific outcome	Probability, correlated fixation model	Probability, equal fixation model
VAKA	0.728	0.636
DGKM	0.105	0.098
DAKV	0.049	0.098
AAKE	0.028	0.030
AGKQ	0.022	0.030
SGKE	0.013	0.030
AAKV	0.014	0.015
AGKM	0.009	0.015
ASKV	0.007	0.015
AGRV	0.013	0.015
GGKV	0.011	0.015
GSKV	0	0

Figure A2.5. The ancestral RH (EGKA) and derived RH (GSKV) can access many SRE-specific outcomes by short paths in AncSR1+11P. a, Concentric rings contain RH genotypes of minimum path length 1, 2, or 3 steps from AncSR1+11P:EGKA (center). The historical outcome (GSKV, boxed, bottom) is accessible through a three-step path (EGKA – GGKA – GGKV – GSKV). Alternative SRE-specific outcomes accessible in three or fewer steps are in green. Lines connect genotypes separated by a single nonsynonymous nucleotide mutation; lines among genotypes in the outer ring are not shown for clarity. Orange arrows indicate paths of significantly increasing SRE mean fluorescence. **b**, For trajectories indicated by orange arrows in (**a**), SRE mean fluorescence is shown versus mutational distance from AncSR1+11P:EGKA (with x-axis jitter to avoid overplotting). Gray lines connect variants separated by single-nucleotide mutations. Error bars, 90% confidence intervals. Green dashed line, activity of AncSR1+11P:GSKV on SRE. **c**, For the SRE-specific outcomes accessed in orange paths in (**a**), the probability of each outcome under models where the probability of taking a step depends on the relative increase in SRE mean fluorescence (correlated fixation model), or where any SRE-enhancing step is equally likely (equal fixation model)(22, 24). **d**, The historical outcome (GSKV) has SRE-specific single-mutant neighbors. Concentric rings contain SRE-specific RH genotypes of path length 1 or 2 steps from AncSR1+11P:GSKV (center). Lines connect genotypes separated by a single nonsynonymous nucleotide mutation; lines among genotypes in the outer ring are not shown for clarity. **e**, The distribution of SRE mean fluorescence of SRE-specific neighbors of AncSR1+11P:GSKV illustrated in (**d**). Error bars, 90% confidence intervals.

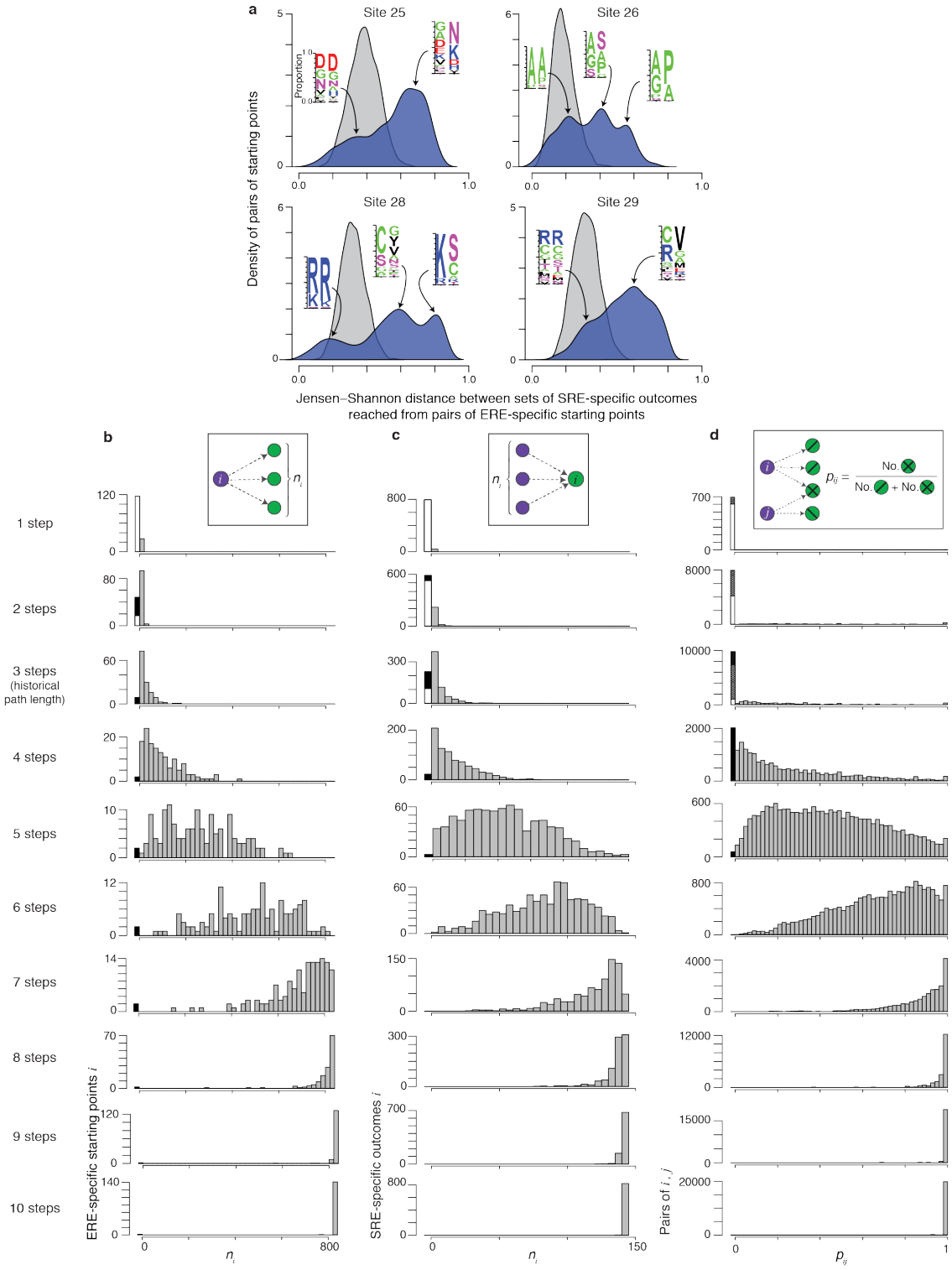


Figure A2.6. Evolvability of SRE specificity in an ancestral sequence space. a, Alternative ERE-specific starting points reach SRE-specific outcomes with very different amino acid states. For each starting point, the frequency profile of amino acid states at each RH site was determined for the set of SRE-specific outcomes reached in ≤ 3 steps; for each pair of starting

(**Figure A2.6, continued**) points, the Jensen-Shannon distance between profiles was calculated. Blue curve, distribution of pairs of starting points by Jensen-Shannon distances of the outcomes they reach; grey, distribution of Jensen-Shannon distances between profiles for randomly sampled sets of SRE-specific variants. In each modal peak, the amino acid frequency profiles for outcomes reached by a representative pair of ERE-specific starting points are shown. **b-d**, Contingency in the accessibility of individual SRE-specific outcomes remains when path lengths longer than the historical trajectory are considered. Plots are equivalent to Figs. 4.2b-d but for trajectories of increasing length.

