



Deep Mutational Scanning of Viral Glycoproteins and Their Host Receptors

Krishna K. Narayanan and Erik Procko*

Department of Biochemistry and Cancer Center at Illinois, University of Illinois, Urbana, IL, United States

OPEN ACCESS

Edited by:

Edoardo Sarti,
Université Pierre et Marie Curie,
France

Reviewed by:

Abbye Elizabeth McEwen,
University of Washington,
United States
Filippo Prischi,
University of Essex, United Kingdom

*Correspondence:

Erik Procko
procko@illinois.edu

Specialty section:

This article was submitted to
Structural Biology,
a section of the journal
Frontiers in Molecular Biosciences

Received: 01 December 2020

Accepted: 18 March 2021

Published: 09 April 2021

Citation:

Narayanan KK and Procko E
(2021) Deep Mutational Scanning
of Viral Glycoproteins and Their Host
Receptors.
Front. Mol. Biosci. 8:636660.
doi: 10.3389/fmolb.2021.636660

Deep mutational scanning or deep mutagenesis is a powerful tool for understanding the sequence diversity available to viruses for adaptation in a laboratory setting. It generally involves tracking an *in vitro* selection of protein sequence variants with deep sequencing to map mutational effects based on changes in sequence abundance. Coupled with any of a number of selection strategies, deep mutagenesis can explore the mutational diversity available to viral glycoproteins, which mediate critical roles in cell entry and are exposed to the humoral arm of the host immune response. Mutational landscapes of viral glycoproteins for host cell attachment and membrane fusion reveal extensive epistasis and potential escape mutations to neutralizing antibodies or other therapeutics, as well as aiding in the design of optimized immunogens for eliciting broadly protective immunity. While less explored, deep mutational scans of host receptors further assist in understanding virus-host protein interactions. Critical residues on the host receptors for engaging with viral spikes are readily identified and may help with structural modeling. Furthermore, mutations may be found for engineering soluble decoy receptors as neutralizing agents that specifically bind viral targets with tight affinity and limited potential for viral escape. By untangling the complexities of how sequence contributes to viral glycoprotein and host receptor interactions, deep mutational scanning is impacting ideas and strategies at multiple levels for combatting circulating and emergent virus strains.

Keywords: viral fusion protein, entry receptor, virus spike, deep mutational scan, selection, viral escape, mutational landscape

INTRODUCTION

The surfaces of enveloped viruses are decorated with glycoproteins that mediate attachment to host cells and fusion of the viral and cell membranes, allowing the viral genetic material to access the cytoplasm. By virtue of being exposed on the virus surface to the immune system and mediating the essential events of attachment and membrane fusion, viral glycoproteins are a primary target of neutralizing antibodies and inhibitors. Viral fusion and attachment glycoproteins possess structural similarities and common mechanisms within families but can have highly variable sequences that impact receptor usage, tissue- and host species-tropism, and antigenic properties (White et al., 2008; Banerjee and Mukhopadhyay, 2016; Murin et al., 2019). This sequence diversity has profound implications for antigenic change and spillover of new strains from animal reservoirs. Understanding the accessible sequence diversity (also referred to as the functional sequence space)

of a viral glycoprotein, especially under selective pressures, helps in understanding and predicting the past and future of a virus's natural evolution. In this review, we introduce and discuss deep mutational scanning as a generalizable methodology for learning about viral glycoproteins and their host receptors, in particular highlighting how the information impacts therapeutic and vaccine development.

DEEP MUTATIONAL SCANNING AS A GENERALIZABLE “BIG DATA” TECHNOLOGY

For decades, understanding the effects of mutations on a protein's structure and function has involved targeted mutagenesis and the individual characterization of unique sequence variants. Much of what the field understands about protein sequence and its impacts on folding, stability, and function, originates from this classical approach. However, targeted mutagenesis is limited in scope by practical considerations, and is generally focused on a small number of candidate mutations that are hypothesized to be critical for activity, often based on similarities or differences with homologs. When critical residues in a protein sequence are unknown, an unbiased approach is needed. This led to the application of amino acid scanning mutagenesis, in which residues in the protein sequence are systematically and sequentially substituted to a specific amino acid. Most often this is by alanine substitutions (Cunningham and Wells, 1989), which illustrates the effect of removing an amino acid's side chain, but other amino acids have also been explored as alternatives for scanning mutagenesis (i.e., using aspartic acid or glutamic acid for finding ligand binding interfaces) (Gray et al., 2017). Scanning mutagenesis has become a staple method for finding important functional or structural sites and has been extensively applied to viral glycoproteins and their receptors. This is well demonstrated by the example of human immunodeficiency virus 1 (HIV-1). The surface glycoprotein spike of HIV-1, Env, is a trimer that binds its primary receptor CD4 on a host cell, triggering conformational changes in Env that expose a binding site for a co-receptor, CCR5 or CXCR4. Co-receptor binding is followed by further conformational changes that mediate fusion of the HIV-1 envelope and host cell membranes (Merk and Subramaniam, 2013; Wang et al., 2020). This system has been extensively interrogated by alanine scanning, both in the viral spike (Lu et al., 2001; Jacobs et al., 2005; Jiang and Aiken, 2007; Walker et al., 2009, 2011; Sen et al., 2010; Li et al., 2011) and its receptors (Ashkenazi et al., 1990; Dragic et al., 1998; Rabut et al., 1998; Chabot et al., 1999), identifying key residues for expression, stability, physical interactions, conformational dynamics, cell entry, and interactions with monoclonal antibodies. Alanine substitution mutants have been quantitatively characterized in exceptional detail, such as determining changes in Env expression, proteolytic processing into its subunits, virus incorporation, receptor binding, and infectivity in culture. However, these scanning mutagenesis experiments do not fully account for the dependency on the chemical properties of an amino acid present at a certain

position (Haddox et al., 2016; Dingens et al., 2017); findings may be very different depending on which of the 20 amino acids is chosen for a substitution. In contrast, deep mutational scanning may test all possible amino acid substitutions at each position, demonstrating how side chain properties and possible steric or electrostatic clashes influence the protein's structure and function. This increase in scale is achieved by forgoing individual characterization of each mutation and instead using *in vitro* selections of variant libraries coupled with next generation sequencing to determine the effects of thousands of mutations in a single experiment.

In the past decade, deep mutational scanning has provided a more complete picture of a protein's mutational landscape under controlled experimental conditions. In a typical deep mutagenesis experiment, a library is constructed that encompasses the mutations of interest, often by incorporating degenerate codons at each position in the cDNA sequence to encode all possible single amino acid substitutions. The variant library is then exposed to *in vitro* selection and changes in variant frequency are measured by deep sequencing. In this way, it becomes possible to characterize thousands of mutations simultaneously from a single selection experiment (Fowler et al., 2010; Fowler and Fields, 2014). Deleterious mutations are depleted while beneficial mutations are enriched, and the enrichment ratio for a given sequence variant acts as a proxy for relative phenotype. This qualitatively defines the phenotypic effects of each mutation at diversified positions in the variant library. If a series of parallel selections are performed with increasing stringency [for example, cells expressing libraries of protein variants are collected in parallel for increasingly higher expression levels (Matreyek et al., 2018) or for ligand binding at increasingly lower ligand concentrations (Adams et al., 2016)], then it is also possible to quantitatively determine phenotypic changes for a mutant protein based on trends going from low to high selection stringency. Deep mutagenesis can also be performed without using an *in vitro* selection, such as when protein activity is linked to transcription of a barcoded reporter (Jones et al., 2020), but these cases are not considered in this review where the focus is on selections of viral glycoproteins and their receptors.

The disadvantage of deep mutagenesis is that scale may be achieved at the expense of data accuracy, especially as the diversity of variants in the library becomes too large to be sufficiently sampled or the selections lack stringency to discriminate between variants of differing activities. This is again highlighted by an example from the study of HIV-1, where a deep mutational scan of Env based on virus infectivity in culture showed partial agreement with previous targeted mutagenesis of the receptor binding sites (Olshevsky et al., 1990; Basmaciogullari et al., 2002), but no correlation with prior alanine scanning in the Env gp41 subunit (Jacobs et al., 2005; Sen et al., 2010). There was also little correlation to natural sequence diversity in circulating strains (Haddox et al., 2016). These issues were due to low selection stringency that allowed an uncharacteristically high number of Env variants to persist during passaging, in addition to high noise in the data emphasized by poor agreement between independent experimental replicates

(Haddox et al., 2016). However, despite these deficits, sequence features were still apparent, such as lower mutational tolerance in epitopes for broadly neutralizing antibodies (Haddox et al., 2016). Comparisons of earlier deep mutational scans of HIV-1 Env (Haddox et al., 2016; Heredia et al., 2019) with more recent investigations of SARS coronavirus 2 (SARS-CoV-2) spike and its receptor (Chan et al., 2020; Starr et al., 2020) have shown dramatic improvements in data quality, due to changes in selection strategies toward surface display technologies and more efficient sampling of smaller libraries. These more recent works withstood extensive validation by targeted mutagenesis of selected individual mutations. Furthermore, as discussed below, confidence in experimental mutational data is improved through the use of algorithms that bring other sources of information to bear on the problem, such as conservation among homologous sequences and consideration of chemico-physical properties of the amino acids being substituted (Weile et al., 2017; Shamsi et al., 2020). The relevance of deep mutagenesis to understanding viral glycoproteins and their receptors is therefore expected to grow as the technology matures.

Different selection strategies open possibilities for deep mutational scanning as a generalizable tool for studying different protein properties and activities, from solubility-activity relationships for protein engineering (Klesmith et al., 2017; Wrenbeck et al., 2017; Gupta and Varadarajan, 2018) to examining the molecular determinants of amyloid- β aggregation (Gray et al., 2019) to exploring pathogenic variants in the human genome (Stein et al., 2019). Moreover, deep sequencing has been used to analyze library selections of increasing sophistication and creativity. This has included selections for folded structure based on protease sensitivity (Rocklin et al., 2017), co-trafficking of subunits in oligomeric complexes (Park et al., 2019), protein-protein interactions both inside and outside the cell (McLaughlin et al., 2012; Procko et al., 2014; McShan et al., 2019), signaling (Jones et al., 2020), and processing of endogenous substrates (McShan et al., 2020), among other examples. In order to better combat current and new strains of circulating viruses, deep mutagenesis studies seek to mimic aspects of virus evolution in a laboratory setting, providing insights into the accessible sequence diversity for genetic drift, immune escape, and drug resistance. Deep mutational scanning reveals amino acid preferences of viral glycoproteins and how those preferences shift in the context of applied selective pressures, such as antibodies or drugs, and can further inform structural understanding and rational design of therapeutic or prophylactic interventions, for example, by revealing vulnerable epitopes.

SELECTION STRATEGIES FOR UNDERSTANDING SEQUENCE DIVERSITY AND EVOLUTION OF VIRAL GLYCOPROTEINS

The requirements for deep mutagenesis are (i) a library of diverse sequence variants, (ii) a suitable host that links phenotype to genotype, and (iii) a selection strategy (Fowler et al., 2010;

Fowler and Fields, 2014). The simple fulfillment of these requirements by using viruses and infected cells as the hosts and virus replication in culture as the selection (**Figure 1**) have propelled deep mutagenesis to the forefront of virology within the past decade. The general approach of tracking replication of virus variants through deep sequencing has been used to extensively characterize the glycoproteins of two viruses highly relevant to public health: HIV-1 (Haddox et al., 2016, 2018; Dingens et al., 2017, 2019a,b) and influenza A (Thyagarajan and Bloom, 2014; Wu et al., 2014, 2017a, 2020; Doud and Bloom, 2016; Canale et al., 2018; Lee et al., 2018). There have also been many efforts within the past few years to apply this selection strategy and others more broadly to other viruses, including murine leukemia virus (MLV) (Salamango et al., 2016), Zika virus (ZIKV) (Gong et al., 2018; Setoh et al., 2019; Sourisseau et al., 2019), and most recently, SARS-CoV-2 (Greaney et al., 2020; Linsky et al., 2020; Starr et al., 2020; Chan et al., 2021). Below, we introduce and discuss examples of selection strategies used for deep mutagenesis of viral glycoproteins.

