2016-05-23

Systematic Experimental Determination of Functional Constraints on Proteins and Adaptive Potential of Mutations: A Dissertation

Li Jiang
University of Massachusetts Medical School

SYSTEMATIC EXPERIMENTAL DETERMINATION OF FUNCTIONAL CONSTRAINTS ON PROTEINS AND ADAPTIVE POTENTIAL OF MUTATIONS

A Dissertation Presented

By

LI JIANG

Submitted to the Faculty of the
University of Massachusetts Graduate School of Biomedical Sciences, Worcester
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

MAY 23RD, 2016
BIOCHEMISTRY AND MOLECULAR PHARMACOLOGY

SYSTEMATIC EXPERIMENTAL DETERMINATION OF FUNCTIONAL CONSTRAINTS ON PROTEINS AND ADAPTIVE POTENTIAL OF MUTATIONS

A Dissertation Presented
By
LI JIANG

The signatures of the Dissertation Defense Committee signify completion and approval as to style and content of the Dissertation

Daniel Bolon, Ph.D., Thesis Advisor

Celia Schiffer, Ph.D., Member of the Committee

Jennifer Wang, M.D., Member of the Committee

William Royer, Ph.D., Member of the Committee

John Logsdon, Ph.D., External Member of the Committee

The Signature of the Chair of the Committee signifies that the written dissertation meets the requirements of the Dissertation Committee

Scot Wolfe, Ph.D., Chair of Committee

The signature of the Dean of the Graduate School of Biomedical Science signifies that the student has met all graduation requirements of the School.

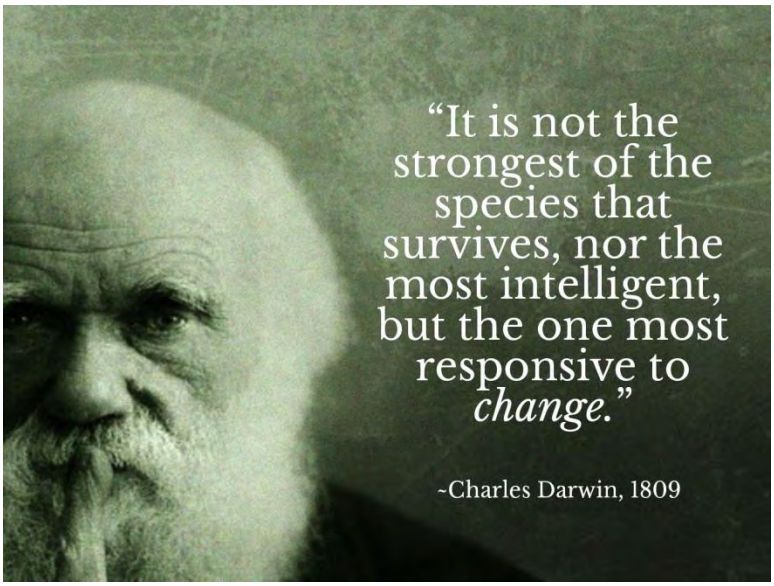
Anthony Carruthers, Ph.D.,
Dean of the Graduate School of Biomedical Sciences

Biochemistry and Molecular Pharmacology

May 23rd, 2016

Dedication

This is work is dedicated to my wife Ying Li (李盈), and my mother and father, Liujing Yang (杨柳静) and Hesheng Jiang (蒋和胜)



Acknowledgement

Before starting graduate school, I thought that pursing a doctorate degree was a lonely path towards bringing another tiny piece of knowledge to the ever-lasting intellectual advancement of humanity. However, after nearly six years at the graduate school of the University of Massachusetts Medical School, I realize that to obtain a doctorate degree, you need motivation, perseverance, diligence and even fortune. More importantly, I find out that I would not be able to finish this quest without generous support, help and encouragements from many people.

First and foremost, I would like to thank my thesis advisor Dr. Dan Bolon. Dan has been always supportive, patient, and encouraging throughout my thesis research training. Every time I am facing problems either in my scientific research or career development, Dan's advice has always been guiding me to the right track. Dan not only cares about my scientific progress, but also spends much effort nurturing my maturation as a responsible scientist and growth as a good person. He is also very understanding and considerate to help me study in the United States as an international student. I would also like to thank members of my qualifying exam, thesis research advisory and thesis defense committees: Dr. Celia Schiffer, Dr. Scot Wolfe, Dr. Jennifer Wang, Dr. Paul Clapham, Dr. Nick Rhind, Dr. William Royer, and Dr. John Logsdon, for their constructive advice, kind support, and inspiring encouragement. I would like to thank Dr. Guangping Gao for valuable advice and motivating encouragement on my career development.

I am also grateful to the members of the Bolon Lab, Dr. Parul Mishra, Dr. Jeffrey Boucher, Dr. Julia Flynn, Aneth Laban, Pamela Cote, Gila Schneider-Nachum as well as past members including Dr. Ryan Hietpas, Dr. Benjamin Roscoe, Ammeret Rossouw and Lester Pullen, for their spiritual and technical support as well as great scientific discussions. Particularly, I would like to thank Parul for being a great collaborator on our joint project, listening to my struggles and encouraging me to carry on scientific endeavor. I would also like to thank Aneth for cheering me up and drawing all the medals and stickers to make me feel like an important person with some accomplishments. I would like to thank Jeff for being my go-to person for questions about job application, presentation slides and manuscripts editing. I would also like to thank Julia, Pam and Ammeret for their kind help and suggestions on my research and life. For Ryan and Ben, I would like to thank them for their guidance, support, and funny jokes in my first 3-4 years in the Bolon lab. In particular for Ryan, thanks for harvesting mid-night time point samples for yeast growth.

I would also like to thank all my collaborators. Thanks to Dr. Robert Finberg and Dr. Jennifer Wang for granting me the opportunity to use their laboratory space and resources to conduct influenza virus drug resistance research and guiding me through the process, Ping Liu for teaching me cell culture and influenza virus techniques, Dr. Melanie Trombly for manuscript editing, Debra Poliquin for meeting scheduling help, and other members of the Finberg/Wang lab for helping and/or tolerating me in the laboratory.

Special thanks to Dr. Paul Clapham and Dr. Maria Jose Duenas-Decamp for their great conceptual insights, experimental work and insightful discussion in our collaboration on mapping HIV evolutionary potential. Thanks to Dr. Jeffrey Jensen and Dr. Claudia Bank for their statistical analysis support and education. Thanks to Dr. Celia Schiffer, Dr. Nese KurtYilmaz and Dr. Kristina Prachanronarong for their inputs and help on structural analysis and modeling as well as manuscript editing. Thanks to Dr. Konstatin Zeldovich and Dr. Sergey Venev for great answers for my bioinformatics and statistics questions. Thanks to Dr. Timothy Kowalik and Dr. Nick Renzette for intellectual input and technical support to influenza viral growth assays and RNA extraction.

I am also grateful to the Department of Biochemistry and Molecular Pharmacology which is invested in research training and career development of its graduate students. Special thanks to the constant help supplied by Dr. William Kobertz, Luca Leone, Irene Couture, Karen Welch, and Betty Ann Hoyt. Thanks to Yvonne Chan from the Matthews lab and Ashiwini Sunkavalli from the Finberg/Wang lab for help on my experiments and comforting chatting about difficulties in research and graduate school.

Although this work is dedicated to my wife and parents, I would like to express my gratitude again to my wife, Ying Li, for her genuine and patient support and accompany, for her tolerance of my sometimes demanding experiments, for her great

cooking that keeps me energized, for her introducing me to great movies and TV series that help maintain my work-life balance and prevent me from becoming a dull boy. I would also like to express my gratitude again to my parents, Liujing Yang and Hesheng Jiang, for their emotional support and inspiring encouragement. They have stood behind me in all my endeavors and provide me with soothing space for a break before starting another adventure. Thank you.

ABSTRACT

Sequence-function relationship is a fundamental question for many branches of modern biomedical research. It connects the primary sequence of proteins to the function of proteins and fitness of organisms, holding answers for critical questions such as functional consequences of mutations identified in whole genome sequencing and adaptive potential of fast evolving pathogenic viruses and microbes. Many different approaches have been developed to delineate the genotype-phenotype map for different proteins, but are generally limited by their throughput or precision. To systematically quantify the fitness of large numbers of mutations, I modified a novel high throughput mutational scanning approach (EMPIRIC) to investigate the fitness landscape of mutations in important regions of essential proteins from the yeast or RNA viruses. Using EMPIRIC, I analyzed the interplay of the expression level and sequence of Hsp90 on the yeast growth and revealed latent effect of mutations at reduced expression levels of Hsp90. I also examined the functional constraint on the receptor binding site of the Env of Human Immunodeficiency Virus (HIV) and uncovered enhanced receptor binding capacity as a common pathway for adaptation of HIV to laboratory conditions. Moreover, I explored the adaptive potential of neuraminidase (NA) of influenza A virus to a NA inhibitor, oseltamivir, and identified novel oseltamivir resistance mutations with distinct molecular mechanisms. In summary, I applied a high throughput functional genomics approach to map the sequence-function relationship in various systems and examined the evolutionary constraints and adaptive potential of essential proteins ranging from molecular chaperones to drug-targetable viral proteins.

Table of Contents

List of Tables

List of Figures

List of copyrighted materials produced by the author

Chapter II represents work previously published previously and is presented in accordance with copyright law.

Jiang L, Mishra P*, Hietpas RT, Zeldovich KB, Bolon DNA."Latent effects of Hsp90 mutants revealed at reduced expression*-levels*". PLoS Genet. 2013;9:e1003600. (* equal contribution)*

Chapter IV represents work previously published previously and is presented in accordance with copyright law.

Jiang L, Liu P, Bank C, Renzette N, Prachanronarong K, Yilmaz LS, et al. A balance between inhibitor binding and substrate processing confers influenza drug resistance. J Mol Biol. 2016;428[1]:538-53.

List of Electronic Table Files

Tables that are too large to be incorporated into the text of this dissertation are listed below and supplied as Microsoft Excel format files. Tables are named by the chapter in which they are referred, preceding the table number within the chapter.

Table 2.2 - Table of experimentally measured selection coefficients for Hsp90 Mutants

Table 2.3 - Table of Hsp90 mutant function estimates

Table 2.5 – Table of sequencing counts of all mutations at each time point

Table 3.1 - Datasets of amino acid fitness effect

Table 3.7 - Datasets of sequencing counts and frequency observations

Table 3.8 - Datasets of codon fitness effects

Table 4.2 - Datasets of sequencing counts and frequency observations

Table 4.3 - Datasets of fitness estimates of mutants in the presence or absence of 1 micromolar oseltamivir

Table 4.4 - Datasets of fitness estimates of oseltamivir responsive mutants

Table 4.6 - Datasets of sequencing counts, frequency observations, and fitness estimates for mutations encoding amino acids 291-300 at 0.25 and 4 μM oseltamivir

Preface

The work presented in Chapter II has been published previously as *Jiang L*, Mishra P*, Hietpas RT, Zeldovich KB, Bolon DNA."Latent effects of Hsp90 mutants revealed at reduced expression*-levels*". PLoS Genet. 2013;9:e1003600. (* equal contribution)*

The work presented in Chapter III has been submitted previously to *Nature Structural & Molecular Biology as Duenas-Decamp M, Jiang L, Bolon DN and Clapham PR. Saturation mutagenesis of the HIV-1 Envelope CD4 binding loop reveals residues controlling distinct trimer conformations.*

The work presented in Chapter IV has been published previously as *Jiang L, Liu P, Bank C, Renzette N, Prachanronarong K, Yilmaz LS, et al. A balance between inhibitor binding and substrate processing confers influenza drug resistance. J Mol Biol. 2016;428[1]:538-53.*

Chapter I – General Introduction

Sequence-function relationship: an essential question in protein science

Proteins are essential and versatile functional units that enable growth and evolution of all kinds of organisms, from viruses to humans. The primary amino acid sequence of a protein determines its secondary, tertiary and quaternary structure; the structure of a protein dictates its function. Mapping this sequence-structure-function relationship has been a fundamental question that is constantly pursued by many branches of biomedical research. For example, numerous human diseases including cancer, immunological and neurological disorders, metabolic diseases and cardiovascular diseases can be traced back to variations in protein sequence, which cause aberrations in protein functions and thus dysfunctions on the cell and organ level [2]. In addition, progress in mapping a protein sequence to its function will greatly facilitate protein engineering and design. Improvements on proteins with known functions and *de novo* design of proteins with neo-functions have been central goals for this field with numerous breakthroughs and successful cases [3]. However, due to the remarkable complexity of physical and chemical constraints on protein structure and function, we are still far from mastering the art of creating or improving protein functions from scratch using available rule-based designs or blind selections that capture limited enhancing mutations [4]. Elucidation on sequence-function relationship of proteins will likely contribute to our understanding of biophysical constraints on protein structure and requirements on

specific protein functions as well as provide abundant information to train new advanced protein design algorithms.

According to the central dogma, DNA serves as the template for RNA transcription; RNA then serves as the template for protein translation [5]. Changes in DNA or protein sequence are termed mutations. Most mutations in protein sequences result from non-synonymous mutations in coding regions of DNA or RNA because of replication errors of DNA or RNA polymerase. From a biochemical perspective, mutations in protein sequences may lead to changes in protein structure with an impact on protein folding, thermodynamic stability and solubility that alter protein functions. From an evolutionary perspective, inheritable mutations in DNA sequences generate mutations in protein sequences that affect organismal fitness, providing raw material for selection pressure to act upon. This iterative process provides the molecular basis for evolution and shapes the world full of different organisms with various adaptation strategies. Interestingly, evolution can be regarded as a massive experiment on genotype-phenotype relationships that nature has conducted for billions of years. The result manifests as a big network of extant sequence spaces with mutational connections between them and collections of proteins with distinct properties and functions. Their linkage yields a genotype-phenotype map and provides rich mechanistic insights into protein sequence-function relationships. This entire mutagenesis followed by a natural selection process can be reproduced in the laboratory, though on a much smaller scale and shorter timeframe. Manually generating mutations of either known or unknown

identities in predetermined or random positions in proteins and interrogating the mutational effect on protein functions to understand sequence-function relationship of proteins has, therefore, become one of the most fundamental and powerful tools to characterize all aspects of proteins and associated cellular and organismal level properties.

Mutagenesis approach to map sequence-function relationships

Mutagenesis approach originally started as a technical cornerstone for the field of molecular genetics and evolution and then adopted by other disciplines of biology. The underlying logic is the structure, function, and regulation of a protein can be inferred from functional consequences of perturbation to the protein or elimination of the protein. Classic mutagenesis experiments adopt a forward genetics methodology, i.e. random mutagenesis and screening of specific phenotypes, followed by mapping, cloning, sequencing, and annotating the mutant genes [6]. Experimental geneticists applied x-rays [7] and chemical mutagens [8] to induce random mutations in purified DNA or organisms and investigated mutant phenotypes, *e.g.* eye pigmentation in *Drosophila melanogaster* [9] and auxotrophy in *Neurospora crassa* [10]. However, this approach is inefficient due to the randomness and low probability of mutations of interest and confounding effects from irrelevant secondary mutations. An improved and more targeted forward genetics mutagenesis approach is transposon mutagenesis, which results in one unique insertion mutation per genome, followed by phenotypic screening and subsequent transposon-tagging to identify and clone genes of interest [11].

With advancements in recombinant DNA technology [12], it became feasible to precisely and efficiently introduce desired mutations into specific loci of DNA and therefore in the protein. Single-strand DNA repair by error-prone polymerase coupled with nucleotide analog and biased pool of nucleotides generates localized random mutagenesis that results in multiple mutations [13, 14]. Development of Polymerase Chain Reaction (PCR) [15] and advancements in oligonucleotide synthesis [16] opened new avenues for engineering single mutation at specific loci of the gene. For example, the Quikchange method [17], which allowed site-directed mutagenesis in one day with >80% efficiency, used synthetic oligonucleotides to amplify the whole plasmid in a thermocyling reaction to induce a single mutation; cassette mutagenesis constructed annealed oligonucleotides with single mutations (cassettes) and ligated them back to the gene with part of the wild-type sequence removed [18]. Conversely, error-prone PCR enables generation of mutations in a random fashion, but the average number of mutations per genome can be controlled through altering the PCR reaction condition [19]. These single mutations or combinations of mutations are then subject to a broad range of assays to measure defects or enhancements in stability [20], catalytic domains [21], binding interfaces [22] and phosphorylation [23], either in a purified system or model organism. The *in vivo* environment provides an endogenous condition where the protein naturally functions, including trafficking, post-translational modification, buffers and interacting partners. Protein in purified form enables careful characterization of the biophysical and biochemical properties of the protein in a highly controlled and reproducible *in vitro* environment.

One essential approach that analyzes mutational effects on protein function and genotype-phenotype relationships is protein display, which revolutionarily links the expressed protein and its encoding nucleotide sequence. Protein display presents a single mutant (*e.g.* introduced by site-directed mutagenesis) or a collection of diverse mutants (*e.g.* introduced by random mutagenesis or error-prone PCR) on the surface. These mutants are subject to functional screening, followed by sequencing of the selected encoding sequence, which is linked to the most optimized protein. This approach traditionally enables selection of the most optimized protein variants from a diverse library ribosome [24], phage [25], bacterial [26], yeast [27] and cellular [28] displays. Protein display is coupled to fluorescence activated cell sorting (FACS) to mediate high throughput quantitative screening with reduced background [29].

A special case of the site-directed mutagenesis that increases mutation screening throughput is alanine scanning [30]. Alanine scanning replaces wild-type amino acid with alanine and probes the impact of this mutation on protein stability or function to identify essential residues *in vivo* or *in vitro*. Alanine is chosen because its side chain terminates at the beta carbon, which results in little effect on the main chain conformation and limited electrostatic or steric restrictions; in addition, it is a frequent amino acid observed both in buried and exposed positions and all secondary structures [30]. Alanine scanning has become a prototype for all next generation mutational scanning approaches and

useful for identifying critical residues that are necessary for maintaining protein stability or function [31, 32].

Directed evolution not only selects for mutations with enhanced or altered activities, but also reveals mutational effects on protein structure and function with associated mechanistic details. Directed evolution couples multiple rounds of mutagenesis on protein sequence to selection or screening for specific functions or properties. The process starts with a library of mutants and the pool of mutations in any round inherits mutations that pass selection/screening from the previous round. This is similar to an evolutionary process: the genotype walks the fitness landscape after each round of diversification and selection, which update the available sequence space, and climbs to the peak of fitness; when reaching the peak of fitness, this iterative process yields highly optimized proteins for certain selection pressures. Biophysical or biochemical characterization of any intermediate points during and the endpoint uncovers a combination of positions that modulate protein function [33, 34], stability [35, 36] and protein-protein interaction [37]. Directed evolution complements alanine scanning in analyzing protein binding, for example, alanine scanning on evolved receptors from directed evolution unlocked residues that were necessary for enhanced binding [37].

Although the existing mutagenesis approaches have been generally successful in revealing sequence-function relationships in protein properties and uncovered many

biochemical or biophysical processes, they possess several inherent limitations. For common site-directed mutagenesis that converts one residue into another, throughput is usually quite limited due to its labor-intensive nature. Moreover, prior information like structural analysis or biochemical characterization is required for determining which residue to mutate. Alanine scanning yields mutational effects from only a single side chain substitution, although with an augmented scale of mutagenesis. Position specific biophysical requirements cannot be assessed without delineation of mutational effects from other side chains with distinct polarity, volume, shape or electrostatic effect. Similarly, directed evolution elucidates mutational effects only for a subset of highly optimized mutations and fails to illuminate impacts from the majority of other mutations. Furthermore, the phenotype is frequently derived from a combination of mutations; therefore, critical residues that alter protein stability/function still demand further clarification.

Systematic mutational scanning – a high throughput functional genomics approach

The success of the human genome project (HGP) and the availability of the complete map of the human genomic sequence [38] facilitated unparalleled progress in development of new platforms of DNA sequencing technology and a boost in DNA sequencing capability [39]. The HGP accelerated development and commercialization of next generation sequencing platforms able to produce millions of short reads (50-500bp) in a single run [39]. This revolutionary technical advancement allows quick and accurate

decoding of the diversity of complex mixtures of DNA molecules, which results in fast and efficient genome sequencing as well as quantification of a relative abundance of individual unique DNA mutations. However, increasing human genome sequencing accentuates a long-standing problem: functional consequences of the prevailing genomic variability among the human population. Each individual is estimated to have, on average, approximately 300 rare but protein-encoding mutations with largely unknown functions [40], so functional annotation of these mutations will provide invaluable information to prompt timely diagnostics of a variety of diseases, accelerated developments of novel therapeutics and accurate prediction of prognosis. Increasingly accumulated sequencing of cancer cells, pathogenic bacteria and viruses has been revealing novel mutations that may affect pathogenicity, drug resistance and immune response escape.

Currently available site-directed mutagenesis approaches are impractical for inspecting the phenotype changes of rapidly growing numbers of mutations mainly due to limited throughput. Computational approaches that predict impact of mutations on protein function or pathogenicity have limited accuracy [41]. Condel [42], GERP [43] , PolyPhen-2 [44], SIFT [45] and CADD [46] are examples of prediction algorithms that employ statistical analysis on conservation/ diversification of natural sequences or physicochemical constraints on amino acid residues. They make relatively accurate predictions on average, but fail in about half of the individual cases, which render them incapable for making reliable predictions of mutational effects on protein function [47, 48]. Therefore, there is an immense necessity for high throughput functional genomics

methodologies that are able to efficiently interrogate effects of hundreds or thousands of mutations on protein function and phenotype changes in parallel, in particular for disease-related mutations either in human or pathogen proteins.

The massive power of next generation sequencing enables accurate determination of sequence identity of each individual DNA molecule in an ultra complex and heterogeneous pool of millions of DNA molecules. When coupled with saturation mutagenesis and bulk competition, the entire procedure provides a massive parallel estimate of functional consequences of mutations in proteins that mediate biochemical (*e.g.* binding to a partner protein) or physiological (*e.g.* cell proliferation) processes in a single experiment. This general experimental pipeline diversifies into two independently developed high throughput mutational scanning approaches—deep mutational scanning (DMS) [47] and exceedingly methodical and parallel investigation of randomized individual codons (EMPIRIC) [49]. The massive mutagenesis that generates most single amino acid substitutions and occasionally double/triple amino acid mutations can be realized through chemical DNA synthesis [47], error-prone PCR [50], or randomized cassette ligation [51]. All pooled mutations are analyzed in the same physical sample during bulk competition, which ensures equivalent experimental conditions for each mutation and therefore improves measurement precision. Selection pressure imposed during bulk competition can be any environmental perturbations that affect enzymatic function [52], protein-protein interaction [53], growth rate of organisms [54], including

antibiotics [55], varied salinity [56], and novel ligands [57]. Deep sequencing of the pool of mutations provides frequency estimates of each mutation both before and after selection and the change in frequency represents a direct measurement of mutational effects on protein function or organismal fitness compared to other variants and the wild-type in the bulk competition.

As a pioneer in developing deep mutational scanning (DMS), Fowler *et al.* combined phage display of a library of ~600,000 variants (including the majority of single mutations and a fraction of double/triple mutations) of a human WW domain and bulk selection on the binding affinity of each mutation to the peptide ligand [58]. Deep sequencing of the pool of mutations prior to and post selection yielded frequency change of mutations, which corresponded to their binding affinities [58]. They were able to construct a complete map of functional constraints on each position of the WW domain that revealed amino acid preferences at each position [58]. DMS was then adopted to screen for stabilizing mutations in the WW domain [59], identify mutations with enhanced auto-ubiquitination activity in a murine E3 ligase [60], construct a sequence-function map and epistatic interaction network of RRM2 domain of a yeast polyA binding protein [61], predict homology-direct DNA repair and tumor suppression activity of mutations in the RING domain of BRCA1 [62].