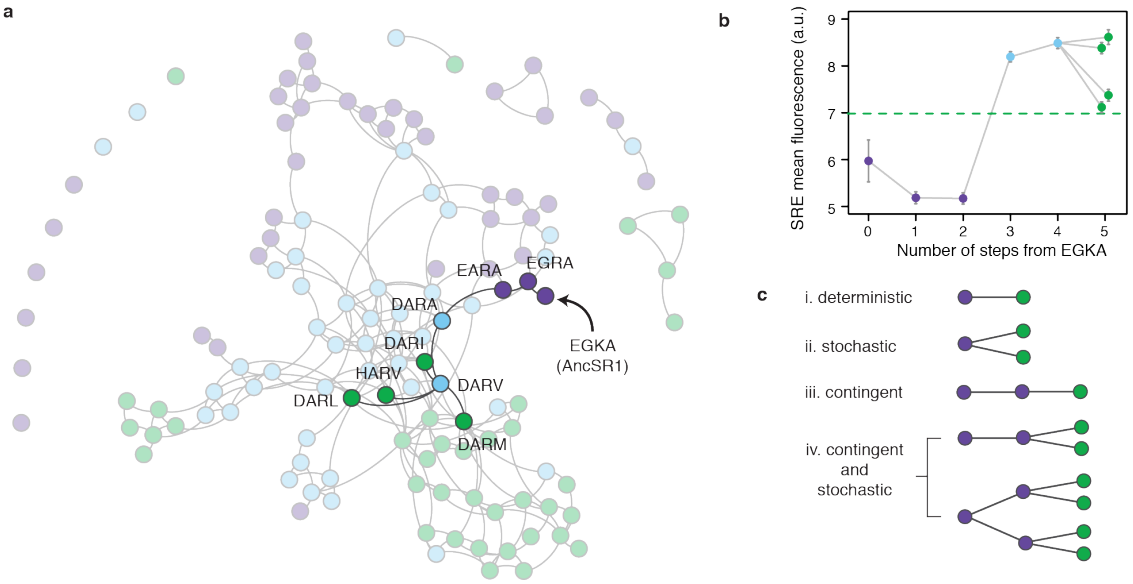


Figure A2.7. The historical starting point cannot access the derived function without permissive mutations. a, AncSR1 RH functional network layout as in Fig. 4.3c, with the shortest paths from AncSR1:EGKA to SRE specificity highlighted. The ancestral RH (EGKA) can access SRE specificity. However, all trajectories are at least 5 steps long, require permissive RH changes that confer no SRE activity (e.g. K28R and G26A) and proceed through promiscuous intermediates. **b**, For paths highlighted in (**a**), SRE mean fluorescence is shown versus mutational distance from AncSR1:EGKA; gray lines connect variants separated by single-nucleotide mutations. Error bars, 90% confidence intervals. Green dashed line, activity of AncSR1+11P:GSKV on SRE. EGKA was represented by only 7 cells in the SRE library, so its FACS-seq SRE mean fluorescence estimate is unreliable (and its classification was thus inferred by the predictive model). In isolated flow cytometry experiments, its SRE mean fluorescence was indistinguishable from null alleles; the decrease in SRE mean fluorescence from step 0 to step 1 suggested by this figure is therefore more likely a flat line (no change in SRE activity). **c**, Stochasticity and contingency in trajectories of functional change. Diagrams illustrate paths from a purple starting point (left) to possible green outcomes (right). In a deterministic trajectory (*i*), a particular genotype encoding the green function will evolve deterministically if selection favors acquisition of the green function and only that genotype is accessible. The outcome of evolution is stochastic (*ii*) if multiple outcomes are accessible, so which one occurs is random. An outcome is contingent (*iii*) if its accessibility depends on the prior occurrence of some step that cannot be driven by selection for that outcome. Contingency and stochasticity can occur independently (*ii* and *iii*), or they can co-occur in serial (*iv*).

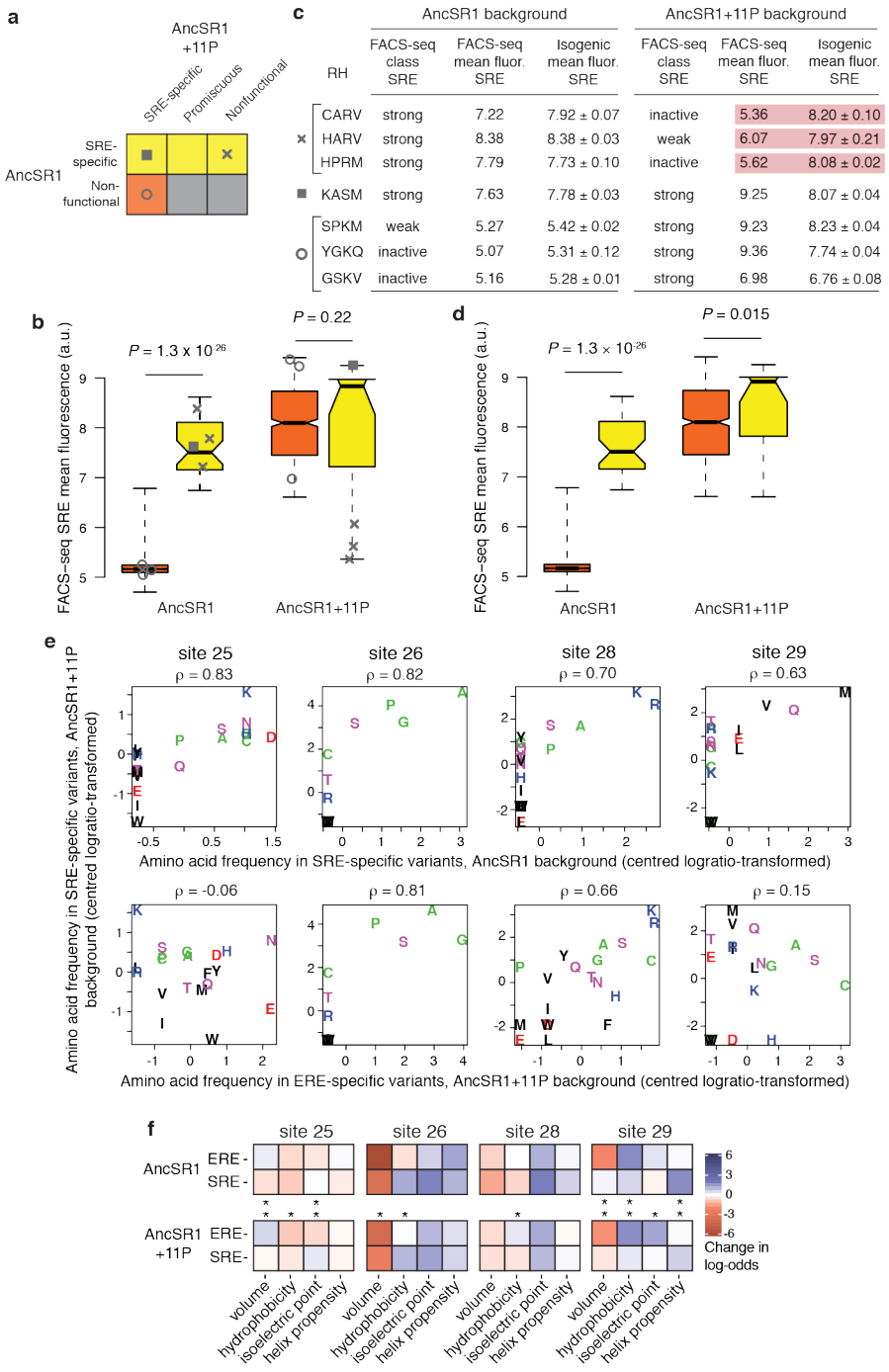


Figure A2.8. The effect of historical permissive substitutions is mediated by nonspecific increases in affinity. a-d, 11P nonspecifically increase transcriptional activity as measured by FACS-seq, consistent with FoldX predictions of effects on binding affinity. **a**, Classification of SRE-specific variants as 11P-dependent (orange) and 11P-independent (yellow) based on their functions in AncSR1 and AncSR1+11P backgrounds. Icons for individual variants specifically assessed in (**b**) and (**c**) are shown. **b**, FACS-seq mean fluorescence estimates for 11P-dependent (orange) and 11P-independent (yellow) RH variants in the AncSR1 (left) and AncSR1+11P (right) backgrounds, shown as box-and-whisker plots as in Fig. 4.4a. Icons represent variants

