Published in final edited form as:

Nat Methods. 2007 December ; 4(12): 995–997. doi:10.1038/nmeth1207-995.

Directed evolution approaches to the study of proteins

Courtney M. Yuen and David R. Liu

Howard Hughes Medical Institute and Department of Chemistry and Chemical Biology, Harvard University

Introduction

The amino acid sequence of a protein encodes its three-dimensional structure and determines its biological function. Although researchers can readily generate proteins with altered amino acid sequences in the laboratory, achieving an understanding of how individual amino acids contribute to the structure and function of a protein of interest often remains a challenge. This understanding can be valuable in several ways— it enhances our fundamental understanding of the relationship between linear sequence and specific three-dimensional structures, it reveals how differences in amino acid sequence endow related proteins with different biological functions, and it can enable researchers to engineer, or even design *de novo*, proteins with tailor-made structural or functional properties.

Much of the difficulty of determining the contributions of individual residues to protein structure and function stems from the need to generate and evaluate a large number of mutants in order to draw meaningful conclusions. The contributions of specific residues can be cooperative or context-dependent, and different aspects of side-chain identity, including shape, charge, size, or polarity, are important at different positions. In order to generate and analyze the large number of mutants necessary for this type of characterization, some researchers have adopted a “directed evolution” approach that parallels the process by which natural proteins evolve.

During natural protein evolution, spontaneous mutations result in variants of a protein within a population of organisms. Those protein variants with beneficial properties promote the survival of their host organisms and thereby facilitate the propagation of the genetic information encoding the most beneficial proteins. Surviving organisms containing improved proteins undergo subsequent mutation and natural selection. Similarly, the directed evolution of proteins entails multiple rounds of diversifying (mutating) the sequence of a protein, selecting or screening for mutants that exhibit a desired property, and subjecting the surviving mutants to subsequent rounds of diversification and selection or screening. During or after the directed evolution process, researchers can isolate and characterize individual mutants from the population of surviving molecules.

A true directed evolution process is distinct from a simpler set of experiments in which a protein is diversified into a collection (or “library”) of mutants, which are then screened or

selected. The latter process simply generates and evaluates a number of mutants of a protein of interest. In contrast, a bona fide evolutionary process exploits iterated rounds of diversification and selection or screening. Because the mutants generated in any given round of directed evolution inherit those mutations that arose and survived screening or selection in previous rounds, the directed evolution process searches a protein's sequence space in a highly guided manner (Figure 1). A number of studies have provided both theoretical and experimental evidence that this guided search through sequence space typically accesses more highly functional areas of sequence space than can be readily accessed in a simple mutagenesis-and-selection approach.^{i,ii,iii,iv}

While directed evolution can produce proteins with improved or altered activities^{v,vi} it can also be used to answer basic questions of protein science, particularly those that probe relationships between individual amino acid residues and protein structure or function. In this Commentary, we will focus on this application, highlighting recent studies in which researchers have used directed evolution to answer questions about protein stability and protein-protein interactions.

Protein stability

Protein stability is determined by interactions among individual residues, but not all residues contribute equally or in the same way. Determining the contributions of individual residues to the stability of a protein requires making and characterizing a pertinent number of mutants. Ideally, one would like to be able to mutate each residue to a variety of different amino acids, and assess the effects of these mutations both alone and in combination with other mutated residues. Protein library sizes, however, are typically limited by the transformation efficiency of bacteria, and the largest libraries comprising $\sim 10^9$ members can exhaustively cover the mutagenesis of only six to seven positions.

Directed evolution can overcome the consequences of this limitation by enabling beneficial mutations to accumulate over several successive rounds. Because each round begins with the generation of a new library of mutants based on the most “fit” mutants isolated from the previous round, evolution guides the search through a protein's vast sequence space towards beneficial sets of mutations. The iterative process of diversification and selection or screening therefore increases the likelihood of finding desirable mutants within each library,^{iv} and enables the characterization of a much greater number of residues of potential importance compared with either site-directed or non-iterative mutation and screening methods (Figure 1).