Hietpas *et al*. independently developed a distinct approach coined EMPIRIC that explored the functional constraints on a nine-amino-acid region (a putative binding site) of yeast Hsp90 [51]. A saturation mutagenesis library consisting of every possible codon substitution (including single, double and triple nucleotide mutations) was assembled through randomized cassette ligation, transformed into a yeast temperature sensitive strain and subject to bulk growth competition at elevated temperatures. Samples were harvested at a series of time points and processed for deep sequencing. Changes in frequency of mutants over the time course were analyzed to estimate the impact of Hsp90 mutations on yeast growth rate, which was termed (experimental) fitness effect in evolutionary biology. Hietpas *et al.* thoroughly mapped the sequence-function relationship for this nine amino acid region of yeast Hsp90, discovered correspondence and discordance between experimental and natural evolution and related their experimental fitness measurements to classic evolutionary theories about fitness landscapes. EMPIRIC-type systematic mutagenesis has been utilized in multiple studies: analyzing the impact of shifted environmental conditions (*e.g.* elevated salinity) on fitness landscape of yeast Hsp90 [56]; illuminating a intragenic epistatic landscape in yeast Hsp90 [63]; uncovering the mutation tolerant face, which comprises mostly core residues, and the mutation sensitive face, which encompasses mainly surface residues responsible for binding interactions in yeast ubiquitin [54]; characterizing functional effects of mutations in yeast ubiquitin on E1 activation [53]; mapping mutational sensitivity and adaptive potential for binding to a novel ligand for coevolved protein sector residues in a PDZ domain [57]; and inspecting fitness effect of single mutations in

TEM-1 β-lactamase under purifying selection on ampicillin resistance and for a positive selection on cefotaxime resistance [55]. As a comparison, DMS is more often utilized to analyze the impact of mutations on protein stability, enzymatic function or binding affinity *in vitro*, whereas the EMPIRIC-type approach is commonly used for *in vivo* evaluation of the impact of mutations on organismal growth rate mediated by alterations in protein properties. DMS frequently employs large-scale randomized mutagenesis to cover most single mutations and a fraction of double or higher-order mutations while EMPIRIC generally implements systematic mutagenesis to cover every possible single amino acid substitution.

Applying high throughput mutational scanning to analyze viral evolution

High throughput mutational scanning has proved successful in dissecting evolutionary pathways of viruses, in particular RNA viruses. The high mutation rate [64], and large population size of RNA viruses [65] enables sampling of nearly all possible single mutations in a single round of viral replication [66]. This rapid evolution of RNA viruses results in viral adaption to constantly changing environmental conditions and spread of beneficial mutations under therapeutic selection pressures. Clinical relevant examples include viral escape from immune response [67], viral drug resistance [68] and zoonotic viral infection transmitted to humans [69, 70]. Therefore, analyzing the evolutionary trajectory of pathogenic RNA viruses will provide insights into control and treatment of viral infection, and inspection of mutational effects on viral protein functions

that affect viral replication and/or transmission capability (fitness) will unlock the consequences of the most common viral evolutionary steps in nature.

As the first large scale mutagenesis study to measure purifying selection on RNA viruses, Loeb *et al.* applied a PCR based random mutagenesis strategy to measure HIV-1 protease activity for a fraction of possible single mutations (20-55%) at each position and identified three separate regions that were sensitive to mutations, indicating their functional importance [1]. When high throughput mutational scanning became available, Wu *et al.* adapted this approach to measure fitness of mutations in the neuraminidase gene of H1N1 influenza virus and identified three permissive mutations that partially restored the fitness defect of the predominant drug resistance mutation H275Y (N1 numbering system) [50]. Heaton *et al.* utilized a transposon dependent random insertion mutagenesis approach to screen the entire genome of influenza for regions tolerant of mutations and revealed inherent elevated mutation acceptance capacity of hemagglutinin (HA) and nonstructural protein 1 (NS1) of influenza virus [71].

The first two studies that applied high throughput mutational scanning to characterize mutational effect in influenza virus presented exciting new methods, but lacked stringent statistical analysis and comparison of fitness effects of mutations in experimental conditions and in nature. Subsequent studies applied more advanced statistical approaches and often compared experimental fitness measurements with

frequency of mutations in isolates from patients. Bloom constructed a complete amino acid sequence preference map of influenza nucleoprotein and used experimentally determined positional amino acid preferences to build an evolutionary model that outperformed other computational approaches in phylogeny fit [72]. Thyagarajan and Bloom utilized the same approach to estimate amino acid preferences at each position of influenza HA and elucidated the inherent mutation tolerance of antigenic sites [73, 74]. Wu *et al.* coupled experimental fitness measurements and protein stability prediction for the majority of single mutations in influenza PA polymerase subunit and identified influenza type-specific functional constraints on many naturally non-conserved residues among different influenza types [74]. Wu *et al.* also performed high throughput mutational scanning on NS1 of influenza to measure fitness of most single mutants in the presence or absence of type I interferon and identified detrimental mutations that were sensitive to type I interferon [75]. Qi *et al*. first adopted systematical mutational scanning to determine drug effects on fitness of viral mutants: they tested sensitivity of all possible single amino acid mutations in the active site of nonstructural protein 5A (NS5A) of Hepatitis C to daclatasvir and identified critical residues that mediated drug resistance [76].

The mutational scans of whole viral genes generated remarkably detailed sequence-function maps of viral proteins of clinical significance and exciting insights into possible sequence space of viruses under diverse selection pressures. However, the

reproducibility of these scans, estimated by correlation of functional/fitness measurements on single mutants between biological replicates of the bulk competition or between bulk competition and individually generated clones, was generally modest. Bottleneck effect during transfer of libraries of massive numbers of mutations probably contributes to this suboptimal reproducibility. For example, influenza HA is composed of about 580 amino acids, so a saturation library of every possible single amino acid substitution will encompass at least 11,020 (580 x 19) mutations [73, 77]. The complexity of mutation libraries are likely to trigger a bottleneck effect that leads to stochastic changes in frequency of mutations during generation of a viral library by transfecting the plasmid library into mammalian cell lines and/or randomly sampling from the original viral library to initiate infection experiments. This moderate precision in measuring fitness effect of individual mutations is sufficient for estimating mutational tolerance and strength of selection on each position by averaging across different amino acid substitutions, but becomes suboptimal in precise determination of functional or fitness measurements on specific mutations, which is required for isolation of adaptive mutations with moderate benefit. Although including more experimental replicates and combining them to obtain average fitness estimates on individual mutations partially overcomes this technical challenge, increased precision in fitness estimates will benefit identification of adaptive mutations to therapeutic selection, for which small differences may result in quite distinct evolutionary outcomes. For that reason, high throughput approaches that facilitate systematical and precise functional or fitness measurements of large numbers of mutations are still in great demand.

Protein function as a product of protein expression and activity

Expression (number of properly folded protein molecules) and activity (defined as activity per protein molecule) collectively determine the total function of a protein. The activity of a protein is mostly dependent on the primary amino acid sequence of the protein; thus, many mutations in the primary sequence alter the protein activity [78]. The impact of sequence variation on protein function or organismal fitness is well documented as discussed previously. On the other hand, protein expression is determined by many factors: promoter strength [79], non-translation regions [80], translational regulators [81], epigenetic factors [82], and gene dosages (either duplications or deletions) [83]. Several studies have looked into the effect of changes in expression level on protein function: one well known example is that expression level of Agouti protein regulates coat coloration of mice and drives adaption of mice to varied predation patterns [84]. However, most studies to date have been focused on effects of changes to either protein sequence or expression on protein function or organismal fitness individually, but these two factors are interdependent in shaping protein function and organismal fitness [85].

Changes in protein expression modulate organismal fitness. Direct measurements of *E. coli* growth rate with varied expression of the Lac operon over a wide range of lactose concentrations demonstrated an optimal expression of the Lac operon that maximized the fitness of *E. coli* with different concentrations of lactose [86]. Several *in*

silico studies that simulated *E. coli* growth also suggested the maximization of fitness by tuning of protein expression levels [87-89]. However, significant reduction in expression of many essential proteins confers quite limited defects in fitness [90-93]. For example, heterozygotes with only one functional allele are often highly fit, which indicates even only half of total functional protein sustains growth [94]. This implies a non-linear relationship between protein expression and fitness: at close to wild-type expression levels, the decrease in protein expression causes much less reduction in fitness (the slope is less than one), while at significantly reduced expression levels (in the transition range of the function), reduction in protein expression levels leads to a significant decrease in fitness [95]. This apparent contradiction was addressed by a cost-benefit model, which proposes a balance between the cost of using cellular resources for protein synthesis and the benefit of the specific protein function to improve fitness of the organism [86]. Of note, since total protein function is a product of its expression and activity, the effect of reduction in protein expression (with protein sequence unchanged) on protein function is equivalent to the effect of reduction in protein activity conferred by deleterious mutations (with protein expression unchanged). When combined with elasticity function, this points to a testable prediction: intermediate deleterious effects of mutations on fitness can be hidden at high protein expression levels. Therefore, experimental examination of fitness effect of mutations at varied expression levels are likely to uncover latent detrimental effects of mutations in highly expressed protein and provide a plausible explanation for the mechanism behind this elasticity function, *i.e.* seemingly over-expression of essential proteins may buffer negative effects of deleterious mutations on fitness.

Hsp90: a molecular chaperone that enhances evolvability of other proteins

Heat shock protein 90kDa (Hsp90), also known as Hsp82 in *Saccharomyces cerevisiae*, is a conserved and highly expressed molecular chaperone with a large network of interacting partners. It was first discovered in *Drosophila melanogaster* as a highly expressed protein after exposing the fly larvae to elevated temperatures [96]. Hsp90 is one of the most conserved proteins; even distantly related eukaryote species like yeast and human share at least 50% amino acid sequence identity [97] and eukaryotic Hsp90 shares about 40% amino acid sequence identity with its counterpart in *E. coli* [98]. Hsp90 is highly expressed in the cytoplasm and other cellular compartments, constituting about 1-2% of total proteins in the cytoplasm [91]. Its expression is further up-regulated to 5-6% under stress conditions, such as heat shock [99]. In response to stresses, a transcription factor known as Heat shock factor 1 (HSF1) dissociates from Hsp90 and translocates into the nucleus, where it becomes transcriptionally activated and drives elevated Hsp90 expression [100]. Other factors such as nuclear factor-κB subunit p65 [101], nF-Il6β [102] and STAT3 [103] also induce expression of Hsp90 following different stimulation.

Hsp90 is a homodimer comprising of two 709 amino acid monomers, dimerized by their C terminals (Figure 1.1). Each monomer is composed of three domains: N-terminal domain (NTD), middle domain (MD) and C-terminal domain (CTD) [104]. The

Figure 1.1: Hsp90 is a homodimer and each monomer encompasses three domains

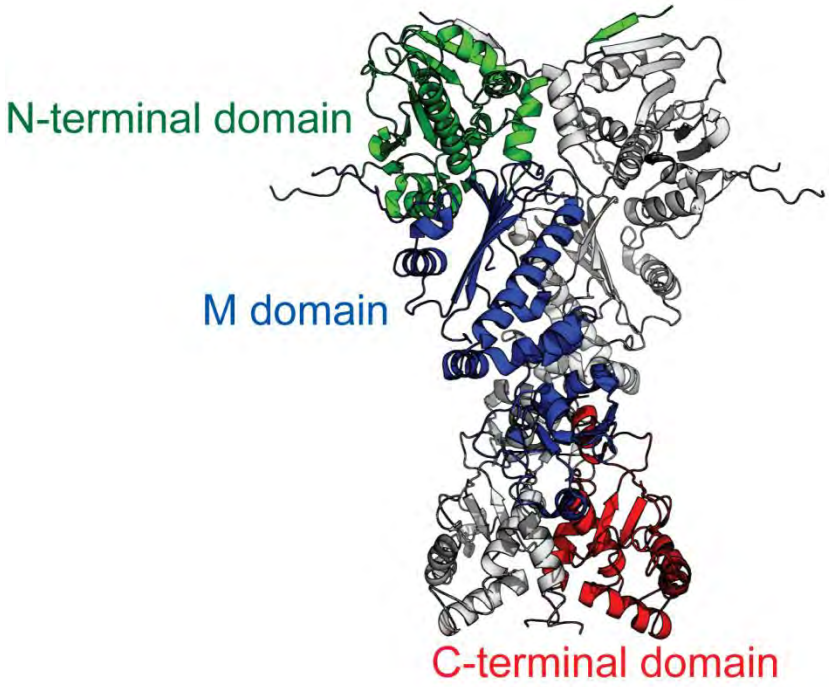


Figure 1.1 Hsp90 is a homodimer and each monomer encompasses three domains. Molecular image based on the PDB ID: 2CG9 structure of Hsp90 [104] shows the two identical monomers of Hsp90. For one monomer, the N-terminal domain is colored green; Middle domain is colored blue; C-terminal domain is colored red [105].

NTD encompasses the highly conserved ATP binding site that binds to and hydrolyzes ATP subsequent to clients and co-chaperones binding [106]. Of note, a 'lid' formed by several conserved residues closes over and covers the nucleotide binding site when bound to ATP [107]. The MD is connected to the NTD within the monomer through a flexible and charged linker that affects Hsp90 function [108]. In the MD, two αβα motifs connect with each other by several α-helices [107]. The MD associates with the NTD to facilitate ATP hydrolysis [109], and heavily involves in interactions with co-chaperones [110] and client recognition and maturation [107, 111]. The CTD comprises a mixed α- and β-structures and mediates dimerization of the homodimers with a four-helix bundle dimer interface formed by two α-helices from each monomer [112]. The five C-terminal residues form a motif (Met-Glu-Glu-Val-Asp; MEEVD) that recognizes a tetratricopeptide repeat (TPR) binding domain and interact with many co-chaperones and clients [113].

Hsp90 undergoes a conformational change driven by hydrolysis of ATP [106]. Although the structural and mechanistic details are still under investigation, a generally agreed model was derived from accumulated biophysical and structural data. Binding of ATP to NTD (mediated by D79 of NTD) followed by closure of a molecular lid gradually shifts the conformational equilibrium of open and closed state to a more closed and active conformation. This closed conformation enables dimerization of two NTD monomers and intramonomer contacts between NTD and MD. Hydrolysis of ATP (mediated by E33 in NTD and R380 in MD) to ADP leads to a second, but unclear state. Then dissociation of

ADP “resets” the conformation to its originally open state [114]. In the ATP cycle, multiple co-chaperones individually or collectively affect functions of Hsp90 by organizing interactions between Hsp90 and other chaperones (HOP) [115], activating (Aha1) or inhibiting (Cdc37 and p23) ATP hydrolysis [116-118], and recruiting clients (Cdc37) [119].

Hsp90 facilitates maturation and proper folding of its clients. More specifically, it keeps the often inherently unstable client (such as kinases and transcription factor) poised for activation until stabilized by conformational changes [120]. Presumably, Hsp90 clients are recognized by their unstable structural features, sequestered by Hsp90 and protected from exposure and aggregation, so that they have higher probability to fold into its native conformation and become activated or functional [97]. Hsp90 also cooperates with E3 ligase to target clients that stay unfolded after multiple rounds of chaperone interaction to ubiquitin-mediated proteasomal degradation [121]. The mechanistic detail of client recognition and the possible role of the ATP cycle on client maturation are still mostly unknown and require further study.

Hsp90 is a central hub in the intracellular protein network and has direct interactions with 3% of the yeast proteome estimated by yeast two-hybrid and SGA analysis [122, 123], although the biological significance of the majority of these interactions are unknown partially due to their transient nature [124]. However, owing to this massive interactome network, the clients of Hsp90 fall into diverse categories,

including signal transduction, innate immunity, steroid and calcium signaling, protein trafficking, viral infection and telomere maintenance [97].

Hsp90 is also known to play mixed roles in the evolution of cryptic mutations. On the one hand, Hsp90 functions as a capacitor that buffers morphological or phenotypic variations, as shown in *Drosophila melanogaster* and *Arabidopsis thaliana* [125, 126]. These studies demonstrated that Hsp90 masks effects of existing variations on morphology or phenotype in nature, while compromised Hsp90 function unleashes the morphological or phenotypic variations associated with these cryptic variations. These traits continue to express under selection, even after restoration of Hsp90 function. On the other hand, Hsp90 potentiates evolution of new traits, in particular drug resistance, by instantly revealing the phenotypic consequences of new mutations [127, 128]. When the Hsp90 mediated protein-folding reservoir is compromised, traits from potentiator-dependent variants disappear, while traits from capacitor-dependent variants reveal [128]. In *Saccharomyces cerevisiae*, Hsp90 appears to act as a capacitor as frequently as a potentiator [128]. In summary, Hsp90 imposes distinct control on the emergence of new traits, followed by selection driven and Hsp90 independent fixation of new traits.

As a capacitor, Hsp90 is predicted to promote fixation of adaptive mutations under stress conditions and confers robustness to deleterious mutations under normal conditions. By buffering traits from cryptic mutations, Hsp90 protects normal

development from detrimental effects of random mutations [125]. Although the buffering capacity of Hsp90 is regarded as a byproduct of its chaperoning function, it facilitates evolution under stress conditions (e.g. increased temperature) that reduce the capacitor function of Hsp90 and thus reveal selective advantages of previously invisible mutations [129]. For example, multiple cryptic mutations exhibit resistance to antibiotics with pharmacologically inhibited Hsp90 [128]. Moreover, accumulation of multiple cryptic mutations enables exploring phenotype of combinations of mutations in a single step [129]. The capacitor or buffering function of Hsp90 has also been demonstrated in zebra fish with a highly selective mechanism that only affects certain traits [130]. Taken together, these studies showed that buffering function of Hsp90 is associated with strength and timing of Hsp90 inhibition, strains (with different pre-existing genetic variants) and developmental stages of organisms and dependence of phenotypes on Hsp90 function. Although more specific targets of Hsp90 buffering system have been mapped, the detailed molecular mechanism remains largely unknown.

Human immunodeficiency virus type 1: discovery, life cycle, pathogenesis and treatment

Human immunodeficiency virus type 1 (HIV-1) was first transmitted from chimpanzees to local residents in sub-Saharan Africa as a zoonotic disease in the 1920s [131, 132] and by the 1970s HIV-1 (referred to as HIV) had been introduced to North America. In 1981, five strange medical cases, termed Acquired Immunodeficiency Syndrome (AIDS), were reported in the United States [133]. It wasn't until 1983 that

HIV was isolated from AIDS patient biopsies [134, 135], and confirmed as the causative agent of AIDS one year later from serological tests and large scale sero-epidemiological studies . The nucleotide sequence of the HIV genome was determined in 1985 [136, 137], which facilitated PCR based measurement of plasma HIV titer [138] and the monitoring of HIV evolution patterns globally. Both serological and molecular based methods contributed to the rapid and accurate diagnosis of HIV infection. With more than 70 million reported cases and 25 million deaths, HIV is one of the most widely disseminated and deadly pandemics of the 20th century.

The replication cycle of HIV starts with the surface trimeric spike (Env) binding to the CD4 receptor on host cells. Env then binds to a co-receptor (either CCR5 or CXCR4) and mediates fusion of the viral and host membranes. Upon membrane fusion, the viral genome and associated proteins are released into the cytosol. Viral RNA is reverse transcribed into cDNA, which is then transported into the nucleus. In the nucleus, viral integrase enables stable integration of cDNA into the human genome, most frequently within the introns of actively expressed genes. The integrated viral genome is expressed in a highly coordinated fashion to produce viral proteins that are trafficked to the cell surface, packaged into budding viral membrane, and cleaved by protease for mature virus assembly.

Many HIV proteins, including reverse transcriptase, integrase and protease, are validated drug targets with potent inhibitors in clinical use. However, due to the high mutation rate and the large population size of HIV, these viral proteins rapidly develop resistance to individual agents [139]. Combinations of drugs targeting different HIV proteins, *e.g.* (non-)nucleoside analog reverse transcriptase inhibitors and protease inhibitors [140-142], significantly raise the barrier for evolved drug resistance and are able to clear more than 99% of circulating viruses within two weeks [143]. However, integration of the HIV genome into resting $CD4^+$ T cells produces a latent reservoir of HIV that is inaccessible to antiviral therapies and can re-establish infection upon disruption of therapy. Highly effective antiviral therapy has converted HIV-1 from an acute, life-threatening disease into a chronic condition that requires continuous therapy throughout a person's lifetime [144, 145]. An imminent consequence is approximately 37 million chronically infected people and two million newly infected cases in 2014 alone, which imposes a heavy burden on the healthcare systems. Thus, HIV vaccines that would prevent the spread and reduce HIV-associated economic burden are in great need.

Vaccines are responsible for the eradication/control of many once deadly diseases, such as smallpox, polio, and measles [146-149]. HIV vaccine development has been focused on the sole surface protein, Env [150]. Unfortunately, vaccine trials based on Env have been mostly unsuccessful [150] due to the unique properties of Env, such as extensive glycosylation, masked receptor binding sites and steric exclusion from

trimerization [151]. However, the moderate success from the recently completed clinical trial (RV144) has revitalized the interest to develop immunogens based on trimeric Env to induce protective antibodies [152]. The regime of RV144 combined vector-expressed trimeric Env and recombinant monomeric gp120, while previous unsuccessful trials only used recombinant monomeric gp120 [152]. RV144 exhibited 60% efficacy against HIV acquisition at 12 months and 31.2% efficacy even at 42 months. Neutralizing antibodies targeting V2 of Env with antibody-dependent cell-mediated cytotoxicity was isolated from RV144 and correlated with vaccine efficacy. The success of RV144 highlighted the importance of using the native trimeric Env as immunogen for vaccine trial.

The surface trimeric Env spike

The trimeric Env spike, which mediates viral entry, is the sole protein on the surface of HIV. It is composed of three identical exterior envelope glycoproteins gp120 and three identical trans-membrane subunits gp41 (Figure 1.2). The precursor gp160 is cleaved to form gp120 and gp41, which are non-covalently attached to each other [153]. gp120 is responsible for binding to both the primary receptor (CD4) [154, 155] and the co-receptor (CCR5 or CXCR4) [156, 157], while gp41 anchors the trimeric spike in the viral membrane. A multiple sequence alignment of gp120s from a large collection of natural isolates identified 5 variable loops (V1-V5) that modulate antigenicity along with several conserved regions that form the discontinuous quaternary structure important for CD4 binding [158]. Both the variable loops and the conserved regions are heavily

Figure 1.2: Env is a homotrimer and each monomer encompasses two non-covalently linked subunits: gp120 and gp41

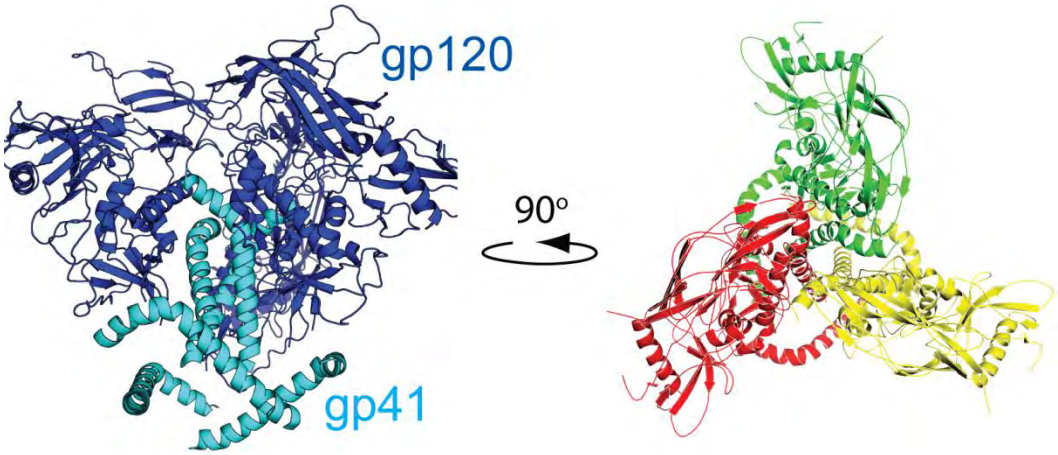


Figure 1.2 Env is a homotrimer and each monomer encompasses two non-covalently linked subunits: gp120 and gp41. Molecular image based on the PDB ID: 4NCO [159] structure of Env shows gp120, which points away from the viral membrane, and gp41, which is embedded into the viral membrane. In the right panel, after 90 degree rotation, the three identical monomers are colored differently (green, red and yellow).

glycosylated [158]. While glycosylation helps to mediate immune escape, broadly neutralizing antibodies can be solicited to target the glycans [160]. This extensive glycosylation and the inherent conformational flexibility impeded efforts to determine the gp120 structure by crystallography; however, in a landmark study by Kwong *et al*., the first high quality crystal structure of gp120 was determined through the use of a deglycosylated gp120 core with the inherently disordered variable loops removed [161]. This crystal structure revealed that gp120 folds into two parallel aligned major domains: an inner domain and an outer domain. The inner domain is composed of a small five-stranded β-sandwich stacked on a two-helix and two-strand bundle that encompasses V1-V2 loops; the outer domain is comprised of a parallel double barrel that encompasses the V3-V5 loops and the CD4 binding site (Figure 1.3). Crystal structures based on gp120 monomers with variable loops or trimers that resemble the native spike further reveal different conformation states with considerable details and resolution [159, 162-166]. Crystal structures of gp120 monomers shared a similar conformation as those observed for gp120 within the native trimer [159]. The detailed structural information of Env enables dissection of the conformational changes of Env, which are critical for its function and immune escape.