(**Figure A2.8, continued**) validated in (**c**). *P*-values, Wilcoxon rank sum test with continuity correction. The mean fluorescence of 11P-independent genotypes is significantly higher in the AncSR1 background but not in AncSR+11P. **c**, Validation of apparently restrictive effect of 11P on some genotypes. For three variants nonfunctional in AncSR1+11P but SRE-specific in AncSR1 FACS-seq assays (×), we measured mean fluorescence of isogenic cultures by flow cytometry. We also assayed variants that are SRE-specific in AncSR1+11P and SRE-specific (square) or nonfunctional (open circle) in AncSR1, as validation controls. Isogenic mean fluorescence is represented as mean ± SEM from three replicate transformations and inductions analyzed via flow cytometry. All FACS-seq classifications were validated except for the three apparently restricted variants in AncSR1+11P (highlighted in red), which are in fact strong SRE-activators in this background. Each of these variants was predicted to be a strong SRE-binder based on its genotype, but had an artificially low FACS-seq mean fluorescence estimate, perhaps due to a strong growth defect in inducing conditions. **d**, After removing the three genotypes with inaccurate FACS-seq fluorescence measurements (×), 11P-independent genotypes have significantly higher mean fluorescence than 11P-dependent genotypes in the AncSR1+11P background, consistent with a nonspecific permissive mechanism via affinity. *P*-values, Wilcoxon rank sum test with continuity correction. **e**, 11P do not alter the genetic determinants of SRE specificity. Each plot shows, for a variable site in the library, the frequency of every amino acid state in two functionally defined sets of variants. Spearman's rho for each correlation is shown. The top row shows that the determinants of SRE specificity are similar in AncSR1 and AncSR1+11P libraries; bottom row shows a much weaker relationship between the determinants of SRE and ERE specificity within the AncSR1+11P library. **f**, Biochemical determinants of ERE and SRE specificity in the AncSR1 (top) and AncSR1+11P (bottom) backgrounds. A multiple logistic regression model predicts the probability that a variant is RE-specific from the biochemical properties of its amino acid state at each of the four variable RH sites. The coefficients of this model represent the change in log-odds of being ERE-specific or SRE-specific per unit change in each property. Asterisks indicate site-specific determinants that differ significantly between ERE and SRE specificity in each background (*Z*-test, $P < 0.05$).

Table A2.1. Library sampling statistics. Sample sizes and sequence read/coverage statistics are shown at various stages of the experimental pipeline for each protein library, yeast reporter strain, and replicate. For details, see Chapter 4 Methods.

					FACS					sequencing					coverage, all variants		coverage, coding variants	
		bacterial transfor-mation yield (cfu)	yeast transfor-mation yield (cfu)	smallest bottleneck during FACS induction (cfu)	bin 1 count (cfu)	bin 2 count (cfu)	bin 3 count (cfu)	bin 4 count (cfu)	total number cells recovered post-sort (cfu)	bin 1 read count	bin 2 read count	bin 3 read count	bin 4 read count	read:cfu > 1 for all bins?	median number of cells sampled	fraction variants with >15 cells	median number of cells sampled	fraction variants with >15 cells
AncSR1 +11P +RH lib	ERE, rep1	2.3e7	6.1e6	3.2e8	2.0e7	3.4e7	3.2e6	4.7e5	5.8e7	2.8e7	3.6e7	3.3e6	2.0e7	yes	55.4	0.780	61.1	0.797
	ERE, rep 2		1.1e7	4.4e8	1.7e7	1.6e7	2.9e6	1.1e5	3.6e7	3.5e7	3.7e7	7.9e6	1.7e6	yes	64.0	0.830	70.5	0.843
	ERE, pooled		1.7e7	*Not applicable*					9.4e7	*Not applicable*					127.1	0.913	140.3	0.921
	SRE, rep 1		7.1e6	1.0e8	1.6e7	1.3e7	1.8e6	1.9e5	3.1e7	3.0e7	1.6e7	2.6e6	1.1e6	yes	70.7	0.811	78.2	0.826
	SRE, rep 2		1.8e7	2.0e8	2.0e7	3.2e7	4.9e6	4.1e5	5.7e7	2.3e7	6.0e7	1.1e7	3.8e6	yes	64.7	0.836	71.9	0.851
	SRE, pooled		2.5e7	*Not applicable*					8.8e7	*Not applicable*					143.6	0.924	158.9	0.931

AncSR1 +RH lib	ERE, rep1	2.3e7	8.3e6	3.4e8	2.0e7	3.2e7	5.3e6	2.2e5	5.8e7	2.4e7	5.1e7	6.4e6	5.5e5	yes	57.5	0.812	61.3	0.822
	ERE, rep 2		8.6e6	2.6e8	1.6e7	1.6e7	2.8e6	1.6e5	3.5e7	2.5e7	2.9e7	3.5e6	1.1e6	yes	37.1	0.734	39.6	0.748
	ERE, pooled		1.7e7	*Not applicable*					9.3e7	*Not applicable*					104.5	0.907	111.7	0.912
	SRE, rep 1		2.0e7	2.5e8	2.1e7	3.6e7	5.4e6	1.3e5	6.2e7	3.3e7	9.3e7	6.6e6	4.3e5	yes	178.4	0.958	191.3	0.961
	SRE, rep 2		2.1e7	2.9e8	2.0e7	3.1e7	5.5e6	5.9e5	5.7e7	3.1e7	5.5e7	2.0e7	1.6e6	yes	82.7	0.873	89.1	0.881
	SRE, pooled		4.1e7	*Not applicable*					1.2e8	*Not applicable*					289.8	0.979	312.1	0.980

Table A2.2. Robustness of inferences to scheme for classification of variants. Each row represents an inference reported in Figs. 4.2 and 4.3; each column is a scheme for functionally classifying variants from FACS-seq data and FACS-seq-trained predictive models. For details of schemes, see Chapter 4 Methods.