To determine how mutations impact replicative fitness, viral libraries are first prepared by transfecting a virus production line with plasmids encoding variants of the spike protein and transducing with any helper virus as necessary. The produced virus particles in the supernatant are then passaged through a permissive cell line expressing the relevant host receptors (**Figure 1**). Infectious virus variants are enriched, while deleterious ones are depleted. This is the selection strategy used in the aforementioned study of HIV-1 Env, finding residues crucial for expression, folding, and receptor binding despite data noise and low selection stringency (Haddox et al., 2016). Similarly, the observed phenotypes of influenza A hemagglutinin (HA) variants using the same selection strategy of virus replication in culture were consistent with reported phenotypes of previously characterized mutants (Wu et al., 2014). For influenza A, spikes composed of three HA subunits on the viral envelope bind sialylated glycans on host cells to mediate endocytosis, followed by acid-induced conformational changes of the HA trimer within the endosome that drive fusion of viral and cell membranes (Dou et al., 2018; Russell et al., 2018). Whether through single-nucleotide mutations (Wu et al., 2014) or through the full breadth of mutations available at each codon (Thyagarajan and Bloom, 2014), the studies elucidated that influenza HA possesses a high inherent mutational tolerance which, in conjunction with external selective pressures, drives the rapid evolution of the virus. The selection strategy illuminates the constraints on a glycoprotein's inherent evolutionary capacity, at least in the context of replication in cell culture, creating specific sequence-fitness links at each site. However, virus replication requires the surface glycoproteins to be properly folded, incorporate into virus particles, bind host entry receptors, and undergo complex dynamic changes that drive fusion of viral and host cell membranes, and it is difficult to know which process is disrupted by a deleterious mutation.

Multiple selection strategies can be implemented for the same deep mutagenesis library to determine the effects of mutations separately on different properties of a viral fusion protein or glycoprotein. Salamango et al. (2016) utilized three different

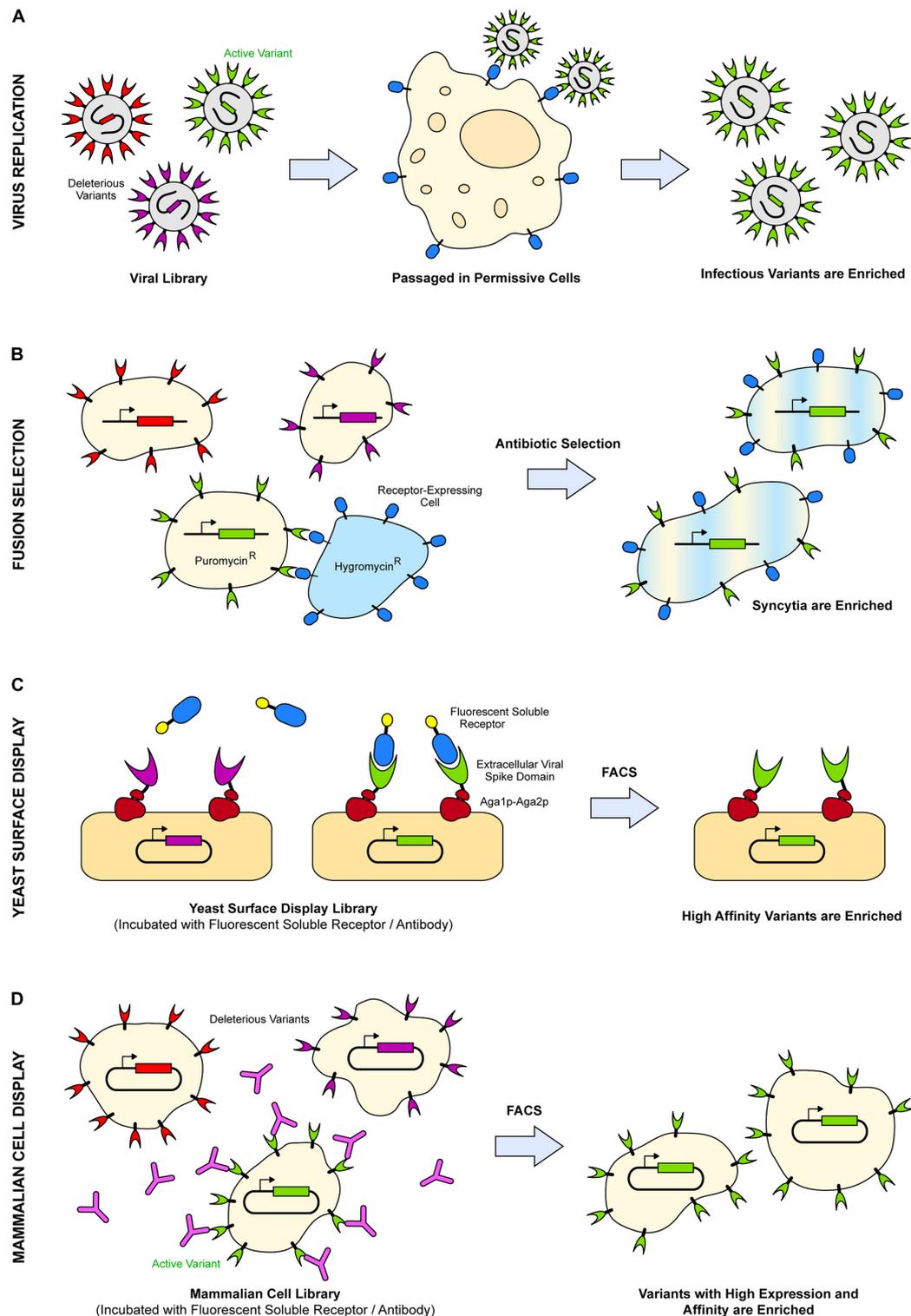


FIGURE 1 | Examples of selection strategies used in deep mutational scans of viral spike proteins. **(A)** Infectious virus variants (green) are enriched after passing through a permissive cell line expressing host target receptors (blue). **(B)** Syncytia are enriched upon fusion of cells expressing variants of viral fusion protein (pale yellow cells) and cells expressing the entry receptor (pale blue cells) in the presence of two different antibiotics. **(C)** Soluble extracellular domains of the viral spike are displayed on the yeast cell wall via Aga1p–Aga2p (dark red). After fluorescence activated cell sorting (FACS), these viral proteins (green) are enriched if they possess high binding affinity and/or expression to fluorescent partners, such as soluble receptors (blue). **(D)** After FACS, mammalian cells expressing viral spike proteins (green) are enriched if they are highly expressed and bind with high affinity to fluorescently labeled antibodies (pink) or soluble receptors.

selection strategies on deep mutational scans of MLV Env for infectivity, Env fusion activity, and incorporation into viral particles. The mutational landscapes for the strategies were compared to filter out how specific sites and their mutations influence each selected property. Since the fusion selection strategy was decoupled from the formation and passaging of virus particles, mutations enriched for Env fusion may arise that do not correlate with infectious, well-formed particles. For the fusion selection, a library of mammalian cells expressing Env variants were incubated with a cell line expressing the entry receptor. The two cell populations had resistance to two different antibiotics, such that the fused syncytia were selected in the presence of both antibiotics (**Figure 1**). Mutations in MLV Env were discovered with defective infectivity because of poor incorporation, despite the Env mutants remaining active for membrane fusion. These MLV Env mutants were poorly incorporated into pseudotyped HIV-1 particles, possibly due to changes in lipid interactions at assembly sites (Salamango et al., 2016). Pseudotyping allows for the manipulation of cell type tropism through the incorporation of envelope glycoproteins from one virus into alternative virus backgrounds. It allows for the ability to study fusion proteins from highly virulent strains in backgrounds of reduced safety risk. In addition to a diverse range of clinical and non-clinical applications, pseudotyped viruses can be used to validate predictions from deep mutagenesis studies, as was done with the mutational landscape of SARS-CoV-2 spike glycoprotein S (Greaney et al., 2020; Starr et al., 2020).

Another selection strategy decoupled from virus replication is yeast surface display. It involves the expression of a soluble protein fragment, such as an extracellular domain of a viral spike, fused with yeast adhesion factor, Aga2p, which complexes with Aga1p for display on the yeast cell wall (Gai and Wittrup, 2007). Yeast transformed with libraries encoding protein variants are fluorescently labeled for detection of expression or activity and sorted by fluorescence-activated cell sorting (FACS) (**Figure 1**). For viral protein variants, surface expression levels can be detected using epitope tags, while binding activity is assessed with fluorescently labeled antibodies or soluble receptors. Yeast display has a fast turnaround for generating data, and unless there is a pressing need to understand the effects of viral glycoprotein mutations on infectivity, it presents a powerful approach for the rapid development and testing of therapeutics (Gaiotto and Hufton, 2016; Linsky et al., 2020; Starr et al., 2020). The benefits of yeast surface display for quickly assessing mutational tolerance were demonstrated in the months following the emergence and global spread of SARS-CoV-2. SARS-CoV-2 spike protein S forms a trimeric fusion protein that engages angiotensin converting enzyme 2 (ACE2) on host cells as an entry receptor (Hoffmann et al., 2020; Lan et al., 2020; Wrapp et al., 2020; Yan et al., 2020; Zhou et al., 2020). The binding site is located on a receptor-binding domain (RBD) of S, which was shown through a deep mutational scan of the RBD expressed on the yeast surface to have high mutational tolerance with respect to both expression and ACE2 binding affinity (Starr et al., 2020). The findings were confirmed by a later deep mutational scan of the RBD in the context of full-length trimeric S expressed on human cells (Chan et al., 2021) and align with comparative bioinformatics

of S proteins from SARS-associated betacoronaviruses isolated from bats, where diversity within the ACE2-binding site of the RBD is surprisingly high (Frank et al., 2020). This diversity is possibly due to an “arms race” between S and ACE2 co-evolution; ACE2 is highly polymorphic across different bat species (Guo et al., 2020).

Mutations at several sites in the SARS-CoV-2 RBD, but especially N501, were predicted by deep mutational scanning and confirmed by targeted mutagenesis to enhance binding to the ACE2 receptor (Starr et al., 2020; Chan et al., 2021), potentially creating opportunities for the virus to become more infectious or partially resistant to therapeutics blocking the RBD-ACE2 interaction. The N501 site has since gained notoriety for emerging in at least three separate variant lineages with increased transmissibility, B.1.351 in South Africa (Tegally et al., 2020), P.1 in Brazil (Faria et al., 2021), and B.1.1.7 in England (Leung et al., 2021; Volz et al., 2021). Even though much is still unconfirmed about the characteristics of these variants, their emergence lends support to the predictive power of deep mutational scanning. The deep mutagenesis data, in combination with the observed natural diversity among SARS-related coronaviruses, raise concerns that the spike sequence has the capacity to drift substantially, potentially causing changes in dominant antigenic epitopes and escape from immunity (Wibmer et al., 2021). SARS-CoV-2 has a moderate mutation rate estimated at 10^{-3} substitutions per site per year (Candido et al., 2020), but has shown rapid accumulation of mutations in farmed mink (Oude Munnink et al., 2021) and an immunocompromised patient (Choi et al., 2020).

