Using Evolution to Guide Protein Engineering: The Devil IS in the Details

Liskin Swint-Kruse[1],[*]
[1]Department of Biochemistry and Molecular Biology, University of Kansas Medical Center, Kansas City, Kansas

ABSTRACT   For decades, protein engineers have endeavored to reengineer existing proteins for novel applications. Overall, protein folds and gross functions can be readily transferred from one protein to another by transplanting large blocks of sequence (i.e., domain recombination). However, predictably fine-tuning function (e.g., by adjusting ligand affinity, specificity, catalysis, and/or allosteric regulation) remains a challenge. One approach has been to use the sequences of protein families to identify amino acid positions that change during the evolution of functional variation. The rationale is that these nonconserved positions could be mutated to predictably fine-tune function. Evolutionary approaches to protein design have had some success, but the engineered proteins seldom replicate the functional performances of natural proteins. This *Biophysical Perspective* reviews several complexities that have been revealed by evolutionary and experimental studies of protein function. These include 1) challenges in defining computational and biological thresholds that define important amino acids; 2) the co-occurrence of many different patterns of amino acid changes in evolutionary data; 3) difficulties in mapping the patterns of amino acid changes to discrete functional parameters; 4) the nonconventional mutational outcomes that occur for a particular group of functionally important, nonconserved positions; 5) epistasis (nonadditivity) among multiple mutations; and 6) the fact that a large fraction of a protein's amino acids contribute to its overall function. To overcome these challenges, new goals are identified for future studies.

Since the dawn of recombinant DNA technology, a significant effort has been made to engineer new protein functions. The holy grail of protein engineering is de novo rational design of novel sequences and functions. However, success in this area has been limited to small proteins (1–3). Thus, to engineer proteins with more complex functions, researchers have developed strategies to modify naturally evolved proteins. One strategy directly incorporates evolutionary information. Since sequence alignments can be used to reverse engineer functional changes that occurred during protein evolution (e.g., (4,5)), the same information should be useful for forward engineering novel protein functions.

Several groups have made efforts to bridge the fields of evolutionary biology and protein chemistry. Harms and Thornton (5) described historical disparities that hobbled this effort, along with efforts to reconcile them. The relationships among evolution, biophysics, and structure were recently reviewed (6). In parallel, dozens of computer algorithms have been developed to analyze amino acid (a.a.) changes in protein sequence alignments, with the rationale that nonrandom patterns of change reflect evolutionary

constraints at important positions (7–16). One class of algorithms identifies positions for which a.a. changes correlate with phylogeny. A second class identifies positions that change together (pairwise or higher-order coevolution). A third class predicts the outcomes of individual a.a. substitutions, and often includes sequence alignments as input.

Despite their varied mathematical bases and desired outputs, many of these algorithms share common assumptions about a.a. substitutions. This *Biophysical Perspective* reviews studies that were designed to explicitly test these assumptions. The results illuminate several factors that currently confound predictive protein engineering.

Definitions and context

Despite the efforts that have been made to bridge disciplines, many communication gaps persist. Thus, the text below defines our use of the terms "conservation" and "neutrality" to provide a context for the subject of this *Biophysical Perspective*: a.a. positions that are *not* conserved and for which mutations are *not* neutral. Likewise, Fig. 1 provides a context for the type of mutational outcomes that are covered here.

Amino acid conservation is identified by two orthogonal methods. First, for each column in a sequence alignment, the degree of conservation can be quantified by its sequence
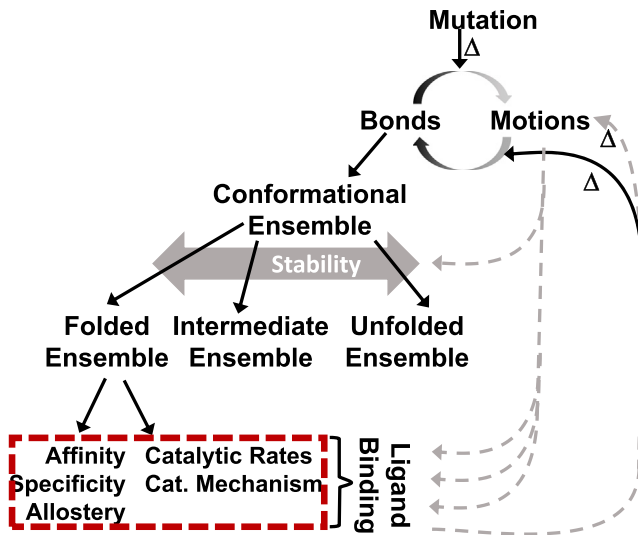
CrossMark

FIGURE 1 Simple hierarchy of the protein structure-function relationship. When a protein is mutated, bonds and/or motions change. Motion changes can alter many factors in this hierarchy (*gray arrows*). Bond changes (*black arrows*) can change the conformational ensemble, which includes subensembles of folded (i.e., functional), intermediate, and unfolded structures; redistribution among these species is detected as altered stability. Folded structures carry out function (*dashed box*). Functional parameters are only detected upon ligand binding, which also alters bonds and motions within the protein molecule. In some proteins, particularly intrinsically disordered proteins, stability is tightly coupled to functional parameters. In many of the studies reviewed here, this does not appear to be so. To see this figure in color, go online.

entropy, which accounts for the number and frequency of distinct a.a. side chains at each position (17). Some studies use all 20 a.a. to calculate sequence entropy, whereas others group a.a. by chemically similar side chains. Our studies used all 20 a.a. and a stringent value for sequence entropy to define conserved positions (7,18,19); this context is used in this *Biophysical Perspective*. Second, conservation can be identified via the use of a.a. substitution matrices (e.g., BLOSUM-62 (20)). These matrices are empirically determined from groups of related sequences, and different input sequences produce different matrix values.