Conformational change and epitope exposure of the Env complex

gp120 has substantial conformational flexibility and undergoes a conformational change upon binding to CD4 (Figure 1.4) [167]. Before binding to CD4, gp120 maintains

Figure 1.3: gp120 of Env encompasses three domains

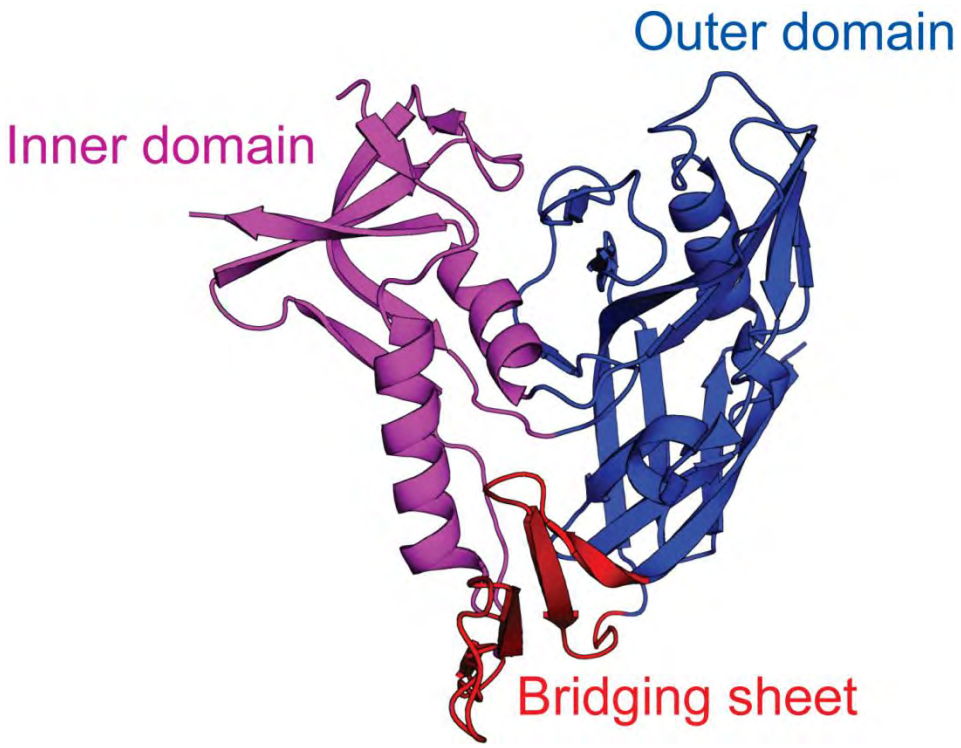


Figure 1.3 gp120 of Env encompasses three domains. Molecular image based on the PDB ID: 1gc1 [161] structure of gp120 highlights an inner domain (purple), an outer domain (blue) and bridging sheet [38].

Figure 1.4: Env undergoes conformational change before and after binding to CD4

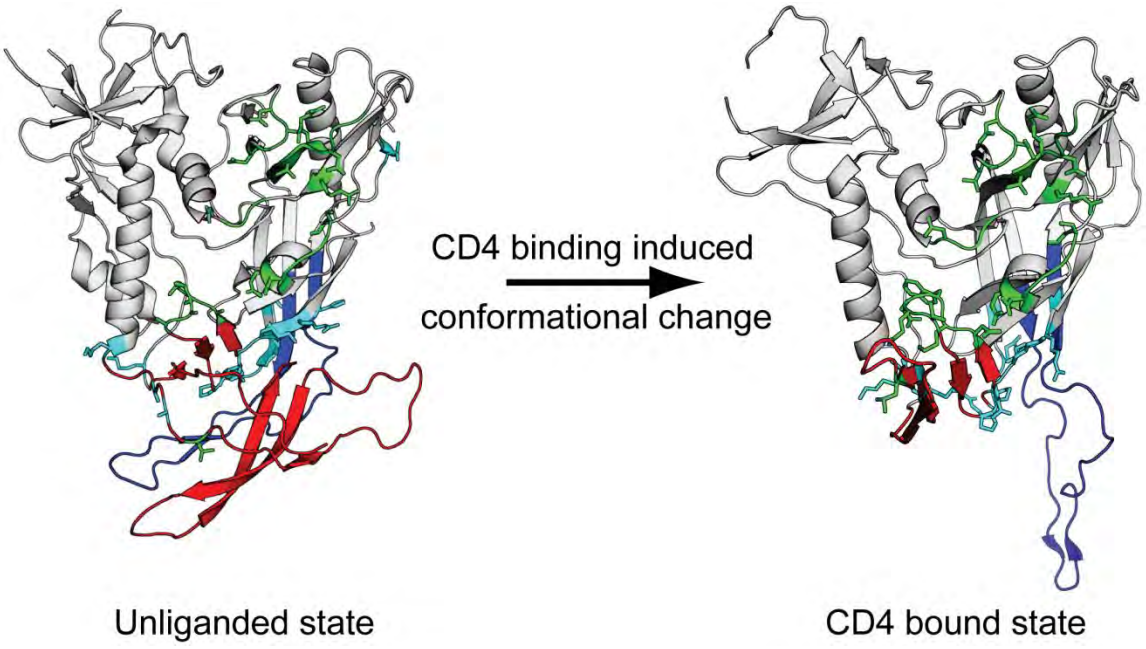


Figure 1.4 Env undergoes conformational change before and after binding to CD4. Molecular image based on the PDB ID: 4NCO structure of Env [159] shows the conformation of one gp120 monomer before binding to CD4. In particular, the bridging sheet (colored red) and co-receptor binding site (colored cyan) are not completely formed; V3 loop (colored blue) is masked by V1V2 loop. After binding to CD4, Env undergoes conformational change. The molecular image based on the PDB ID: 2b4c [163] shows that the bridging sheet and co-receptor binding site are now well formed and the V3 loop is exposed.

a closed ground state conformation that facilitates evasion from antibody recognition; upon binding to CD4, gp120 switches to an open state that unmasks the co-receptor binding site (either CCR5 or CXCR4). Once bound to both CD4 and the co-receptor, gp120 changes conformation to a post-fusion state that presents gp41 with the penetrating peptide to initiate membrane fusion [165]. In the closed state, the trimer resembles the shape of a "mushroom" or a "three-blade propeller" [159, 165]. The V1/V2 loops shield the CD4 binding site, while the V2/V3 loops shield the co-receptor binding site. The apex of V3 loop is covered by the glycan of N197 from the V2 loop, so it is not exposed in the closed state. Of note, the V1, V2 and V3 loops together constitute the trimer associate domain [168] at the membrane-distal apex of the spike [169]. The trimeric Env spike in the closed state is generally resistant to neutralizing antibody binding. This resistance is due to both the extraordinary density of surface glycans, which prevents epitope accessibility [170], and a conformational barrier, which stimulates entropy penalty for antibody-binding induced conformational change [171]. Notably, only 2% of the gp120 surface has significant sequence conservation and therefore remains accessible to antibody recognition. However, the majority of these conserved residues are located at the base of gp120, proximal to the viral membrane and therefore sterically occluded from antibody binding [165].

Upon binding to CD4, the trimeric Env spike undergoes a major conformational change: the TAD dissociates at the trimeric apex due to movements of V1-V3 loops [172].

The V1 and V2 loops separate and shift away from the inner domain, unmasking the V3 crown and orienting it towards the host cell [161, 163]. An antiparallel, four-stranded minidomain, coined “bridging sheet”, is formed between the outer and inner domains, which is critical for CD4 binding (Figure 1.3 and 1.4) [161]. The exposed V3 crown and part of the bridging sheet form the co-receptor binding site (Figure 1.4) [161]. It is the interaction between gp120 and co-receptor that pulls the trimeric Env spike to the host cell membrane to initiate membrane fusion [165]. Compared to the closed state, the CD4-bound state exhibits a considerably higher level of conserved, non-glycosylated surface residues that are better targets for neutralizing antibodies [165]. The major conformational change of Env exposes sites critical for not only viral entry, but also for neutralizing antibody recognition.

Antibodies that target HIV surface receptor

Antibodies that target gp120 either as a trimer or a monomer recognize highly specific epitopes, many of which exist only in the closed or CD4-bound intermediate state. These antibodies can be utilized as molecular probes to inspect the conformational state of the trimeric Env spike. Of note, broadly neutralizing antibodies, which are able to neutralize diverse primary isolates, frequently bind to the highly conserved regions in the closed state, whereas strain-specific antibodies, or non-neutralizing antibodies, often target the gp120 surface that is only accessible in the monomer [173]. For example, VRC01, a broadly neutralizing antibody, binds to an epitope in the CD4 binding site that

exists in both monomeric and trimeric gp120 and neutralizes more than 90% of primary isolates, whereas b6 and b12, strain-specific antibodies, bind to an epitope in the CD4 binding site of monomeric gp120 and neutralize less than 30% of primary isolates (Figure 1.5A). Of note, comparison of epitopes between strain-specific and broad neutralizing antibodies uncovers a general strategy of HIV escape from humoral immunity: HIV viral particles releases shedding gp120 subunit as decoys that display masked epitopes in the trimeric spike and induces antibodies that mostly fail to neutralize trimeric Env spike in the closed state [174].

Although strain-specific neutralizing antibodies are less therapeutically valuable, they can be used to probe the conformational state of the Env complex. For instance, PGT121-131 and 135-137 recognize conserved V3 glycans (N301 and N322), while 447-52D recognizes the V3 apex (Figure 1.5B); PG9/16 and PGT 145 targets V1V2 glycans that is properly presented in the intact TAD, but not disrupted TAD (Figure 1.5C) [173]. Therefore, sensitivity of HIV to b6/b12 gauges exposure of the CD4 binding site and transition to the CD4 bound conformational state. Sensitivity of PG9/16 or PGT145 reveals integrity of the TAD and openness of the trimeric spike. Sensitivity of PGT135-137 reflects glycan patterns at the V3 loop, while 447-52D probes exposure of the V3 tip that forms the co-receptor site.

Figure 1.5: Neutralizing antibodies bind to different epitopes of Env

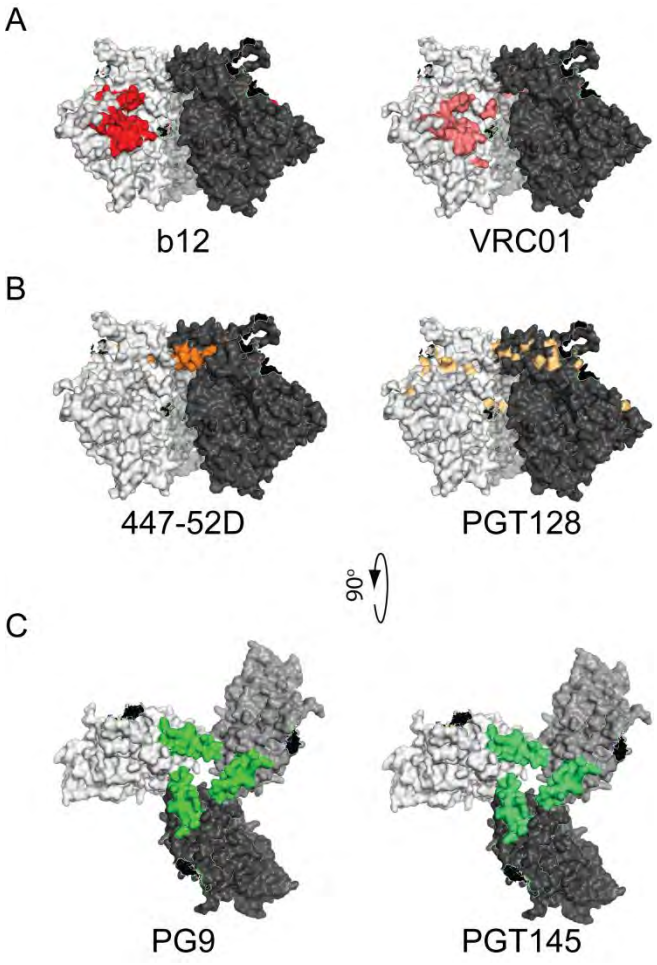


Figure 1.5 Neutralizing antibodies bind to different epitopes of Env. (A) Epitopes of two antibodies that bind to the CD4 binding site of Env: b12 and VRC01. (B) Epitopes of two antibodies that bind to the V3 loop: 447-52D and PGT128. (C) Epitopes of two antibodies that bind to the V2 loop and trimer association domain: PG9 and PGT145. For each panel, the former one is narrow neutralizing antibody and the latter one is broad neutralizing antibody. All molecular images are based on the PDB ID: 4NCO structure of Env [159].

Env: CD4 interaction and the CD4 binding site

The binding of the trimeric Env spike to CD4 is the initial step in the entry stage of HIV. Characterizing the residues that form the interfaces between these proteins yielded insightful mechanistic details of the early events in HIV entry and inform potential therapeutic interventions. The crystal structure of gp120 bound to CD4 by Kwong *et al.* provided a high-resolution snapshot of this interaction [161]. In gp120, the outer domain, inner domain and the bridging sheet form the binding interface (Figure 1.3). Residues that constitute the CD4 binding site (CD4BS) are from six different segments of gp120 (Figure 1.6A). In particular, the interactions between residues 365-371 and 425-430 (HxB82 numbering) of gp120 [175] and residue 43 of CD4 account for 57% of the total interaction between gp120 and CD4 (Figure 1.6B). The core of the CD4BS is the "CD4 binding loop" composed of residues 366-370 (GGDPE). The CD4 binding loop and Trp427 make multiple contacts with Phe43 and Arg59 of CD4 (Figure 1.6B). These contacts make up a large fraction of the total interactions between gp120 and CD4. At the contacting surface, an imprecise complementing electrostatic interaction (positively charged CD4 and negatively charged gp120) stabilizes binding and a mismatch in surface topology creates two interfacial cavities: a large cavity that is solvent accessible and a small cavity with limited solvent accessibility. The large cavity is predominantly lined by hydrophilic residues, half of which are gp120 derived and the other half are CD4 derived. The lining residues from gp120 are mostly adjacent to CD4 contact residues and show significant sequence variation, which suggests a tolerance to mutation and possible role in viral escape from antibody neutralization. The small cavity is called "The Phe43

Figure 1.6: Env has a defined binding site for CD4

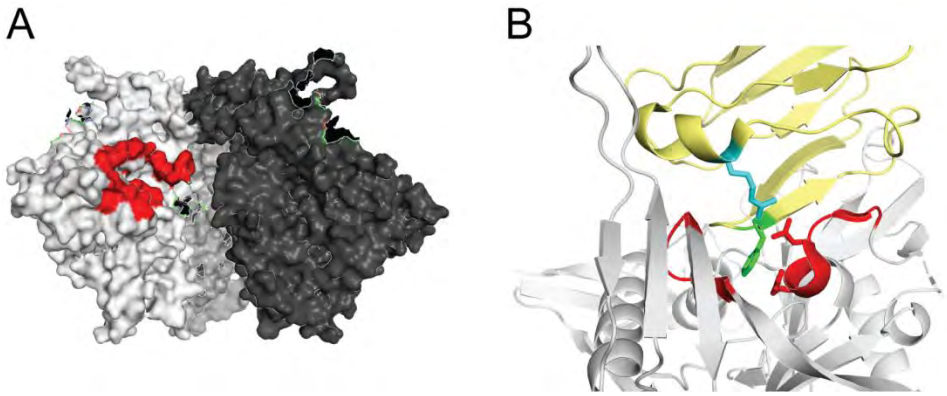


Figure 1.6 Env has a defined binding site for CD4. (A) Molecular image based on the PDB ID: 4NCO structure of Env trimer shows the CD4 binding site [38]. (B) Molecular image based on the PDB ID: 1gc1 structure of gp120 (grey) monomer bound to CD4 (yellow) [161] shows the two regions of gp120 (365-371 and 425-430) with most interactions with major contact residues of CD4: Phe43 and (green) and Arg59 (cyan).

cavity", because Phe43 of CD4 protrudes into it. The Phe43 cavity is deeply rooted in the hydrophobic core of gp120, at the intersection of the inner, outer domain and the bridging sheet. It is mainly lined with highly conserved, hydrophobic residues that favor small substitutions with little steric hindrance, indicating functional constraints. Indeed, many mutations in the Phe43 cavity affect the gp120 and CD4 interaction [176].

The flanking regions of the CD4 binding loop have varied sequence conservation and biological functions. Many residues of the N-terminal flanking region, including residues 362-365, are located in close proximity to the large interfacial cavity and exhibit significant sequence variability [161]. Mutations at residues 361-364 modulate HIV infection/tropism in macrophages through modulating the exposure of the CD4 contact residues (Asp368 and Glu370) [177]. In contrast, many residues in the C-terminal flanking region (*e.g.* residues 375-377) reside in the Phe43 cavity and are more conserved across primary isolates [161]. This conservation is likely due to the direct role these residues play in the gp120:CD4 interaction. Of note, aromatic mutations (*e.g.* Trp, Tyr, Phe and His) at residue Ser375 appear to occupy the Phe43 cavity and drive the conformation of gp120 closer to a CD4-bound state with a large decrease in entropy. These mutations also cause reduced sensitivity to b12 and the entry inhibitor BMS-599793, although they are more sensitive to soluble CD4 (sCD4) [176, 178, 179]. S375H cooperates with layer1 of the inner domain to mediate increased binding to CD4 and CCR5, contributing to adaptation of SIV to infect the macrophages of non-human

primates, which express lower levels of CD4 and CCR5 [179]. The flanking regions of the CD4 binding loop harbor residues that are part of the epitopes of neutralizing CD4BS antibodies (*e.g.* b12 and VRC01) [166, 180]. In fact, the CD4 binding site represents a supersite that is the target of diverse broad neutralizing antibodies isolated from chronically infected individuals [181]. Mutations at residue 386 in the C terminal flanking region of CD4 binding loop also affects neutralization sensitivity to 2G12, a glycan-specific antibody [182]. In summary, the CD4 binding loop and its flanking regions modulate the interactions between gp120 and CD4 and the sensitivity of gp120 to various neutralization antibodies.

Influenza A virus and pandemics

As a member of the *Orthomyxoviridae* family, influenza virus can be classified into three distinct serological types (A, B and C) that differ in their host ranges and pathogenicity [183]. Only influenza A and B virus are pathogenic in humans, and influenza A virus (IAV) is more pathogenic and evolvable [183]. IAV infects a wide range of animals including birds, pigs, horses and dogs, while aquatic birds appear to serve as the natural reservoir. Close and frequent contact between humans and these animals result in multiple cross-species transmissions [184]. IAV can be classified based on the antigenicity of its surface glycoproteins hemagglutinin (HA) and neuraminidase (NA). HA and NA co-determine the subtype of IAV, *e.g.* H1N1 and H3N2. IAV causes seasonal influenza outbreaks every winter that results in approximately 23,000 associated

deaths on average (data from CDC) and 10.4 billion dollars in costs for hospitalization and treatment [185]. H3N2 and H1N1 are currently co-circulating in the human population. Interestingly, the seasonal strain is usually the strain from the last pandemic.

IAV also causes intermittent global pandemics that give rise to millions of infections, higher mortality and heavy burdens on the medical care system. The most notorious pandemic is the 1918 “Spanish Flu” that killed approximately 50 million people worldwide (675000 in the United States): the mortality rate exceeded 2.5%, much higher than the usual rate of less than 0.1% [186]. Most people died of concomitant bacterial/viral pneumonia. Notably, besides infants under the age of four years and senior people above the age of 75 years, this pandemic also resulted in unusually high mortality rates in young people aged 20-35 years who typically possess strong immunity to influenza infection [183]. Analysis of the viral genome isolated from preserved samples demonstrated that “Spanish flu” was caused by an avian-like H1N1 strain that adapted to infecting epithelial cells lining the human upper respiratory tract, indicating a direct transmission from birds [187]. Reverse genetically reconstructed “Spanish flu” strain elicited aberrant innate immune responses in animal models and provided a plausible explanation for the high mortality rate in young people [188-190]. Of note, the direct descendants of the 1918 H1N1 strains have been circulating in the human and swine population and contributing genes to new IAV genetic reassortment [183].

The second pandemic was the 1957 “Asian flu” that originated in China, disseminated in East and Southeast Asia and then spread to North America and Europe. This pandemic caused about two million influenza related deaths worldwide and 70,000 deaths in the United States alone [183]. Genomic analysis showed that avian-flu genes combined with seasonal human influenza viral genes to create a human/avian H2N2 influenza reassortant that caused this pandemic [191, 192]. The third pandemic was the 1968 “Hong Kong flu” that started in Hong Kong and spread to the rest of world, causing pandemics in the winter of 1968-1970. This pandemic killed around one million people globally (33,800 deaths in the United States alone) [183]. The responsible IAV was an H3N2 reassortant from the previous pandemic strain (H2N2) and an HA gene of avian-flu origin [191, 192]. The most recent pandemic was the 2009 “Swine flu” that originated in Mexico, spread to the United States and then the rest of the world. It was caused by an H1N1 reassortant between a North America H1N2 “triple” reassortant and an H1N1 Eurasian avian-like swine virus that was transmitted from pigs to humans [193]. Owing to epidemiological control and availability of antibiotics that treat pneumonia, the total number of global respiratory deaths and related cardiovascular diseases (approximately 0.3 million) was lower than all the past pandemics [194].

