Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Deep sequencing methods for protein engineering and design

Emily E Wrenbeck[1], Matthew S Faber[2] and Timothy A Whitehead[1,3]

The advent of next-generation sequencing (NGS) has revolutionized protein science, and the development of complementary methods enabling NGS-driven protein engineering have followed. In general, these experiments address the functional consequences of thousands of protein variants in a massively parallel manner using genotype-phenotype linked high-throughput functional screens followed by DNA counting via deep sequencing. We highlight the use of information rich datasets to engineer protein molecular recognition. Examples include the creation of multiple dual-affinity Fabs targeting structurally dissimilar epitopes and engineering of a broad germline-targeted anti-HIV-1 immunogen. Additionally, we highlight the generation of enzyme fitness landscapes for conducting fundamental studies of protein behavior and evolution. We conclude with discussion of technological advances.

Addresses
[1] Department of Chemical Engineering and Materials Science, Michigan State University, East Lansing, MI 48824, United States
[2] Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824, United States
[3] Departments of Biosystems and Agricultural Engineering, Michigan State University, East Lansing, MI 48824, United States

Corresponding author: Whitehead, Timothy A (taw@egr.msu.edu)

Current Opinion in Structural Biology 2017, **45**:36–44

This review comes from a themed issue on **Engineering and design: New trends in designer proteins**

Edited by **Niv Papo** and **Julia Shifman**

For a complete overview see the Issue and the Editorial

Available online 22nd November 2016

http://dx.doi.org/10.1016/j.sbi.2016.11.001

Introduction

Researchers have been engineering proteins for almost 4 decades. Early endeavors involved generation of a handful of point mutations followed by low-throughput assays for function; the 'search space' a protein scientist could feasibly explore was miniscule.

As demonstrated by the seminal works of Fowler *et al.* [1] and Hietpas *et al.* [2], the advent of next-generation sequencing (NGS) has presented protein engineers with the ability to economically observe *entire populations* of molecules before, during, and after a high-throughput screen or selection for function (HTS) (Figure 1). A typical NGS run provides sufficient sequencing data to permit the study of tens of thousands of protein variants. Thus, when coupled to HTS, NGS significantly expands the accessible mutational search space. In this way a researcher can test all possible point mutations or combinations of mutations, for example, and remove the duty of having to design small focused libraries that may miss unpredictable beneficial mutations. As a testimonial to the accessibility of these methodologies, experiments can be performed in a beginning graduate-level course [3].

The intent of this review is to highlight examples where deep sequencing has been applied in different areas of protein engineering and design. As such, we will not provide a comprehensive review of directed evolution or of deep mutational scanning (excellent reviews can be found here [4,5]). We will discuss the use of NGS for engineering protein molecular recognition, membrane proteins, and enzymes, highlight recent technological advances, and offer a perspective on the shape of the field over the next several years.

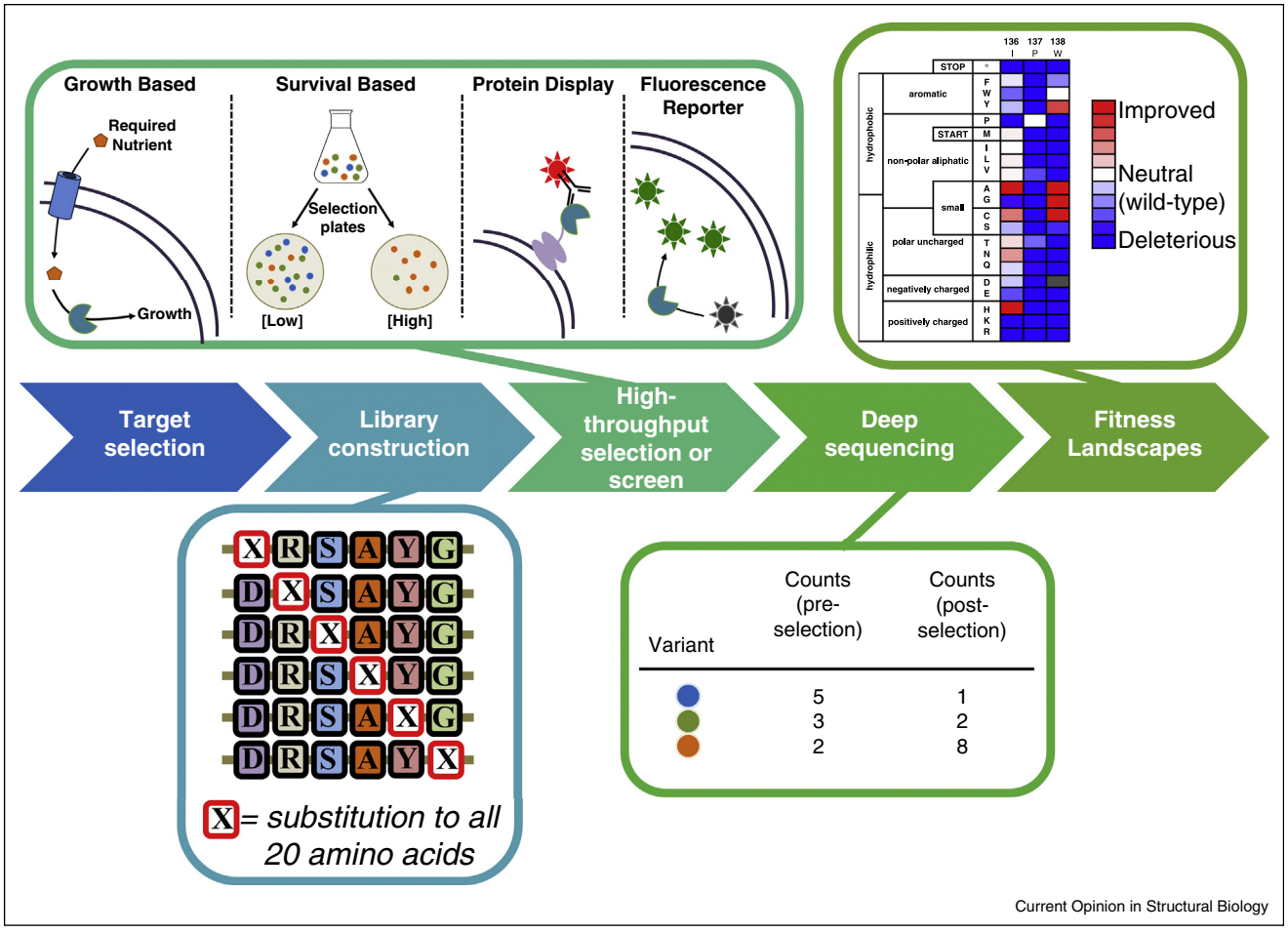
Engineering protein molecular recognition