	Classification Scheme										
Inference	Main text	(A) Use FACS-seq ML estimate for AncSR1/ERE	(B) Increase equivalence margin from 20% to 50%	(C) Classify as functional if weak or strong activity	(D) Classify as functional if ML fluorescence >0.8× that of ancestral reference	(E) Classify as functional only if ML fluor within 20% on either side of ancestral reference	(F) Classify all variants based on predictions from genotype	(G) No predictions; classify undetermined variants as inactive	(H) Classify based on prediction or experiment, whichever assigns stronger function	(I) Keep only classifications identical between replicates	(J) Use per-variant estimate of standard error to classify
# ERE-specific, AncSR1	43	138	108	444	67	36	27	39	47	11	47
# promiscuous, AncSR1	45	94	84	158	58	38	45	44	60	30	46
# SRE-specific, AncSR1	41	41	58	213	45	19	31	38	40	39	40
# ERE-specific, AncSR1+11P	144	326	264	619	212	114	101	108	133	76	123
# promiscuous, AncSR1+11P	378	525	554	719	464	254	319	341	459	282	358
# SRE-specific, AncSR1+11P	829	832	1206	2728	956	296	670	768	899	809	837
AncSR1:EGKA requires permissives to access SRE-specificity?	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Shortest path length from EGKA to SRE-specificity in AncSR1	5	5	3	2	5	6	5	5	5	no paths	4
# SRE-specific outcomes accessed in 3 steps from AncSR1+11P:EGKA	65	66	89	136	77	10	58	53	72	71	65
Proportion ERE-specific starting points unable to access SRE-specificity in 3 steps, AncSR1+11P	0.063	0.037	0.008	0.066	0.014	0.252	0.050	0.139	0.053	0.026	0.089
Proportion SRE-specific outcomes not accessed from any ERE-specific starting point in 3 steps, AncSR1+11P	0.276	0.108	0.118	0.071	0.150	0.571	0.378	0.388	0.276	0.425	0.280
Proportion pairs of ERE-specific starting points with no shared outcomes in 3 steps, AncSR1+11P	0.542	0.530	0.426	0.229	0.501	0.836	0.543	0.611	0.529	0.390	0.541
Fraction ERE-specific variants with no path to SRE-specificity, AncSR1	0.279	0.058	0.505	0.054	0.176	0.378	0.321	0.350	0.250	0.083	0.104
Fraction ERE-specific variants with no path to SRE-specificity, AncSR1+11P	0.014	0.021	0.004	0.066	0.005	0.470	0.010	0.056	0.015	0	0.033
Average shortest path length to SRE-specificity from all connected ERE-specific variants, AncSR1	4.193	4.191	3.796	2.309	4.054	4.304	4.158	4.889	4.278	4.545	4.163
Average shortest path length to SRE-specificity from all connected ERE-specific variants, AncSR1+11P	2.183	2.122	1.867	1.336	1.986	2.885	2.270	2.333	2.206	2.158	2.294
Fraction ERE-specific variants with permissive shortest path, AncSR1	0	0.035	0.059	0.242	0	0	0	0	0	0	0
Fraction ERE-specific variants with permissive shortest path, AncSR1+11P	0.290	0.218	0.225	0.191	0.235	0.136	0.207	0.234	0.210	0.140	0.214
Fraction ERE-specific variants with promiscuous shortest path, AncSR1	0.483	0.461	0.381	0.370	0.381	0.445	0.548	0.361	0.594	0.634	0.524
Fraction ERE-specific variants with promiscuous shortest path, AncSR1+11P	0.413	0.462	0.403	0.133	0.441	0.458	0.504	0.426	0.538	0.524	0.475

(**Table A2.2, continued**)

Fraction ERE-specific variants with permissive and promiscuous shortest path, AncSR1	0.517	0.481	0.530	0.191	0.619	0.555	0.452	0.639	0.406	0.366	0.476
Fraction ERE-specific variants with permissive and promiscuous shortest path, AncSR1+11P	0.108	0.120	0.065	0.002	0.082	0.241	0.149	0.164	0.106	0.165	0.135
Fraction ERE-specific variants with direct shortest path AncSR1	0	0.023	0.030	0.198	0	0	0	0	0	0	0
Fraction ERE-specific variants direct shortest path, AncSR1+11P	0.190	0.201	0.308	0.673	0.242	0.165	0.14	0.176	0.145	0.171	0.176

Bibliography

1. Maynard Smith J (1970) Natural selection and the concept of a protein space. *Nature* 225(5232):563–564.

2. Arnold FH (2011) The Library of Maynard-Smith: my search for meaning in the protein universe. *Microbe* 6(7):316.

3. Hopf TA, et al. (2017) Mutation effects predicted from sequence co-variation. *Nat Biotechnol* 35(2):128–135.

4. Whitehead TA, et al. (2012) Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nat Biotechnol* 30(6):543–548.

5. Arnold FH (2018) Directed evolution: bringing new clife. *Angew Chem Int Ed Engl* 57(16):4143–4148.

6. Lässig M, Mustonen V, Walczak AM (2017) Predicting evolution. *Nat Ecol Evol* 1(3):77.

7. Harms MJ, Thornton JW (2013) Evolutionary biochemistry: revealing the historical and physical causes of protein properties. *Nat Rev Genet* 14(8):559–571.

8. Huang P-S, Boyken SE, Baker D (2016) The coming of age of de novo protein design. *Nature* 537(7620):320–327.

9. Poelwijk FJ, Kiviet DJ, Weinreich DM, Tans SJ (2007) Empirical fitness landscapes reveal accessible evolutionary paths. *Nature* 445(7126):383–386.

10. Fowler DM, et al. (2010) High-resolution mapping of protein sequence-function relationships. *Nat Methods* 7(9):741–746.

11. Fowler DM, Fields S (2014) Deep mutational scanning: a new style of protein science. *Nat Methods* 11(8):801–807.

12. Araya CL, Fowler DM (2011) Deep mutational scanning: assessing protein function on a massive scale. *Trends Biotechnol* 29(9):435–442.

13. Thyagarajan B, Bloom JD (2014) The inherent mutational tolerance and antigenic evolvability of influenza hemagglutinin. *eLife* 3. doi:10.7554/eLife.03300.

14. Bloom JD (2014) An experimentally determined evolutionary model dramatically improves phylogenetic fit. *Mol Biol Evol* 31(8):1956–1978.

15. Roscoe BP, Thayer KM, Zeldovich KB, Fushman D, Bolon DNA (2013) Analyses of the effects of all ubiquitin point mutants on yeast growth rate. *J Mol Biol* 425(8):1363–1377.

16. Mishra P, Flynn JM, Starr TN, Bolon DNA (2016) Systematic mutant analyses elucidate general and client-specific aspects of Hsp90 function. *Cell Rep* 15(3):588–598.

17. Firnberg E, Labonte JW, Gray JJ, Ostermeier M (2014) A comprehensive, high-resolution map of a gene's fitness landscape. *Mol Biol Evol* 31(6):1581–1592.

18. Doud MB, Ashenberg O, Bloom JD (2015) Site-specific amino acid preferences are mostly conserved in two closely related protein homologs. *Mol Biol Evol* 32(11):2944–2960.

19. Haddox HK, Dingens AS, Hilton SK, Overbaugh J, Bloom JD (2018) Mapping mutational effects along the evolutionary landscape of HIV envelope. *eLife* 7:124.

20. Lee JM, et al. (2018) Deep mutational scanning of hemagglutinin helps predict evolutionary fates of human H3N2 influenza variants. *bioRxiv*:1–16.

21. Olson CA, Wu NC, Sun R (2014) A comprehensive biophysical description of pairwise epistasis throughout an entire protein domain. *Curr Biol* 24(22):2643–2651.

22. Podgornaia AI, Laub MT (2015) Pervasive degeneracy and epistasis in a protein-protein interface. *Science* 347(6222):673–677.

23. Aakre CD, et al. (2015) Evolving new protein-protein interaction specificity through promiscuous intermediates. *Cell* 163(3):594–606.

24. Wu NC, Dai L, Olson CA, Lloyd-Smith JO, Sun R (2016) Adaptation in protein fitness landscapes is facilitated by indirect paths. *eLife* 5:e16965.