Viral proteins and the replication cycle of IAV

IAVs are negative-strand RNA viruses that have segmented genomes, compared to a single cohesive genome of HIV that encompasses multiple genes. IAV is an enveloped virus that utilizes a lipid membrane from host cells to enclose viral proteins

and the genome. IAV has eight gene segments that encode at least 11 open reading frames. Three of them are embedded membrane proteins: hemagglutinin (HA), neuraminidase (NA) and matrix 2 (M2). The matrix 1 protein (M1) that lies beneath the membrane interacts with the surface proteins as well as ribonucleoproteins (RNPs), so it serves as a structural scaffold for a viral particle. Each RNP consist of a viral RNA strand, the polymerase complex heterotrimers including polymerase basic protein 2 (PB2), polymerase basic protein 1 (PB1) and polymerase acidic protein (PA) and nucleoprotein (NP) [37]. As a RNA binding protein, NP serves as the scaffold for the viral RNA. The polymerase heterotrimers bind to short hairpins in the 5' and 3' untranslational regions (UTRs) of each RNA segment. There are two non-structural proteins (NS1 and NS2): NS1 plays a variety of functions including enhancing viral mRNA translation, impairing host mRNA translation, and mediating type I interferon antagonism; NS2 mainly mediates nuclear export of viral RNPs.

During viral entry, HA recognizes sialic acid (SA, N-acetyl neuraminic acid) and binds to SA-associated host glycoproteins. After the initial binding, the viral particle is internalized into the endosome. An acidic pH environment in the endosome triggers conformational change in HA, which mediates protrusion of the hidden fusion peptide into the endosomal membrane. Subsequent structural rearrangements in HA facilitates fusion of the viral and endosomal membrane. Meanwhile, M2 forms an ion channel to facilitate proton influx and acidification of the interior of viral particles, assisting dissociation of the M1 from RNP complexes. Viral RNP is released into the cytoplasm

and exported into the nucleus. The polymerase heterotrimers enable viral RNA to "hijack" host cell machinery for its own replication and transcription: PB2 binds to the 5' cap structure of host mRNAs; PA cleaves the cap structure through its endonuclease activity and seizes the cap for priming viral mRNA synthesis (namely "cap snatching") by the RNA-dependent polymerase activity of PB1. Two viral mRNAs (M and NS) are spliced to yield M1/2 and NS1/2 respectively. Viral mRNA is then exported into the cytoplasm for translation. Of note, "cap-snatching" appears to deplete host mRNA of their caps, targeting host mRNA for rapid degradation. "Cap-snatching" and several NS1 related "tricks" down-regulate the translation of host mRNAs and redirect the host translational apparatus to preferentially translate viral mRNAs in the cytoplasm. The newly synthesized polymerase subunits and NP are then imported back into the nucleus for catalyzing another round of replication, transcription of viral RNA and formation of RNP, while M1 and NS1/2 are imported into the nucleus to assist export of RNP. Notably, NP regulates the trafficking of RNPs between the nucleus and cytoplasm. Surface proteins HA, NA and M2 are transported to the cell membrane through endoplasmic reticulum, undergoing glycosylation and additional post-translational modifications. At the site of viral budding on the cell membrane where HA and NA are enriched, the eight viral RNPs are incorporated through viral RNA segment-specific packaging signals, followed by M2-dependent membrane scission. NA cleaves SA on cell as well as viral surface through its sialidase activity to allow efficient release of viral progenies and to prevent aggregation of newly produced viral particles.

HA and NA: substrate binding/processing and functional balance

HA is a type I membrane glycoprotein (single transmembrane pass with N-terminus exposed on the exterior side) integrated into the viral membrane. Mature HA is a homotrimer and each monomer encompasses two subunits (HA1 and HA2) connected by disulfide-bonds (Figure 1.7A). Proteolytic cleavage to generate disulfide-bonds linked HA1 and HA2 from the progenitor HA0 is necessary for HA activation. The cleavage of HA is generally carried out by exogenous serine proteases that recognize Q/E-X-R motif [195]. However, insertional mutations in the HA cleavage site of H5 and H7 subtype enable a furin-like (R-X-R/K-R) recognition, a polybasic cleavage site that broadens specificity of protease, enhancing intracellular cleavage activation and systemic replication [198]. HA recognizes disaccharides made up of SAs and underlying sugars including N-Acetylgalactosamine (GalNAc) or galactose. SA is linked to Gal/GalNAc with either α2,6 or α2,3-linkage, i.e. SAα2,6Gal or SAα2,3Gal. SAα2,6Gal is predominantly expressed in the human upper respiratory tract including the trachea, so IAV adapted to infecting humans preferentially expresses SAα2,6Gal. SAα2,3Gal is predominantly expressed in the lower intestinal tract of birds, so IAV adapted to infecting birds preferentially expresses SAα2,3Gal. A panel of amino acids in the receptor-binding site of HA (mainly amino acid position 130-139, 190-199 and 220-229) determines the binding specificity and mutations in these positions lead to a switch in binding specificity and thus host preference. IAV is largely asymptotic in its natural reservoir, wild birds. Transmission to domestic poultry increases the likelihood for secondary transmission to mammals including human and swine, and occasionally leads to a highly pathogenic

Figure 1.7: HA is a homotrimer while NA is a homotetramer

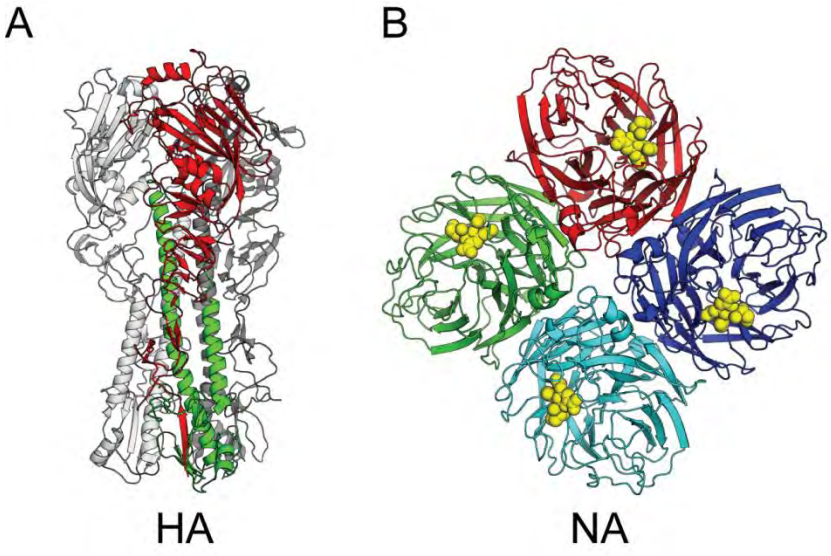


Figure 1.7 HA is a homotrimer while NA is a homotetramer. (A) Molecular image based on the PDB ID: 1ruz structure of HA [196] shows the three identical monomers of HA. In one monomer, the HA1 subunit is colored red while HA2 subunit is colored green. (B) Molecular image based on the PDB ID: 3B7E structure of NA [197] shows the four identical monomers of NA. Each monomer binds to a NA competitive inhibitor, zanamivir (yellow) at its substrate binding pocket.

avian strain [198]. Sporadic avian influenza transmission from wild birds and domestic poultry to mammals reiteratively introduces novel influenza strains into humans and pigs. Expression of both 2,3 and 2,6 glycosidic bonds in epithelial cells lining the upper respiratory tract of pigs makes pigs susceptible to infections with both avian and human strains, so pigs become "mixing-vessels" or intermediate media for potential incubation of human pandemic influenza strains [199]. Most influenza infection in humans is restricted to the upper respiratory tract and causes mild to intermediate respiratory syndrome. However, highly pathogenic avian strains are able to reach the lower respiratory tract and lung (bronchiole and alveolar wall) lined with epithelial cells that express SAα2,3Gal and cause life-threatening pneumonia and systematic infection [200, 201].

NA is a type II transmembrane protein (single transmembrane pass with C-terminus exposed on the exterior side) that cleaves the α2,3 or α2,6 glycosidic bonds in the SA and Gal/GalNAc disaccharide to facilitate release of viral progenies and prevent aggregation of virions. Mature NA is a homo-tetramer with a well-defined substrate binding site (Figure 1.7B). The crystal structures of NA in apo- or substrates/inhibitors-bound state have been solved as monomer, dimer or tetramer in diverse subtypes, revealing the biophysical details of NA structure and substrate recognition [197, 202-205]. Although up to 50% sequence variation exists between different subtypes (*e.g.* N1 and N2), residues that constitute the substrate binding sites are highly conserved in IAV, ensuring a conserved three-dimensional structure. The mushroom-shaped NA tetramer

encompasses four identical monomers in a square-planar arrangement and is anchored to the viral membrane by a long and thin stalk at its N-terminus [197, 205]. The large binding pocket is located on the upper surface of each subunit, surrounded by densely clustered charged residues. Eight functional residues (including R118, D151, R152, R224, E276, R292, R371 and Y406) contact the substrate (SA) and form electrostatic-force-based hydrogen bonding network to poise the substrate in the appropriate angle for catalytic cleavage reaction [206]. Of note, the carboxylate group of C-1 of SA forms a salt bridge with R371 as well as charge-charge interactions with R118 and R292. These interactions make the greatest contribution to binding of SA to NA [206]. Moreover, 11 framework residues (including E119, R156, W178, S179, D198, I222, E227, H274, E277, N294 and E425) that are also highly conserved in IAV contribute to stabilization of the substrate binding site [207]. Catalytic cleavage by NA consists of four major steps [206]. (1) R118, R292 and R371 forces SA to undergo conformational change and become activated for hydrolysis. The involved energy loss is partially compensated by the hydrogen bond network of NA. (2) R151 and R152 activate a water molecule that donates protons to the carboxylate groups of SA and induces formation of a transition-state. The endocyclic sialosyl cation transition-state intermediate is stabilized by E277 [208]. SA is (3) formed through a SN1 mechanism and (4) released along with a transition from α-anomer to a more stable β-anomer. The detailed information about the substrate binding pocket, catalytic steps and active site residues has made NA an ideal drug target.

HA and NA co-determine the subtype of IAV and exhibit a delicate functional balance. Eighteen HA subtypes and eleven NA subtypes have been identified up to now, resulting in 198 possible HA-NA combinations. A large fraction of them have been isolated in birds, so they may "splash" to humans after accumulating adapting mutations. Phylogenetic evidence suggests birds as the original source of evolved novel IAV strains in mammals [183]. HA and NA share the same substrate (SA) and play nearly opposing roles (binding of HA to SA versus removal of SA by NA), so there must be a precise balance between the relative function of HA and NA, coordinating the relative strength of their functions to ensure successful infection [209]. Perturbation to this balance is likely to cause either delayed viral entry or insufficient release of viral progenies. Indeed, much experimental evidence has demonstrated the importance of HA:NA functional balance. Deletions in the stalk region or mutations in the active site frequently reduce enzymatic activity of NA and impede propagation of influenza virus in chicken eggs; mutations in the receptor binding site of HA that reduces SA binding activity of HA is able to restore growth of influenza virus in eggs [210, 211]. Meanwhile, deglycosylating mutations in HA often enhance substrate binding, causing impaired viral growth kinetics in susceptible cell lines; NA with elongated stalk region or stronger enzymatic activity is able to rescue the growth defect [212, 213]. Of note, HA and NA activity is unbalanced in the swine precursor of the 2009 pandemic strain, whereas reassortment of a new HA into the precursor restores the HA:NA balance, allowing for efficient air-borne transmission of the 2009 pandemic strain in human population [214, 215]. Despite these studies providing exciting discoveries, they are dependent on limited number of individual

mutations, so further studies are warranted to systematically map the coordinating or balancing effect between HA and NA and to quantitatively explore the optimal ratio between the activity of HA and NA for efficient IAV infection.

Evolutionary pathways of IAV

IAV rapidly adapts to varied selection pressures through accumulation of mutations or reassortment of gene segments. The mutation rate of influenza is estimated to be 10^{-6} to 10^{-5} mutations per nucleotide per infectious cycle [216, 217], which is ~200-fold higher than DNA viruses and ~10,000-fold higher than bacteria and eukaryotes [183]. Such high mutation rates, coupled with the large population size of influenza, allow for sampling of every possible single nucleotide mutation even in a single replication cycle and rapid fixation of spontaneous mutations that provide growth advantage [68]. Humoral immunity against IAV depends on pre-existing neutralizing antibodies that bind to antigenic sites of HA to thwart viral infections [218]. Antibodies that bind to NA do not prevent infection, but reduce illness duration and/or severity [219, 220]. However, IAV is able to evade pre-existing immunity through two mechanisms: accumulation of amino acid mutations in the antigenic site of HA that impedes recognition of HA and reassortment of gene segments encoding HA and NA to create novel combinations of surface antigenic glycoproteins. In H3N2, accumulation of mutations in HA (genetic evolution) facilitates punctuated dramatic changes in viral sensitivities to antibody neutralization (antigenic evolution) and escape from immune response [221, 222]. Phylogenetic evidence based on the genetic sequences of HA from H3N2 strains, which

were harvested at different dates and locations, highlighted a single central trunk in the phylogenetic tree that indicated massive lineage extinction, limited genetic diversity at any time point and a single successful lineage that continuously evades human immunity [223-225]. Other influenza viral proteins also exhibited high rates of substitutions, which indicates functional constraints on possible combinations of influenza gene segments, optimal compatibility to accommodate escaping mutations in HA and a potential buffer of fitness loss conferred by HA mutations [224].

Antiviral agents against IAV

Two types of inhibitors have been developed for the prophylaxis and treatment of influenza infection: M2 ion channel inhibitors and NA inhibitors. M2 ion channel inhibitors include Amantadine [226], rimantadine [227] and adamantane derivatives [228]. These agents block the ion channel pore of the M2 helix bundle and prevent H+ ions from being imported into the endosome with internalized influenza virions. Inhibition of ion transport impedes pH drop and associated conformational change of HA, which is necessary for uncoating of viral particles [229, 230]. The usage of amantadine and rimantadine has been limited due to rapid emergence and transmission of replication competent and pathogenic drug resistance mutations, mainly M2 mutation S31N [231, 232]. In addition, adamantanes cause severe central nervous system (CNS) side effects, which further limit its use [227]. NA inhibitors include FDA approved oseltamivir, zanamivir and peramivir; laninamivir is approved in Japan and still under clinical trial in the United States. All of these inhibitors compete with the natural substrate sialic acid

(SA) in NA binding, inhibiting NA enzymatic activity to obstruct release of viral progenies. Both zanamivir and oseltamivir are highly specific and potent inhibitors of NA with IC50 ≤ 1ng/ml. They inhibit viral replication *in vitro* and in animal models (mice and ferrets). Both are well tolerated and highly effective in prophylaxis (reducing the number of illnesses) and treatment (shortening illness duration) of influenza infection in humans [233]. In particular, early administration of NA inhibitors has been shown to more effectively reduce the duration of illness, the severity of syndromes (fever and sinusitis) and complications (bronchitis and pneumonia), the burden on healthcare systems (elevated demand on antibiotics and hospitalization), and the transmission among healthcare or household contacts [234, 235]. Post-exposure usage of oseltamivir has also been shown to prevent household transmission [236].

Competitive inhibitor of NA

The first available NA inhibitor is zanamivir (trade name: Relenza), which was approved by the FDA in 1999. The discovery and development of zanamivir serve as one of the most successful examples of structural based drug design. Two derivatives of sialic acid (Neu5Ac, Figure 6.1), 2-deoxy-2,3-didehydro-*N*-acetylneuraminic acid (Neu5Ac2en, DANA) and 2-deoxy-2,3-didehydro-N-trifluoroacetylneuraminic acid (FANA) inhibit NA function at micromolar concentrations and have become the core leads for development of more potent inhibitors [237]. The Crystal structure of NA bound to SA with improved resolution [238] has enabled computational chemistry based structural modification of the micromolar potent leads to obtain nanomolar potent compounds [239].

Particularly, the interactions between residues within the NA substrate binding pocket and functional groups of inhibitors are refined to be energetically favorable (Figure 1.8) [240]. The key discovery was the capability of the C-4 hydroxyl group of DANA to accommodate a larger basic group. Replacement of the hydroxyl group by a guanidinyl group markedly enhances the affinity of the inhibitor to NA (Figure 6.1) [239]. The increased affinity is estimated to be a result of fully occupying the active site by the larger basic moiety, which interacts with E119 and E228 of NA[206, 241]. Zanamivir is 10000 fold more potent compound *in vitro* and *in vivo* [242]. In addition, zanamivir is highly specific for influenza NA, exhibiting limited affinity to sialidase from other sources [244] including the human sialidase, which reduces the likelihood of severe side effects [245]. Zanamivir is formulated as an inhaled powder (10mg twice daily) due to its highly polar nature and limited oral bioavailability. Inhalation delivers the drug directly to the upper respiratory system where the infection predominantly occurs. Intravenous delivered zanamivir was developed and approved during the 2009 swine flu pandemic for patients unable to take inhaled zanamivir.

The second FDA approved NA inhibitor is oseltamivir (trade name: tamiflu), an orally administered drug for treating IAV. It is more commonly used in the clinic and stockpiled by many countries in preparation for future pandemics [246]. Based on cyclohexenes, oseltamivir is designed to mimic the predicted transition state sialosyl cation using extensive medicinal chemistry optimization [247]. Moreover, the glycerol group at C-6 of SA and zanamivir is replaced with a highly hydrophobic pentyloxy

Figure 1.8: Differential binding interactions of oseltamivir and zanamivir with NA

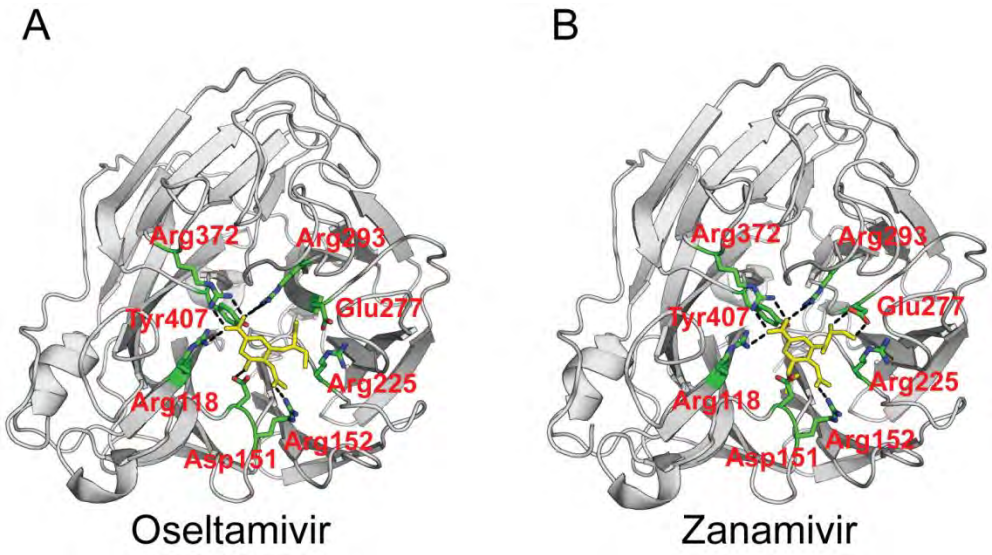


Figure 1.8 Differential binding interactions of oseltamivir and zanamivir with NA. Molecular image of NA bound to oseltamivir (A, PDB ID: 3TI6) and zanamivir (B, PDB ID: 3TI5) [243]. The competitive inhibitors are colored yellow in the center of the binding pocket. All contact residues of NA with the substrate are highlighted as sticks, with each amino acid labeled. The hydrogen bonds formed are shown as black dashes. Hydrogen bonds are formed between oseltamivir/ zanamivir and Arg118, Arg151, Arg152, Arg293, and Arg372. For oseltamivir, its C6 pentolxy group interacts with Arg225 and fits into the hydrophobic pocket near Glu277. For zanamivir, its C6 glycerol group forms hydrogen bonds with Glu277.

substituent to improve lipophilicity (Figure 6.1) [247]. Owing to this hydrophobic group at C-6 of oseltamivir, binding of oseltamivir to NA induces conformational change. E277 orients away from the active site; the hydrophobic group at C-6 interacts with R225, generating a hydrophobic pocket to accommodate itself (Figure 1.8A) [247, 248]. This is distinct from hydrogen bond interactions between E277 and the glycerol group of zanamivir or SA [238]. The oseltamivir induced conformational change potentially exposes an Achilles heel for evolution of drug resistance. Interestingly, replacing the hydrophilic glycerol group with a hydrophobic pentyloxy group does not provide enough oral availability, so a more oral available prodrug was synthesized by adding an ethyl ester group at the termini of the carboxyl group at C-3. The ethyl ester group is cleaved by endogenous esterases to generate the active form *in vivo*. Oseltamivir is orally administered (75mg or 150mg twice a day).

Resistance mutations of influenza to NA inhibitors

Multiple mutations with reduced sensitivity to NA inhibitors (NAI) imposes a significant threat to sustained use of NAI for treatment and prophylaxis of influenza infection. These mutations are broadly categorized as drug resistance mutations. Detection of drug resistance mutations relies on *in vitro* drug titration experiments. Briefly, influenza virus is isolated from patient samples and subject to biochemical assays (fluorescent, chemiluminescent, colorimetric) to measure the enzymatic activity of NA as a function of increasing dose of NA inhibitors [249]. The 50% inhibitory concentrations (IC50) can be determined from the dose-response curve: samples with IC50 greater than

the normal range of the sensitive wild-type strain show drug resistance potential. WHO established a relatively more quantitative metric: samples with IC50 10-100 fold above the normal range are classified as showing reduced inhibition and samples with IC50 greater than 100-fold above the normal range are classified as showing highly reduced inhibition. Samples that show (highly) reduced inhibition are subject to Sanger sequencing to identify resistance mutations that confer the reduction in drug sensitivity. Both oseltamivir and zanamivir exhibited low frequency of resistance in early clinical trials in adults (less than 1%) and higher frequency in children (approximately 4%) [250, 251].

In spite of the low frequency of resistance for oseltamivir and zanamivir, the concern for the emergence and circulation of influenza mutants with resistance to NAI continued. Of note, binding of oseltamivir for NA induces a conformational change of NA that is not necessary for natural substrate binding, so it appears to be more prone to drug resistance. For the first few years, oseltamivir resistance mutations were only identified by reverse genetics experiments in cell culture [252]. Resistance substitutions in HA reduced the substrate binding affinity of HA, attenuating the dependency of IAV on NA for release of viral progenies [253], whereas resistance substitutions in NA directly decreased the affinity of NA to oseltamivir [254]. However, a 2004 study conducted in Japan, which sequenced HA and NA of H3N2 IAV from 50 pediatric samples, revealed mutations in NA associated with drug resistance in 9 samples, indicating that about 18% of patients treated with oseltamivir developed drug resistance

mutations [255]. This was the first time that a large number of drug resistance mutations were isolated from patients. Before 2008, only Japan and the US had intensive usage of oseltamivir. Sporadic cases of drug resistance mutations were identified in oseltamivir treated patients, mainly in immunosuppressed patients that under prolonged treatments [246]. However, during the 2007-2008 influenza season, a strain that harbored a known drug resistance mutation became the dominant strain that circulated globally [256, 257]. For example, high frequency of resistant strains was reported in Norway despite negligible usage of oseltamivir [257]. Fortunately, the frequency of resistant strains fell significantly the next influenza season and stayed relatively stable between 1-10% after that, but resistance mutations have been constantly isolated from untreated people in several geographically separate countries such as the US and Australia [258-260]. Because of increasing reports of oseltamivir resistance mutations, considerable efforts have been invested to characterize various properties of isolated drug resistant mutations using interdisciplinary approaches including virology, animal model, structural biology, biochemistry, experimental evolution and bioinformatics.