Dozens of studies over the past five years have used deep sequencing to identify and engineer protein–ligand interactions. Rapid adoption of deep sequencing by this field is a direct result of mature display-based technologies that can be used to screen very large initial libraries. For example, in the study of protein–protein binding interactions a library of protein variants can be displayed on the surface of yeast using yeast surface display (Figure 1). Yeast cells are labeled with a fluorescently conjugated binding partner, and FACS can be used to screen cells by fluorescence intensity.

Deep sequencing for screening protein binder libraries

NGS is now frequently used in the evaluation of synthetic or natural libraries to identify antigen-specific binders. Advances in pairing V_H and V_L sequences from individual B cells [6] allows one to identify antigen-specific antibodies directly from sequencing, including panels of antibodies targeting Ebola virus [7] and ricin [8]. Methodological details and limitations associated with identification of rare clones and evaluation of library diversity are presented in a recent review [9].

Figure 1



Overview of the steps involved in deep mutational scanning. A library of protein variants is generated. Often this is a comprehensive single-site saturation mutagenesis library. The library is subjected to a high-throughput selection or screen for function. Examples of commonly used selections and screens include survival or competitive growth-based selections, protein binding screens like phage or yeast surface display, and fluorescence reporter-based screens. Variants are quantified in the pre-selection and post-selection populations with counting via deep sequencing. These pre-selection and post-selection counts are transformed to a normalized functional score and are used to generate fitness landscapes of the target protein.

As an emerging area, engineers now use NGS to refine protein binder libraries [10,11•]. In a notable advance, Woldring *et al.* screened a hydrophilic fibronectin domain library to bind various protein targets [11•]. The researchers exploited the site-specific amino acid preferences from an initial library to develop a more focused second library depleted in mutations at the periphery of the binder paratope. Compared to other libraries, this library design afforded far superior performance in isolation of high affinity, stable binders.

Paratope optimization for affinity and specificity

NGS can be used to rapidly improve the affinity and specificity of the binding paratope (Figure 2) [12,13]. A crucial advantage enabled by NGS is the ability to discriminate very small beneficial changes in binding — on the order of 0.1 kcal/mol or about a 20% improvement in dissociation constant. These small-scale beneficial mutations can be additive, allowing one to 'leapfrog' over potential affinity maturation bottlenecks by combining mutations.

Whitehead *et al.* provide the first example of paratope engineering for affinity and specificity using deep sequencing [14]. The researchers screened a comprehensive single-site saturation mutagenesis library of two de novo designed Influenza Hemagglutinin (HA) binders against H1 and H5 HA subtypes. Engineering specificity was demonstrated by comparing site-specific preferences for H1 to the H5 subtype. A single point mutation was identified that gave over a 30-fold specificity switch from the parental designed protein. For affinity maturation, site-specific preferences were encoded into a second library and sorted to improve affinity against both