Neutrality has alternative meanings in different fields. The term can refer to mutations that do not change the overall protein structure (21). Alternatively, neutrality can refer to a.a. variants that function like wild-type protein (as in this *Biophysical Perspective*). Kimura (22) discussed functional neutrality in 1968, when he hypothesized that most individual a.a. *substitutions* are neutral, else catastrophe would preclude protein evolution. More recently, some analyses implicitly expanded this idea to include neutral *positions* in sequence alignments, i.e., positions at which *any* a.a. variant functions like a wild-type protein.

Nonneutral mutations can have a wide range of outcomes (Fig. 1). Simplistically, they are divided into two levels: conserved and nonconserved. Conserved positions convey overall protein folds and gross functions, which can be fairly

easily transferred by transplanting large blocks of sequence (i.e., domain recombination); within these blocks, mutagenesis of individual conserved positions is often catastrophic. Nonconserved positions can be mutated to fine-tune functional parameters without disrupting the overall structure or function. Detailed parameters include the binding affinity, ligand specificity, magnitude of the allosteric response, catalytic rate, catalytic mechanism, and stability. Setting aside stability as distinct, this *Biophysical Perspective* focuses primarily on our ability to predictably modify functional parameters of the folded ensemble (Fig. 1, *dashed box*) by changing a.a. at nonconserved positions.

Justification for this focus comes from comparing paralogs, which exhibit functional variation without a significant change in the overall protein fold. For example, the LacI and PurR transcription regulators have the same function of binding DNA under allosteric regulation by a small molecule. Superimposition of LacI (1efa (23)) and PurR (1wet (24)) monomer structures yields a C^α root mean-square deviation of 1.7 Å. However, these two proteins have evolved sequence differences that fine-tune this function for different biological purposes: they bind different DNA ligands and thereby regulate different operons, they bind different small molecules, and they have inverted allosteric responses (25).

Finally, explicit definitions are needed for the functional parameters in Fig. 1. For most protein functions, the primordial step is a binding event, quantified as the binding affinity. Two or more binding affinities are needed to define the allosteric response (the change in binding for one ligand in the presence of a second one; cooperativity is a subtype of allostery) and ligand specificity (the rank order of binding affinities for all ligands). For enzymes, the chemical rates for catalytic steps can be quantified, whereas the order and types of steps that comprise the catalytic mechanism are descriptive.

A brief history of evolution-guided protein redesign

Evolution-based protein redesign has been approached from two directions. The first uses evolutionary information to rationally identify positions that can be mutated to achieve a desired functional change. The second—directed evolution—relies on selection or screening of random changes to identify a desired function. These approaches have also been combined with structural and energetic calculations to yield the desired functional variation. (The extensive field of structure-based redesign is beyond the scope of this *Biophysical Perspective*.) The successes and shortcomings of the evolutionary approaches are briefly reviewed here.

Rationalization

As soon as the first homologs were identified, their sequences were compared to predict which a.a. changes

convey functional variation. The thought was that a few key a.a. could then be exchanged between homologs to exchange their functions. However, exchange experiments have had modest success: designed proteins showed some transfer of the desired function, but performance seldom (if ever) reached that of natural proteins.

Exchange experiments produced complicated results even between closely related proteins. For example, in an early study by Park and Plapp (26), various combinations of a.a were exchanged between two isozymes of alcohol dehydrogenase that differed by 10 a.a. widely distributed over the protein structure. The results provided an early example that changes that do not occur directly in the active site can alter function. Further, many of the intermediate alcohol dehydrogenase variants (those with a subset of the 10 possible changes) had worse function than either of the natural proteins. In the years since that early work, similar studies using a wide range of proteins have suggested that these two results are rules rather than exceptions. Indeed, the fact that long-range effects are difficult to identify from structural comparisons provides a strong motivation for using evolutionary information to guide protein redesign.

For proteins with a large number of a.a. differences, more complicated sequence comparisons have been devised to identify key a.a. for exchange. Yin and Kirsch (27) used Venn diagrams to predict five positions that differentiate malate dehydrogenase and lactate dehydrogenase (LDH). When the five LDH positions were exchanged into malate dehydrogenase, the desired substrate specificity improved by nearly nine orders of magnitude. However, the catalytic efficiency was still three orders of magnitude below that of natural LDH proteins (27). In another example, Rodriguez et al. (28) used evolutionary trace analysis to identify key positions that differ among G-protein-coupled receptors with alternative specificities for dopamine or serotonin. Individually swapping high-scoring positions conveyed enhanced serotonin binding to the dopamine receptor, whereas swapping low-scoring positions did not. However, when the set of substitutions were combined into one construct, the resulting protein could not be expressed. Thus, the high-scoring positions did not capture all of the key differences between the two proteins, and a.a. contributions to function were not cleanly parsed from those to stability.

Directed evolution

Given the partial success of rationalization, the strategy of directed evolution was developed. The rationale is that when an organism is subjected to the appropriate environmental stress, it will generate mutations that convey the desired functional change (e.g., (29)). A variation of this approach is to build a library of protein mutants and screen or select for the desired function (e.g., (30,31)). Both approaches generate proteins with functional efficiencies similar to those of natural proteins, but they are limited by the number of mutants that can be sampled: a library of 10^{10} variants only samples complete a.a. diversity at a maximum of seven to eight positions. Further, some protein functions are not easily adapted for high-throughput screening or the biological fitness assays that are needed for directed evolution.

Several strategies have been developed to enhance the success of directed evolution. For example, structural comparisons were used to exclude mutations that were predicted to disrupt structure (32). Other studies iterated between rationalization (sequence or structure based) and selection/screening (e.g., (33)). For example, a heroic effort was recently made to redesign LacI so that DNA binding could be regulated by four nonnatural inducers (34). However, this required the combined efforts of several large labs and the generation and characterization of several thousand variants. Many of the best-performing variants required 100 mM ligand to elicit a strong allosteric response, whereas wild-type LacI had a strong response at 0.1 mM isopropyl β-D-1-thiogalactopyranoside and the strongest response at 10 mM. Thus, even a combination of many strategies requires expensive resources and person hours.