A number of researchers have used directed evolution as a technique to increase the stability of a target protein,^{vii} but some have also used directed evolution to gain insights into the mechanisms underlying protein stability.^{viii,ix,x,xi} The results of these studies support the general principles that increasing conformational rigidity has a stabilizing effect,^{viii,ix}. The mechanisms by which rigidity augments stability include restricting the conformational flexibility of disordered termini^{8,9} and increasing interactions between loop residues.⁹ In addition, mutations at specific surface locations can significantly increase stability.^{x,xi} presumably by decreasing the propensity of these locations to initiate local unfolding. In general these studies show that protein stability can be increased either by strengthening

interactions between domains,^x or by rigidifying individual domains while maintaining interdomain flexibility^{viii}

Hecky and Muller^x used directed evolution to address the specific mechanistic question of whether the effects of individual mutations on a protein's stability are dependent on each other. The authors first made destabilizing N-terminal truncations to a β -lactamase, then used directed evolution to restore the activity of the truncated proteins. Because they could not know *a priori* which residues of the protein could restore stability through mutation, the researchers adopted an unbiased evolutionary approach in which they iterated mutagenesis of the entire protein and selection for the most active clones.

After three rounds of evolution, common mutations in the most active selected clones were broadly distributed across or near the protein surface, but none occurred near the site of truncation. The two mutations found in all third-round clones were located at interdomain boundaries, and the authors propose that both mutations stabilize the protein by strengthening the interactions between domains. One of the two mutations (M182T) is a known suppressor of missense mutations,^{xii} and crystal structures of M182T mutants suggest that the mutant Thr residue makes multiple hydrogen bonds with surrounding residues.^{xiii} The location of the other mutation (A224V) occurs at an interdomain boundary near a cluster of hydrophobic residues. The authors suggest that mutation to Val may allow better interaction with this hydrophobic cluster, stabilizing the domain interface.

Hecky and Muller observed that the stability of the evolved truncated proteins was in some cases greater than that of wild-type β -lactamase, and that beneficial mutations in the evolved truncated proteins also increased the stability of the full-length protein. Collectively, these results suggest that proteins can have multiple regions that independently promote instability, and that stabilizing mutations in one region can compensate for destabilizing mutations in another. Furthermore, this study demonstrates the ability of directed evolution to identify both known and novel changes that influence a protein's stability. Because none of the most frequently observed stabilizing mutations were close to the truncation site in either primary sequence or three-dimensional space, they would not have been easily found in a targeted search strategy. Such an approach can in principle be applied to reveal stability determinants in any protein compatible with genetic selection or high-throughput screening.

Protein-protein interactions

Like protein stability, protein-protein interactions are potentially affected by large numbers of residues. Although structural information, when available, can be used to identify candidate residues that may participate in protein-binding interfaces, these residues are often sufficiently numerous that their exhaustive mutation cannot be covered by any single protein library. Researchers frequently avoid this limitation by using alanine scanning, a technique that enables a limited analysis of a large number of interface residues without requiring large libraries.^{xiv} Amino acids at a protein-binding surface are individually mutated to alanine, and residues whose mutation abrogates binding most severely are hypothesized to contribute most to binding affinity.

Alanine scanning cannot reveal effects of altering more than one residue at a time, and only reveals the impact of mutations to alanine. Directed evolution approaches have enabled more detailed characterization of how individual residues contribute to binding affinity. Two recent studies demonstrate how the results of directed evolution-based characterization both differ from and complement the results of alanine scanning-based characterization.

Thom *et al.*^{xv} sought to identify mutations that increase the affinity of the antibody BAK1 for its target protein, and to elucidate the relationship between those residues that are necessary for binding and those that modulate binding affinity. The binding affinity of an antibody for its target antigen is largely determined by the amino acids in the complementary-determining regions (CDR's) of the binding surface. The authors created a library by mutating the residues of the centrally located heavy chain CDR3, which they predicted would play the largest role in determining binding affinity, then subjected this library to three rounds of directed evolution in which they diversified all six of the antibody's CDR's and selected high-affinity binders. The majority of the evolved antibodies had mutations not only in the heavy chain CDR3, but also in the heavy chain CDR2 and light chain CDR1. These results highlight the important ability of directed evolution approaches to access solutions outside of the sequence space of an initial library.