Major limitations of yeast display are (i) that only soluble extracellular domains are expressed that lack transmembrane regions, (ii) that some complex proteins do not properly fold on the yeast surface, and (iii) that yeast lack terminally sialylated *N*-glycans found on human membrane proteins (Wildt and Gerngross, 2005), which can impact interactions with glycan-dependent antibodies (Cohen et al., 2015). Although yeast display is advantageous for rapid characterization of viral glycoproteins, it notably excludes considerations required for expression and folding in human cells. A selection strategy that accounts for these considerations is mammalian cell surface display. Genes encoding spike variants are expressed in mammalian cells using transfection or transduction conditions that yield one protein variant per cell (Forsyth et al., 2013; Steichen et al., 2016; Bruun et al., 2017; Kulp et al., 2017; Matreyek et al., 2017, 2020; Heredia et al., 2018; Procko, 2020). This links the cell phenotype to a single variant genotype. Cells are incubated with fluorescently labeled antibodies or soluble receptors and sorted by FACS, so cells expressing spike variants with high expression and high binding affinity to the fluorescent partner are enriched (**Figure 1**). In addition to deep mutational scanning of the isolated SARS-CoV-2 RBD by yeast display, it has been scanned in the context of full-length S expressed at the plasma membrane in human cells (Chan et al., 2021). The effects of mutations were qualitatively similar to results from yeast surface display, and the two methods therefore reinforce the major conclusion that the viral spike is able to tolerate high mutational diversity while maintaining expression and

ACE2 binding activity (Starr et al., 2020; Chan et al., 2021). Conflicts between the data sets were mostly confined to a small number of RBD residues that are buried in the major prefusion conformation of S yet are exposed when the isolated RBD is expressed on yeast, as well as higher mutational tolerance for receptor binding in the human cell data set, possibly due to differences in avid binding of dimeric ACE2 receptors between the two systems. Both studies ignore mutations in the viral spike outside the RBD that may influence escape from antibodies or modulate receptor binding through allostery or epistasis, for example, by increasing dynamic exposure of the RBD for receptor recognition as occurs in the D614G virus variant (Ozono et al., 2021; Xu et al., 2021). Future work should be dedicated toward understanding the mutational landscape of the entire S glycoprotein for folding/expression, ACE2 binding, infectivity, and interactions with monoclonal antibodies targeting domains other than the RBD. Indeed, previous work with HIV-1 Env has shown how mutations distal from the binding interface impact receptor recognition through conformational effects (Heredia et al., 2019).

PREDICTING VIRUS ESCAPE FROM ANTIBODIES AND DRUGS THROUGH DEEP MUTATIONAL SCANNING

External selective pressures influence variant fitness and spike protein adaptation and are, thus, required for recapitulating aspects of natural evolution in deep mutagenesis studies. Furthermore, viruses can adapt to prevailing selective pressures from therapeutics or the immune system, posing a challenge to therapeutic and vaccine development. By performing selections *in vitro* on comprehensive variant libraries, one can search for escape mutations accessible to the virus. Expression systems of viral proteins, whether it be viruses (Doud and Bloom, 2016; Dingens et al., 2017, 2019a,b; Wu et al., 2017a; Phillips et al., 2018; Sourisseau et al., 2019), yeast (Gaiotto and Hufton, 2016; Greaney et al., 2020; Linsky et al., 2020), or mammalian cell surface display (Chan et al., 2021), are incubated with therapeutics before passaging through cell culture for the former and before FACS sorting for the latter two (Figure 1). However, escape mutations predicted by surface display strategies may not account for a loss of replicative fitness, unlike selections with live virus, although they permit a more comprehensive assessment of mutational effects without the constraint of viral replication. For example, while the influenza A H1 HA stem is moderately conserved in studies with live virus, yeast displayed HA identified mutations at and around the fusion peptide in the HA stem that preserved expression but led to a reduction in nanobody binding. The fusion peptide is an essential motif for mediating virus-cell membrane fusion, so mutations in the vicinity of this site, while reducing nanobody binding, also reduce infectivity (Gaiotto and Hufton, 2016). An approach to address the pleiotropic effects of mutations has been to test surface displayed libraries of viral spikes for binding to host receptors versus antibodies (or antibody-like molecules), thereby isolating putative escape variants that

selectively lose antibody affinity while maintaining tight receptor recognition (Greaney et al., 2020; Linsky et al., 2020; Chan et al., 2021; Starr et al., 2021). This strategy has isolated mutations in S of SARS-CoV-2 that mediate escape from monoclonal antibodies used clinically for the treatment of COVID-19 (Starr et al., 2021). Epitopes that are well-established as sites for antibody neutralization may have a high capacity to tolerate mutations. Therefore, mapping escape mutations to atomic-resolution structures defines important sites for the improvement of a therapeutic's design and identifies virus variants that might yet emerge in nature if the therapeutic is widely used.

Deep mutational scans in the presence of different external pressures can reveal conserved sites for the rational design of therapeutics, such as universal antibodies (Doud and Bloom, 2016; Gaiotto and Hufton, 2016; Phillips et al., 2018), and can support creative treatment strategies to restrict a virus's capability for escape. For example, by comparing potential escape mutations specific to different antibodies, mixtures of antibodies can be chosen that bind non-competing epitopes and do not share common escape mutations. These antibody cocktails exploit the orthogonality of escape mutations to suppress the emergence of resistance (Gaiotto and Hufton, 2016; Dingens et al., 2019b; Greaney et al., 2020). While escape mutations to antibodies are often found at the epitope, deep mutational scanning also identifies escape mutations at sites that do not interact directly with an antibody but rather influence binding through allostery and conformational shifts. A deep mutational scan focused on two loops of HIV-1 Env that interact with CD4 found mutations that increase virus fitness in culture by changing Env conformational dynamics (Duenas-Decamp et al., 2016). Trimeric Env exists in an equilibrium of conformational isomers and undergoes a complex dynamic process of structural changes during fusion. Mutations that alter these dynamics provide a mechanism for escape from conformation-dependent antibodies (Dingens et al., 2019b). Mutations to residues that are separated by a small number of structural contacts to the epitope, or mutations that alter glycosylation patterns, also provide opportunities for escape from bnAbs (Dingens et al., 2019b). In contrast to antibodies, engineered decoy receptors resemble the host receptors, aside from mutations to improve binding or specificity to the viral glycoproteins of interest. The viral spikes may not be able to develop escape mutations without a simultaneous loss of affinity to their membrane-bound target receptors. This has already been demonstrated in human cells expressing a library of RBD-focused variants in full-length SARS-CoV-2 S, sorted in the presence of competing wildtype and engineered ACE2 decoy receptors. Escape mutations in S discriminating against an engineered, high affinity, soluble decoy receptor weren't found, and if used as a therapeutic, the virus is unlikely to become resistant (Chan et al., 2021). However, viruses are not limited to single substitutions for escape mutations, and epistatic relationships can be complex when multiple sites are mutated together. These kinds of studies are therefore best considered as supplementing, rather than replacing, classical selections for escape variants, in which viruses are passaged in the presence of antiviral agents (e.g., antibodies, drugs) to promote

the emergence of resistance on an accelerated time scale *in vitro* (Fellinger et al., 2019; Baum et al., 2020; Higuchi et al., 2020).

CONSIDERATIONS OF EPISTASIS, PHYLOGENY, AND CELL TYPE IN DEEP MUTATIONAL SCANNING

Single substitutions in deep mutational scanning of viral proteins cannot capture all the possibilities for immune escape and natural evolution. Early deep mutagenesis studies of influenza HA primarily focused on single amino acid substitutions and did not account for epistasis when multiple sites are mutated in combination, which can lead to shifts in amino acid preferences and alter the tolerance of sites (Hilton and Bloom, 2018). These scans reveal many single substitutions that cause deleterious effects on expression or receptor binding, but functionality and fitness might recover when substitutions are combined. For example, mutations in a loop at the receptor binding site of HA show extensive epistasis, including cases where two mutations that are deleterious on their own are neutral when combined together (Wu et al., 2017a). These epistatic effects substantially increase the functional sequence space compared to what additive effects of single mutations would predict. Epistasis between residues demonstrates a fundamental limit to focusing on single substitutions in deep mutational scanning to accurately predict how viruses might respond to new pressures, but this focus is often experimentally necessary to reduce library diversity for sufficient sampling of variants during selection. In the study by Wu et al. (2017a), library diversity was instead constrained by only allowing combinations of mutations at a small subset of residues.

As amino acid substitutions accumulate in viral glycoproteins in response to selective pressures, epistatic interactions can modulate the local mutational landscape and may shift amino acid preferences, stabilizing previously deleterious mutations (Wu et al., 2020). This may lead to the entrenchment of these substitutions and may even cause wildtype reversion to become unfavorable (Haddox et al., 2018). If strains are separated over a period of time, the accumulation of sequence differences can lead to substantial divergence between their mutational landscapes (Wu et al., 2020). Deep mutational scans of Env proteins from different HIV-1 strains demonstrate, despite substantial overlap, that amino acid preferences at some sites are strain-specific and therefore highly dependent on the background sequence (Haddox et al., 2018; Heredia et al., 2019). Strain-specific differences shape both antigenicity and mutational tolerance, so escape mutations from antibodies may be strain-specific as well. In the case of rapidly mutating viruses, like influenza A, understanding mutational tolerance of the glycoprotein may be context dependent because of evolutionary differences between subtypes. Within the phylogenetic tree of influenza HA, all subtypes share a highly conserved structure and perform the same function of binding to sialic acids on host cells. However, amino acid sequence identity between subtypes can be as low as 38% (Lee et al., 2018) and the subtypes have distinct amino acid preferences (Canale et al., 2018; Hilton and Bloom, 2018). For instance, in a deep mutational scan of HA subtype

H1, the head domain is more mutationally tolerant than the stalk/stem domain, but a deep mutational scan of HA subtype H3 reveals the reverse (Lee et al., 2018). The implication is that studying the molecular evolution of one subtype is less useful for understanding the evolution and predicting escape mutations in others, an important caveat that may extend into comparing variants within the same subtype.

By carefully choosing the cell line that acts as the host for virus replication, deep mutagenesis can identify mutations in viral proteins that influence host interactions and adaptation (Ashenberg et al., 2017; Setoh et al., 2019; Shirleen Soh et al., 2019; Sourisseau et al., 2019). Zika virus (ZIKV) has received significant interest in recent times due to its broad tissue tropism that permits placental transmigration and neurological defects in a developing fetus (Castanha and Marques, 2020). Compared to HIV-1 Env and influenza HA, much less is known about the natural evolution of ZIKV envelope (E) protein, and host entry receptors and attachment factors remain unclear (Sirohi and Kuhn, 2017). A deep mutational scan of ZIKV E protein has indicated that its surface-exposed regions, except for the fusion loop and hypervariable glycan loop, display higher mutational tolerance in the experimental scan compared to alignments of natural sequence variants. ZIKV E protein is likely less tolerant of mutations in nature where it faces stronger selective pressures during replication in hosts that actively mount immune responses, in addition to undergoing part of its replication cycle in *Aedes* species mosquitoes (Sourisseau et al., 2019). Furthermore, most natural ZIKV variants were sequenced from human patients in just the last few years and are therefore closely related on a short evolutionary timescale; the available functional sequence space will therefore be under-sampled in natural isolates. Other deep mutational scans have identified E variants that specifically enhance ZIKV replication in either mosquito or primate cell lines, possibly due to the effects of mutations on temperature-sensitive structural transitions (replication in insect cells occurs at a lower temperature) (Setoh et al., 2019) or the removal of glycans that specifically aid human cell infection via the lectin DC-SIGN (Gong et al., 2018). Performing deep mutagenesis studies of viruses with zoonotic origins in multiple cell types can inform our understanding of how adaptation and evolution leads to spillover from animal reservoirs.