25. Poelwijk FJ, Socolich M, Ranganathan R (2017) Learning the pattern of epistasis linking genotype and phenotype in a protein. *bioRxiv*:1–31.

26. Domingo J, Diss G, Ben Lehner (2018) Pairwise and higher-order genetic interactions during the evolution of a tRNA. *Nature*:1–26.

27. Kondrashov DA, Kondrashov FA (2015) Topological features of rugged fitness landscapes in sequence space. *Trends Genet* 31(1):24–33.

28. Povolotskaya IS, Kondrashov FA (2010) Sequence space and the ongoing expansion of the protein universe. *Nature* 465(7300):922–926.

29. Pollock DD, Thiltgen G, Goldstein RA (2012) Amino acid coevolution induces an evolutionary Stokes shift. *P Natl Acad Sci USA* 109(21):E1352–9.

30. Shah P, McCandlish DM, Plotkin JB (2015) Contingency and entrenchment in protein evolution under purifying selection. *P Natl Acad Sci USA* 112(25):E3226–E3235.

31. Breen MS, Kemena C, Vlasov PK, Notredame C, Kondrashov FA (2012) Epistasis as the primary factor in molecular evolution. *Nature* 490(7421):535–538.

32. Usmanova DR, Ferretti L, Povolotskaya IS, Vlasov PK, Kondrashov FA (2015) A model of substitution trajectories in sequence space and long-term protein evolution. *Mol Biol Evol* 32(2):542–554.

33. Lunzer M, Golding GB, Dean AM (2010) Pervasive cryptic epistasis in molecular evolution. *PLOS Genet* 6(10):e1001162.

34. Gong LI, Suchard MA, Bloom JD (2013) Stability-mediated epistasis constrains the evolution of an influenza protein. *eLife* 2:e00631.

35. Risso VA, et al. (2015) Mutational studies on resurrected ancestral proteins reveal conservation of site-specific amino acid preferences throughout evolutionary history. *Mol Biol Evol* 32(2):440–455.

36. Gould SJ (1990) *Wonderful Life: The Burgess Shale and the Nature of History* (W. W. Norton & Company).

37. Blount ZD, Borland CZ, Lenski RE (2008) Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli. *P Natl Acad Sci USA* 105(23):7899–7906.

38. Harms MJ, Thornton JW (2014) Historical contingency and its biophysical basis in glucocorticoid receptor evolution. *Nature* 512:203–207.

39. de Visser JAGM, Krug J (2014) Empirical fitness landscapes and the predictability of evolution. *Nat Rev Genet* 15(7):480–490.

40. Salverda MLM, et al. (2011) Initial mutations direct alternative pathways of protein evolution. *PLOS Genet* 7(3):e1001321.

41. Kaltenbach M, Jackson CJ, Campbell EC, Hollfelder F, Tokuriki N (2015) Reverse evolution leads to genotypic incompatibility despite functional and active site convergence. *eLife* 4:e06492.

42. Natarajan C, et al. (2016) Predictable convergence in hemoglobin function has unpredictable molecular underpinnings. *Science* 354(6310):336–339.

43. Dickinson BC, Leconte AM, Allen B, Esvelt KM, Liu DR (2013) Experimental interrogation of the path dependence and stochasticity of protein evolution using phage-assisted continuous evolution. *P Natl Acad Sci USA* 110(22):9007–9012.

44. Woods R, Schneider D, Winkworth CL, Riley MA, Lenski RE (2006) Tests of parallel molecular evolution in a long-term experiment with Escherichia coli. *P Natl Acad Sci USA* 103(24):9107–9112.

45. Kryazhimskiy S, Rice DP, Jerison ER, Desai MM (2014) Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science* 344(6191):1519–1522.

46. Wagner A (2011) Genotype networks shed light on evolutionary constraints. *Trends Ecol Evol* 26(11):580–587.

47. Alexander PA, He Y, Chen Y, Orban J, Bryan PN (2009) A minimal sequence code for switching protein structure and function. *P Natl Acad Sci USA* 106(50):21149–21154.

48. Anderson DP, Whitney DS, Hanson-Smith V (2016) Evolution of an ancient protein function involved in organized multicellularity in animals. *eLife*. doi:10.7554/eLife.10147.001.

49. Bloom JD, Labthavikul ST, Otey CR, Arnold FH (2006) Protein stability promotes evolvability. *P Natl Acad Sci USA* 103(15):5869–5874.

50. Bershtein S, Segal M, Bekerman R, Tokuriki N, Tawfik DS (2006) Robustness–epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature* 444(7121):929–932.

51. Jacquier H, et al. (2013) Capturing the mutational landscape of the beta-lactamase TEM-1. *P Natl Acad Sci USA* 110(32):13067–13072.

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

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

54. Phillips PC (2008) Epistasis--the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet* 9(11):855–867.

55. Doud MB, Ashenberg O, Bloom JD (2015) Site-specific amino acid preferences are mostly conserved in two closely related protein homologs. *Mol Biol Evol* 32(11):2944–2960.

56. Ashenberg O, Gong LI, Bloom JD (2013) Mutational effects on stability are largely conserved during protein evolution. *P Natl Acad Sci USA* 110(52):21071–21076.

57. Ortlund EA, Bridgham JT, Redinbo MR, Thornton JW (2007) Crystal structure of an ancient protein: evolution by conformational epistasis. *Science* 317(5844):1544–1548.

58. Lynch VJ, May G, Wagner GP (2011) Regulatory evolution through divergence of a phosphoswitch in the transcription factor CEBPB. *Nature* 480(7377):383–386.

59. Natarajan C, et al. (2013) Epistasis among adaptive mutations in deer mouse hemoglobin. *Science* 340(6138):1324–1327.

60. Peisajovich SG, Tawfik DS (2007) Protein engineers turned evolutionists. *Nat Methods* 4(12):991–994.

61. Wagner A (2008) Neutralism and selectionism: a network-based reconciliation. *Nat Rev Genet* 9(12):965–974.

62. Romero PA, Arnold FH (2009) Exploring protein fitness landscapes by directed evolution. *Nat Rev Mol Cell Biol* 10(12):866–876.

63. Gumulya Y, Reetz MT (2011) Enhancing the thermal robustness of an enzyme by directed evolution: least favorable starting points and inferior mutants can map superior evolutionary pathways. *Chembiochem* 12(16):2502–2510.

64. Bridgham JT, Carroll SM, Thornton JW (2006) Evolution of hormone-receptor complexity by molecular exploitation. *Science* 312(5770):97–101.

65. Bloom JD, Romero PA, Lu Z, Arnold FH (2007) Neutral genetic drift can alter promiscuous protein functions, potentially aiding functional evolution. *Biol Direct* 2:17.

66. Fasan R, Meharenna YT, Snow CD, Poulos TL, Arnold FH (2008) Evolutionary history of a specialized P450 propane monooxygenase. *J Mol Biol* 383:1069–1080.

67. Bershtein S, Goldin K, Tawfik DS (2008) Intense neutral drifts yield robust and evolvable consensus proteins. *J Mol Biol* 379(5):1029–1044.

68. Field SF, Matz MV (2010) Retracing evolution of red fluorescence in GFP-like proteins from Faviina corals. *Mol Biol Evol* 27(2):225–233.

69. Bloom JD, Gong LI, Baltimore D (2010) Permissive secondary mutations enable the evolution of influenza oseltamivir resistance. *Science* 328(5983):1272–1275.