Factors that confound the rational design of functional variation

The limitations of directed evolution and library screening justify continued efforts to improve rational protein design. One avenue is to examine the common assumptions that underlie computational analyses of protein families. The results described below come from many types of experiments and computations, but together they identify current limitations and specify new directions that must be explored before we can fully understand the protein sequence-structure-function relationship.

Data thresholding greatly influences results

One of the biggest hurdles in our studies, both computational and experimental, has been to identify thresholds that discriminate between significant and insignificant for each type of results.

First, to use evolutionary information to guide engineering, one must choose a sequence identity threshold to define the protein family. A common threshold has been ~40%, although dozens of studies have used the LacI/GalR paralogs with a threshold of ~15% (19). However, depending on the threshold, various sequence analyses generate entirely different results. For example, some positions that are nonconserved in the full LacI/GalR family (15% threshold) are conserved in the subfamilies (40% threshold). The recognition of different thresholds allows nested analyses that identify family- and subfamily-specific positions,

thereby identifying a greater number of important positions (18,28).

Second, thresholding is necessary to interpret output scores from computer algorithms. For example, algorithms that detect nonrandom patterns of a.a. changes calculate a score for every position (or pair of positions), and top scores presumably indicate the most important a.a. positions. However, we have never seen a natural break in the scores that delineates the difference between important and nonimportant (Fig. 2). A conservative solution to this problem is to take the top percent of scores or to set a Z-score threshold, but this carries the risk of truncating important results. More information can be gleaned by assessing the results as a function of the threshold cutoff. For example, in a comparison of two sets of scores, Jaccard analyses can quantify the similarities across all possible thresholds (18). A second approach is to defer thresholding until the final step in the analysis, which prevents the loss of top positions that only emerge from downstream calculations (7).

Ideally, the choice of a computational threshold should be guided by biology: a significant sequence change should cause a functional change large enough to alter the host organism. However, and contrary to many expectations, this can be surprisingly small. For example, <2-fold differences

in the K_m for the tetracycline resistance protein proved adaptive at clinically relevant drug concentrations (35). In humans, various normal phenotypes have been predicted to arise through combinations of weakly nonneutral protein variants (36).

Further, biological thresholds can be difficult to approximate in the lab because in real life, various conditions impose different thresholds (37). For example, de Vos et al. (38) showed that the same mutation can have different outcomes in alternative environments, subsequently leading to different trajectories of evolutionary change. Steinberg and Ostermeier (39) showed that transient environmental changes early in protein evolution can allow a temporary tolerance to deleterious mutations, which in turn allows subsequent mutations that become beneficial in the final environment. Rockah-Shmuel et al. (40) showed that neutral substitutions in DNA methyl-transferase are actually deleterious under some conditions.

Indeed, the influence of experimental thresholds and assay resolution has perhaps been underappreciated in computational studies. For example, many researchers do not realize that in the large LacI mutational data set (41), wild-type variants have a repression range of ~40-fold. The influence of biological thresholds is especially important to consider with regard to the newly popular approach of deep mutational scanning, which uses biological competition and next-generation sequencing to infer functional changes from massively mutated libraries (42). The influences of threshold on such experiments were recently summarized (43).

Based on these observations, the ideal experiments for benchmarking algorithms would directly report protein function, with minimal contributions from other cellular or environmental processes. One could then use the resulting data in a variety of settings after applying the relevant biological threshold(s). Computer algorithms should strive to predict mutation outcomes at this same level of resolution. If large sets of mutational variants are assessed with in vivo functional assays, they should be benchmarked to in vitro studies. For example, LacI/GalR studies monitored in vivo repression under conditions in which repression was highly correlated with the K_d for DNA binding (44). In another example, Firnberg et al. (43) performed deep mutational scanning in 13 different environments to avoid thresholds that would have truncated the functional results.
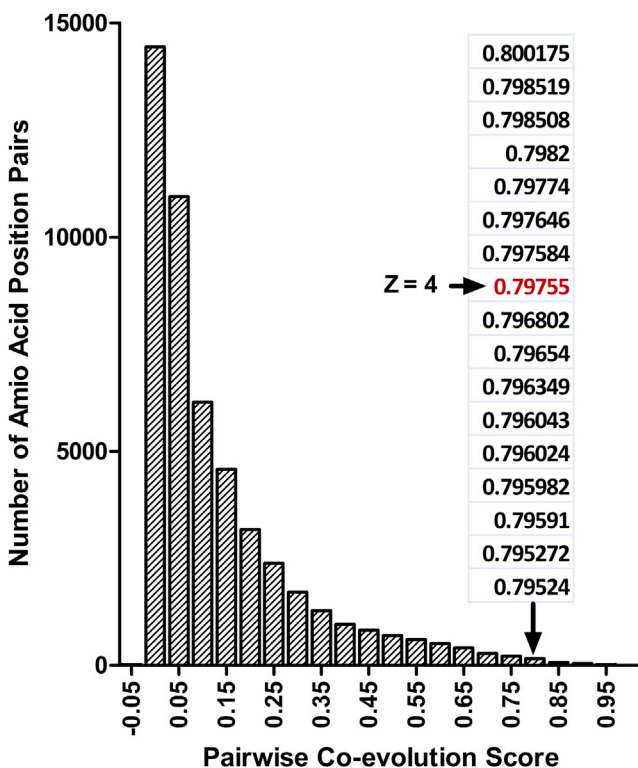

FIGURE 2 Computational scores do not contain obvious thresholds. The histogram shows the distribution of pairwise coevolution scores determined for the aldolase family (7) using the McBASC algorithm (11). A few scores are excerpted for the bar centered at 0.8 and the score at Z = 4 is highlighted. No natural break occurs around this or any other potential threshold. To see this figure in color, go online.