DEEP MUTAGENESIS-GUIDED IMMUNOGEN ENGINEERING

Deep mutational scanning of viral glycoproteins not only provides insight into possible evolutionary pathways the virus may take, especially in the presence of drugs and antibodies, but it may also guide the engineering of the glycoproteins as optimized immunogens to promote broad and effective immunity. Subunit vaccines are composed of purified viral glycoproteins or their components, which stimulate the clonal expansion, affinity maturation, and immunoglobulin class switching of naive B cells expressing low affinity receptors. It is often critical that the immunogenic glycoproteins are conformationally stable and pure, and if necessary, bind rare germline B cell receptors that are able to mature into the most potent and broad

neutralizing antibodies (Medina-Ramírez et al., 2017; Stamatatos et al., 2017; Ringel et al., 2018). These concepts have been particularly well demonstrated in the deep mutational scanning and engineering of immunogens for HIV-1, where vaccines have thus far failed to elicit a broadly protective humoral response against diverse virus strains.

Human immunodeficiency virus 1 Env adopts a range of conformational states, with most broadly neutralizing antibodies (bnAbs) targeting a so-called “closed” conformation in which strain-specific epitopes are hidden (Medina-Ramírez et al., 2017; Wang et al., 2020). Researchers have therefore placed priority on engineering soluble Env proteins that fold as stable closed trimers for optimum presentation of epitopes that may elicit broad protection. By having selections based on high affinity binding to conformation-dependent bnAbs, deep mutational scans of Env expressed on human cells have been able to identify mutations that stabilize the closed trimeric state and optimally present bnAb epitopes (Steichen et al., 2016; Kulp et al., 2017; Heredia et al., 2019).

Broadly neutralizing antibodies targeting HIV-1 Env are extensively mutated from their germline antibody genes, which have low affinity for Env and are therefore poorly primed by standard Env immunogens (Stamatatos et al., 2017). To overcome this barrier toward eliciting bnAbs in a naive individual, deep mutagenesis of Env trimers expressed on human cells (Steichen et al., 2016; Kulp et al., 2017) or of a highly engineered Env fragment displayed on yeast (Jardine et al., 2016) have aided the engineering of immunogens that tightly bind and prime naive B cells expressing relevant germline antibody precursors. One of these engineered immunogens is being evaluated in a phase I clinical trial (ClinicalTrials.gov Identifier: NCT03547245). Indeed, it is possible to engineer a series of immunogens that coax antibody maturation down a lineage toward potent and broad HIV-1 neutralization (Escolano et al., 2016; Steichen et al., 2016). The first immunogens in the vaccine regimen have the highest number of mutations from native Env and bind tightly to the necessary germline B cell receptors. Subsequent immunogens have successively fewer mutations and more closely resemble native Env. At each step, B cells are activated that bind engineered immunogens that ultimately match Env on native virus particles.

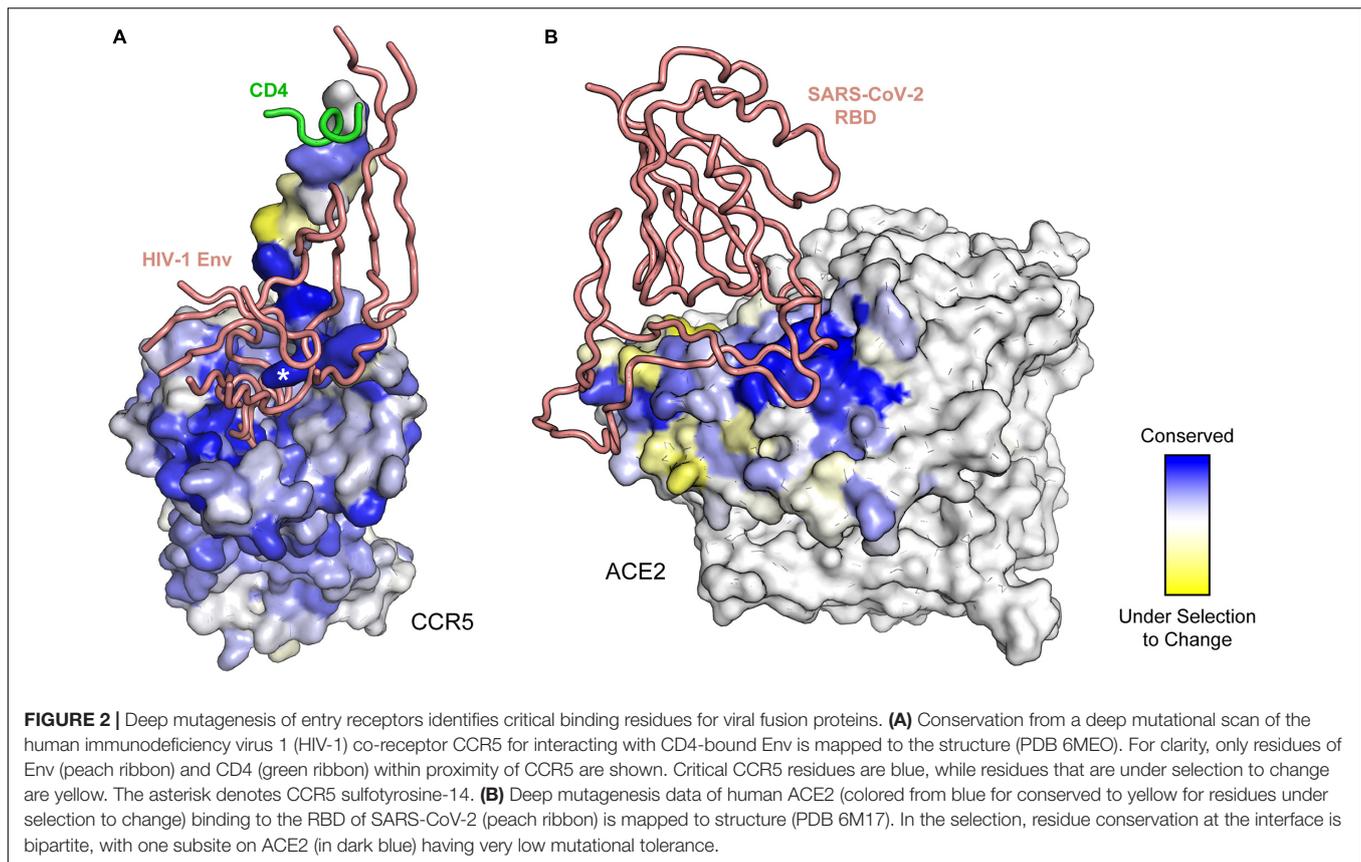
DEEP MUTAGENESIS OF HOST RECEPTORS: IMPLICATIONS OF HUMAN POLYMORPHISMS AND ENGINEERED RECEPTORS AS THERAPEUTICS

In examinations of viral fitness, one cannot exclude the essential roles that host proteins play and how host genetic diversity may also impact infection and disease. Less attention has been given to deep mutational scanning of host proteins due to experimental challenges, since the underlying selection strategies can no longer be based on passaging of viral libraries. Instead, sequence diversity must be encoded in libraries of the host proteins that are expressed in mammalian cells or

at the surface in yeast or mammalian cell display platforms. Host proteins with dedicated roles fighting viruses, such as antiviral restriction factors, have undergone co-evolution with viruses, a challenging feat due to vast differences in host versus virus evolutionary timescales. A deep mutational scan of TRIM5 α , an antiviral restriction factor, has shown that its viral capsid-binding loop exhibits a very high mutational tolerance, which supports the main purpose of this unstructured loop to rapidly evolve against emergent retroviruses and hinder their propagation (Tenthorey et al., 2020). By comparison, host entry receptors generally have important physiological functions and are therefore often highly conserved. Instead of understanding their natural evolution, deep mutagenesis of virus entry or attachment receptors serves other important research objectives. How does receptor sequence diversity influence species tropism? Do some receptor polymorphisms make the host more or less susceptible to infection? Can mutational landscapes of receptors guide structural modeling and therapeutic engineering? These are questions that have only begun to be explored and offer rich areas for future investigation.

Two deep mutational scans of human ACE2, one from full-length protein expressed in human cells (Chan et al., 2020) and the second from surface display of the isolated ACE2 protease domain on yeast (Heinzelman and Romero, 2020), provide a near-comprehensive overview of how single substitutions contribute to SARS-CoV-2 S protein affinity. The deep mutational scans confirmed the structurally characterized S binding interface (Lan et al., 2020; Yan et al., 2020) and identified distal sites that affect binding, implicating conformational dynamics associated with ACE2 enzymatic activity in S affinity (Heinzelman and Romero, 2020). The mutational landscapes may contribute to our understanding of how ACE2 polymorphisms affect susceptibility to COVID-19 and for identifying at risk groups. ACE2 has a number of allelic variations within or near to the viral spike binding site. Because it is an X-linked gene, ACE2 polymorphisms may be of particular consequence in males (Gemmati et al., 2020). However, any predictions will need to be validated clinically by genetic studies.

Deep mutational scanning can be used to map protein-protein interfaces (Figure 2) and guide structural modeling. Several groups have developed methods to glean structural information from deep mutagenesis and have shown how the data can be used to model both monomeric and oligomeric proteins (Park et al., 2019; Rollins et al., 2019; Schmiedel and Lehner, 2019; Fantini et al., 2020). These methods may be particularly useful for modeling complexes between viral glycoproteins and their host receptors when individual structures are known but the assembled complex is not, although we are unaware of any such studies at this time. Deep mutational scanning of CCR5, a co-receptor for HIV-1 Env, mapped critical receptor residues for Env binding (Heredia et al., 2018), but the data were of insufficient quality to enable accurate modeling. However, a retrospective comparison of the deep mutagenesis data to the cryo-EM structure, which was later solved (Shaik et al., 2019), demonstrates excellent agreement (Figure 2A). In particular, an exposed sulfonated tyrosine (sTyr-14) side chain from the receptor becomes buried in a deep pocket of Env; sTyr-14 is one of the most highly conserved residues for binding in



the mutational scan. In another example, the structure for a trimeric glycoprotein complex from human cytomegalovirus (HCMV) binding its receptor, PDGFR α , was unknown beyond exceedingly low resolution cryo-EM images (Kabanova et al., 2016). Deep mutational scanning illustrates that HCMV trimer binding is largely resistant to single amino acid substitutions within two domains of the PDGFR α receptor (Park et al., 2020). Consequently, the PDGFR α mutational landscape failed to unambiguously highlight a surface epitope for modeling. A recently solved high-resolution structure of PDGFR-bound HCMV trimer shows why; the binding interface is extensive and spread over three receptor domains, such that disruption of interactions by any one receptor domain has minimal impact on HCMV trimer binding (Kschonsak et al., 2021).

An application of biochemical insights from deep mutagenesis is the optimization and engineering of soluble receptors as potent antivirals. Soluble decoy receptors present a unique opportunity for neutralizing viruses with little opportunity for the emergence of escape mutations. However, soluble decoy receptors can have problems with specificity and affinity that limit their efficacy *in vivo*. To again use the example of the HCMV glycoprotein trimer and its receptor PDGFR α , soluble PDGFR α ectodomain potentially blocks virus entry at low nanomolar concentrations, yet the receptor also competitively binds four endogenous host factors involved in growth factor signaling, limiting its safety and efficacy as a therapeutic (Kabanova et al., 2016; Stegmann et al., 2017; Wu et al., 2017b). Using deep mutational scanning

based on a competition selection between HCMV trimer and endogenous PDGFR α ligands, receptor mutations were screened that maintain tight virus binding and potent neutralization, while eliminating unwanted off-target interactions. These have been termed “orthogonal” receptors, as they are virus-specific and no longer participate in (i.e., are orthogonal to) the receptor’s normal biology (Park et al., 2020). This solves the specificity problem in the absence of an atomic resolution structure that could otherwise guide rational protein engineering.