This study also reveals an important difference between the information obtained by directed evolution and the information obtained by the non-evolution-based technique of alanine scanning. An alanine scan of the CDR's of BAK1 showed that mutation of the residues in an 11 amino acid patch in the center of the binding surface had the greatest effect on binding affinity. However, the frequently mutated positions in the evolved antibodies occurred not within this central region, but around the periphery of the binding surface. The authors suggest that because the residues identified by alanine scanning are the most important contributors to binding energy and tend to have the most side chain contact with the ligand, they will be resistant to mutation during the directed evolution process. Therefore, while alanine scanning identifies the residues that are necessary for binding, directed evolution identifies those that modulate binding affinity.

Directed evolution and alanine scanning can also be used as complementary methods to characterize the contributions of individual amino acids to protein-binding affinity. Buonpane *et al.*^{xvi} used directed evolution to increase the binding affinity of a T-cell receptor domain to its antigen by 10,000-fold. In order to identify the residues responsible for the significant decrease in binding off-rate, the authors performed alanine scanning mutagenesis on 20 binding surface residues of one of the final evolved receptors, C10. They identified four positions whose mutation preserved high binding affinity, but significantly increased the binding off-rate of the receptor. Alanine mutation at these positions rendered the off-rate of C10 comparable to those of the three parental receptors from the previous round of evolution.

Combining the results of the alanine scan with sequence analysis of the evolved receptors enabled the authors to propose an explanation for the decreased off-rate of C10. In the wild-type receptor, two of the residues identified by alanine scanning are located on a complementary-determining region loop, while the other two are in the neighboring

framework region (Figure 2). In the mutant C10, the net charge of the loop is positive, and the net charge of the two framework region residues is negative. In contrast, while the parental receptors collectively contain all the individual mutations of C10, none of the parental receptors modulates the net charge of these two regions in the same fashion. Thus, the researchers suggest that charge modulation in these two discrete regions of the protein is responsible for the decreased off-rate observed in mutant C10. The complexity of this analysis demonstrates the ability of directed evolution to reveal combinatorial contributions of individual amino acids to binding.

Conclusion

Although directed evolution is commonly viewed as a way to generate proteins with improved or altered activities, the studies reviewed here demonstrate its value for studying protein structure and interactions. The contributions of individual amino acid residues to structure and protein complex formation can be interdependent, and their characterization can require the generation and analysis of a large number of mutations without prior knowledge of which changes will prove to be the most revealing. Directed evolution approaches, can search sequence space in a guided, iterative manner in which the production and evaluation of new mutants builds upon the functionality of promising parental variants. As a result, directed evolution can focus an otherwise intractable search on those regions of sequence space that best reveal how individual residues contribute to a protein's unique functional properties.

Acknowledgments

We thank Kevin Esvelt for helpful discussions. C.M.Y. is supported by an Eli Lilly Organic Chemistry Graduate Fellowship. D.R.L. gratefully acknowledges support from the Howard Hughes Medical Institute.

References

- i. Voigt CA, Kauffman S, Wang ZG. Rational evolutionary design: the theory of in vitro protein evolution. *Adv. Protein Chem.* 2000; 55:79–160. [PubMed: 11050933]
- ii. Chen K, Arnold FH. Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proc. Natl. Acad. Sci. USA.* 1993; 90:5618–22. [PubMed: 8516309]
- iii. Aita T, Husimi Y. Adaptive walks by the fittest among finite random mutants on a Mt. Fuji-type fitness landscape. *J. Theor. Biol.* 1998; 193:383–405. [PubMed: 9735268]
- iv. Matsuura T, et al. Nonadditivity of mutational effects on the properties of catalase I and its application to efficient directed evolution. *Protein Eng.* 1998; 11:789–795. [PubMed: 9796828]
- v. Yuan L, Kurek I, English J, Keenan R. Laboratory-directed protein evolution. *Microbiol. Mol. Biol. Rev.* 2005; 69:373–392. [PubMed: 16148303]
- vi. Johannes TW, Zhao H. Directed evolution of enzymes and biosynthetic pathways. *Curr. Op. Microbiol.* 2006; 9:261–267.
- vii. Roodveldt C, Aharoni A, Tawfik DS. Directed evolution of proteins for heterologous expression and stability. *Curr. Op. Struct. Biol.* 2005; 15:50–56.
- viii. Hoseki J, et al. Increased rigidity of domain structures enhances the stability of a mutant enzyme created by directed evolution. *Biochemistry.* 2003; 42:14469–14475. [PubMed: 14661958]
- ix. Acharya P, Rajakumara E, Sankaranarayanan R, Rao NM. Structural basis of selection and thermostability of laboratory evolved *Bacillus subtilis* lipase. *J. Mol. Biol.* 2004; 341:1271–1281. [PubMed: 15321721]