The sequences of protein families show multiple patterns of a.a. change

When using evolutionary information to understand protein function, one assumes that a.a. changes at important positions will be constrained (nonrandom) during evolution. Many algorithms have been developed to identify nonrandom patterns, including (with representative citations) overall conservation/sequence entropy (8), a.a. changes that mirror

the branching of phylogenetic trees (9,45), combinations of conservation and phylogeny (10), pairwise and higher-order coevolving positions (11–13,46), and coevolutionary network centrality (7).

These patterns are often treated as alternative and competing options, with the supposition that one pattern predicts important positions better than the others do. Alternatively, each pattern may contain distinct information, with the combination of multiple patterns giving rise to the emergent protein function. This idea is supported by experiment: when calculations were compared with experimental mutation outcomes in LacI, different algorithms identified different top-scoring positions, but all of the top-scoring positions were mutationally sensitive (7,18,19).

Furthermore, the distinct patterns may or may not comprise overlapping positions. For example, high-scoring positions for the patterns of network centrality and phylogeny were shown to have significant overlap in the LacI/GalR family (7). In contrast, for the aldolase family, positions in the two patterns were largely separate (7). To our knowledge, the ramifications of pattern overlap have only been considered for coevolution and phylogeny, with phylogeny treated as a contaminant in coevolutionary signals (47,48). However, another possibility is that some coevolving positions *track with* phylogeny due to biologically meaningful constraints. Since both coevolution and phylogenetic patterns identify mutationally sensitive positions, discerning between these alternatives will be difficult.

Another complication is that a common structural scaffold can be plastic with respect to the locations of constrained positions. For example, conserved and coevolving positions in the full LacI/GalR family were compared with those in six LacI/GalR subfamilies (18). As expected, all subfamilies had a common set of constrained positions. In addition, each subfamily had uniquely constrained positions that were unconstrained in other subfamilies. Other protein families show similar plasticity in the locations of constrained positions (10), and thus plasticity must be widespread.

Individual evolutionary patterns may not correlate with specific functional parameters

Some studies have attempted to match specific patterns of a.a. change with specific functional parameters. For example, coevolutionary patterns were proposed to identify allosteric positions, but this approach is unlikely to generalize to a wide range of allosteric proteins (49). In another example, paralog sequences were analyzed for specificity determinants, and nonconserved positions that correlate with phylogeny were proposed to alter ligand specificity (45). However, ligand specificity is not the sole parameter that changes during paralog evolution. Enzyme evolution can change the catalytic mechanism and/or the final chemistry (50), the evolution of allosterically regulated proteins can alter the degree of the allosteric response (51,52), and

if host organisms thrive at different temperatures, mutations can be required to alter protein stability (53). Changes in affinity may also be present in the evolutionary record. For example, one paralog may have 10-fold stronger binding affinities for its ligands compared with another paralog. Indeed, for several LacI/GalR homologs, mutations at specificity-determinant positions most often altered the overall affinity, and changes in ligand specificity occurred for <25% of the variants (52,54).

Nor should changes in binding affinities (or other parameters) be treated as contamination in the evolutionary record. Affinity changes in both paralog and ortholog evolution can allow organisms to adapt to new environments. Affinity changes are also needed in protein engineering: even in domain recombination, affinity modulation is often needed to compensate for the far- and mid-range influences of nonconserved positions on domain-domain or subunit-subunit interfaces (55–57).

Mutations at some nonconserved positions do not follow conventional substitution rules

In many computational studies, the results were validated according to the criterion that important nonconserved positions were mutationally sensitive. Further, mutations at these positions were expected to follow the conventional rules (Box 1) that arose from our collective laboratory experiences. However, most laboratory experiments have been biased to mutations at conserved positions (58), which are expected to abolish function (toggle off) unless the substitution is chemically similar (e.g., Ser for Thr; toggle on). Notably, nonconserved positions have seldom been the subject of experiment, and therefore we must consider whether the rules apply to these positions.

In experiments to directly test these rules, multiple a.a. were substituted at nonconserved positions in LacI/GalR proteins (59). Instead of partitioning variants into the expected on/off pattern, the mutational outcomes ranged progressively over orders of magnitude (Fig. 3), which was described as a rheostatic pattern of change (59). This rheostat mutational behavior was also observed for variants of Bcl-2 family proteins (60), a PDZ domain (61), an E3 ubiquitin ligase (62), and pyruvate kinase (51). Thus, rheostat positions are probably widespread in the protein universe,

BOX 1	Conventional Rules for Mutating Important Positions

Most substitutions damage function or structure
A few amino acids allow normal function
   Physicochemically similar
   Presence allowed during evolution
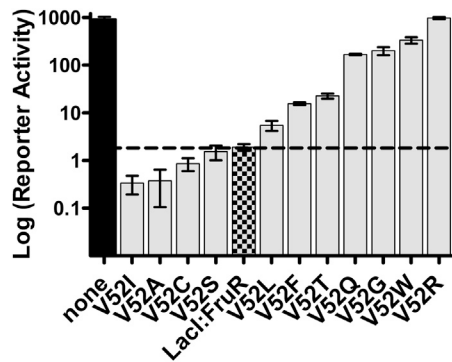Substitution = same outcome in any homolog

FIGURE 3   Example of an evolutionary rheostat. Data are shown for 12 variants of a synthetic LacI/GalR transcription repressor (LacI:FruR) (59). The checkered bar and dashed line represent the activity of the parent protein. The gray bars represent the activities for variants at position 52. Mutations with increased repression fall below the dashed line; those with diminished repression fall above the line. The black bar depicts reporter gene activity in the absence of repressor protein.

and a variety of protein functions appear to have the potential for rheostatic changes.

Notably, rheostat mutations in the LacI/GalR family do not follow any of the conventional rules for mutational outcomes (59). Nor is the rule-breaking limited to this family: in one study, evolutionary frequency did not correlate with mutation outcomes in Hsp90 (63), and in another, mutational outcomes in human growth hormone did not correlate with a.a. chemistry (64). The idea of conservative substitutions based on chemical similarities was challenged in a study that compared a.a. distributions on the structures of natural proteins (65).