In contrast to the specificity problem of soluble PDGFR α , the endogenous activity of soluble ACE2 acts as a negative regulator of the renin-angiotensin system to protect against lung injury in animal models of inflammation and infection (Imai et al., 2005; Trembl et al., 2010; Zou et al., 2014). Soluble ACE2 catalyzes the proteolytic turnover of vasoconstrictive peptide hormones, which may offer direct relief from COVID-19 symptoms, and it is under evaluation in a phase II clinical trial (ClinicalTrials.gov Identifier: NCT04335136). However, wildtype soluble ACE2 binds S of SARS-CoV-2 with only moderate nanomolar affinity (Shang et al., 2020), significantly lower than affinity matured antibodies. Optimization of soluble ACE2 as a therapeutic is therefore not a problem of specificity but of affinity. Deep mutational scanning identified multiple mutations in ACE2 that enhance S binding, which in combination allow soluble ACE2 to achieve picomolar affinity (Chan et al., 2020). The neutralization potency of the engineered decoy receptor rivals monoclonal antibodies (Chan et al., 2020) and broadly binds with tight affinity to the RBDs of

human SARS-CoV-1 and -2 as well as related bat coronaviruses (Chan et al., 2021). By resembling the natural receptor, the soluble decoy therefore achieves breadth for SARS-associated viruses.

In a similar vein, deep mutagenesis of computationally-designed proteins has been used to optimize their affinities and specificities for broad and potent antiviral activity. Small, hyperstable proteins can be designed that mimic natural entry receptors or monoclonal antibodies, but due to the immense freedom to explore structural and sequence space, the designed proteins may barely resemble the natural complexes from which they are inspired. In theory, proteins can be designed to bind any exposed surface on the viral target protein, and design can be focused to vulnerable epitopes that are conserved or essential for virus infection (Fleishman et al., 2011). Deep mutational scanning improves upon the computationally-designed proteins, which generally have only moderate affinity and require optimization (Whitehead et al., 2012). With this approach, one can optimize inhibitory “designer” proteins for high affinity and specificity to viral glycoproteins as well as orthogonality to host factors (Whitehead et al., 2012; Procko et al., 2014; Cao et al., 2020; Linsky et al., 2020). Moreover, in contrast to antibodies and soluble receptors, the designed proteins do not require expression in mammalian cells for proper folding and glycosylation, but can be generated in large quantities from bacteria (Cao et al., 2020; Linsky et al., 2020). Deep mutagenesis has been used to improve designed inhibitors targeting the proteins of influenza A (Whitehead et al., 2012), Epstein-Barr virus (Procko et al., 2014), and SARS-CoV-2 (Cao et al., 2020; Linsky et al., 2020).

LIMITATIONS OF DEEP MUTAGENESIS AND AVENUES FOR FUTURE ADVANCEMENT

Deep mutagenesis data sets have inherent noise from multiple sources: insufficient experimental sampling of variants in the library, errors with accurately replicating collection gates associated with FACS-based selections, low signal-to-noise, and/or selections that are not properly stringent to discriminate between mutants of differing activities. Furthermore, epistatic interactions are often only assessed between a small number of sites to keep library diversity manageable (Wu et al., 2017a, 2020; Zhang et al., 2020) or are missed entirely in mutational scans based on single amino acid substitutions. As an alternative, statistical and computational methods are increasingly capable of accurately predicting mutational effects. Most popular are probabilistic models for describing sequence variation and fitness based on the Potts model from statistical physics, which incorporate both site-specific constraints on amino acid identity plus all of the possible pairwise constraints that describe covariation or coupling between pairs of positions (i.e., the degree to which amino acid identities at two positions are co-dependent) (Hopf et al., 2017; Levy et al., 2017). This information is extracted from alignments of homologous sequences, with the assumption that pairwise constraints between two positions are predictive of higher order couplings up to the entire sequence. The model can be improved if additional structure

in sequence families, due to higher order epistasis not captured by pairwise constraints, is considered (Riesselman et al., 2018). The methods have accurately (and impressively) captured aspects of viral protein evolution in the clinic, especially for HIV-1, including drug resistance and escape by HIV-1 proteins from cellular and humoral immunity (Ferguson et al., 2013; Mann et al., 2014; Barton et al., 2016; Flynn et al., 2017; Louie et al., 2018; Biswas et al., 2019; Zhang et al., 2020). The pairwise couplings are critical to model accuracy, and the likelihood of an escape mutation occurring is heavily influenced by epistatic interactions with the background sequence (Barton et al., 2016). For example, if the background sequence has residues that are negatively coupled to an escape mutation, then more time is required for escape to occur as compensatory mutations must also accumulate. While surface glycoproteins have been less studied, modeling of the fitness landscape of HIV-1 Env has shown that, when considering both single site diversity and pairwise couplings, the protein surface is sparsely populated by sites that incur a large fitness penalty when mutated (Louie et al., 2018). This means antigenic epitopes tend to contain few residues that are truly conserved hidden amongst variable positions, emphasizing the difficulties for an antibody to achieve breadth.

The power of these statistical models derives from the enormous wealth of sequence information available from which to infer site-specific constraints and couplings. Where they triumph over deep mutational scanning is in their consideration of epistatic networks, yet they are also biased by sequence variation that occurs during natural evolutionary processes and may not provide mechanistic insights. However, models for predicting the effects of mutations, including but not limited to Potts probabilistic models, can be benchmarked or trained with experimental data, leading to more accurate predictions (Weile et al., 2017; Gray et al., 2018; Otwinowski et al., 2018; Riesselman et al., 2018; Saito et al., 2018; Wu et al., 2019; Yang et al., 2019; Shamsi et al., 2020). Limited experimental information from deep mutagenesis of one protein can be transferred through machine learning algorithms, which when combined with statistical models can better predict the effects of mutations within protein families (Gray et al., 2018; Shamsi et al., 2020). Overall, the use of statistical models offers solutions to problems with experimental mutational scans, namely limited information on epistasis and experimental noise in the data. This will continue to be a rich area for future development and improvements in the interpretation of deep mutational scans.

CONCLUSION

Within the past decade, deep mutational scanning has changed how researchers approach topics of virus evolution and diversity. One is no longer limited to analyses of natural sequences, targeted mutagenesis, or isolating small numbers of clones from directed evolution in tissue culture, and instead, many thousands of mutations within viral proteins can be comprehensively assessed experimentally. This technique, whether used with live virus libraries passed through cell culture, or expression of

protein variants by yeast surface display or in mammalian cells, allows for the residue-level mapping of functional interaction sites, structural modeling, and prediction of escape mutations in response to selective pressures such as antibodies, small molecule drugs, and other therapeutics. Deep mutagenesis of viral glycoprotein spikes and host receptors also supports efforts to engineer immunogens to elicit broadly protective immunity as well as develop new treatment options, such as antibody cocktails and engineered decoy receptors. We are likely only beginning to see the tip of the iceberg for what the technology can accomplish, especially as novel selection strategies are implemented that are more quantitative. We foresee many more applications for deep mutational scanning in the future

as it becomes a staple tool for exploring protein sequence landscapes.

AUTHOR CONTRIBUTIONS

Both authors drafted, edited the manuscript together, contributed to the article, and approved the submitted version.

FUNDING

This work was supported by NIH award R01AI129719 to EP.

REFERENCES

- Adams, R. M., Mora, T., Walczak, A. M., and Kinney, J. B. (2016). Measuring the sequence-affinity landscape of antibodies with massively parallel titration curves. *Elife* 5:e23156. doi: 10.7554/eLife.23156
- Ashenberg, O., Padmakumar, J., Doud, M. B., and Bloom, J. D. (2017). Deep mutational scanning identifies sites in influenza nucleoprotein that affect viral inhibition by MxA. *PLoS Pathog.* 13:e1006288. doi: 10.1371/journal.ppat.1006288
- Ashkenazi, A., Presta, L. G., Marsters, S. A., Camerato, T. R., Rosenthal, K. A., Fendly, B. M., et al. (1990). Mapping the CD4 binding site for human immunodeficiency virus by alanine-scanning mutagenesis. *Proc. Natl. Acad. Sci. U.S.A.* 87, 7150–7154. doi: 10.1073/pnas.87.18.7150
- Banerjee, N., and Mukhopadhyay, S. (2016). Viral glycoproteins: biological role and application in diagnosis. *Virusdisease* 27, 1–11. doi: 10.1007/s13337-015-0293-5
- Barton, J. P., Goonetilleke, N., Butler, T. C., Walker, B. D., McMichael, A. J., and Chakraborty, A. K. (2016). Relative rate and location of intra-host HIV evolution to evade cellular immunity are predictable. *Nat. Commun.* 7:11660. doi: 10.1038/ncomms11660
- Basmaciogullari, S., Babcock, G. J., Van Ryk, D., Wojtowicz, W., and Sodroski, J. (2002). Identification of conserved and variable structures in the human immunodeficiency virus gp120 glycoprotein of importance for CXCR4 binding. *J. Virol.* 76, 10791–10800. doi: 10.1128/jvi.76.21.10791-10800.2002
- Baum, A., Fulton, B. O., Wloga, E., Copin, R., Pascal, K. E., Russo, V., et al. (2020). Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. *Science* 369, 1014–1018. doi: 10.1126/science.abd0831
- Biswas, A., Haldane, A., Arnold, E., and Levy, R. M. (2019). Epistasis and entrenchment of drug resistance in HIV-1 subtype B. *Elife* 8:e50524. doi: 10.7554/eLife.50524
- Bruun, T. H., Grassmann, V., Zimmer, B., Asbach, B., Peterhoff, D., Kliche, A., et al. (2017). Mammalian cell surface display for monoclonal antibody-based FACS selection of viral envelope proteins. *MAbs* 9, 1052–1064. doi: 10.1080/19420862.2017.1364824
- Canale, A. S., Venev, S. V., Whitfield, T. W., Caffrey, D. R., Marasco, W. A., Schiffer, C. A., et al. (2018). Synonymous mutations at the beginning of the influenza A virus hemagglutinin gene impact experimental fitness. *J. Mol. Biol.* 430, 1098–1115. doi: 10.1016/j.jmb.2018.02.009
- Candido, D. S., Claro, I. M., de Jesus, J. G., Souza, W. M., Moreira, F. R. R., Dellicour, S., et al. (2020). Evolution and epidemic spread of SARS-CoV-2 in Brazil. *Science* 369, 1255–1260. doi: 10.1126/science.abd2161
- Cao, L., Goresnik, I., Coventry, B., Case, J. B., Miller, L., Kozodoy, L., et al. (2020). De novo design of picomolar SARS-CoV-2 miniprotein inhibitors. *Science* 370, 426–431. doi: 10.1126/science.abd9909
- Castanha, P. M. S., and Marques, E. T. A. (2020). A glimmer of hope: recent updates and future challenges in Zika vaccine development. *Viruses* 12:1371. doi: 10.3390/v12121371
- Chabot, D. J., Zhang, P.-F., Quinnan, G. V., and Broder, C. C. (1999). Mutagenesis of CXCR4 identifies important domains for human immunodeficiency Virus Type 1 X4 isolate envelope-mediated membrane fusion and virus entry and reveals cryptic coreceptor activity for R5 isolates. *J. Virol.* 73, 6598–6609. doi: 10.1128/jvi.73.8.6598-6609.1999
- Chan, K. K., Dorosky, D., Sharma, P., Abbasi, S. A., Dye, J. M., Kranz, D. M., et al. (2020). Engineering human ACE2 to optimize binding to the spike protein of SARS coronavirus 2. *Science* 369, 1261–1265. doi: 10.1126/SCIENCE.ABC0870
- Chan, K. K., Tan, T. J. C., Narayanan, K. K., and Procko, E. (2021). An engineered decoy receptor for SARS-CoV-2 broadly binds protein S sequence variants. *Sci. Adv.* 7:eabf1738. doi: 10.1126/sciadv.abf1738
- Choi, B., Choudhary, M. C., Regan, J., Sparks, J. A., Padera, R. F., Qiu, X., et al. (2020). Persistence and evolution of SARS-CoV-2 in an immunocompromised host. *N. Engl. J. Med.* 383, 2291–2293. doi: 10.1056/NEJMc2031364
- Cohen, Y. Z., Lavine, C. L., Miller, C. A., Garrity, J., Carey, B. R., and Seaman, M. S. (2015). Glycan-dependent neutralizing antibodies are frequently elicited in individuals chronically infected with HIV-1 Clade B or C. *AIDS Res. Hum. Retroviruses* 31, 1192–1201. doi: 10.1089/aid.2015.0135
- Cunningham, B., and Wells, J. (1989). High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science* 244, 1081–1085. doi: 10.1126/science.2471267
- Dingens, A. S., Arenz, D., Overbaugh, J., and Bloom, J. D. (2019a). Massively parallel profiling of HIV-1 resistance to the fusion inhibitor enfuvirtide. *Viruses* 11:439. doi: 10.3390/v11050439
- Dingens, A. S., Arenz, D., Weight, H., Overbaugh, J., and Bloom, J. D. (2019b). An antigenic atlas of HIV-1 escape from broadly neutralizing antibodies distinguishes functional and structural epitopes. *Immunity* 50, 520–532.e3. doi: 10.1016/j.immuni.2018.12.017
- Dingens, A. S., Haddox, H. K., Overbaugh, J., and Bloom, J. D. (2017). Comprehensive mapping of HIV-1 escape from a broadly neutralizing antibody. *Cell Host Microbe* 21, 777–787.e4. doi: 10.1016/j.chom.2017.05.003
- Dou, D., Revol, R., Ostbye, H., Wang, H., and Daniels, R. (2018). Influenza A virus cell entry, replication, virion assembly and movement. *Front. Immunol.* 9:1581. doi: 10.3389/fimmu.2018.01581
- Doud, M. B., and Bloom, J. D. (2016). Accurate measurement of the effects of all amino-acid mutations on influenza hemagglutinin. *Viruses* 8:155. doi: 10.3390/v8060155
- Dragic, T., Trkola, A., Lin, S. W., Nagashima, K. A., Kajumo, F., Zhao, L., et al. (1998). Amino-terminal substitutions in the CCR5 coreceptor impair gp120 binding and human immunodeficiency virus type 1 entry. *J. Virol.* 72, 279–285. doi: 10.1128/jvi.72.1.279-285.1998
- Duenas-Decamp, M., Jiang, L., Bolon, D., and Clapham, P. R. (2016). Saturation mutagenesis of the HIV-1 envelope CD4 binding loop reveals residues controlling distinct trimer conformations. *PLoS Pathog.* 12:e1005988. doi: 10.1371/journal.ppat.1005988
- Escolano, A., Steichen, J. M., Dosenovic, P., Kulp, D. W., Golijanin, J., Sok, D., et al. (2016). Sequential immunization elicits broadly neutralizing anti-HIV-1 antibodies in Ig knockin mice. *Cell* 166, 1445–1458.e12. doi: 10.1016/j.cell.2016.07.030
- Fantini, M., Lisi, S., De Los Rios, P., Cattaneo, A., and Pastore, A. (2020). Protein structural information and evolutionary landscape by in vitro evolution. *Mol. Biol. Evol.* 37, 1179–1192. doi: 10.1093/molbev/msz256