Common oseltamivir resistance mutations

The predominant oseltamivir resistance mutation in the N1 subtype is H275Y (N1 numbering, Figure 1.9). This histidine to tyrosine mutation at position 275 of NA was first reported in 2001: healthy volunteers experimentally infected with an H1N1 IAV and treated with oseltamivir developed low frequency of H275Y (~4%) [261]. However, this mutation was not detected in patients treated with oseltamivir until the influenza season

Figure 1.9: Residues with known oseltamivir resistance mutations are clustered in the substrate binding site

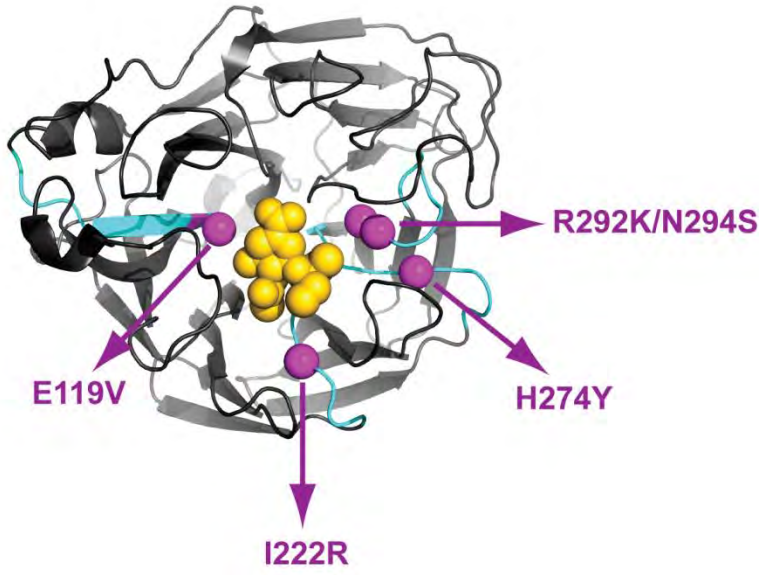


Figure 1.9 Residues with known oseltamivir resistance mutations are clustered in the substrate binding site. Molecular image based on the PDB ID: 3CL0 structure of NA [264] shows the binding substrate, oseltamivir (yellow) and residues with known oseltamivir resistance mutations (purple).

of 2005-2006. H275Y exhibits highly reduced inhibition by oseltamivir (IC50 > 400 fold than the sensitivity wild-type virus) [262]. H275Y has only been associated with oseltamivir resistance in N1 subtype [263]. The frequency of H275Y has been primarily modest, ranging from 1-10% in the tested population, though it is the only drug resistance mutation with greater than 1% frequency in humans [258-260]. One plausible explanation is that H275Y reduces the surface enzymatic activity of NA, leading to attenuated replication and transmission of the mutant IAV, which was confirmed by several *in vitro* and *in vivo* studies [252, 262, 265]. However, the frequency of the H275Y mutation was greater than 50% during the influenza season of 2007-2008 and caused tremendous concern that this mutation would circulate globally and render oseltamivir ineffective [256, 257]. The H1N1 IAV with H275Y exhibits very little defect in replication, transmission and virulence compared to the wild-type [266-270], resulting in its sustained circulation in the human population. Collins *et al.* solved the crystal structure of N1 with H275Y and uncovered the biophysical underpinnings of reduced affinity of H275Y to oseltamivir. The bulkier side chain of Y275 pushes the carboxyl group of E277 further into the active site; protrusion of the charged group into the binding site disrupts the hydrophobic pocket to accommodate the hydrophobic pentyloxyl group at C-6 of oseltamivir, severely impeding the binding affinity of oseltamivir for oseltamivir [202]. This has little effect on binding of zanamivir to NA because movement of Y275 and E277 does not interfere with the hydrogen bond interactions between the glycerol group of zanamivir and the carboxyl group of E277 [202]. Moreover, Y253 lies beneath residue 275 and disrupts the hydrophobic binding pocket together with Y275 in N1. However,

T253, with a smaller side chain in N2, preserves the binding pocket and therefore enables accommodation of H275Y in N2 [202]. Taken together, the combined steric effects at residues 275 and 253 determine the binding affinity of H275Y for oseltamivir in a subtype specific manner.

E119V and R293K are N2 subtype specific oseltamivir resistance mutations (Figure 1.9). They were first isolated from children in Japan under treatment with oseltamivir in 2004. Since that time they have showed a low prevalence in H3N2 IAV infected patients treated with oseltamivir [255]. Both mutations highly reduced sensitivity of NA to oseltamivir (IC50 > 500 fold compared to the sensitive wild-type strain) in H3N2 subtype. However, they imposed different impacts on replication and transmission capacity of H3N2 IAV. R293K caused severely compromised replication and virulence in *in vitro* viral growth experiments and *in vivo* mice experiments [21, 265, 271]. R293K also showed attenuated transmission in animal models [265, 271, 272]. In contrast, E119V maintained comparable replication and transmission fitness compared to wild-type virus [265, 271]. R293K appeared to be completely lethal in N1 subtype, so its sensitivity to oseltamivir was not evaluated in the N1 [262]. E119V was lethal in lab adapted H1N1 strains (e.g. WSN) [262], but was viable in the 2009 pandemic H1N1 strain (pH1N1) [273]. Although E119V displayed considerable resistance to oseltamivir in pH1N1 (approximately 60 fold increase in IC50), strong fitness cost probably impeded its spread in humans [273]. Crystal structural analysis on group1 NA (including N1) and group2 NA (including N2) shed light on the structural basis of reduced binding of R293K

for oseltamivir in N2, but not N1 [203]. Apparently, K293 abolishes a hydrogen bond between the side chain of wild-type arginine and the carboxylate group at the C-4 of oseltamivir and changes the conformation of E277 to push away the pentyloxyl group at C-6 of oseltamivir. However, a conserved tyrosine at residue 348 in N1 forms a hydrogen bond with the carboxyl group of oseltamivir and partially compensates for the loss of interactions due to R293K mutations. No such compensations are observed in N2.

N295S confers resistance to oseltamivir in multiple subtypes, though its frequency is very low in humans (Figure 1.9). It was first isolated from H3N2 infected children treated with oseltamivir [255], later from H5N1 infected patients [274] and then from pandemic H1N1 infected patients [275]. N295S induces intermediate to strong resistance in these subtypes (IC50 increases by 20-300 fold) [255, 263, 273, 276]. N295S only causes mild defect to viral replication and virulence in lab-adapted H1N1 strains [263], but exhibits comparable replication capability in pH1N1 and H5N1[273, 276]. There has been no study on the effect of N295S on IAV transmission in animal models, so it is not clear whether N295S impairs transmission of IAV. The structural mechanism of reduced binding of N295S for oseltamivir has been solved using the crystal structure of N1 bound to oseltamivir [202]. The side chain of the wild-type asparagine stabilizes the conformation of Y348, which coordinates the bound calcium ion and forms a hydrogen bond with the carboxyl group at C-4 of oseltamivir; substituted serine instead forms a hydrogen bond with E277 and drives Y348 away, which leads to a weaker

hydrogen bond interaction between Y348 and oseltamivir, reducing the binding of N295S for oseltamivir.

Position 223 of NA is a hotspot for mutations with reduced sensitivity to oseltamivir (Figure 1.9). The wild-type residue, isoleucine, is part of a smaller pocket that contributes to accommodation of the pentyloxyl group at C-6 of oseltamivir [277]. I223R was the first isolated mutation with reduced binding of NA for oseltamivir (45-48 fold increase in IC50) [277-279]. In the pH1N1, I223R showed comparable *in vitro* replication kinetics and transmission capacity among ferrets, although its virulence was slightly milder compared to that of wild-type in a ferret pathogenic model [279]. Biochemical assays showed a two-fold increase in K_M for I223R, indicating slightly reduced binding of NA for the natural substrate [277]. The crystal structure of N1 with I223R bound to oseltamivir demonstrated that R223 shrinks the small hydrophobic pocket that accommodates the pentyloxyl group of oseltamivir and interacts with residue S247 to hinder binding of oseltamivir [277]. Other mutations at this residue also convey oseltamivir resistance in different subtypes. For example, I223T and I223M exhibited reduced oseltamivir binding in a H5N1 strain [280]; I223K and I223T showed decreased sensitivity to oseltamivir in a pH1N1 strain [281]. In summary, mutations at residue 223 appear to change local conformation and disrupt the smaller pocket that is important for oseltamivir binding. There are also sporadic cases of other spontaneous mutations that confer reduced sensitivity to oseltamivir, although with very low frequency [249, 282,

283]. The vast majority of spontaneous mutations exhibit attenuated replication fitness *in vitro* viral and/or *in vivo*, indicating that they are unlikely to circulate in human populations.

Standing Questions and the Scope of this Dissertation

The sequence-function relationship of proteins has been under intensive investigation since the dawn of molecular biology and maturation of DNA sequencing. It is a central question that holds keys to unlock breakthroughs for many branches of biological studies such as the effect of protein functions on organismal fitness and evolutionary trajectories. Numerous methods have been developed to approach this question, including mutagenesis, protein surface display, structural analysis, directed evolution and computational prediction, yielding countless exciting findings, which substantially enrich and advance our understanding of protein sequence-function relationship and impact of protein function on organismal fitness. However, it remains an open question owing to the tremendous sequence space and inherent difficulty in measuring biochemical functions of many proteins.

Mutagenesis coupled with functional selection is a powerful and widely used tool in mapping the sequence-function relationship, but a major technological caveat is its throughput. The development of high throughput mutagenesis approaches that couple bulk competition of systematically engineered mutations with deep sequencing to determine the fitness of each mutation enables parallel functional characterization of

large numbers of mutations (on the magnitude of hundreds to millions) in a short time window. The high throughput mutagenesis approaches open up new avenues to efficiently map sequence to fitness relationship, incorporate additional factors (*e.g.* protein expression) into mutagenesis analysis, identify adaptive mutations to distinct selection pressures (*e.g.* beneficial mutations in RNA virus under therapeutic selection pressure) and elucidate unexplored sequence space available for protein evolution.

The work presented in this dissertation focused on applying high throughput systematic mutagenesis approaches to understand the sequence-function relationship and explore sequence space for molecular adaptation in diverse proteins and organisms. I first applied the EMPIRIC approach to investigate a fundamental question of the interplay of the relationship of protein sequence-function relationship between protein expression and sequence on organismal fitness, which I describe in Chapter II. Chapter III and IV describe optimization of the EMPIRIC approach: Chapter III reports mapping the functional constraint on the CD4 binding loop and flanking regions of the trimeric env complex of HIV; Chapter IV describes using this approach to quantify fitness of mutations in the active site and proximal regions of NA in IAV in the presence or absence of a NA competitive inhibitor (oseltamivir) to screen for beneficial mutations that adapt to drug selection. A summary and discussion of my results are presented in Chapter V.

Chapter II

The impact of protein expression or sequence on protein total function and organismal fitness has been separately characterized, but their integrated effects have not been systematically interrogated. Several studies have demonstrated that the expression of essential proteins have been optimized for maximal growth, while other studies showed that even half reduction in protein expression led to negligible effects on organismal growth rate. The apparent contradiction may be reconciled by systematic quantification of fitness effect of mutations with varied expression levels to probe the interaction between protein expression and sequence. Moreover, stability has been proposed as the predominant determinant of distribution of fitness effect with the assumption that mutations that directly affect protein activity are rare. Effect of mutations on protein activity can be derived from mutational fitness estimated at altered protein expression levels and used to examine this critical assumption as well as prevalence of mutations that directly affect protein activity.

Chapter III

The CD4 binding loop (GGDPE) of the HIV trimeric spike mediates a large proportion of the binding interactions between CD4 and the trimeric spike. The CD4 binding loop is highly conserved across different clades of HIV, but its N- and C-terminal flanking regions harbor variable sequences. Previous structural and functional studies have identified residues of functional importance in the flanking region of the CD4 binding loop. For example, a number of N-terminal residues modulate exposure of the

trimeric spike to CD4, while several C-terminal residues are part of the Phe43 cavity as well as epitopes of neutralizing antibodies. These residues contribute to the determination of the affinity of the trimer complex to CD4 and several neutralizing antibodies. However, there is an absence of systematic studies delineating the detailed biophysical constraints on these regions. EMPIRIC is optimized for investigating mutational effects on HIV replication and enables comprehensive study of functional constraints on the CD4 binding loop and its flanking regions.

Chapter IV

Drug resistance has been a daunting problem that severely reduces efficacy of precious drugs developed with the expense of billions of dollars. This is especially true for oseltamivir, the only orally available drug that competitively inhibits NA enzymatic function and helps treatment and control of influenza infection. Comprehensive isolation and characterization of mutations with reduced sensitivity to oseltamivir will facilitate clinical monitoring of oseltamivir resistance, appropriate administration of alternative NA competitive inhibitors to minimize selection of pre-existing oseltamivir resistance mutations, and the rationale for the design of new inhibitors and/or improvement of existing ones. The current standard approach to identify oseltamivir resistance mutations is biochemically measuring the sensitivity of the virus isolated from persons treated with oseltamivir, and Sanger sequencing of viruses to pinpoint mutations with reduced sensitivity resistance. There are two major caveats associated with this approach. Firstly it only evaluates the sensitivity of mutations to drug inhibition, which fails to take into

account the fitness of mutations in the presence or absence of drugs. In this chapter, I argue that the relative fitness of mutations under oseltamivir selection pressure, which integrate various factors beyond just biochemical sensitivity to drugs, should more accurately determine if these mutations out-compete the drug sensitive wild-type, while the relative fitness of mutations without oseltamivir selection should determine the likelihood of fixation and spread of these mutations in treatment-naïve populations. Secondly, manual isolation, biochemical characterization and Sanger sequencing of individual patient samples is tedious and time-consuming, significantly delaying the investigation of oseltamivir resistance mutations. To more systematically identify oseltamivir resistance mutations and interrogate the underlying biochemical mechanisms, we optimized EMPIRIC to determine the fitness effects of large number of mutations in the active site of NA, where oseltamivir adaptive mutations are most likely to occur, in the presence and absence of oseltamivir. Drug adaptive mutations are then subjected to biochemical assays to examine various features that may contribute to reduced sensitivity to oseltamivir.

Chapter II - Latent effects of Hsp90 mutants revealed at reduced expression levels

This chapter has been published previously as *Jiang L*, Mishra P*, Hietpas RT, Zeldovich KB, Bolon DNA."Latent effects of Hsp90 mutants revealed at reduced expression*-levels*". PLoS Genet. 2013;9:e1003600. (* equal contribution)*

The work presented in this chapter was a collaborative effort. I performed the yeast growth competition, lysis of yeast cells and DNA extraction, DNA sequencing library preparation and sequencing, and sequence analysis. Dr. Parul Mishra performed growth rate analysis of wild-type yeast strain and biochemical analyses including western blots, FACS analysis, and circular dichroism. Dr. Ryan T. Hietpas contributed to the yeast growth competition experiments and cloned individual Hsp90 mutations for protein expression, purification and circular dichroism analysis. Dr. Konstantin Zeldovich performed Rosetta analysis to estimate ΔΔG of individual Hsp90 mutations. Dr. Daniel N.A. Bolon constructed the model and computed the estimate of protein function per molecule. I, Dr. Parul Mishra, Dr. Ryan T Hietpas, Dr. Konstantin Zeldovich, and Dr. Daniel N. A. Bolon analyzed the data and prepared the manuscript.

Abstract

In natural systems, selection acts on both protein sequence and expression level, but it is unclear how selection integrates over these two dimensions. We recently developed the EMPIRIC approach to systematically determine the fitness effects of all possible point mutants for important regions of essential genes in yeast. Here, we systematically investigated the fitness effects of point mutations in a putative substrate binding loop of yeast Hsp90 (Hsp82) over a broad range of expression strengths. Negative epistasis between reduced expression strength and amino acid substitutions was common, and the endogenous expression strength frequently obscured mutant defects. By analyzing fitness effects at varied expression strengths, we were able to uncover all mutant effects on function. The majority of mutants caused partial functional defects, consistent with this region of Hsp90 contributing to a mutation sensitive and critical process. These results demonstrate that important functional regions of proteins can tolerate mutational defects without experimentally observable impacts on fitness.

Introduction

Genetic changes that alter protein sequence or expression level can lead to adaptation, suggesting these protein properties are central to evolutionary processes. Many studies have individually investigated the effects of changes to either protein sequence or expression level. For example, protein sequences have been optimized under selective pressure using *in vitro* evolution[284]. In addition, changes in protein sequence relative to synonymous substitutions are a hallmark of positive selection in natural populations[285, 286]. The influence of protein expression level on fitness has also been well documented[287]. For example, changes to the expression level of the Agouti protein (but not its sequence) have been shown to affect fitness in wild mice by modulating coat coloration[84]. In addition, experiments in *E. coli* demonstrate that expression from the *lac* operon is rapidly tuned for optimal growth over a wide range of lactose concentrations[86]. While most studies to date have focused individually on either expression level or protein sequence, in principle the fitness effects of these two protein properties are interdependent[85, 288]. Here, we systematically investigate selection on the sequence and expression level of yeast Hsp90 (Hsp82).

We recently developed an approach termed EMPIRIC[51], which is a genetic screen that provides fitness measurements of all possible amino acid substitutions in short regions of important genes in yeast. By sampling across the variety of different amino acid substitutions, EMPIRIC provides detailed information about the physical constraints on protein function. We previously reported a bimodal distribution of fitness effects

(DFE) for an evolutionarily conserved region of the yeast Hsp90 gene[51], an essential chaperone required for the maturation of many kinases[289-291]. Bimodal DFEs, where most mutants have fitness effects close to either null or wild type (WT), appear common in nature as they have been observed in many other fitness studies[93, 292-295].

Bi-modal DFEs are consistent with a recently proposed model where the impacts of mutations on protein stability have a dominant impact on fitness[296]. This model is founded on two concepts: positions that contribute directly to rate-limiting steps in protein function are rare; and the natively folded structure is required for function. Under these conditions, selection results in stably-folded proteins[297, 298], such that modestly destabilizing mutations can be tolerated without dramatic changes to the fraction of natively folded protein molecules and hence function. Because protein folding is cooperative there is a narrow range of stability where both the folded and unfolded state are highly populated, consistent with relatively few mutations having intermediate function. In this stability-dominated model, mutations to critical functional positions (*e.g.* catalytic sites in enzymes) destroy activity, but are presumed rare and so do not contribute greatly to the DFE. Of note, the prevalence of positions in proteins that directly contribute to rate-limiting steps in protein function and the fragileness of these positions to mutation have not been thoroughly investigated.

The effects of mutations on protein function can be investigated based on fitness effects; however, fitness effects need not correspond directly to functional effects. For

example, many essential proteins can be dramatically reduced in net function (defined here as the product of expression level and function per molecule) without dramatic reductions of fitness[90, 91, 93, 299, 300]. Heterozygotes with one null allele are often highly fit, indicating that 50% reductions in net function can be tolerated[94]. The relationship between fitness and the net function of a protein is formally an elasticity function[299]. Around the wild type net function, the elasticity function often has a slope less than one indicating that reductions in net function have dampened impacts on fitness[89, 301]. Experimental analyses of fitness effects are also constrained by experimental measurement precision, which is currently on the order of 1%[302]. In natural systems, the resolution of selection depends upon the inverse of effective population size and is on the order of 10^{-7} for yeast[303, 304]. Thus, the effects of mutations on function that are important in natural selection can be hidden to experimental fitness analyses. For example, the net function of lysozyme in phage T4 must be reduced about 30-fold before experimentally measurable impacts on growth are observed[93]. At the endogenous expression level in this system, large defects in per molecule function are hidden to experimental fitness analyses.

We searched for hidden fitness effects in Hsp90 by examining the Hsp90 elasticity function. We varied the expression level of the native protein sequence and monitored effects on yeast growth rate. Determining theHsp90 elasticity function enabled us to estimate mutant effects on per molecule function from fitness measurements. The elasticity function was non-linear such that at the endogenous expression level, mutant

defects up to 79% in per molecule function were hidden to experimental fitness analyses. To reveal potentially hidden functional defects of mutants, we repeated EMPIRIC analyses at reduced expression strengths, which systematically varied fitness sensitivity to amino acid substitutions in Hsp90. Using this approach, we were able to construct a full distribution of mutant effects on function for a region of Hsp90. Structural analyses suggest that the region we chose to analyze is a putative substrate binding loop[305]. Our experimental fitness analyses at the wild type expression level resulted in a bimodal DFE, which is a hallmark of a scaffolding region with stability dominated effects on fitness[296]. By analyzing fitness at varied expression strengths, we found that the majority of Hsp90 point mutants had intermediate (10-90%) defects in per molecule function that were hidden to our analyses at wild type expression level. These observations indicate the region of Hsp90 we analyzed is involved in a rate-limiting step in function, and supports its putative role in binding to substrates[305]. Because many mutant defects may be hidden to experimental measurement at the wild type expression level, our results suggest that rate-limiting functional sites in proteins may be more prevalent than previously appreciated, and provides a useful guide for interpreting the growing field of systematic mutant analyses[51, 57-59, 306-309].

Results and Discussion

While our initial EMPIRIC study [51] was performed with a temperature sensitive allele of Hsp90 co-expressed with all mutants; here, we report results in an Hsp90 shutoff strain where mutants were analyzed without potential co-expression artifacts. We developed a yeast shutoff strain (DBY288) where the only chromosomal copy of Hsp90 is regulated by a strictly galactose-dependent promoter [310]. In galactose media, the DBY288 strain expressed Hsp90 at endogenous levels and grew robustly. When switched to dextrose media, the DBY288 strain stalled in growth with Hsp90 levels rapidly dropping below detection (Figure 2.1). This strain enabled plasmid encoded Hsp90 variants to be maintained and amplified under non-selective conditions (galactose media). Switching to dextrose media then applied selective pressure on the plasmid encoded Hsp90 variants.

We analyzed the fitness effects of Hsp90 point mutants by performing a bulk competition in the DBY288 strain. A library of plasmids containing all possible single codon substitutions at amino acid positions 582-590 (Figure 2.2A) was transformed into a single batch of yeast. These experiments used a plasmid and promoter construction previously shown to match the endogenous expression level of Hsp90 [311]. Transformed yeast cells were preferentially amplified in galactose media that allowed all mutations including null alleles to propagate. The bulk culture was transferred to shutoff conditions to initiate selection on the mutant library. The beginning of strong selection on the mutant library was estimated from the growth plateau of control cells harboring a null rescue plasmid (Figure 2.1). After the initiation of selection on the mutant libraries, samples

Figure 2.1: Hsp90 shutoff strain

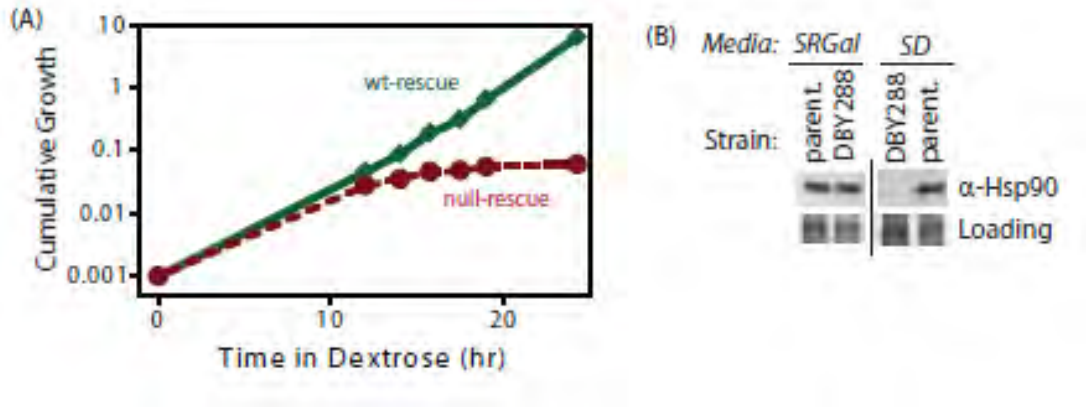


Figure 2.1 Hsp90 shutoff strain. (A) The DBY288 strain grows robustly in dextrose when provided with a rescue plasmid that constitutively expresses Hsp90, but stalls in growth with a null-rescue plasmid. (B) Expression level of Hsp90 in DBY288 is near-endogenous in media with galactose (SRGal), but below Western blot detection after 19 h in dextrose media (SD).