Thus, the existence of rheostat positions has important ramifications for computational predictions of a.a. changes. As far as we can tell, most evolutionary algorithms and/or their interpretation explicitly or implicitly assume the rules in Box 1. As such, they will incorrectly handle mutations at rheostat positions. To improve predictions, new rules must be devised to explain mutational outcomes at rheostat positions. One challenge will be to distinguish nonconserved positions at which similar a.a. are interchangeable from rheostat positions that do not follow this rule. It also remains to be seen whether rheostat behaviors correlate with evolutionary patterns or with specific functional parameters. For the LacI/GalR proteins, rheostat positions might correlate with phylogenetic patterns (59). For these and other proteins, we expect that changes in affinity, catalytic rates, and stability can be incrementally modified by mutating various rheostat positions. The behavior would then propagate to any derived parameters, such as the magnitude of the allosteric response.

Epistasis among multiple a.a. positions

In every analysis that uses a protein sequence alignment, an explicit assumption is that the a.a. present in a column pro-

vide information about the importance of that position in individual homologs. The interpretative corollary is that when a mutational outcome is known for one homolog, similar outcomes are expected for all other homologs. As with the other rules listed in Box 1, this assumption was largely derived from mutations at conserved positions. However, for nonconserved positions, this need not be true.

To test this possibility, mutations must be created in multiple proteins at analogous nonconserved positions. Such a study was carried out using 10 members of the LacI/GalR family (55,66,67). The results showed that changes at nonconserved positions had protein-specific outcomes: commonly, a given a.a. substitution was detrimental in one protein, neutral in another, and enhancing in yet another. Disparate outcomes were common even among closely related homologs.

This behavior (i.e., the same variant causing different mutational outcomes in different sequence contexts) is the hallmark of epistasis. Epistasis has confounded many attempts to design proteins. For example, epistasis was observed during efforts to transplant protein-protein interfaces among PDZ domains (68). Epistasis is now recognized to be a dominant feature in the evolution of natural proteins such the influenza A surface proteins (69) and TEM-1 β-lactamase (70). Since a huge number of mutated proteins are needed to explore epistasis experimentally, this will be one of the most significant challenges to the field of rational protein design.

A high percentage of amino acids contribute to a protein's function

As far as we are aware, attempts to identify a discrete subset of amino acids for functional exchange have never yielded a protein that functions as well as natural proteins. One reason may be that numerous a.a. positions contribute to overall function. This notion is based on the aggregate results of many different mutational experiments. One example comes from a screen for novel aminotransferase activity, which yielded 12 variants with desirable function (71). Together, these proteins had more than 40 substitutions, seven of which were common to all. When only these seven a.a. were mutated on the parent protein, the resulting protein had only 60% of the functionality of the selected proteins (71). Thus, the other 30-plus positions must contribute to the final function.

Another example comes from a study involving whole-protein mutagenesis of wild-type LacI, in which 12–13 a.a. were substituted at nearly every position (41). The resulting data set has been used in dozens of studies to verify that high scores from computational algorithms identify important positions. However, an examination of the full data set shows that at least 50% of all LacI positions contribute to function. Indeed, when the full data set was compared with three algorithms, mutationally sensitive

positions were found to occur over the full range of scores (see Fig. 5 in (19)). Thus, no single evolutionary pattern has captured all of the relevant functional information for this family.

Future directions

The results described above have led us to consider what kinds of studies are needed in the future to advance protein engineering. An immediate and economical approach would be to glean more information from existing directed-evolution experiments by retrospectively comparing the mutations with analyses of sequence alignments. However, one must remember that such experiments were designed to generate "winners" and do not yield much information about "losers."

Another approach would be to expand the number of experimental variants available for comparison with sequence analyses by characterizing multiple a.a. substitutions at each position studied. For example, as noted above, no correlation has yet been documented between specific patterns and specific functional parameters. A more fruitful approach could be to correlate an orthogonal classification of mutation outcomes (e.g., toggle, rheostat, and neutral behaviors) with patterns of evolutionary change. Toggle behavior strongly correlates with mutations at conserved positions. Rheostat behavior may correlate with positions that show phylogenetic patterns of change; this should be tested in other protein families. And although it is commonly assumed that positions that show unconstrained, random changes during evolution are functionally neutral, to my knowledge, this has never been tested.

Rheostat behavior, epistasis, and the fact that many a.a. positions contribute to protein function are likely to affect structure-based, de novo, and evolution-based engineering. This issue could be addressed by tying functional changes back to other areas shown in Fig. 1. To that end, the biophysical community has begun useful studies. Several have explored the apparent trade-off between function and protein stability (e.g., (72,73)), and a compelling link was noted between changes in protein dynamics during the evolution of functional variation (74,75). It will be particularly interesting to understand the biophysics that underlies mutational outcomes at evolutionary rheostat positions, and this simply requires that common tools be systematically applied to a series of a.a. at each rheostat position. To illuminate epistatic relationships, researchers will need to generate all possible (pairwise through *n*-wise) mutational combinations, which would be orders of magnitude larger than the number generated in studies completed to date. Deep mutational scanning (42) has been used to compare functions for large sets of protein variants, but the characterization strategies must be scaled up to yield more detailed functional and biophysical information.

Another area to consider is the different classes of protein structures (globular soluble, integral membrane, and intrinsically disordered). To date, most attempts to integrate protein design with evolutionary information have been carried out on soluble proteins, and only a few studies have focused on membrane proteins. However, these three classes have been subjected to fundamentally different structural constraints during evolution, which may give rise to different correlations between evolutionary patterns of change and mutational outcomes (76).