- Faria, N. R., Claro, I. M., Candido, D., Franco, L. A. M., Andrade, P. S., Thais, M., et al. (2021). *Genomic Characterisation of an Emergent SARS-CoV-2 Lineage in MANAUS: Preliminary Findings*. *Virological*. Available online at: <https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586> (accessed February 12, 2021).
- Fellinger, C. H., Gardner, M. R., Weber, J. A., Alfant, B., Zhou, A. S., and Farzan, M. (2019). eCD4-Ig Limits HIV-1 escape more effectively than CD4-Ig or a broadly neutralizing antibody. *J. Virol.* 93, e443–e19. doi: 10.1128/jvi.00443-19
- Ferguson, A. L., Mann, J. K., Omarjee, S., Ndung'u, T., Walker, B. D., and Chakraborty, A. K. (2013). Translating HIV sequences into quantitative fitness landscapes predicts viral vulnerabilities for rational immunogen design. *Immunity* 38, 606–617. doi: 10.1016/j.immuni.2012.11.022
- Fleishman, S. J., Whitehead, T. A., Ekiert, D. C., Dreyfus, C., Corn, J. E., Strauch, E. M., et al. (2011). Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science* 332, 816–821. doi: 10.1126/science.1202617
- Flynn, W. F., Haldane, A., Torbett, B. E., and Levy, R. M. (2017). Inference of epistatic effects leading to entrenchment and drug resistance in HIV-1 protease. *Mol. Biol. Evol.* 34, 1291–1306. doi: 10.1093/molbev/msx095
- Forsyth, C. M., Juan, V., Akamatsu, Y., DuBridg, R. B., Doan, M., Ivanov, A. V., et al. (2013). Deep mutational scanning of an antibody against epidermal growth factor receptor using mammalian cell display and massively parallel pyrosequencing. *MAbs* 5, 523–532. doi: 10.4161/mabs.24979
- Fowler, D. M., Araya, C. L., Fleishman, S. J., Kellogg, E. H., Stephany, J. J., Baker, D., et al. (2010). High-resolution mapping of protein sequence-function relationships. *Nat. Methods* 7, 741–746. doi: 10.1038/nmeth.1492
- Fowler, D. M., and Fields, S. (2014). Deep mutational scanning: a new style of protein science. *Nat. Methods* 11, 801–807. doi: 10.1038/nmeth.3027
- Frank, H., Enard, D., and Boyd, S. (2020). Exceptional diversity and selection pressure on SARS-CoV and SARS-CoV-2 host receptor in bats compared to other mammals. *bioRxiv* [Preprint]. doi: 10.1101/2020.04.20.051656
- Gai, S. A., and Wittrop, K. D. (2007). Yeast surface display for protein engineering and characterization. *Curr. Opin. Struct. Biol.* 17, 467–473. doi: 10.1016/j.sbi.2007.08.012
- Gaiotto, T., and Hufton, S. E. (2016). Cross-neutralising nanobodies bind to a conserved pocket in the hemagglutinin stem region identified using yeast display and deep mutational scanning. *PLoS One* 11:e0164296. doi: 10.1371/journal.pone.0164296
- Gemmati, D., Bramanti, B., Serino, M. L., Secchiero, P., Zauli, G., and Tisato, V. (2020). COVID-19 and individual genetic susceptibility/receptivity: Role of ACE1/ACE2 genes, immunity, inflammation and coagulation. might the double x-chromosome in females be protective against SARS-COV-2 compared to the single x-chromosome in males? *Int. J. Mol. Sci.* 21:3474. doi: 10.3390/ijms21103474
- Gong, D., Zhang, T. H., Zhao, D., Du, Y., Chapa, T. J., Shi, Y., et al. (2018). High-throughput fitness profiling of Zika virus E protein reveals different roles for glycosylation during infection of mammalian and mosquito cells. *iScience* 1, 97–111. doi: 10.1016/j.isci.2018.02.005
- Gray, V. E., Hause, R. J., and Fowler, D. M. (2017). Analysis of large-scale mutagenesis data to assess the impact of single amino acid substitutions. *Genetics* 207, 53–61. doi: 10.1534/genetics.117.300064
- Gray, V. E., Hause, R. J., Luebeck, J., Shendure, J., and Fowler, D. M. (2018). Quantitative missense variant effect prediction using large-scale mutagenesis data. *Cell Syst.* 6, 116–124. doi: 10.1016/j.cels.2017.11.003
- Gray, V. E., Sitko, K., Ngako Kamani, F. Z., Williamson, M., Stephany, J. J., Hasle, N., et al. (2019). Elucidating the molecular determinants of Ab aggregation with deep mutational scanning. *G3 Genes Genomes Genet.* 9, 3683–3689. doi: 10.1534/g3.119.400535
- Greaney, A. J., Starr, T. N., Gilchuk, P., Zost, S. J., Binshtein, E., Loes, A. N., et al. (2020). Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition. *Cell Host Microbe* 29, 44–57. doi: 10.1016/j.chom.2020.11.007
- Guo, H., Hu, B.-J., Yang, X.-L., Zeng, L.-P., Li, B., Ouyang, S., et al. (2020). Evolutionary arms race between virus and host drives genetic diversity in bat severe acute respiratory syndrome-related coronavirus spike genes. *J. Virol.* 94, e00902–e00920. doi: 10.1128/JVI.00902-20
- Gupta, K., and Varadarajan, R. (2018). Insights into protein structure, stability and function from saturation mutagenesis. *Curr. Opin. Struct. Biol.* 50, 117–125. doi: 10.1016/j.sbi.2018.02.006
- Haddox, H. K., Dingens, A. S., and Bloom, J. D. (2016). Experimental estimation of the effects of all amino-acid mutations to HIV's envelope protein on viral replication in cell culture. *PLoS Pathog.* 12:e1006114. doi: 10.1371/journal.ppat.1006114
- Haddox, H. K., Dingens, A. S., Hilton, S. K., Overbaugh, J., and Bloom, J. D. (2018). Mapping mutational effects along the evolutionary landscape of HIV envelope. *Elife* 7:e34420. doi: 10.7554/eLife.34420
- Heinzelman, P., and Romero, P. A. (2020). Discovery of human ACE2 variants with altered recognition by the SARS-CoV-2 spike protein. *bioRxiv* [Preprint]. doi: 10.1101/2020.09.17.301861
- Heredia, J. D., Park, J., Brubaker, R. J., Szymanski, S. K., Gill, K. S., and Procko, E. (2018). Mapping interaction sites on human chemokine receptors by deep mutational scanning. *J. Immunol.* 200, 3825–3839. doi: 10.4049/jimmunol.1800343
- Heredia, J. D., Park, J., Choi, H., Gill, K. S., and Procko, E. (2019). Conformational engineering of HIV-1 Env based on mutational tolerance in the CD4 and PG16 bound states. *J. Virol.* 93, e219–e219. doi: 10.1128/jvi.00219-19
- Higuchi, Y., Suzuki, T., Arimori, T., Ikemura, N., Kirita, Y., Ohgihara, E., et al. (2020). High affinity modified ACE2 receptors prevent SARS-CoV-2 infection. *bioRxiv* [Preprint]. doi: 10.1101/2020.09.16.299891
- Hilton, S. K., and Bloom, J. D. (2018). Modeling site-specific amino-acid preferences deepens phylogenetic estimates of viral sequence divergence. *Virus Evol.* 4:vey033. doi: 10.1093/ve/vey033
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., et al. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 181, 271–280. doi: 10.1016/j.cell.2020.02.052
- Hopf, T. A., Ingraham, J. B., Poelwijk, F. J., Schärfe, C. P. I., Springer, M., Sander, C., et al. (2017). Mutation effects predicted from sequence co-variation. *Nat. Biotechnol.* 35, 128–135. doi: 10.1038/nbt.3769
- Imai, Y., Kuba, K., Rao, S., Huan, Y., Guo, F., Guan, B., et al. (2005). Angiotensin-converting enzyme 2 protects from severe acute lung failure. *Nature* 436, 112–116. doi: 10.1038/nature03712
- Jacobs, A., Sen, J., Rong, L., and Caffrey, M. (2005). Alanine scanning mutants of the HIV gp41 loop. *J. Biol. Chem.* 280, 27284–27288. doi: 10.1074/jbc.M414411200
- Jardine, J. G., Kulp, D. W., Havenar-Daughton, C., Sarkar, A., Briney, B., Sok, D., et al. (2016). HIV-1 broadly neutralizing antibody precursor B cells revealed by germline-targeting immunogen. *Science* 351, 1458–1463. doi: 10.1126/science.aad9195
- Jiang, J., and Aiken, C. (2007). Maturation-dependent human immunodeficiency virus type 1 particle fusion requires a carboxyl-terminal region of the gp41 cytoplasmic tail. *J. Virol.* 81, 9999–10008. doi: 10.1128/jvi.00592-07
- Jones, E. M., Lubock, N. B., Venkatakrishnan, A. J., Wang, J., Tseng, A. M., Paggi, J. M., et al. (2020). Structural and functional characterization of G protein-coupled receptors with deep mutational scanning. *Elife* 9:e54895. doi: 10.7554/eLife.54895
- Kabanova, A., Marcandalli, J., Zhou, T., Bianchi, S., Baxa, U., Tsybovsky, Y., et al. (2016). Platelet-derived growth factor- α receptor is the cellular receptor for human cytomegalovirus gHgLgO trimer. *Nat. Microbiol.* 1:16082. doi: 10.1038/nmicrobiol.2016.82
- Klesmith, J. R., Bacik, J. P., Wrenbeck, E. E., Michalczyk, R., and Whitehead, T. A. (2017). Trade-offs between enzyme fitness and solubility illuminated by deep mutational scanning. *Proc. Natl. Acad. Sci. U.S.A.* 114, 2265–2270. doi: 10.1073/pnas.1614437114
- Kschonsak, M., Rougé, L., Arthur, C. P., Hoangdung, H., Patel, N., Kim, I., et al. (2021) Structures of HCMV trimer reveal the basis for receptor recognition and cell entry. *Cell* 184, 1232–1244. doi: 10.1016/j.cell.2021.01.036
- Kulp, D. W., Steichen, J. M., Pauthner, M., Hu, X., Schiffrer, T., Liguori, A., et al. (2017). Structure-based design of native-like HIV-1 envelope trimers to silence non-neutralizing epitopes and eliminate CD4 binding. *Nat. Commun.* 8:1655. doi: 10.1038/s41467-017-01549-6
- Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., et al. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 581, 215–220. doi: 10.1038/s41586-020-2180-5
- Lee, J. M., Huddleston, J., Doud, M. B., Hooper, K. A., Wu, N. C., Bedford, T., et al. (2018). Deep mutational scanning of hemagglutinin helps predict evolutionary fates of human H3N2 influenza variants. *Proc. Natl. Acad. Sci. U.S.A.* 115, E8276–E8285. doi: 10.1073/pnas.1806133115