Figure 2.2: Fitness effects of Hsp90 amino acid substitutions

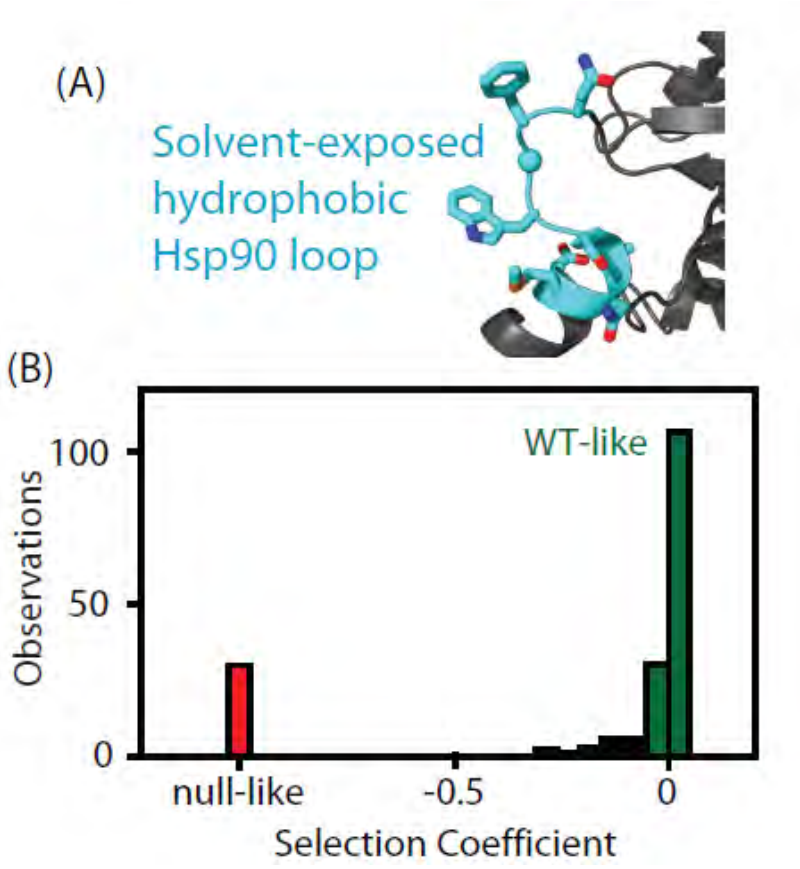


Figure 2.2 Fitness effects of Hsp90 amino acid substitutions. (A) The fitness effects of all possible amino acid substitutions were analyzed for the region highlighted in cyan. This representation is based on the crystal structure of yeast Hsp90[104]. (B) At endogenous expression strength, the distribution of fitness effects was bimodal with many mutations resulting in either WT-like (green) or null-like [38] growth rates.

were harvested over the following 36 hours and the relative abundance of each mutant quantified using focused deep sequencing. By comparing the trajectory of each mutant relative to wild type, we directly determined competitive advantage or disadvantage of each amino-acid substitution as an effective selection coefficient (s) that represents the competitive asexual growth advantage/disadvantage of each mutant in a defined environment [302]. We have previously demonstrated that the EMPIRIC approach provides highly reproducible measures of fitness effects that strongly correlates with the growth rate of individual mutants grown in monoculture [312]. Consistent with our previous work, effective selection coefficients were highly reproducible (R^2=0.96) in a full experimental repeat (Figure 2.3). At the endogenous expression strength, the distribution of fitness effects for this region of Hsp90 was bi-modal (Figure 2.2B, Table 2.1), with peaks near wild type and null. Bi-modal fitness distributions are predicted based on a model where fitness effects are dominated by the impact of mutations on protein stability [296]. Thus, our fitness analyses at wild type expression level are consistent with this region of Hsp90 serving a primarily scaffolding purpose.

To further probe the relationship between the net function of Hsp90 and fitness, we varied expression level of the WT sequence and analyzed impacts on growth rate (Figure 2). To vary expression level, we swapped both promoter and terminator (3' untranslated) sequences. Closely following the start of strong shutoff selection (19 hours in dextrose), we observed a 2-fold range in growth rate with these constructs (Figure 2.4A) and a 100-fold range in expression level (Figure 2.4B). We quantified expression level using a Western blot assay directed against an 6xHis epitope tag only present on the rescue copy

Figure 2.3: Correlation between effective selection coefficients measured in a full experimental repeat at endogenous expression level

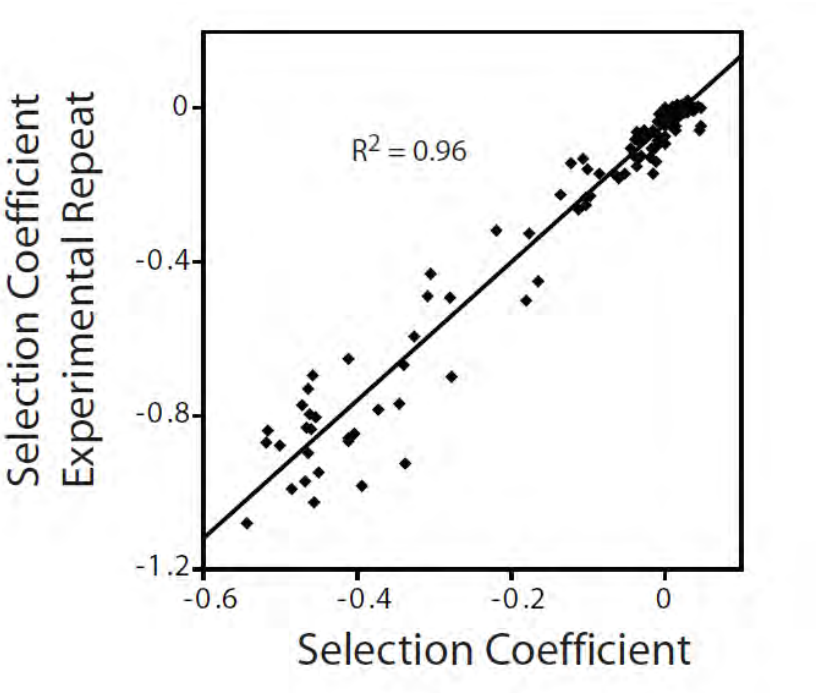


Figure 2.3 Correlation between effective selection coefficients measured in a full experimental repeat at endogenous expression level. Strongly deleterious mutants rapidly deplete in bulk competition and are not monitored as precisely as more fit mutants. Relative selection coefficients are strongly reproduced in these full experimental repeats performed on different days. The slope of this correlation is 1.7, likely due to a linear influence from estimates of the WT growth rate in these separate experiments.

Table 2.1: Relative Expression Measurements

Relative Expression Measurements

Construct	GFP/FACS	Western	Average
GPD	1	1	1
TEF	0.31	0.32	0.32
TEFΔter	0.12	0.067	0.094
CYC	0.055	0.032	0.044
ADH	0.014	0.014	0.014
CYCΔter	0.026	0.013	0.019
ADHΔter	0.014	0.001	0.008

Figure 2.4: Effect of reduced Hsp90 expression on yeast growth

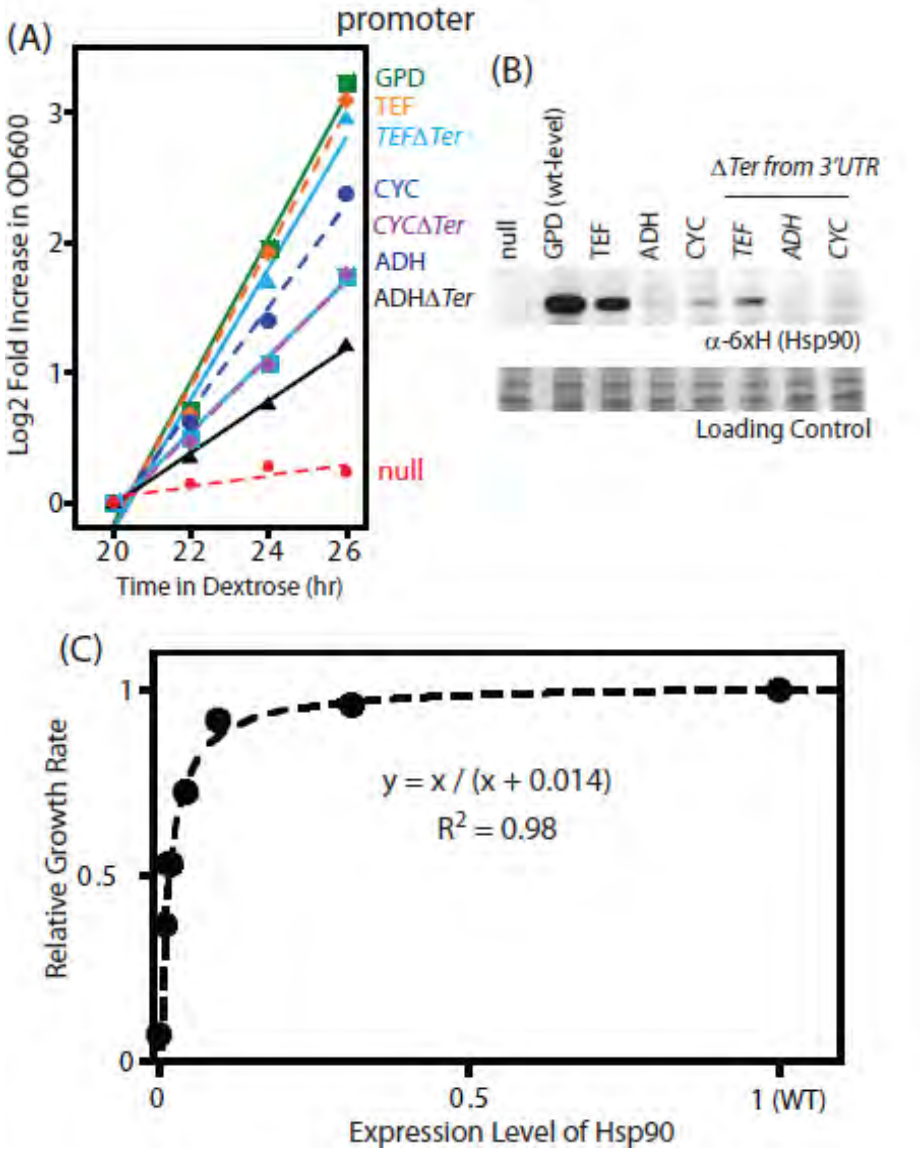


Figure 2.4 Effect of reduced Hsp90 expression on yeast growth. (A) Growth of Hsp90 shutoff yeast harboring rescue Hsp90 plasmids with varied promoters with or without a terminator in the 3' UTR. Yeast were grown at 30 °C and monitored by optical density at 600 nm. (B) Hsp90 expression in cells grown in dextrose for 19 hours was monitored by Western blotting. (C) Relationship between observed Hsp90 expression level and growth rate.

of Hsp90 that we had previously optimized to yield a linear response [313]. These expression level measurements were performed after 19 hours in dextrose, where the second copy of Hsp90 driven by the galactose regulated promoter was undetectable (Figure 2.1). To further investigate expression level, we developed an Hsp90-GFP fusion construct that we monitored by flow cytometry. Across all promoter constructs, the Hsp90-GFP fusion supported similar yeast growth rates to non-GFP tagged versions (Figure 2.5). These findings indicate that the GFP fusion has minimal impacts on Hsp90 function. The expression levels determined by GFP and flow cytometry were in close agreement with those measured by Western blotting and the average of both measures was used to estimate expression levels (Table 2.1).

Both the Western and GFP experiments demonstrate that the expression level of Hsp90 can be reduced dramatically (15-fold) without major impacts on growth rate, which is consistent with previous reports [91, 314]. The growth rate to Hsp90 expression level profile that we determined has the shape of a binding curve (Figure 2.4C), and can be fit to a binding equation that represents the elasticity function for Hsp90. This elasticity function defines how yeast growth rate varies with the net Hsp90 function and enabled us to calculate per molecule function of mutants from fitness measurements.

The non-linear elasticity function for Hsp90 describes the coupling of mutant effects on function and fitness. For example, when expressed at endogenous levels, an Hsp90 amino acid substitution would need to reduce per molecule function by 79% in order to result in a readily measureable growth defect of 5%. Thus the bimodal DFE that we

Figure 2.5: Hsp90-GFP fusions

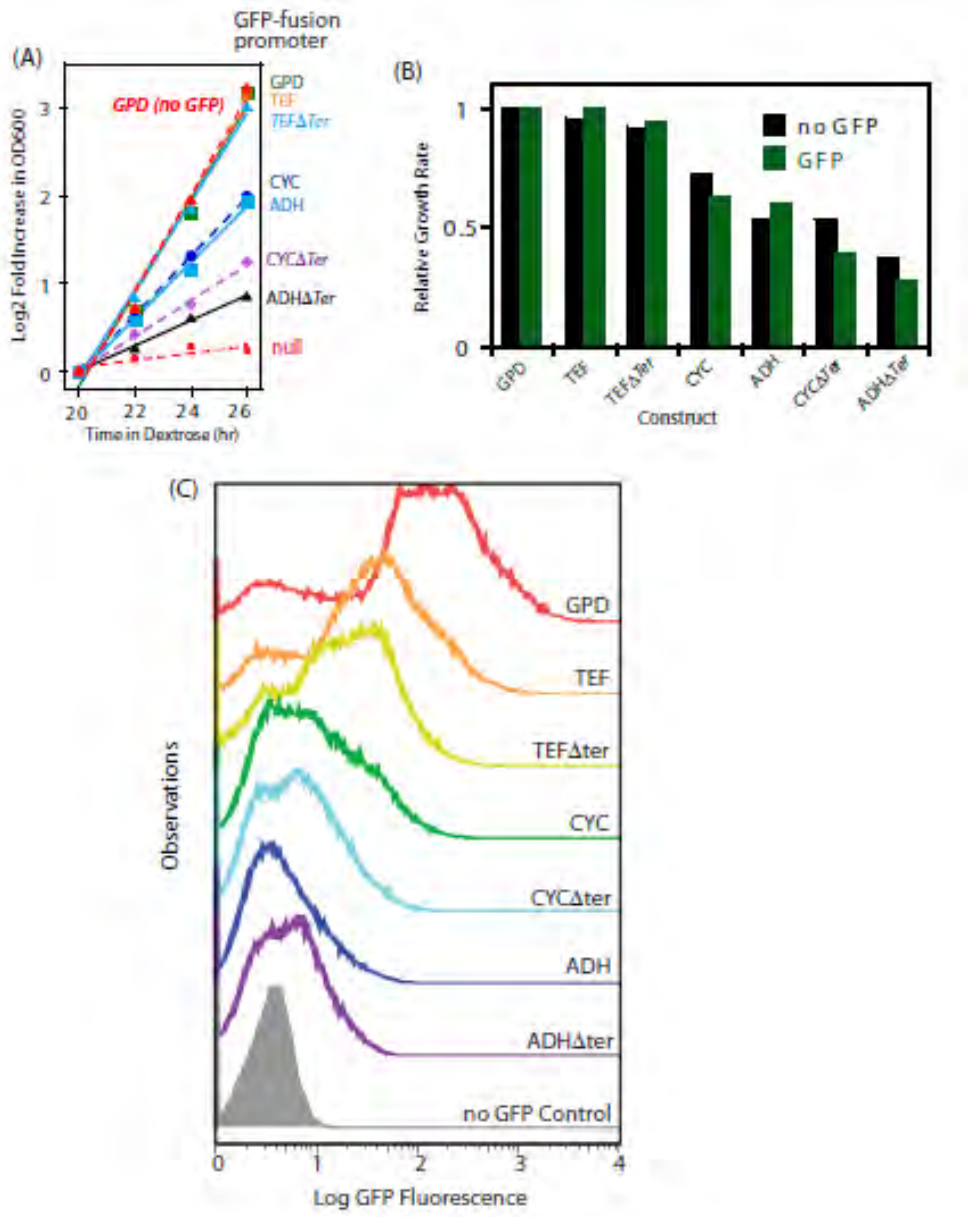


Figure 2.5 Hsp90-GFP fusions. (A) Growth rate supported by Hsp90-GFP fusion in shutoff yeast. (B) Comparison of growth rates observed with and without GFP fusion. (C) GFP levels observed by flow cytometry. To estimate bulk expression comparable to the Western analyses in Figure 2.4, mean fluorescence was calculated and corrected for the auto fluorescence from cells lacking GFP.

observe for Hsp90 (Figure 2.2B) does not necessarily imply a bimodal distribution of mutant effects on function. In particular, the fitness analyses do not provide detailed information on mutants with up to 79% defects in function. Due to the shape of the Hsp90 elasticity curve, the bimodal DFE is consistent with either a bimodal distribution of function as predicted by the stability dominated fitness model [296], or a primarily unimodal distribution of functional effects (Figure 2.6). To distinguish between these possibilities we sought to reveal effects on function that could be hidden at wild type expression strength.

To reveal the latent function of Hsp90 mutants, we analyze fitness effects at reduced expression strengths (Figure 2.7, Table 2.2). The population in all bulk competitions was managed such that the population size at constriction points was always in gross excess to library diversity (Figure 2.8). Because there is selection pressure to increase expression in these experiments, we examined the expression level of the wild type Hsp90 sequence over time in shutoff conditions using Hsp90-GFP fusions (Figure 2.9). Cells respond to selection by increasing expression from weak promoters over time. As predicted by the elasticity function (Figure 2.4), the increased expression from weak promoters results in an increase in growth rate (Figure 2.10). The observed increase in growth rate closely matches predictions based on the expression increase we observed by flow cytometry and the elasticity function, indicating that the underlying model is sound. To minimize the impact of time dependent changes in expression on fitness analyses of coding sequence mutations, we performed bulk competition of Hsp90 mutants over a short time window, 12-48 hours in dextrose (Figure 2.8). We performed simulations to investigate how the

Figure 2.6: Mutant effects on protein function can be hidden to fitness analyses

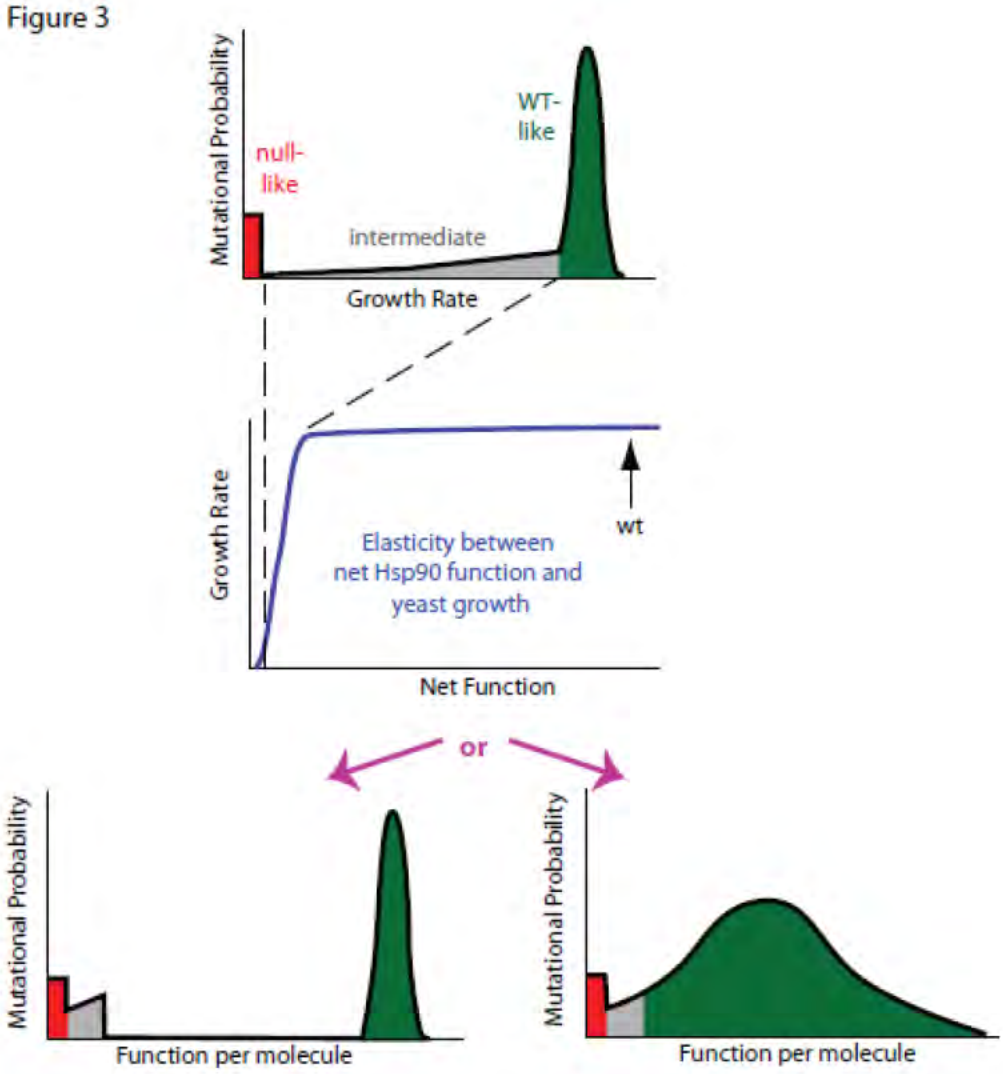


Figure 2.6 Mutant effects on protein function can be hidden to fitness analyses. Fitness effects (top panel) are a property of the elasticity relationship between net protein function and growth (middle panel) and the impact of mutations on function per molecule (bottom panels). For Hsp90, the non-linear elasticity relationship could mask defects, making multiple distributions of functional effects (bottom panels) indistinguishable to fitness analyses. In the top and bottom panels, green represents mutants with WT-like, red null-like, and grey intermediate fitness effects.

Figure 2.7: Distribution of observed fitness effects

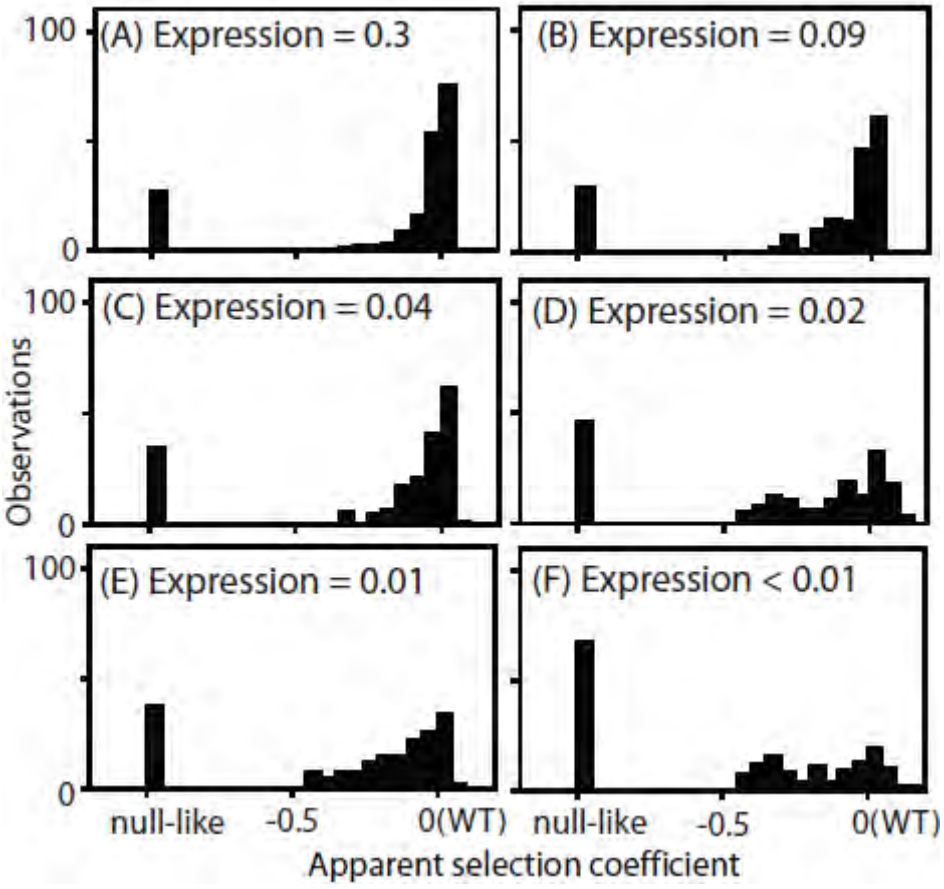


Figure 2.7 Distribution of observed fitness effects. EMPIRIC results for Hsp90 mutants with varied promoters with and without terminator sequences in the 3' UTR: (A) TEF promoter, (B) TEF promoter without a terminator, (C) CYC promoter, (D) CYC promoter without a terminator, (E) ADH promoter, and (F) ADH promoter without a terminator.