- x. Hecky J, Muller KM. Structural perturbation and compensation by directed evolution at physiological temperature leads to thermostabilization of beta-lactamase. *Biochemistry*. 2005; 44:12640–12654. [PubMed: 16171379]
- xi. Miyazaki K, et al. Thermal stabilization of *Bacillus subtilis* family-11 xylanase by directed evolution. *J. Mol. Biol.* 2006; 281:10236–10242.
- xii. Huang W, Palzkill T. A natural polymorphism in beta-lactamase is a global suppressor. *Proc. Natl. Acad. Sci. USA*. 1997; 94:8801–8806. [PubMed: 9238058]
- xiii. Orenica MA, Yoon JS, Ness JE, Stemmer WP, Stevens RC. Predicting the emergence of antibiotic resistance by directed evolution and structural analysis. *Nat. Struct. Biol.* 2001; 8:238–242. [PubMed: 11224569]
- xiv. Cunningham BC, Wells JA. High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science*. 1989; 244:1081–1085. [PubMed: 2471267]
- xv. Thom G, et al. Probing a protein-protein interaction by in vitro evolution. *Proc. Natl. Acad. Sci. USA*. 2006; 103:7619–7624. [PubMed: 16684878]
- xvi. Buonpane RA, Moza B, Sundberg EJ, Kranz DM. Characterization of T Cell receptors engineered for high affinity against toxic shock syndrome toxin-1. *J. Mol. Biol.* 2005; 353:308–321. [PubMed: 16171815]

Figure 1.

A comparison of non-evolutionary and evolutionary approaches to library creation and selection. The black rectangle represents the sequence space surrounding a protein of interest (black star). Each colored star represents a member of a library of mutants, and is characterized by its similarity to the original protein (black star) in sequence space (horizontal axis), and by its activity (vertical axis). Each panel represents a discrete round of selection or screening. The library member with the highest activity is represented by the solid-colored star. (a) A non-evolutionary strategy can identify a mutant with the highest activity from within a single library. (b) In an evolutionary strategy, each successive library is based not on the original protein, but on the highest activity mutant(s) of the preceding library. After three rounds, the best mutant isolated in the evolutionary approach is both higher in activity and farther in sequence space from the original protein compared with the best mutant to arise from the non-evolutionary approach, even if the total number of mutants generated and screened in the two approaches is the same.

	CDR loop	Framework region
WT	...MLNATSNEGSKATYEQGVKDKFLIN...	...MLNATSNEGSKATYEQGVKDKFLIN...
R9	...MLNATS RID FHATYEQGVKDKFLIN...	...MLNATSNEGSKATYEQGVKDKFLIN...
R17	...MLNATS RLW DSATYEQGV V KDKFLIN...	...MLNATSNEGSKATYEQGVKDKFLIN...
R18	...MLNATS HQGF NAIYEQGVKDKFLIN...	...MLNATSNEGSKATYEQGVKDKFLIN...
C10	...MLNATS RID FHATYEQGV V KDKFLIN...	...MLNATSNEGSKATYEQGVKDKFLIN...

Clone	Net charge of CDR loop	Net charge of critical framework region residues	K_D (M)
WT	neutral	neutral	2.3×10^{-6}
R9	positive	neutral	1.3×10^{-8}
R17	neutral	positive	
R18	positive	neutral	
C10	positive	positive	3.4×10^{-10}

Figure 2.

Sequence analysis and alanine scanning suggest a mechanism underlying the increased affinity of a mutant T-cell receptor generated by directed evolution (adapted from Ref. 17). Ecolved receptor C10 has higher affinity for the antigen than receptors R9, R17 and R18 from the previous round of evolution. Alanine scanning of the binding surface residues of C10 identified four residues (boxed in orange) responsible for this gain in affinity. The authors suggest that these critical residues mediate a change-charge interaction at two locations (green and purple). Mutated residues (blue) of C10 change of both these regions from neutral to positive.