70. McKeown AN, et al. (2014) Evolution of DNA specificity in a transcription factor family produced a new gene regulatory module. *Cell* 159(1):58–68.

71. Melero C, Ollikainen N, Harwood I, Karpiak J, Kortemme T (2014) Quantification of the transferability of a designed protein specificity switch reveals extensive epistasis in molecular recognition. *P Natl Acad Sci USA* 111(43):15426–15431.

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

73. Araya CL, et al. (2012) A fundamental protein property, thermodynamic stability, revealed solely from large-scale measurements of protein function. *P Natl Acad Sci*

USA 109(42):16858–16863.

74. Melamed D, Young DL, Gamble CE, Miller CR, Fields S (2013) Deep mutational scanning of an RRM domain of the Saccharomyces cerevisiae poly(A)-binding protein. *RNA* 19(11):1537–1551.

75. Bank C, Hietpas RT, Jensen JD, Bolon DNA (2015) A systematic survey of an intragenic epistatic landscape. *Mol Biol Evol* 32(1):229–238.

76. O'Maille PE, et al. (2008) Quantitative exploration of the catalytic landscape separating divergent plant sesquiterpene synthases. *Nat Chem Biol* 4(10):617–623.

77. Kondrashov AS, Sunyaev S, Kondrashov FA (2002) Dobzhansky-Muller incompatibilities in protein evolution. *P Natl Acad Sci USA* 99(23):14878–14883.

78. Kulathinal RJ, Bettencourt BR, Hartl DL (2004) Compensated deleterious mutations in insect genomes. *Science* 306(5701):1553–1554.

79. Soylemez O, Kondrashov FA (2012) Estimating the rate of irreversibility in protein evolution. *Genome Biol Evol* 4(12):1213–1222.

80. Xu J, Zhang J (2014) Why human disease-associated residues appear as the wild-type in other species: genome-scale structural evidence for the compensation hypothesis. *Mol Biol Evol* 31(7):1787–1792.

81. Ferrer-Costa C, Orozco M, Cruz X de L (2007) Characterization of compensated mutations in terms of structural and physico-chemical properties. *J Mol Biol* 365(1):249–256.

82. Barešić A, Hopcroft LEM, Rogers HH, Hurst JM, Martin ACR (2010) Compensated pathogenic deviations: analysis of structural effects. *J Mol Biol* 396(1):19–30.

83. Jordan DM, et al. (2015) Identification of cis-suppression of human disease mutations by comparative genomics. *Nature* 524(7564):225–229.

84. Naumenko SA, Kondrashov AS, Bazykin GA (2012) Fitness conferred by replaced amino acids declines with time. *Biol Letters* 8(5):825–828.

85. Zou Z, Zhang J (2015) Are convergent and parallel amino acid substitutions in protein evolution more prevalent than neutral expectations? *Mol Biol Evol* 32(8):2085–2096.

86. Goldstein RA, Pollard ST, Shah SD, Pollock DD (2015) Nonadaptive amino acid convergence rates decrease over time. *Mol Biol Evol* 32(6):1373–1381.

87. McCandlish DM, Rajon E, Shah P, Ding Y, Plotkin JB (2013) The role of epistasis in protein evolution. *Nature* 497(7451):E1–2– discussion E2–3.

88. Socolich M, et al. (2005) Evolutionary information for specifying a protein fold. *Nature* 437(7058):512–518.

89. Russ WP, Lowery DM, Mishra P, Yaffe MB, Ranganathan R (2005) Natural-like function in artificial WW domains. *Nature* 437(7058):579–583.

90. Marks DS, et al. (2011) Protein 3D structure computed from evolutionary sequence variation. *PLOS ONE* 6(12):e28766.

91. Morcos F, et al. (2011) Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *P Natl Acad Sci USA* 108(49):E1293–E1301.

92. Hopf TA, et al. (2012) Three-dimensional structures of membrane proteins from genomic sequencing. *Cell* 149(7):1607–1621.

93. Ovchinnikov S, et al. (2015) Large scale determination of previously unsolved protein structures using evolutionary information. *eLife* 4:e09248.

94. Skerker JM, et al. (2008) Rewiring the specificity of two-component signal transduction systems. *Cell* 133(6):1043–1054.

95. Ovchinnikov S, Kamisetty H, Baker D (2014) Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *eLife* 3:e02030.

96. Hopf TA, et al. (2014) Sequence co-evolution gives 3D contacts and structures of protein complexes. *eLife* 3:e03430.

97. Figliuzzi M, Jacquier H, Schug A, Tenaillon O, Weigt M (2015) Coevolutionary landscape inference and the context-dependence of mutations in beta-lactamase TEM-1. *Mol Biol Evol*. doi:10.1093/molbev/msv211.

98. Hopf TA, et al. (2015) Quantification of the effect of mutations using a global probability model of natural sequence variation. *arXiv*.

99. Süel GM, Lockless SW, Wall MA, Ranganathan R (2002) Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat Struct Biol* 10(1):59–69.

100. Morcos F, Jana B, Hwa T, Onuchic JN (2013) Coevolutionary signals across protein lineages help capture multiple protein conformations. *P Natl Acad Sci USA* 110(51):20533–20538.

101. Sutto L, Marsili S, Valencia A, Gervasio FL (2015) From residue coevolution to protein conformational ensembles and functional dynamics. *P Natl Acad Sci USA* 112(44):13567–13572.

102. Draghi JA, Plotkin JB (2013) Selection biases the prevalence and type of epistasis along adaptive trajectories. *Evolution* 67(11):3120–3131.

103. Potapov V, Cohen M, Schreiber G (2009) Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details. *Protein Eng Des Sel* 22(9):553–560.

104. Halabi N, Rivoire O, Leibler S, Ranganathan R (2009) Protein sectors: evolutionary units of three-dimensional structure. *Cell* 138(4):774–786.

105. Yokoyama S, et al. (2014) Epistatic adaptive evolution of human color vision. *PLOS Genet* 10(12):e1004884.

106. 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.

107. Bridgham JT, Ortlund EA, Thornton JW (2009) An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature* 461(7263):515–519.

108. Dellus-Gur E, et al. (2015) Negative epistasis and evolvability in TEM-1 β-lactamase—the thin line between an enzyme's conformational freedom and disorder. *J Mol Biol* 427(14):2396–2409.

109. Huang W, Palzkill T (1997) A natural polymorphism in beta-lactamase is a global suppressor. *P Natl Acad Sci USA* 94(16):8801–8806.

110. Sideraki V, Huang W, Palzkill T, Gilbert HF (2001) A secondary drug resistance mutation of TEM-1 beta-lactamase that suppresses misfolding and aggregation. *P Natl Acad Sci USA* 98(1):283–288.

111. Bloom JD, et al. (2005) Thermodynamic prediction of protein neutrality. *P Natl Acad Sci USA* 102(3):606–611.

112. Tokuriki N, Tawfik DS (2009) Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature* 459(7247):668–673.

113. Lunzer M, Miller SP, Felsheim R, Dean AM (2005) The biochemical architecture of an ancient adaptive landscape. *Science* 310(5747):499–501.

114. Tokuriki N, Tawfik DS (2009) Stability effects of mutations and protein evolvability. *Curr Opin Struc Biol* 19(5):596–604.