Figure 2.8: Population management during bulk competitions

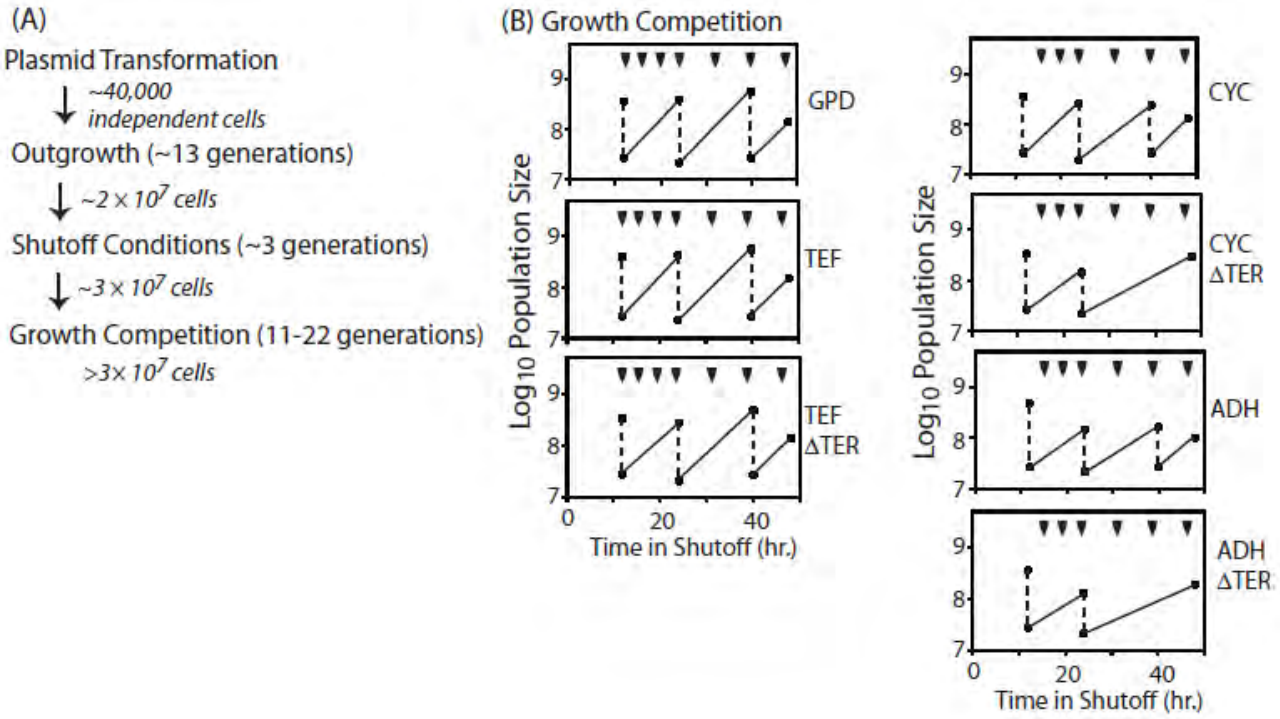


Figure 2.8 Population management during bulk competitions. (A) Outline of experiment from transformation through selection. For all steps, the smallest population bottleneck is indicated. These bottlenecks were managed so that they were always in gross excess to the diversity of engineered mutations (64 codons x 9 positions = 576). (B) Population management during selective growth competition (after 12 hours in dextrose). Dashed lines represent dilutions to maintain cells in logarithmic growth, and arrows at the top indicate when samples were harvested for sequencing.

Figure 2.9: Expression of Hsp90-GFP fusions as a function of time in shutoff

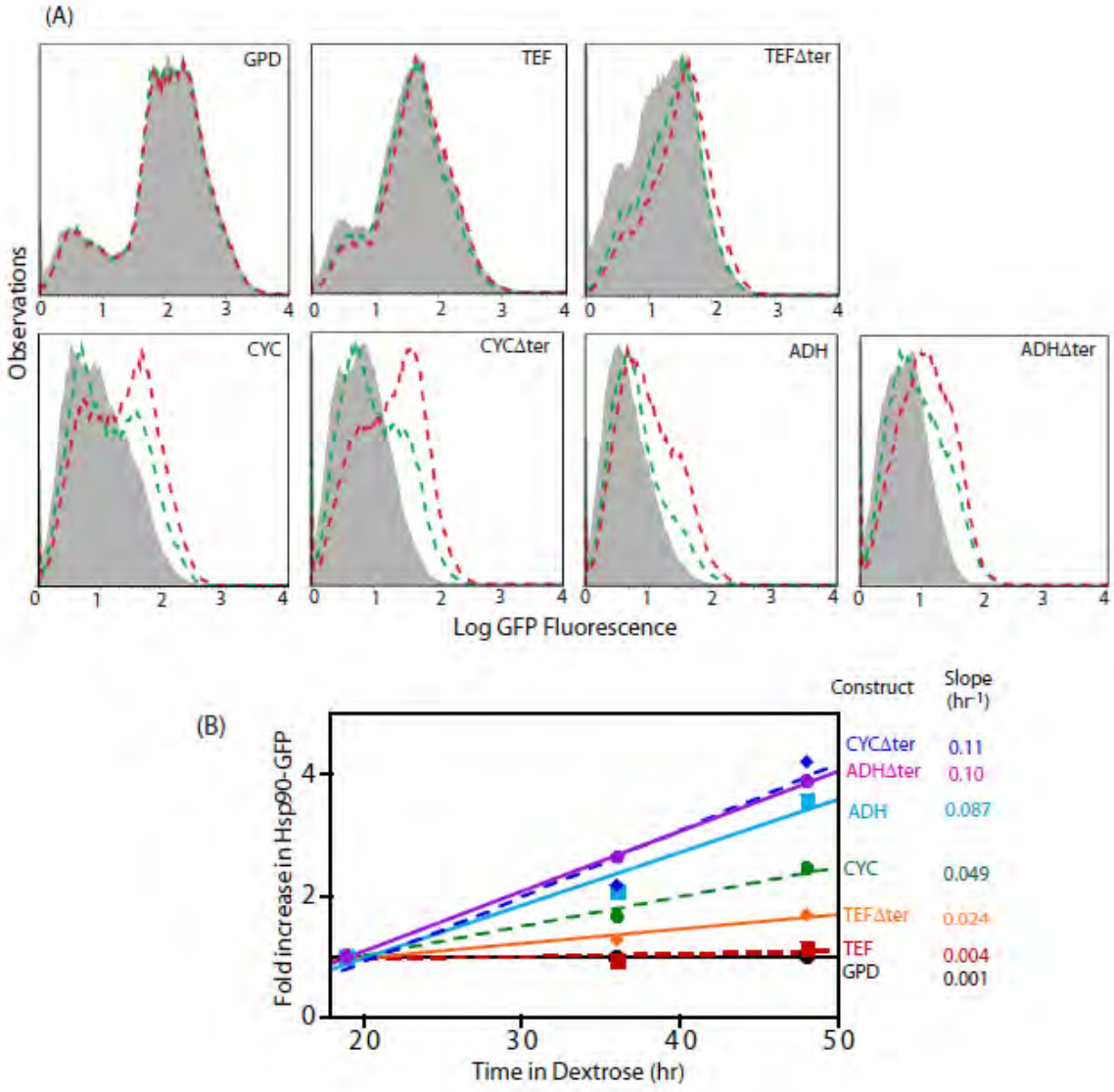


Figure 2.9 Expression of Hsp90-GFP fusions as a function of time in shutoff conditions. (A) GFP levels observed by FACS after 19 hours (grey filled), 36 hours (green dashed line), or 48 hours (red dashed lines) in dextrose for each promoter strength construct. To estimate the bulk expression level, mean fluorescence was calculated and corrected for the autofluorescence of cells lacking GFP. While the observed distribution of expression level among populations of cells was complex and will be interesting to examine in future studies, straightforward analyses of the population mean provided useful estimates for this study. (B) The fold increase relative to the 19 hour time point for each construct was plotted and fit to a linear model.

Figure 2.10: Influence of time in dextrose on growth rates

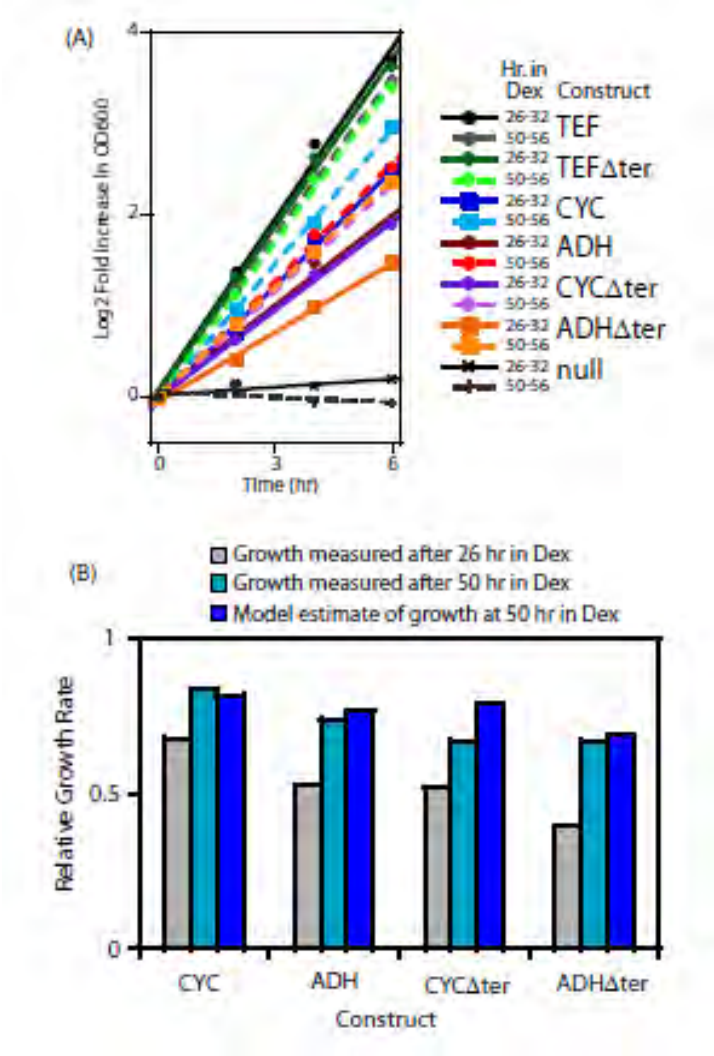


Figure 2.10 Influence of time in dextrose on growth rates. (A) Shutoff yeast harboring rescue plasmids with the WT Hsp90 coding sequence under different promoter strengths were monitored after 26 or 50 hours in dextrose. The TEF and TEFΔter constructs exhibit less than a 5% change in observed growth rate, consistent with the robust growth observed for these constructs immediately after selection begins in dextrose. (B) Relative growth rate of strains exhibiting growth defects upon selection in dextrose. Growth estimates at 50 hours were estimated from the growth rate at 26 hours, the elasticity function, and the observed increase in expression of Hsp90-GFP fusions.

observed increase in expression level over time in shutoff conditions would impact competition trajectories (Figure 2.11). The impact of increasing expression level has a minor impact on competition trajectories and indicates that constant expression models provide estimates of sufficient quality to interpret general features of the distribution of mutant effects on fitness and function, which is the focus of this study.

The DFEs that we observed exhibited a consistent trend as expression strength was reduced. At high expression strength, the majority of mutants had WT-like growth rates, with very few mutants of intermediate effect. As expression strength was reduced, the WT-like peak decreased and the prevalence of mutants with intermediate effects increased. In terms of epistasis, the fitness effects of amino acid substitutions displayed pervasive negative epistasis with expression strength (Figure 2.12). In terms of function, these results strongly indicate that the DFE at endogenous expression strength (Figure 2.2B) does not mirror the underlying effects of point mutations on Hsp90 function.

We estimated mutant effects on Hsp90 function (Figure 2.13, Table 2.3) based on fitness measurements at distinct expression strengths and the elasticity function. As described in the methods section, we employed the elasticity function to calculate per molecule function from fitness taking into account bounds on measurement and calculation precision. For example, at the endogenous expression strength, mutants with activity defects of up to 79% were obscured to fitness analyses and were demarcated as such (functional efficiency > 0.21). Because a distinct range of function is revealed to selection at each expression strength (Table 2.4), our integrated analyses provided

Figure 2.11: Models of time-dependent changes in expression level

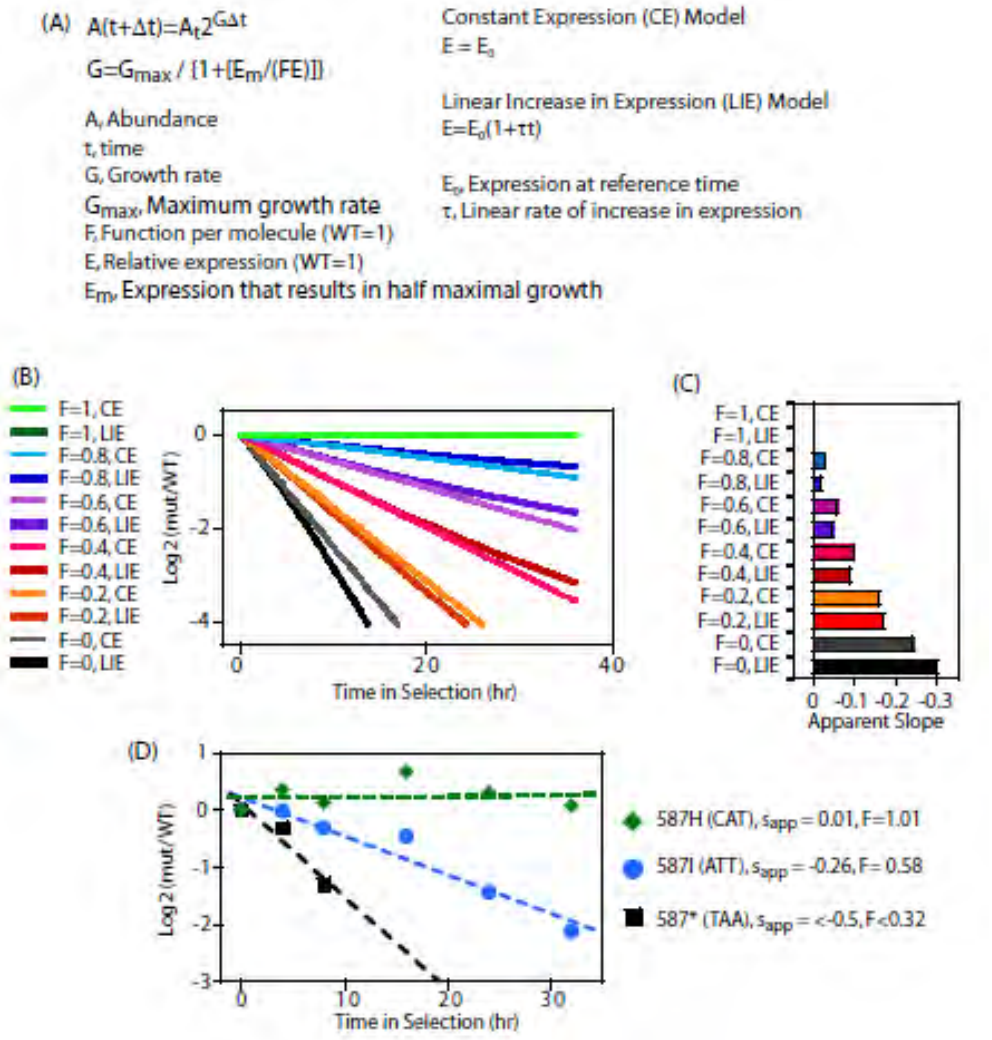


Figure 2.11 Models of time-dependent changes in expression level. (A) Mathematical descriptions of numerical integration models relating abundance to function per molecule and expression level. (B) Trajectories of theoretical competitions of mutants with different impacts on function and WT. Trajectories simulations use Dt of 0.1 hours and growth and expression parameters from the CYCΔter construct (E_0=0.016 - calculated from observed growth rate and elasticity function, E_m=0.014 - from the elasticity function, $\tau = 0.11\ hr^{-1}$ - based on flow cytometry of Hsp90-GFP fusions over time, $G_{max} = 0.45\ hr^{-1}$ - based on observed growth curves with the GPD construct). Simulations were performed for other constructs and exhibited less variation between CE and LIE models, indicating that the model is most sensitive to changes in E around E_m. (C) Comparison of the slope of linear fits to the plots in panel B. (D) Representative data for competition trajectories of mutants in the CYCΔter constructs with linear fits.

Figure 2.12: Epistasis between expression strength and amino acid substitutions

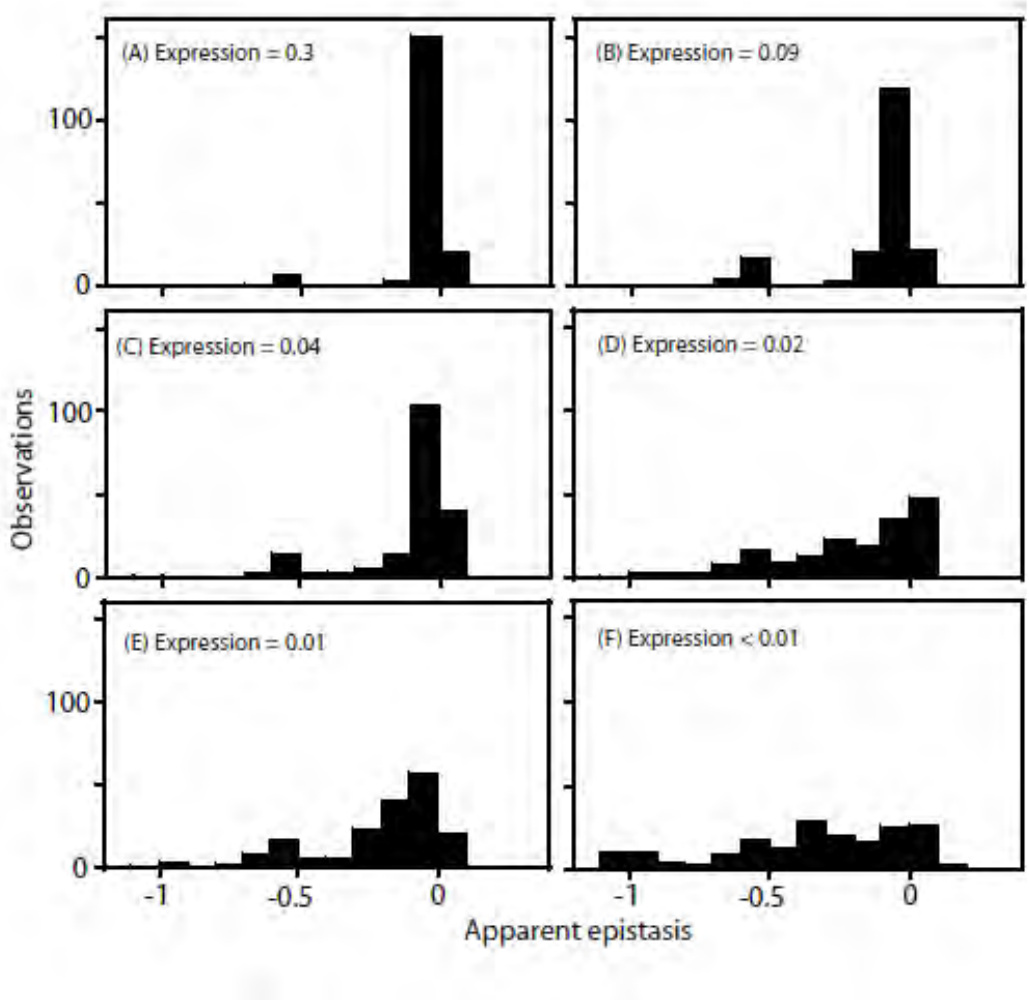


Figure 2.12 Epistasis between expression strength and amino acid substitutions. Fitness effects of point mutants under reduced expression strength compared to endogenous expression strength. Results observed for Hsp90 mutants with varied promoters with and without terminator sequences in the 3' UTR: (A) TEF promoter, (B) TEF promoter without a terminator, (C) CYC promoter, (D) CYC promoter without a terminator, (E) ADH promoter, and (F) ADH promoter without a terminator.

Figure 2.13: Effects of mutations on Hsp90 function

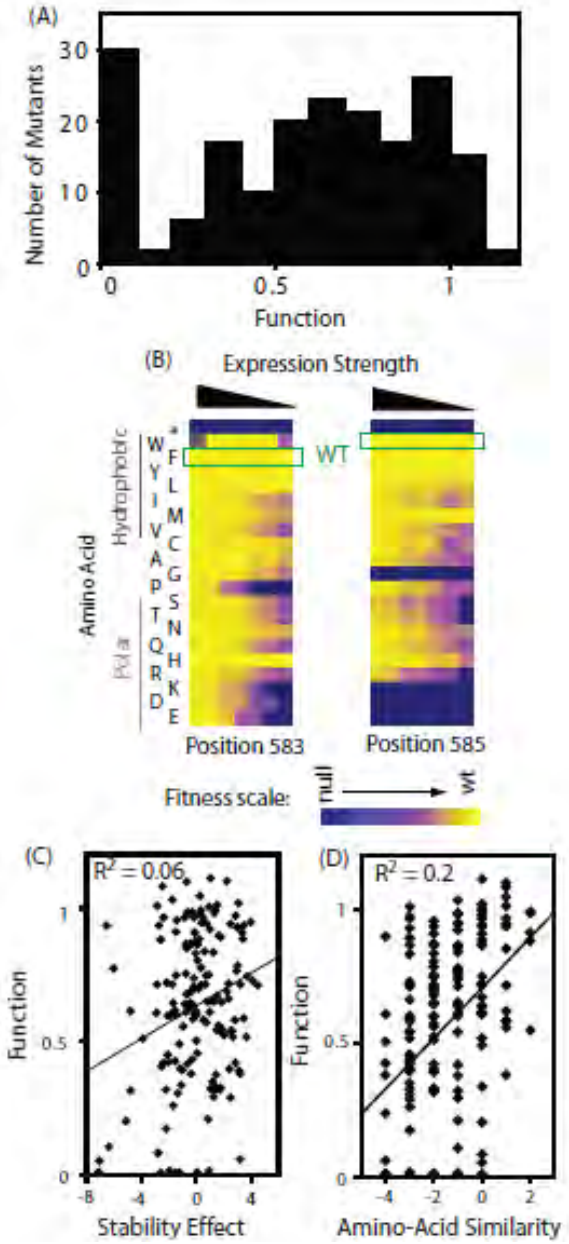


Figure 2.13 Effects of mutations on Hsp90 function. (A) Distribution of mutant effects on Hsp90 function calculated from fitness effects at varied expression levels. (B) Impact of mutations at two solvent exposed hydrophobic amino acids in Hsp90 on yeast growth at different expression strengths. (C) Mutant impacts on folding stability (-ΔΔG estimated from structural simulation) related to function. (D) Similarity of amino acid substitutions to wild type (based on Blosum62 matrix) relative to observed functional effects.