- Leung, K., Shum, M. H., Leung, G. M., Lam, T. T., and Wu, J. T. (2021). Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. *Eurosurveillance* 26:2002106. doi: 10.2807/1560-7917.ES.2020.26.1.2002106
- Levy, R. M., Haldane, A., and Flynn, W. F. (2017). Potts Hamiltonian models of protein co-variation, free energy landscapes, and evolutionary fitness. *Curr. Opin. Struct. Biol.* 43, 55–62. doi: 10.1016/j.sbi.2016.11.004
- Li, Y., O'Dell, S., Walker, L. M., Wu, X., Guenaga, J., Feng, Y., et al. (2011). Mechanism of neutralization by the broadly neutralizing HIV-1 monoclonal antibody VRC01. *J. Virol.* 85, 8954–8967. doi: 10.1128/jvi.00754-11
- Linsky, T. W., Vergara, R., Codina, N., Nelson, J. W., Walker, M. J., Su, W., et al. (2020). De novo design of potent and resilient hACE2 decoys to neutralize SARS-CoV-2. *Science* 370:eabe0075. doi: 10.1126/science.abe0075
- Louie, R. H. Y., Kaczorowski, K. J., Barton, J. P., Chakraborty, A. K., and McKay, M. R. (2018). Fitness landscape of the human immunodeficiency virus envelope protein that is targeted by antibodies. *Proc. Natl. Acad. Sci. U.S.A.* 115, E564–E573. doi: 10.1073/pnas.1717765115
- Lu, M., Stoller, M. O., Wang, S., Liu, J., Fagan, M. B., and Nunberg, J. H. (2001). Structural and functional analysis of interhelical interactions in the human immunodeficiency virus type 1 gp41 envelope glycoprotein by alanine-scanning mutagenesis. *J. Virol.* 75, 11146–11156. doi: 10.1128/jvi.75.22.11146-11156.2001
- Mann, J. K., Barton, J. P., Ferguson, A. L., Omarjee, S., Walker, B. D., Chakraborty, A., et al. (2014). The fitness landscape of HIV-1 Gag: advanced modeling approaches and validation of model predictions by in vitro testing. *PLoS Comput. Biol.* 10:e1003776. doi: 10.1371/journal.pcbi.1003776
- Matreyek, K. A., Starita, L. M., Stephany, J. J., Martin, B., Chiasson, M. A., Gray, V. E., et al. (2018). Multiplex assessment of protein variant abundance by massively parallel sequencing. *Nat. Genet.* 50, 874–882. doi: 10.1038/s41588-018-0122-z
- Matreyek, K. A., Stephany, J. J., Chiasson, M. A., Hasle, N., and Fowler, D. M. (2020). An improved platform for functional assessment of large protein libraries in mammalian cells. *Nucleic Acids Res.* 48:e1. doi: 10.1093/nar/gkz910
- Matreyek, K. A., Stephany, J. J., and Fowler, D. M. (2017). A platform for functional assessment of large variant libraries in mammalian cells. *Nucleic Acids Res.* 45:e102. doi: 10.1093/nar/gkx183
- McLaughlin, R. N., Poelwijk, F. J., Raman, A., Gosal, W. S., and Ranganathan, R. (2012). The spatial architecture of protein function and adaptation. *Nature* 491, 138–142. doi: 10.1038/nature11500
- McShan, A., Devlin, C., Morozov, G., Overall, S., Moschidi, D., Akella, N., et al. (2020). TAPBPR promotes antigen loading on MHC-I molecules using a peptide trap. *bioRxiv* [Preprint]. doi: 10.1101/2020.04.24.059634
- McShan, A. C., Devlin, C. A., Overall, S. A., Park, J., Toor, J. S., Moschidi, D., et al. (2019). Molecular determinants of chaperone interactions on MHC-I for folding and antigen repertoire selection. *Proc. Natl. Acad. Sci. U. S. A.* 116, 25602–25613. doi: 10.1073/pnas.1915562116
- Medina-Ramírez, M., Sanders, R. W., and Sattentau, Q. J. (2017). Stabilized HIV-1 envelope glycoprotein trimers for vaccine use. *Curr. Opin. HIV AIDS* 12, 241–249. doi: 10.1097/COH.0000000000000363
- Merk, A., and Subramaniam, S. (2013). HIV-1 envelope glycoprotein structure. *Curr. Opin. Struct. Biol.* 23, 268–276. doi: 10.1016/j.sbi.2013.03.007
- Murin, C. D., Wilson, I. A., and Ward, A. B. (2019). Antibody responses to viral infections: a structural perspective across three different enveloped viruses. *Nat. Microbiol.* 4, 734–747. doi: 10.1038/s41564-019-0392-y
- Olshesky, U., Helseth, E., Furman, C., Li, J., Haseltine, W., and Sodroski, J. (1990). Identification of individual human immunodeficiency virus type 1 gp120 amino acids important for CD4 receptor binding. *J. Virol.* 64, 5701–5707. doi: 10.1128/jvi.64.12.5701-5707.1990
- Otwinowski, J., McCandlish, D. M., and Plotkin, J. B. (2018). Inferring the shape of global epistasis. *Proc. Natl. Acad. Sci. U.S.A.* 115, E7550–E7558. doi: 10.1073/pnas.1804015115
- Oude Munnink, B. B., Sikkema, R. S., Nieuwenhuijse, D. F., Molenaar, R. J., Munger, E., Molenkamp, R., et al. (2021). Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. *Science* 371, 172–177. doi: 10.1126/science.abe5901
- Ozono, S., Zhang, Y., Ode, H., Sano, K., Tan, T. S., Imai, K., et al. (2021). SARS-CoV-2 D614G spike mutation increases entry efficiency with enhanced ACE2-binding affinity. *Nat. Commun.* 12:848. doi: 10.1038/s41467-021-21118-2
- Park, J., Gill, K. S., Aghajani, A. A., Heredia, J. D., Choi, H., Oberstein, A., et al. (2020). Engineered receptors for human cytomegalovirus that are orthogonal to normal human biology. *PLoS Pathog.* 16:e1008647. doi: 10.1371/journal.ppat.1008647
- Park, J., Selvam, B., Sanematsu, K., Shigemura, N., Shukla, D., and Procko, E. (2019). Structural architecture of a dimeric class C GPCR based on co-trafficking of sweet taste receptor subunits. *J. Biol. Chem.* 294, 4759–4774. doi: 10.1074/jbc.RA118.006173
- Phillips, A. M., Doud, M. B., Gonzalez, L. O., Butty, V. L., Lin, Y. S., Bloom, J. D., et al. (2018). Enhanced ER proteostasis and temperature differentially impact the mutational tolerance of influenza hemagglutinin. *Elife* 7:e38795. doi: 10.7554/eLife.38795
- Procko, E. (2020). Deep mutagenesis in the study of COVID-19: a technical overview for the proteomics community. *Expert Rev. Proteomics* 17, 633–638. doi: 10.1080/14789450.2020.1833721
- Procko, E., Berguig, G. Y., Shen, B. W., Song, Y., Frayo, S., Convertine, A. J., et al. (2014). A computationally designed inhibitor of an Epstein-Barr viral Bcl-2 protein induces apoptosis in infected cells. *Cell* 157, 1644–1656. doi: 10.1016/j.cell.2014.04.034
- Rabut, G. E. E., Konner, J. A., Kajumo, F., Moore, J. P., and Dragic, T. (1998). Alanine Substitutions of polar and nonpolar residues in the amino-terminal domain of CCR5 differently impair entry of macrophage- and dualtropic isolates of human immunodeficiency virus Type 1. *J. Virol.* 72, 3464–3468. doi: 10.1128/jvi.72.4.3464-3468.1998
- Riesselman, A. J., Ingraham, J. B., and Marks, D. S. (2018). Deep generative models of genetic variation capture the effects of mutations. *Nat. Methods* 15, 816–822. doi: 10.1038/s41592-018-0138-4
- Ringel, O., Vieillard, V., Debré, P., Eichler, J., Büning, H., and Dietrich, U. (2018). The hard way towards an antibody-based HIV-1 Env vaccine: lessons from other viruses. *Viruses* 10:197. doi: 10.3390/v10040197
- Rocklin, G. J., Chidyausiku, T. M., Goresnik, I., Ford, A., Houliston, S., Lemak, A., et al. (2017). Global analysis of protein folding using massively parallel design, synthesis, and testing. *Science* 357, 168–175. doi: 10.1126/science.aan0693
- Rollins, N. J., Brock, K. P., Poelwijk, F. J., Stiffler, M. A., Gauthier, N. P., Sander, C., et al. (2019). Inferring protein 3D structure from deep mutation scans. *Nat. Genet.* 51, 1170–1176. doi: 10.1038/s41588-019-0432-9
- Russell, C. J., Hu, M., and Okda, F. A. (2018). Influenza hemagglutinin protein stability, activation, and pandemic risk. *Trends Microbiol.* 26, 841–853. doi: 10.1016/j.tim.2018.03.005
- Saito, Y., Oikawa, M., Nakazawa, H., Niide, T., Kameda, T., Tsuda, K., et al. (2018). Machine-learning-guided mutagenesis for directed evolution of fluorescent proteins. *ACS Synth. Biol.* 7, 2014–2022. doi: 10.1021/acssynbio.8b00155
- Salamango, D. J., Alam, K. K., Burke, D. H., and Johnson, M. C. (2016). In vivo analysis of infectivity, fusogenicity, and incorporation of a mutagenic viral glycoprotein library reveals determinants for virus incorporation. *J. Virol.* 90, 6502–6514. doi: 10.1128/jvi.00804-16
- Schmiedel, J. M., and Lehner, B. (2019). Determining protein structures using deep mutagenesis. *Nat. Genet.* 51, 1177–1186. doi: 10.1038/s41588-019-0431-x
- Sen, J., Yan, T., Wang, J., Rong, L., Tao, L., and Caffrey, M. (2010). Alanine scanning mutagenesis of HIV-1 gp41 heptad repeat 1: insight into the gp120-gp41 interaction. *Biochemistry* 49, 5057–5065. doi: 10.1021/bi1005267
- Setoh, Y. X., Amarilla, A. A., Peng, N. Y. G., Griffiths, R. E., Carrera, J., Freney, M. E., et al. (2019). Determinants of Zika virus host tropism uncovered by deep mutational scanning. *Nat. Microbiol.* 4, 876–887. doi: 10.1038/s41564-019-0399-4
- Shaik, M. M., Peng, H., Lu, J., Rits-Volloch, S., Xu, C., Liao, M., et al. (2019). Structural basis of coreceptor recognition by HIV-1 envelope spike. *Nature* 565, 318–323. doi: 10.1038/s41586-018-0804-9
- Shamsi, Z., Chan, M., and Shukla, D. (2020). TLmutation: predicting the effects of mutations using transfer learning. *J. Phys. Chem. B* 124, 3845–3854. doi: 10.1021/acs.jpcc.0c00197
- Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., et al. (2020). Structural basis of receptor recognition by SARS-CoV-2. *Nature* 581, 221–224. doi: 10.1038/s41586-020-2179-y