115. Brown NG, Pennington JM, Huang W, Ayvaz T, Palzkill T (2010) Multiple global suppressors of protein stability defects facilitate the evolution of extended-spectrum TEM β-lactamases. *J Mol Biol* 404(5):832–846.

116. Bloom JD, Nayak JS, Baltimore D (2011) A computational-experimental approach identifies mutations that enhance surface expression of an oseltamivir-resistant influenza neuraminidase. *PLOS ONE* 6(7):e22201.

117. 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(1):e1004058.

118. Weinreich DM, Lan Y, Wylie CS, Heckendorn RB (2013) Should evolutionary geneticists worry about higher-order epistasis? *Curr Opin Genet Dev* 23(6):700–707.

119. Hietpas RT, Jensen JD, Bolon DNA (2011) Experimental illumination of a fitness landscape. *P Natl Acad Sci USA* 108(19):7896–7901.

120. Poelwijk FJ, Krishna V, Ranganathan R (2015) The context-dependence of mutations: a linkage of formalisms. *arXiv*.

121. Sorrells TR, Booth LN, Tuch BB, Johnson AD (2015) Intersecting transcription networks constrain gene regulatory evolution. *Nature* (523):361–365.

122. McLaughlin RN, Poelwijk FJ, Raman A, Gosal WS, Ranganathan R (2012) The spatial architecture of protein function and adaptation. *Nature* 491(7422):138–142.

123. Kachroo AH, et al. (2015) Evolution. Systematic humanization of yeast genes reveals conserved functions and genetic modularity. *Science* 348(6237):921–925.

124. Sarkisyan KS, et al. (2016) Local fitness landscape of the green fluorescent protein. *Nature* 533(7603):397–401.

125. Mendes FK, Hahn Y, Hahn MW (2016) Gene tree discordance can generate patterns of diminishing convergence over time. *Mol Biol Evol* 33(12):3299–3307.

126. Hochberg GKA, Thornton JW (2017) Reconstructing ancient proteins to understand the causes of structure and function. *Annu Rev Biophys* 46(1):247–269.

127. Piper PW, et al. (2003) Yeast is selectively hypersensitised to heat shock protein 90 (Hsp90)-targetting drugs with heterologous expression of the human Hsp90β, a property that can be exploited in screens for new Hsp90 chaperone inhibitors. *Gene* 302(1-2):165–170.

128. Wider D, Péli-Gulli M-P, Briand P-A, Tatu U, Picard D (2009) The complementation of yeast with human or Plasmodium falciparum Hsp90 confers differential inhibitor sensitivities. *Mol Biochem Parasit* 164(2):147–152.

129. Adl SM, et al. (2012) The revised classification of eukaryotes. *J Eukaryot Microbiol* 59(5):429–514.

130. Jiang L, Mishra P, Hietpas RT, Zeldovich KB, Bolon DNA (2013) Latent effects of Hsp90 mutants revealed at reduced expression levels. *PLOS Genet* 9(6):e1003600.

131. McCandlish DM, Shah P, Plotkin JB (2016) Epistasis and the dynamics of reversion in molecular evolution. *Genetics* 203(3):1335–1351.

132. Taylor JW, Berbee ML (2006) Dating divergences in the Fungal Tree of Life: review and new analyses. *Mycologia* 98(6):838–849.

133. Eme L, Sharpe SC, Brown MW, Roger AJ (2014) On the age of eukaryotes: evaluating evidence from fossils and molecular clocks. *CSH Perspect Biol* 6(8):a016139–a016139.

134. Ali MMU, et al. (2006) Crystal structure of an Hsp90–nucleotide–p23/Sba1 closed chaperone complex. *Nature* 440(7087):1013–1017.

135. Starr TN, Thornton JW (2016) Epistasis in protein evolution. *Protein Sci* 25(7):1204–1218.

136. Sailer ZR, Harms MJ (2017) Detecting high-order epistasis in nonlinear genotype-phenotype maps. *Genetics* 205(3):1079–1088.

137. Tsai IJ, Bensasson D, Burt A, Koufopanou V (2008) Population genomics of the wild yeast Saccharomyces paradoxus: Quantifying the life cycle. *P Natl Acad Sci USA* 105(12):4957–4962.

138. Peris D, et al. (2014) Population structure and reticulate evolution of Saccharomyces eubayanus and its lager-brewing hybrids. *Mol Ecol* 23(8):2031–2045.

139. Almeida P, et al. (2015) A population genomics insight into the Mediterranean origins of wine yeast domestication. *Mol Ecol* 24(21):5412–5427.

140. Pantzartzi CN, Drosopoulou E, Scouras ZG (2013) Assessment and reconstruction of novel HSP90 genes: duplications, gains and losses in fungal and animal lineages. *PLOS ONE* 8(9):e73217.

141. Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22(13):1658–1659.

142. Sievers F, et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:1–6.

143. Le SQ, Gascuel O (2008) An improved general amino acid replacement matrix. *Mol Biol Evol* 25(7):1307–1320.

144. Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.

145. Brown MW, et al. (2013) Phylogenomics demonstrates that breviate flagellates are related to opisthokonts and apusomonads. *P R Soc B* 280(1769):20131755–20131755.

146. Brown MW, Spiegel FW, Silberman JD (2009) Phylogeny of the "forgotten" cellular slime mold, Fonticula alba, reveals a key evolutionary branch within Opisthokonta. *Mol Biol Evol* 26(12):2699–2709.

147. Paps J, Medina-Chacón LA, Marshall W, Suga H, Ruiz-Trillo I (2013) Molecular phylogeny of Unikonts: new insights into the position of Apusomonads and Ancyromonads and the internal relationships of Opisthokonts. *Protist* 164(1):2–12.

148. Kurtzman CP, Robnett CJ (2013) Relationships among genera of the Saccharomycotina(Ascomycota) from multigene phylogenetic analysis of type species. *FEMS Yeast Res* 13(1):23–33.

149. Shen X-X, et al. (2016) Reconstructing the Backbone of the Saccharomycotina Yeast Phylogeny Using Genome-Scale Data. *G3* 6(12):3927–3939.

150. Yang Z, Kumar S, Nei M (1995) A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics* 141(4):1641–1650.

151. Cunningham CN, Southworth DR, Krukenberg KA, Agard DA (2012) The conserved arginine 380 of Hsp90 is not a catalytic residue, but stabilizes the closed conformation required for ATP hydrolysis. *Protein Sci* 21(8):1162–1171.

152. Hiatt JB, Patwardhan RP, Turner EH, Lee C, Shendure J (2010) Parallel, tag-directed assembly of locally derived short sequence reads. *Nat Methods* 7(2):119–122.

153. Hietpas RT, Bank C, Jensen JD, Bolon DNA (2013) Shifting fitness landscapes in response to altered environments. *Evolution* 67(12):3512–3522.

154. Hietpas R, Roscoe B, Jiang L, Bolon DNA (2012) Fitness analyses of all possible point mutations for regions of genes in yeast. *Nat Protoc* 7(7):1382–1396.

155. Chevin LM (2011) On measuring selection in experimental evolution. *Biol Letters* 7(2):210–213.

156. Sella G, Hirsh AE (2005) The application of statistical physics to evolutionary biology. *P Natl Acad Sci USA* 102(27):9541–9546.

157. Benaglia T, Chauveau D, Hunter D (2009) mixtools: An R package for analyzing finite mixture models. *J Stat Softw* 32(6).