Finally, some protein structures may have evolved to evolve, that is, the underlying structural scaffold of a protein family can support a wide range of functional variation. For example, the >45 paralog groups of the LacI/GalR family (19) have a common structure that tolerates a large number of a.a. changes to effect wide functional variation. As evolvable scaffolds may be more useful for redesign, it is imperative to understand what features make one structure more amenable to functional variation than another one. Bloom et al. (21,77) found that proteins with enhanced stability can accept a wider range of a.a. mutations. Tóth-Petróczy and Tawfik (78) proposed that highly evolvable proteins have a high fraction of functional positions in flexible regions. It will be interesting to compare efforts to redesign proteins that do and do not have these features. (This will almost certainly pose another gradient/threshold problem, with different protein families tolerating varied mutation loads.) At the same time, it may be harder to predict mutational outcomes in highly evolvable proteins than in proteins with stronger structural, functional, and evolutionary constraints.

CONCLUSIONS

The applications for which proteins are engineered often require complex, nuanced functions. Directed evolution and library screening have limitations that could be circumvented by rational protein design. Likewise, the widely used structure/energy calculations used to guide protein redesign (too extensive to review here) are limited by the need for 1) a protein structure and 2) user input about the targeted region, which is difficult to identify outside of binding sites. Evolution-guided redesign has the potential to circumvent these problems and to complement structural/energetic approaches.

The confounding factors described above suggest several reasons why attempts to guide rational protein engineering with evolutionary information have not been fully successful. Further, these factors impact more than protein engineering: they hamper the ability to predict mutation outcomes for genomic diagnoses, and they hamper the ability to assign functions to uncharacterized proteins.

Although these confounding factors are challenging, I remain hopeful that they can be overcome. Since directed evolution and selection from massive libraries are successful

strategies, the correct mutational answers exist in the protein universe. Rational protein design is not limited by nature, only by our own understanding.

ACKNOWLEDGMENTS

REFERENCES

1. Joh, N. H., T. Wang, …, W. F. DeGrado. 2014. De novo design of a transmembrane Zn^{2+}-transporting four-helix bundle. *Science.* 346:1520–1524.

2. Smith, B. A., and M. H. Hecht. 2011. Novel proteins: from fold to function. *Curr. Opin. Chem. Biol.* 15:421–426.

3. Kuhlman, B., G. Dantas, …, D. Baker. 2003. Design of a novel globular protein fold with atomic-level accuracy. *Science.* 302:1364–1368.

4. Harms, M. J., G. N. Eick, …, J. W. Thornton. 2013. Biophysical mechanisms for large-effect mutations in the evolution of steroid hormone receptors. *Proc. Natl. Acad. Sci. USA.* 110:11475–11480.

5. Harms, M. J., and J. W. Thornton. 2013. Evolutionary biochemistry: revealing the historical and physical causes of protein properties. *Nat. Rev. Genet.* 14:559–571.

6. Liberles, D. A., S. A. Teichmann, …, S. Whelan. 2012. The interface of protein structure, protein biophysics, and molecular evolution. *Protein Sci.* 21:769–785.

7. Parente, D. J., J. C. J. Ray, and L. Swint-Kruse. 2015. Amino acid positions subject to multiple coevolutionary constraints can be robustly identified by their eigenvector network centrality scores. *Proteins.* 83:2293–2306.

8. Landau, M., I. Mayrose, …, N. Ben-Tal. 2005. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* 33:W299–W302.

9. Mihalek, I., I. Res, and O. Lichtarge. 2006. Evolutionary trace report_maker: a new type of service for comparative analysis of proteins. *Bioinformatics.* 22:1656–1657.

10. Gu, X., and K. Vander Velden. 2002. DIVERGE: phylogeny-based analysis for functional-structural divergence of a protein family. *Bioinformatics.* 18:500–501.

11. Dekker, J. P., A. Fodor, …, G. Yellen. 2004. A perturbation-based method for calculating explicit likelihood of evolutionary co-variance in multiple sequence alignments. *Bioinformatics.* 20:1565–1572.

12. Fodor, A. A., and R. W. Aldrich. 2004. Influence of conservation on calculations of amino acid covariance in multiple sequence alignments. *Proteins.* 56:211–221.

13. Brown, C. A., and K. S. Brown. 2010. Validation of coevolving residue algorithms via pipeline sensitivity analysis: ELSC and OMES and ZNMI, oh my! *PLoS One.* 5:e10779.

14. Chakraborty, A., and S. Chakrabarti. 2015. A survey on prediction of specificity-determining sites in proteins. *Brief. Bioinform.* 16:71–88.

15. Grimm, D. G., C. A. Azencott, …, K. M. Borgwardt. 2015. The evaluation of tools used to predict the impact of missense variants is hindered by two types of circularity. *Hum. Mutat.* 36:513–523.

16. de Juan, D., F. Pazos, and A. Valencia. 2013. Emerging methods in protein co-evolution. *Nat. Rev. Genet.* 14:249–261.

17. Valdar, W. S. 2002. Scoring residue conservation. *Proteins.* 48:227–241.

18. Parente, D. J., and L. Swint-Kruse. 2013. Multiple co-evolutionary networks are supported by the common tertiary scaffold of the LacI/GalR proteins. *PLoS One.* 8:e84398.

19. Tungtur, S., D. J. Parente, and L. Swint-Kruse. 2011. Functionally important positions can comprise the majority of a protein's architecture. *Proteins.* 79:1589–1608.

20. Henikoff, S., and J. G. Henikoff. 1992. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. USA.* 89:10915–10919.

21. Bloom, J. D., J. J. Silberg, …, F. H. Arnold. 2005. Thermodynamic prediction of protein neutrality. *Proc. Natl. Acad. Sci. USA.* 102:606–611.

22. Kimura, M. 1968. Evolutionary rate at the molecular level. *Nature.* 217:624–626.

23. Bell, C. E., and M. Lewis. 2000. A closer view of the conformation of the Lac repressor bound to operator. *Nat. Struct. Biol.* 7:209–214.