- Shirleen Soh, Y., Moncla, L. H., Eguia, R., Bedford, T., and Bloom, J. D. (2019). Comprehensive mapping of adaptation of the avian influenza polymerase protein PB2 to humans. *Elife* 8:e45079. doi: 10.7554/eLife.45079
- Sirohi, D., and Kuhn, R. J. (2017). Zika virus structure, maturation, and receptors. *J. Infect. Dis.* 216, S935–S944. doi: 10.1093/infdis/jix515
- Sourisseau, M., Lawrence, D. J. P., Schwarz, M. C., Storrs, C. H., Veit, E. C., Bloom, J. D., et al. (2019). Deep mutational scanning comprehensively maps how zika envelope protein mutations affect viral growth and antibody escape. *J. Virol.* 93, e1291-19. doi: 10.1128/jvi.01291-19
- Stamatatos, L., Pancera, M., and McGuire, A. T. (2017). Germline targeting immunogens. *Immunol. Rev.* 275, 203–216. doi: 10.1111/imr.12483
- Starr, T. N., Greaney, A. J., Addetia, A., Hannon, W. W., Choudhary, M. C., Dingens, A. S., et al. (2021). Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. *Science* 371, 850–854. doi: 10.1126/science.abf9302
- Starr, T. N., Greaney, A. J., Hilton, S. K., Ellis, D., Crawford, K. H. D., Dingens, A. S., et al. (2020). Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell* 182, 1295–1310.e20. doi: 10.1016/j.cell.2020.08.012
- Stegmann, C., Hochdorfer, D., Lieber, D., Subramanian, N., Stöhr, D., Laib Sampaio, K., et al. (2017). A derivative of platelet-derived growth factor receptor alpha binds to the trimer of human cytomegalovirus and inhibits entry into fibroblasts and endothelial cells. *PLoS Pathog.* 13:e1006273. doi: 10.1371/journal.ppat.1006273
- Steichen, J. M., Kulp, D. W., Tokatlian, T., Escolano, A., Dosenovic, P., Stanfield, R. L., et al. (2016). HIV vaccine design to target germline precursors of glycan-dependent broadly neutralizing antibodies. *Immunity* 45, 483–496. doi: 10.1016/j.immuni.2016.08.016
- Stein, A., Fowler, D. M., Hartmann-Petersen, R., and Lindorff-Larsen, K. (2019). Biophysical and mechanistic models for disease-causing protein variants. *Trends Biochem. Sci.* 44, 575–588. doi: 10.1016/j.tibs.2019.01.003
- Tegally, H., Wilkinson, E., Giovanetti, M., Iranzadeh, A., Fonseca, V., Giandhari, J., et al. (2020). Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *medRxiv* [Preprint]. doi: 10.1101/2020.12.21.20248640
- Tenthorey, J. L., Young, C., Sodeinde, A., Emerman, M., and Malik, H. S. (2020). Mutational resilience of antiviral restriction favors primate TRIM5 α in host-virus evolutionary arms races. *Elife* 9:e59988. doi: 10.7554/ELIFE.59988
- Thyagarajan, B., and Bloom, J. D. (2014). The inherent mutational tolerance and antigenic evolvability of influenza hemagglutinin. *Elife* 3:e03300. doi: 10.7554/elife.03300
- Treml, B., Neu, N., Kleinsasser, A., Gritsch, C., Finsterwalder, T., Geiger, R., et al. (2010). Recombinant angiotensin-converting enzyme 2 improves pulmonary blood flow and oxygenation in lipopolysaccharide-induced lung injury in piglets. *Crit. Care Med.* 38, 596–601. doi: 10.1097/CCM.0b013e3181c03009
- Volz, E., Mishra, S., Chand, M., Barrett, J. C., Johnson, R., Hopkins, S., et al. (2021). Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: insights from linking epidemiological and genetic data. *medRxiv* [Preprint]. doi: 10.1101/2020.12.30.20249034
- Walker, L. M., Huber, M., Doores, K. J., Falkowska, E., Pejchal, R., Julien, J. P., et al. (2011). Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature* 477, 466–470. doi: 10.1038/nature10373
- Walker, L. M., Phogat, S. K., Chan-Hui, P.-Y., Wagner, D., Phung, P., Goss, J. L., et al. (2009). Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. *Science* 326, 285–289. doi: 10.1126/science.1178746
- Wang, Q., Finzi, A., and Sodroski, J. (2020). The conformational states of the HIV-1 envelope glycoproteins. *Trends Microbiol.* 28, 655–667. doi: 10.1016/j.tim.2020.03.007
- Weile, J., Sun, S., Cote, A. G., Knapp, J., Verby, M., Mellor, J. C., et al. (2017). A framework for exhaustively mapping functional missense variants. *Mol. Syst. Biol.* 13:957. doi: 10.15252/msb.20177908
- White, J. M., Delos, S. E., Brecher, M., and Schornberg, K. (2008). Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme. *Crit. Rev. Biochem. Mol. Biol.* 43, 189–219. doi: 10.1080/10409230802058320
- Whitehead, T. A., Chevalier, A., Song, Y., Dreyfus, C., Fleishman, S. J., De Mattos, C., et al. (2012). Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nat. Biotechnol.* 30, 543–548. doi: 10.1038/nbt.2214
- Wibmer, C. K., Ayres, F., Hermanus, T., Madzivhandila, M., Kgagudi, P., Lambson, B. E., et al. (2021). SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. *bioRxiv* [Preprint]. doi: 10.1101/2021.01.18.427166
- Wildt, S., and Gerngross, T. U. (2005). The humanization of N-glycosylation pathways in yeast. *Nat. Rev. Microbiol.* 3, 119–128. doi: 10.1038/nrmicro1087
- Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C.-L., Abiona, O., et al. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 367, 1260–1263. doi: 10.1126/science.abb2507
- Wrenbeck, E. E., Faber, M. S., and Whitehead, T. A. (2017). Deep sequencing methods for protein engineering and design. *Curr. Opin. Struct. Biol.* 45, 36–44. doi: 10.1016/j.sbi.2016.11.001
- Wu, N. C., Otwinowski, J., Thompson, A. J., Nycholat, C. M., Nourmohammad, A., and Wilson, I. A. (2020). Major antigenic site B of human influenza H3N2 viruses has an evolving local fitness landscape. *Nat. Commun.* 11:1233. doi: 10.1038/s41467-020-15102-5
- Wu, N. C., Xie, J., Zheng, T., Nycholat, C. M., Grande, G., Paulson, J. C., et al. (2017a). Diversity of functionally permissive sequences in the receptor-binding site of influenza hemagglutinin. *Cell Host Microbe* 21, 742–753.e8. doi: 10.1016/j.chom.2017.05.011
- Wu, N. C., Young, A. P., Al-Mawsawi, L. Q., Olson, C. A., Feng, J., Qi, H., et al. (2014). High-throughput profiling of influenza A virus hemagglutinin gene at single-nucleotide resolution. *Sci. Rep.* 4:4942. doi: 10.1038/srep04942
- Wu, Y., Prager, A., Boos, S., Resch, M., Brizic, I., Mach, M., et al. (2017b). Human cytomegalovirus glycoprotein complex gH/gL/gO uses PDGFR- α as a key for entry. *PLoS Pathog.* 13:e1006281. doi: 10.1371/journal.ppat.1006281
- Wu, Z., Jennifer Kan, S. B., Lewis, R. D., Wittmann, B. J., and Arnold, F. H. (2019). Machine learning-assisted directed protein evolution with combinatorial libraries. *Proc. Natl. Acad. Sci. U.S.A.* 116, 8852–8858. doi: 10.1073/pnas.1901979116
- Xu, C., Wang, Y., Liu, C., Zhang, C., Han, W., Hong, X., et al. (2021). Conformational dynamics of SARS-CoV-2 trimeric spike glycoprotein in complex with receptor ACE2 revealed by cryo-EM. *Sci. Adv.* 7:eabe5575. doi: 10.1126/sciadv.abe5575
- Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., and Zhou, Q. (2020). Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* 367, 1444–1448. doi: 10.1126/science.abb2762
- Yang, K. K., Wu, Z., and Arnold, F. H. (2019). Machine-learning-guided directed evolution for protein engineering. *Nat. Methods* 16, 687–694. doi: 10.1038/s41592-019-0496-6
- Zhang, T. H., Dai, L., Barton, J. P., Du, Y., Tan, Y., Pang, W., et al. (2020). Predominance of positive epistasis among drug resistance-associated mutations in HIV-1 protease. *PLoS Genet.* 16:e1009009. doi: 10.1371/journal.pgen.1009009
- Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., et al. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579, 270–273. doi: 10.1038/s41586-020-2012-7
- Zou, Z., Yan, Y., Shu, Y., Gao, R., Sun, Y., Li, X., et al. (2014). Angiotensin-converting enzyme 2 protects from lethal avian influenza A H5N1 infections. *Nat. Commun.* 5:3594. doi: 10.1038/ncomms4594

Conflict of Interest: EP was the inventor on patent filings by the University of Illinois covering soluble decoy receptors. EP was a co-founder of Orthogonal Biologics, Inc.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Narayanan and Procko. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.