158. Monod J (1972) *Chance and Necessity: An Essay on the Natural Philosophy of Biology* (Vintage Books, New York).

159. Carroll JS, et al. (2006) Genome-wide analysis of estrogen receptor binding sites. *Nat Genet* 38(11):1289–1297.

160. Watson LC, et al. (2013) The glucocorticoid receptor dimer interface allosterically transmits sequence-specific DNA signals. *Nat Struct Mol Biol* 20(7):876–883.

161. Luisi BF, et al. (1991) Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature* 352(6335):497–505.

162. Schwabe JW, Chapman L, Finch JT, Rhodes D (1993) The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements. *Cell* 75(3):567–578.

163. Zilliacus J, Carlstedt-Duke J, Gustafsson JA, Wright AP (1994) Evolution of distinct DNA-binding specificities within the nuclear receptor family of transcription factors. *P Natl Acad Sci USA* 91(10):4175–4179.

164. Bain DL, et al. (2012) Glucocorticoid receptor-DNA interactions: binding energetics are the primary determinant of sequence-specific transcriptional activity. *J Mol Biol* 422(1):18–32.

165. Eick GN, Bridgham JT, Anderson DP, Harms MJ, Thornton JW (2017) Robustness of reconstructed ancestral protein functions to statistical uncertainty. *Mol Biol Evol* 34(2):247–261.

166. Lynch M, Hagner K (2015) Evolutionary meandering of intermolecular interactions along the drift barrier. *P Natl Acad Sci USA* 112(1):E30–E38.

167. Fox JE, Bridgham JT, Bovee TFH, Thornton JW (2007) An evolvable oestrogen receptor activity sensor: development of a modular system for integrating multiple genes into the yeast genome. *Yeast* 24(5):379–390.

168. Mumberg D, Müller R, Funk M (1995) Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene* 156(1):119–122.

169. Gietz RD, Woods RA (2002) Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol* 350:87–96.

170. Muggeo V (2008) Segmented: an R package to fit regression models with broken-line relationships. *R news*.

171. Sluder AE, Mathews SW, Hough D, Yin VP, Maina CV (1999) The nuclear receptor superfamily has undergone extensive proliferation and diversification in nematodes. *Genome Res* 9(2):103–120.

172. Benatuil L, Perez JM, Belk J, Hsieh CM (2010) An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Eng Des*

Sel 23(4):155–159.

173. Scanlon TC, Gray EC, Griswold KE (2009) Quantifying and resolving multiple vector transformants in S. cerevisiae plasmid libraries. *BMC Biotechnol* 9(1):95.

174. Fowler DM, Stephany JJ, Fields S (2014) Measuring the activity of protein variants on a large scale using deep mutational scanning. *Nat Protoc* 9(9):2267–2284.

175. Mir K, Neuhaus K, Bossert M, Schober S (2013) Short barcodes for next generation sequencing. *PLOS ONE*. doi:10.1371/journal.pone.0082933.g001.

176. Peterman N, Levine E (2016) Sort-seq under the hood: implications of design choices on large-scale characterization of sequence-function relations. *BMC Genomics* 17(1):206.

177. Delignette-Muller ML, Dutang C (2015) *fitdistrplus: An R package for fitting distributions* (Journal of Statistical Software).

178. Archer KJ, Williams AAA (2012) L1 penalized continuation ratio models for ordinal response prediction using high-dimensional datasets. *Stat Med* 31(14):1464–1474.

179. Bastian M, Heymann S, Jacomy M (2009) Gephi: an open source software for exploring and manipulating networks. *ICWSM*.

180. Sailer ZR, Harms MJ (2017) Detecting high-order epistasis in nonlinear genotype-phenotype maps. *Genetics*:genetics.116.195214–34.

181. Schymkowitz J, et al. (2005) The FoldX web server: an online force field. *Nucleic Acids Res* 33(Web Server):W382–W388.

182. Luscombe NM, Laskowski RA, Thornton JM (1997) NUCPLOT: a program to generate schematic diagrams of protein-nucleic acid interactions. *Nucleic Acids Res* 25(24):4940–4945.

183. Schymkowitz JWH, et al. (2005) Prediction of water and metal binding sites and their affinities by using the Fold-X force field. *P Natl Acad Sci USA* 102(29):10147–10152.

184. Crooks GE, Hon G, Chandonia J-M, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* 14(6):1188–1190.

185. Abriata LA, Palzkill T, Dal Peraro M (2015) How structural and physicochemical determinants shape sequence constraints in a functional enzyme. *PLOS ONE* 10(2):e0118684.

186. Weinreich DM, Watson RA, Chao L (2005) Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* 59(6):1165–1174.

187. Tufts DM, et al. (2015) Epistasis constrains mutational pathways of hemoglobin adaptation in high-altitude pikas. *Mol Biol Evol* 32(2):287–298.

188. Reetz MT (2013) The Importance of Additive and Non-Additive Mutational Effects in Protein Engineering. *Angew Chem Int Ed* 52(10):2658–2666.

189. Adams RM, Kinney JB, Walczak AM, Mora T (2017) Physical epistatic landscape of antibody binding affinity. *arXiv* q-bio.PE.

190. Posfai A, Zhou J, Plotkin JB, Kinney JB, McCandlish DM Selection for protein stability enriches for epistatic interactions. *biorxiv*.

191. Kauffman S, Levin S (1987) Towards a general theory of adaptive walks on rugged landscapes. *J Theor Biol* 128(1):11–45.

192. Wagner A (2008) Robustness and evolvability: a paradox resolved. *P R Soc B* 275(1630):91–100.

193. Wagner A (2012) The role of robustness in phenotypic adaptation and innovation. *P R Soc B* 279(1732):1249–1258.

194. Draghi JA, Parsons TL, Wagner GP, Plotkin JB (2010) Mutational robustness can facilitate adaptation. *Nature* 463(7279):353–355.

195. Payne JL, Wagner A (2014) The robustness and evolvability of transcription factor binding sites. *Science* 343(6173):875–877.

196. Hayden EJ, Ferrada E, Wagner A (2011) Cryptic genetic variation promotes rapid evolutionary adaptation in an RNA enzyme. *Nature* 474(7349):92–95.

197. Aguilar-Rodríguez J, Peel L, Stella M, Wagner A, Payne JL (2018) The architecture of an empirical genotype-phenotype map. *Evolution* 166:1–19.

198. Poelwijk FJ, Krishna V, Ranganathan R (2016) The context-dependence of mutations: a linkage of formalisms. *PLOS Comput Biol* 12(6):e1004771.

199. Otwinowski J, McCandlish DM, Plotkin J (2018) Inferring the shape of global epistasis. *bioRxiv*:1–26.

200. Sailer ZR, Harms MJ (2017) High-order epistasis shapes evolutionary trajectories. *PLoS Comput Biol* 13(5):e1005541–16.

201. Starr TN, Picton LK, Thornton JW (2017) Alternative evolutionary histories in the sequence space of an ancient protein. *Nature* 549(7672):409–413.

202. Ciliberti S, Martin OC, Wagner A (2007) Innovation and robustness in complex regulatory gene networks. *P Natl Acad Sci USA* 104(34):13591–13596.

203. Aguilar-Rodríguez J, Payne JL, Wagner A (2017) A thousand empirical adaptive landscapes and their navigability. *Nat Ecol Evol* 1(2):45.

204. Wurm MJ, Rathouz PJ, Hanlon BM (2017) Regularized ordinal regression and the ordinalNet R Package. *arXiv* stat.CO.