24. Schumacher, M. A., A. Glasfeld, …, R. G. Brennan. 1997. The X-ray structure of the PurR-guanine-purF operator complex reveals the contributions of complementary electrostatic surfaces and a water-mediated hydrogen bond to corepressor specificity and binding affinity. *J. Biol. Chem.* 272:22648–22653.

25. Swint-Kruse, L., and K. S. Matthews. 2009. Allostery in the LacI/GalR family: variations on a theme. *Curr. Opin. Microbiol.* 12:129–137.

26. Park, D. H., and B. V. Plapp. 1992. Interconversion of E and S isoenzymes of horse liver alcohol dehydrogenase. Several residues contribute indirectly to catalysis. *J. Biol. Chem.* 267:5527–5533.

27. Yin, Y., and J. F. Kirsch. 2007. Identification of functional paralog shift mutations: conversion of Escherichia coli malate dehydrogenase to a lactate dehydrogenase. *Proc. Natl. Acad. Sci. USA.* 104:17353–17357.

28. Rodriguez, G. J., R. Yao, …, T. G. Wensel. 2010. Evolution-guided discovery and recoding of allosteric pathway specificity determinants in psychoactive bioamine receptors. *Proc. Natl. Acad. Sci. USA.* 107:7787–7792.

29. Bloom, J. D., and F. H. Arnold. 2009. In the light of directed evolution: pathways of adaptive protein evolution. *Proc. Natl. Acad. Sci. USA.* 106 (*Suppl 1*):9995–10000.

30. Miersch, S., Z. Li, …, S. S. Sidhu. 2015. Scalable high throughput selection from phage-displayed synthetic antibody libraries. *J. Vis. Exp.* 17:51492.

31. Murtaugh, M. L., S. W. Fanning, …, J. R. Horn. 2011. A combinatorial histidine scanning library approach to engineer highly pH-dependent protein switches. *Protein Sci.* 20:1619–1631.

32. Silberg, J. J., P. Q. Nguyen, and T. Stevenson. 2010. Computational design of chimeric protein libraries for directed evolution. *Methods Mol. Biol.* 673:175–188.

33. Goldsmith, M., and D. S. Tawfik. 2012. Directed enzyme evolution: beyond the low-hanging fruit. *Curr. Opin. Struct. Biol.* 22:406–412.

34. Taylor, N. D., A. S. Garruss, …, S. Raman. 2016. Engineering an allosteric transcription factor to respond to new ligands. *Nat. Methods.* 13:177–183.

35. Walkiewicz, K., A. S. Benitez Cardenas, …, Y. Shamoo. 2012. Small changes in enzyme function can lead to surprisingly large fitness effects during adaptive evolution of antibiotic resistance. *Proc. Natl. Acad. Sci. USA.* 109:21408–21413.

36. Bromberg, Y., P. C. Kahn, and B. Rost. 2013. Neutral and weakly nonneutral sequence variants may define individuality. *Proc. Natl. Acad. Sci. USA.* 110:14255–14260.

37. Soskine, M., and D. S. Tawfik. 2010. Mutational effects and the evolution of new protein functions. *Nat. Rev. Genet.* 11:572–582.

38. de Vos, M. G. J., F. J. Poelwijk, …, S. J. Tans. 2013. Environmental dependence of genetic constraint. *PLoS Genet.* 9:e1003580.

39. Steinberg, B., and M. Ostermeier. 2016. Environmental changes bridge evolutionary valleys. *Sci. Adv.* 2:e1500921.

40. Rockah-Shmuel, L., Á. Tóth-Petróczy, and D. S. Tawfik. 2015. Systematic mapping of protein mutational space by prolonged drift reveals the

deleterious effects of seemingly neutral mutations. *PLOS Comput. Biol.* 11:e1004421.

41. Suckow, J., P. Markiewicz, …, B. Müller-Hill. 1996. Genetic studies of the Lac repressor. XV: 4000 single amino acid substitutions and analysis of the resulting phenotypes on the basis of the protein structure. *J. Mol. Biol.* 261:509–523.

42. Fowler, D. M., and S. Fields. 2014. Deep mutational scanning: a new style of protein science. *Nat. Methods.* 11:801–807.

43. Firnberg, E., J. W. Labonte, …, M. Ostermeier. 2014. A comprehensive, high-resolution map of a gene's fitness landscape. *Mol. Biol. Evol.* 31:1581–1592.

44. Tungtur, S., H. Skinner, …, D. Beckett. 2011. In vivo tests of thermodynamic models of transcription repressor function. *Biophys. Chem.* 159:142–151.

45. Ye, K., G. Vriend, and A. P. IJzerman. 2008. Tracing evolutionary pressure. *Bioinformatics.* 24:908–915.

46. Lee, Y., J. Mick, …, L. J. Beamer. 2012. A coevolutionary residue network at the site of a functionally important conformational change in a phosphohexomutase enzyme family. *PLoS One.* 7:e38114.

47. Dunn, S. D., L. M. Wahl, and G. B. Gloor. 2008. Mutual information without the influence of phylogeny or entropy dramatically improves residue contact prediction. *Bioinformatics.* 24:333–340.

48. Kann, M. G., B. A. Shoemaker, …, T. M. Przytycka. 2009. Correlated evolution of interacting proteins: looking behind the mirrortree. *J. Mol. Biol.* 385:91–98.

49. Livesay, D. R., K. E. Kreth, and A. A. Fodor. 2012. A critical evaluation of correlated mutation algorithms and coevolution within allosteric mechanisms. *Methods Mol. Biol.* 796:385–398.

50. Martínez Cuesta, S., S. A. Rahman, …, J. M. Thornton. 2015. The classification and evolution of enzyme function. *Biophys. J.* 109:1082–1086.

51. Ishwar, A., Q. Tang, and A. W. Fenton. 2015. Distinguishing the interactions in the fructose 1,6-bisphosphate binding site of human liver pyruvate kinase that contribute to allostery. *Biochemistry.* 54:1516–1524.

52. Zhan, H., M. Taraban, …, L. Swint-Kruse. 2008. Subdividing repressor function: DNA binding affinity, selectivity, and allostery can be altered by amino acid substitution of nonconserved residues in a LacI/GalR homologue. *Biochemistry.* 47:8058–8069.

53. Peña, M. I., M. Davlieva, …, Y. Shamoo. 2010. Evolutionary fates within a microbial population highlight an essential role for protein folding during natural selection. *Mol. Syst. Biol.* 6:387.

54. Zhan, H., L. Swint-Kruse, and K. S. Matthews. 2006. Extrinsic interactions dominate helical propensity in coupled binding and folding of the lactose repressor protein hinge helix. *Biochemistry.* 45:5896–5906.

55. Meinhardt, S., M. W. Manley, Jr., …, L. Swint-Kruse. 2012. Novel insights from hybrid LacI/GalR proteins: family-wide functional attributes and biologically significant variation in transcription repression. *Nucleic Acids Res.* 40:11139–11154.

56. Connaghan, K. D., M. T. Miura, …, D. L. Bain. 2013. Analysis of a glucocorticoid-estrogen receptor chimera reveals that dimerization energetics are under ionic control. *Biophys. Chem.* 172:8–17.

57. Perica, T., Y. Kondo, …, S. A. Teichmann. 2014. Evolution of oligomeric state through allosteric pathways that mimic ligand binding. *Science.* 346:1254346.

58. Gray, V. E., K. R. Kukurba, and S. Kumar. 2012. Performance of computational tools in evaluating the functional impact of laboratory-induced amino acid mutations. *Bioinformatics.* 28:2093–2096.

59. Meinhardt, S., M. W. Manley, Jr., …, L. Swint-Kruse. 2013. Rheostats and toggle switches for modulating protein function. *PLoS One.* 8:e83502.

60. DeBartolo, J., S. Dutta, …, A. E. Keating. 2012. Predictive Bcl-2 family binding models rooted in experiment or structure. *J. Mol. Biol.* 422:124–144.

61. McLaughlin, R. N., Jr., F. J. Poelwijk, …, R. Ranganathan. 2012. The spatial architecture of protein function and adaptation. *Nature.* 491:138–142.

62. Starita, L. M., J. N. Pruneda, …, R. E. Klevit. 2013. Activity-enhancing mutations in an E3 ubiquitin ligase identified by high-throughput mutagenesis. *Proc. Natl. Acad. Sci. USA.* 110:E1263–E1272.

63. Hietpas, R. T., J. D. Jensen, and D. N. A. Bolon. 2011. Experimental illumination of a fitness landscape. *Proc. Natl. Acad. Sci. USA.* 108:7896–7901.

64. Pál, G., J. L. Kouadio, …, S. S. Sidhu. 2006. Comprehensive and quantitative mapping of energy landscapes for protein-protein interactions by rapid combinatorial scanning. *J. Biol. Chem.* 281:22378–22385.

65. Jonson, P. H., and S. B. Petersen. 2001. A critical view on conservative mutations. *Protein Eng.* 14:397–402.

66. Meinhardt, S., and L. Swint-Kruse. 2008. Experimental identification of specificity determinants in the domain linker of a LacI/GalR protein: bioinformatics-based predictions generate true positives and false negatives. *Proteins.* 73:941–957.

67. Tungtur, S., S. M. Egan, and L. Swint-Kruse. 2007. Functional consequences of exchanging domains between LacI and PurR are mediated by the intervening linker sequence. *Proteins.* 68:375–388.

68. Melero, C., N. Ollikainen, …, T. Kortemme. 2014. Quantification of the transferability of a designed protein specificity switch reveals extensive epistasis in molecular recognition. *Proc. Natl. Acad. Sci. USA.* 111:15426–15431.

69. Kryazhimskiy, S., J. Dushoff, …, J. B. Plotkin. 2011. Prevalence of epistasis in the evolution of influenza A surface proteins. *PLoS Genet.* 7:e1001301.

70. Dellus-Gur, E., M. Elias, …, D. S. Tawfik. 2015. Negative epistasis and evolvability in TEM-1 β-lactamase—the thin line between an enzyme's conformational freedom and disorder. *J. Mol. Biol.* 427:2396–2409.

71. Rothman, S. C., and J. F. Kirsch. 2003. How does an enzyme evolved in vitro compare to naturally occurring homologs possessing the targeted function? Tyrosine aminotransferase from aspartate aminotransferase. *J. Mol. Biol.* 327:593–608.

72. Dellus-Gur, E., A. Toth-Petrocy, …, D. S. Tawfik. 2013. What makes a protein fold amenable to functional innovation? Fold polarity and stability trade-offs. *J. Mol. Biol.* 425:2609–2621.

73. Couñago, R., C. J. Wilson, …, Y. Shamoo. 2008. An adaptive mutation in adenylate kinase that increases organismal fitness is linked to stability-activity trade-offs. *Protein Eng. Des. Sel.* 21:19–27.

74. Bahar, I., M. H. Cheng, …, S. Zhang. 2015. Structure-encoded global motions and their role in mediating protein-substrate interactions. *Biophys. J.* 109:1101–1109.

75. Marsh, J. A., and S. A. Teichmann. 2014. Parallel dynamics and evolution: protein conformational fluctuations and assembly reflect evolutionary changes in sequence and structure. *BioEssays.* 36:209–218.

76. Siltberg-Liberles, J., J. A. Grahnen, and D. A. Liberles. 2011. The evolution of protein structures and structural ensembles under functional constraint. *Genes (Basel).* 2:748–762.

77. Bloom, J. D., S. T. Labthavikul, …, F. H. Arnold. 2006. Protein stability promotes evolvability. *Proc. Natl. Acad. Sci. USA.* 103:5869–5874.

78. Tóth-Petróczy, A., and D. S. Tawfik. 2014. The robustness and innovability of protein folds. *Curr. Opin. Struct. Biol.* 26:131–138.