Table 2.4: Activity ranges interrogated at each expression-strength

Table 2.4. Activity ranges interrogated at each expression-strength

Construct Name	Expression Strength	Function Range[1]
GPD	1.0	0.034-0.21
TEF	0.32	0.067-0.45
TEFΔ*Ter*	0.094	0.16-0.71
CYC	0.044	0.38 and above
CYCΔ*Ter*	0.019	0.38 and above
ADH	0.014	0.39 and above
ADHΔ*Ter*	0.008	0.43 and above

[1]Function range with informative fitness effects (s_{app}>null-like – where mutants persist in the culture and are more accurately monitored; and observed growth rates at least 5% slower than G_{max} where growth rate and function are strongly coupled.

estimates of the functional effects of all mutants. Estimates of mutant effects on function based on fitness measurements at different expression strengths exhibit a reasonable correlation ($R^2 = 0.75$) (Figure 2.14). The strength of this correlation, despite simplifying assumptions (further discussed in the methods section), indicates that the calculated mutant effects on function are fair estimates.

The distribution of functional effects for a region of a protein provides information about the contributions of that region to biochemical activity. For example, scaffolding regions that are not directly involved in a critical or rate-limiting step in protein function should be hard to break by mutation (due to selection for stability in the wild type protein), but once broken destroy activity [297, 315]. In contrast, regions that contribute to a rate-limiting step should be easy to injure by mutation, with the severity of mutant defects mediated by the rigidness of chemical and physical requirements (*e.g.* catalytic sites in enzymes being ultimately rigid with any mutation destroying activity).

The distribution of functional effects (Figure 2.13A) for the region of Hsp90 we analyzed had one main peak with most mutations exhibiting partial defects relative to wild type. Our finding is consistent with this region of Hsp90 contributing to a critical and rate-limiting step in function. The intermediate functional defect of most mutants indicates that the chemical and physical requirements are flexible, consistent with this region of Hsp90 providing a hydrophobic docking site for binding to substrates, as was inferred from structure [316]. Taking a closer look at the aromatic amino acids at position 583 [38] and 585 (Trp) located on the surface of the Hsp90 structure, most amino acid

Figure 2.14: Mutant effects on function

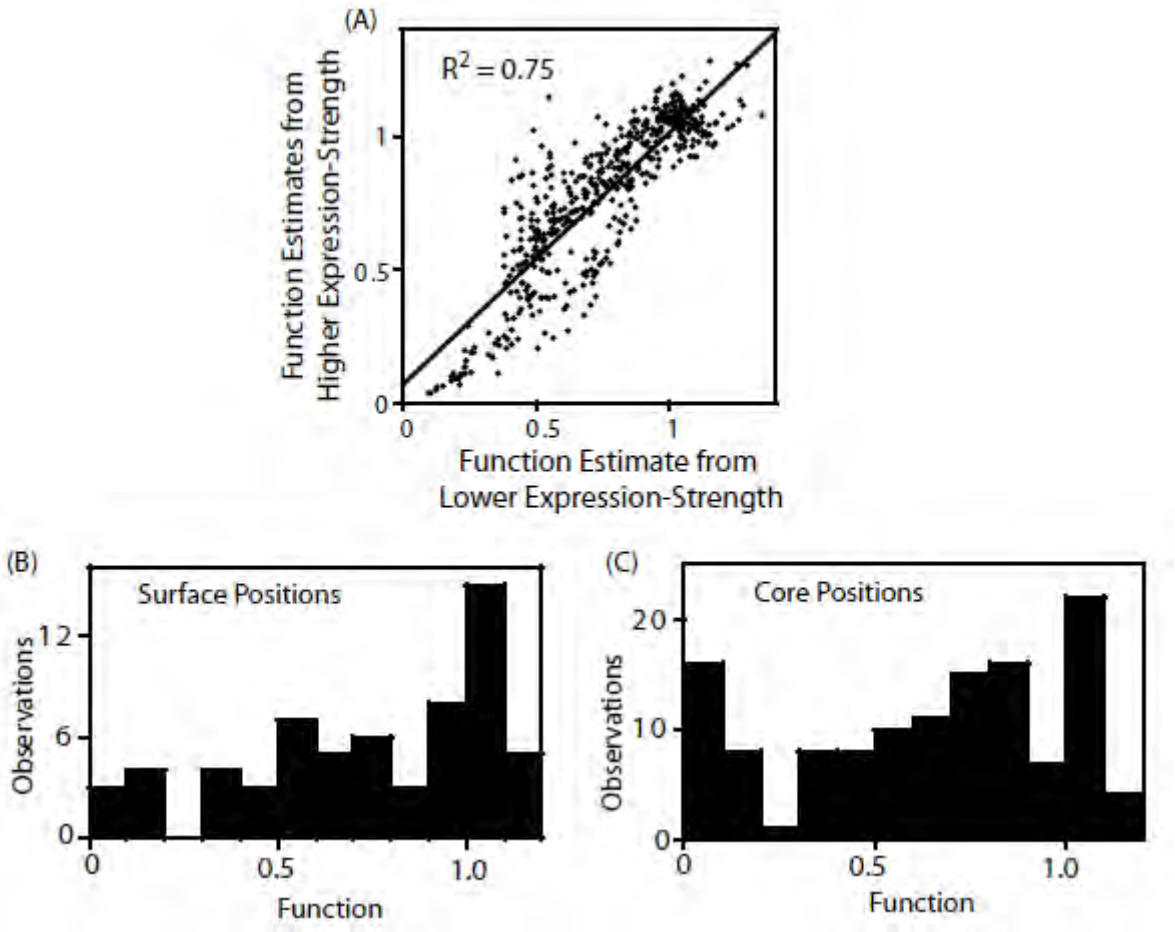


Figure 2.14 Mutant effects on function. (A) Cross-correlation between function estimates generated at different expression-strengths. Comparisons were made between non bounded estimates of function in constructs that neighbored in expression strength (e.g. GPD was compared to TEF., TEF to TEFΔter, etc.) (B and C) Functional estimates for positions oriented towards the protein interior or exterior. Distribution of mutant effects on function for positions 582, 583, and 585 that are oriented towards solvent (B), compared to positions (584, 586-590) that are oriented towards the protein core (C).

substitutions are tolerated when expressed at endogenous levels, but a clear functional preference for hydrophobic amino acids is revealed at reduced expression strengths (Figure 2.13B). Hydrophobic interactions [317] are malleable to slight alterations in geometry and physical composition compared to other physical interactions (*e.g.* hydrogen bonds). Thus, it is reasonable that some substitutions that maintain hydrophobicity would be well tolerated, but that most non-conservative substitutions would result in strong defects.

Our fitness-based estimates of mutant effects on function integrate over all properties that contribute to cell growth including catalysis, binding affinity, as well as the thermodynamic stability of folding to the native state [296, 297, 301, 315, 318]. In terms of stability, the prevalence of intermediate functional defects that we observe is inconsistent with this region of Hsp90 serving a purely scaffolding function, which theory predicts should exhibit a bi-modal distribution [296]. Furthermore, we observed a similar distribution of functional effects for positions located on the protein surface, which should have relatively small impacts on stability [319], as those that orient towards the protein interior (Figure 2.14). This finding suggests that the functional effects of mutants at solvent shielded positions are caused primarily by local structural changes that impact the organization of solvent exposed positions (*e.g.* as required for efficient binding to substrate). We have observed a similar surface-core relationship in ubiquitin [312], and at a lower resolution this type of surface-core association has been postulated based on the slow evolutionary divergence of sites in proteins located proximal to binding sites [320]. Of note, Hsp90 is a dimeric protein and subunit folding and

association are coupled [313]. Thus, decreased expression strength could increase sensitivity to destabilizing mutations. In this case, destabilizing mutations would exhibit larger activity defects at lower expression strength. Across the dataset our functional estimates are largely independent of expression strength (Figure 2.14, Panel A). Thus, the effects of mutations on dimer stability appear to have at most a minor impact on our activity estimates, consistent with the location of this region of Hsp90 far from the dimer interface [104].

To further examine the effect of mutations on stability, we simulated the stability effects of each possible point mutation based on the structure of Hsp90 [104] using Rosetta [321], which accurately predicts the experimental effects of mutations on stability. The simulated stability effects for Hsp90 correlate extremely weakly with activity (Figure 2.13C), consistent with our conclusion that stability is not a dominant contributor to activity for this region of Hsp90. Of note, substitutions of amino acids with similar physical and chemical properties (as estimated by BLOSUM similarity) to the wild type residue tend to be compatible with function (Figure 2.13D). The stronger correlation of function with amino acid similarity compared to stability suggests that the stability simulations do not fully capture all biologically relevant structures. For example, high resolution structures of Hsp90 bound to substrate are not available; but if they were available, might provide a stronger structural explanation for the observed functional effects of mutations.

To further test our model and conclusions, we experimentally investigated the biochemical properties of five non-conservative amino acid substitutions. We chose mutations that dramatically change the hydrophobic binding surface and largely destroy function (F583D and W585D), mutations that disrupt intra-molecular interactions and severely impair function (S586H disrupts a buried hydrogen bond, and A587D introduces a buried charge at a solvent shielded location), and a charge reversal mutation (E590K) on the surface that causes a moderate functional defect. The growth rate of these mutants in monoculture closely matched the fitness effects observed in the bulk competitions (Figure 2.15). As discussed above, our estimates of function integrates over multiple protein properties. For example, a mutation that increases the degradation rate (with the synthesis rate unchanged) should exhibit reduced steady state levels leading to a defect in net function. All of the disruptive individual mutations that we investigated accumulated at similar steady state levels (Figure 2.16A), suggesting that individual mutations do not commonly disrupt Hsp90 protein levels.

We examined the biophysical properties of these non-conservative Hsp90 mutant proteins in purified form. To maximize the sensitivity of these analyses for potential alterations to structure and stability, we generated C-domain constructs. All of the mutations we analyzed are located in the C-domain and do not contact other domains in the Hsp90 structure. The circular dichroism (CD) spectra of all five mutant proteins overlay closely with WT (Figure 2.16B) indicating that all of the mutants fold into native conformations with similar secondary structure content to WT. We investigated the stability of each mutant protein to urea-induced unfolding (Figure 2.16C). Similar concentrations of urea

Figure 2.15: Monoculture growth of individual mutants

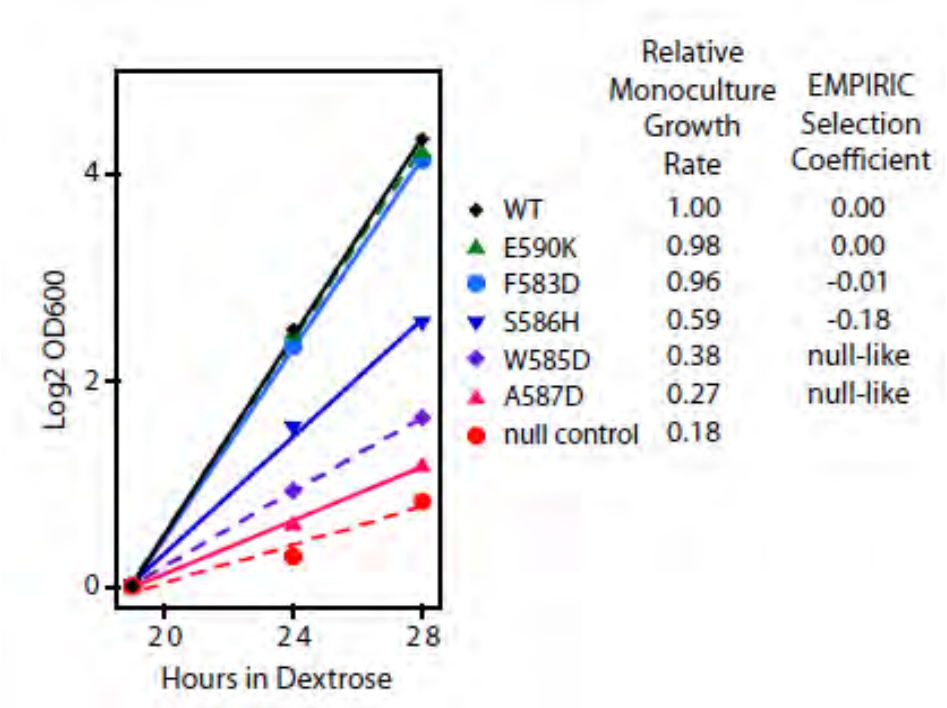


Figure 2.15 Monoculture growth of individual mutants. Individual mutants were generated in the GPD construct, introduced into DBY288 yeast, and growth observed after switching to shutoff conditions. We chose to analyze mutants with large differences in effective selection coefficients from the bulk competitions. Both mutants (W585D and A587D) with null-like EMPIRIC selection coefficients exhibited dramatic growth defects, while both mutants (E590K and F583D) with WT-like effective selection coefficients exhibited robust growth in monoculture. The S586H mutant that had an intermediate fitness defect in the bulk competitions also exhibited an intermediate growth defect in monoculture. For this data, the Spearman's rank correlation was -0.99 using identical EMPIRIC rankings for WT-like variants (WT, E590K, F583D) as well as for null-like variants (W585D and A587D). The negative sign in the correlation indicates that selection coefficients and growth are inversely related by definition.

Figure 2.16: Expression level and stability of five non-conservative mutations

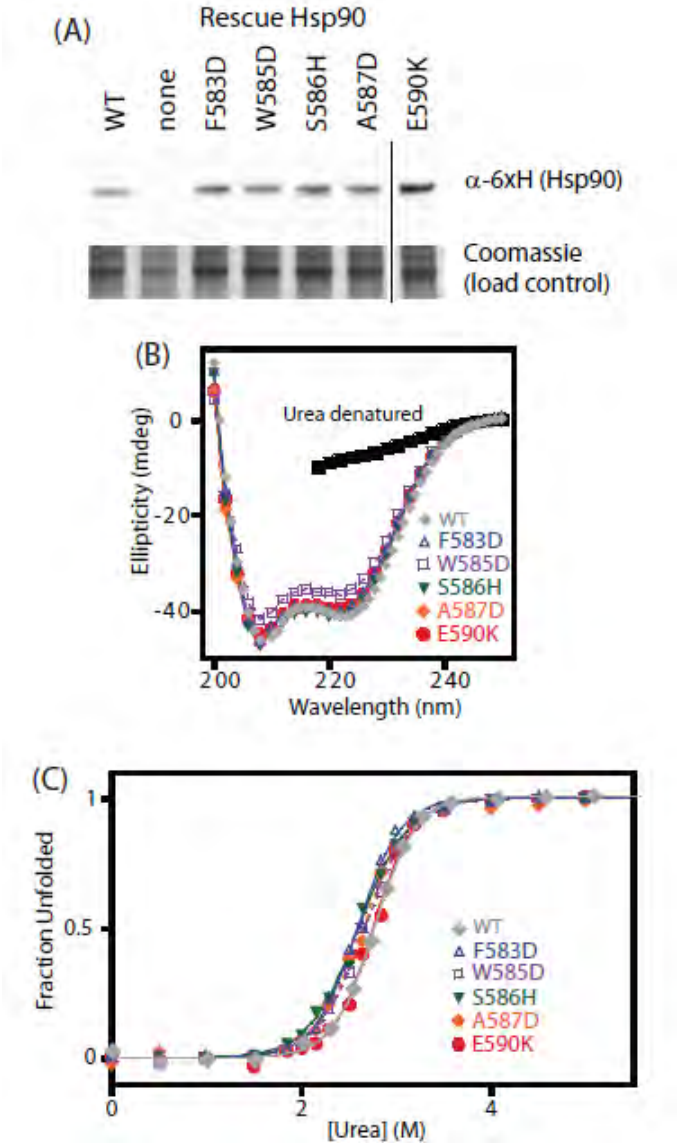


Figure 2.16 Expression level and stability of five non-conservative mutations. (A) Expression level in shutoff yeast analyzed by Western blotting. The vertical line represents intervening lanes that were removed for clarity. (B) Secondary structure of purified C-domain constructs analyzed by circular dichroism. The spectrum of a denatured sample in 5M urea is shown for comparison (below 218 nm absorbance from urea interferes with signal). (C) Urea induced unfolding of purified C-domain constructs. The fraction unfolded was determined based on ellipticity at 222 nm.

were required to unfold all mutants and WT indicating that none of the mutants compromises folding under native conditions. These findings demonstrate that non-conservative mutations in this region of Hsp90 are generally capable of folding to stable native states, and strengthen our conclusions that the 582-590 region of Hsp90 that we analyzed is not critical for folding stability, and is instead a structurally malleable region that forms a critical hydrophobic docking site.

Our studies as well as those of others [90, 92, 299, 301, 322, 323] demonstrate that biochemical flux models and the elasticity function in particular provide a fundamental link between molecular and cellular/organismal properties. Non-linear elasticity functions of the identical form to those described here for Hsp90 have also been observed in *E. coli* for β-galactosidase [323], isopropylmalate dehydrogenase [92], and dihydrofolate reductase (DHFR) [90]. In *E. coli*, DHFR point mutations were commonly observed to impact protein degradation rates leading to fitness effects that were strongly dependent on the level of protein quality control [90]. In addition, flux models can provide a mechanistic explanation for many common fitness features including pleiotropy and epistasis [324].

This study clearly demonstrates that functional defects of mutants can be hidden to experimental fitness measurements due to a non-linear elasticity function. Uncovering these latent effects revealed that the region of Hsp90 we analyzed contributes to a rate-limiting step in Hsp90 function. These findings indicate that critical functional regions in proteins are more prevalent than considered based on fitness analyses performed without

consideration of the elasticity function. The elasticity function relating net function and fitness is critical for a thorough understanding of mutant fitness effects.

Methods

Plasmid and strain construction. For expression analysis, the yeast Hsp90 gene was cloned into the pRS414 plasmid with different promoters and 3' untranslated region (UTR). We used constitutive promoters previously demonstrated to generate a wide variation in expression level [79] including GPD, TEF, ADH, and CYC. Constructs were generated with or without the 3' UTR from the CYC gene, which allowed further variation in expression level [80]. In constructs lacking the CYC terminator, the 3'UTR was composed of sequence from the plasmid vector. All Hsp90 plasmids contained a 6X-His sequence (GGHHHHHHGGH) at the N-terminus to facilitate detection by Western blotting. Point mutant libraries previously generated in p417 plasmids [51] were transferred to the pRS414 promoter variant plasmids using SLIC cloning [325]. Briefly, for each promoter strength construct, we prepared a destination vector with the first and last 30 bases of Hsp90 bracketing a unique SphI restriction site. We excised the Hsp90 library from the original 417GPD plasmid using restriction enzymes that cut immediately upstream and downstream of the Hsp90 gene. We cut destination vectors with SphI. We generated ~30 base complementary overhangs using T4 DNA polymerase in both the destination vectors and the Hsp90 library, annealed the complementary DNA, transformed into competent bacteria, grew in bulk selective (Amp) cultures and prepared plasmid. A small portion of the transformation was plated and the number of independent transformants (~30,000) was in gross excess to the library diversity. In addition, all replication is performed in bacteria where multiple systems ensure high fidelity reducing the probability of undesired secondary mutations. The DBY288 Hsp90 shutoff strain

(can1-100 ade2-1 his3-11,15 leu2-3,12 trp1-1 ura3-1 hsp82::leu2 hsc82::leu2 ho::pgals-hsp82-his3) was generated from the Ecu Hsp90 plasmid swap strain [311] by integration of Hsp90 driven by a GalS [310] promoter together with a HIS3 marker into the HO genomic locus.

Yeast growth rate. DBY288 cells were transformed with pRS414 plasmids and selected on synthetic raffinose and galactose (SRGal) plates lacking tryptophan (-W). Single colonies were then grown in liquid SRGal-W on a rotator at 30°C to late-log phase (OD_{600} ~0.8). Cells were collected by centrifugation, washed with synthetic dextrose (SD) –W media, and then grown in SD-W medium at 30 °C in an orbital shaker. Culture density was maintained in log phase (OD_{600} between 0.1 and 0.8) by periodic dilution. Culture growth was monitored based on increases in OD_{600} taking into account cumulative dilution. The log of OD_{600} versus time was fit to a linear equation to determine growth rate. Analyses were performed on time points in dextrose where control cells lacking a rescue Hsp90 had depleted Hsp90 by Western analyses (Figure 2.4B & Figure 2.1) and had stalled in growth (Figure 2.4A and Figure 2.1).

Analyses of Hsp90 expression level by Western. To analyze expression levels of different promoter constructs, cells were grown for 19 hours in SD -W media, and 10^8 yeast cells were collected by centrifugation, and frozen as pellets at -80 °C. Cell lysates were prepared by vortexing thawed pellets with glass beads in lysis buffer (50mM Tris-HCl pH 7.5, 5mM EDTA and 10mM PMSF), followed by addition of SDS to 2%. Lysed cells were centrifuged at 18,000*g* for 1 minute to remove debris, and the protein

concentration of the supernatants was determined using a BCA assay (Pierce Inc.). Lysates with 15μg of cell protein were resolved by SDS-PAGE, transferred to a PVDF membrane, and Hsp90 probed using α-HisG antibody (Invitrogen Inc.). Importantly, we have previously shown that detection of this 6xHis Hsp90 construct in yeast can be detected with a broad linear range using this antibody and Western blot approach [313].

Analyses of Hsp90 expression level using flow cytometry. Flow cytometry was used as an alternative approach to measure the expression level of Hsp90 at the single cell level in yeast cells. A gene encoding EGFP was inserted into the unstructured tail of Hsp90 after amino acid position 684. This Hsp90-GFP fusion construct was cloned into the variable strength promoter constructs used with non-GFP tagged Hsp90. These plasmids were transformed into DBY288 yeast competent cells and grown on SRGal-W plates. A single colony of each strain was grown for two days at 30°C in SRGal-W media to near saturation. These cultures were diluted 1:50 into SRGal-W media and grown to late log phase (~10^6 cells/ml). Each strain was then further diluted 1:50 in SD-W media for 48 hours at 30°C with dilution every 12 hrs in order to maintain cells in log phase growth. Samples of cells were collected after 19, 36, and 48 hours in dextrose. Collected cells were washed twice in wash buffer (50mMTris, 150mMNaCl, pH 7.6, 0.1% w/v BSA), diluted to 10^7cells/ml in wash buffer and analyzed on a Becton-Dickinson FACSCalibur flow cytometer equipped with a 15mW air cooled 488nm argon-ion laser using a 530 nm high-pass filter. Greater than 100,000 cells were analyzed for each sample. Data were processed and analyzed using FlowJo software. Debris including clumped cells was excluded by gating on the forward and side scatter (excluded less than 5% of points). To

compare with bulk Western measurements, mean fluorescence was calculated using cells without GFP in order to subtract out background due to autofluorescence.

Circular Dichroism. C-domain constructs of Hsp90 bearing an N-terminal 6xHis tag were generated in a bacterial over-expression plasmid, expressed, purified, and analyzed by circular dichroism (CD) as previously described [313]. Briefly, CD spectra were obtained using a 1 mm path length cuvette at a protein concentration of 20μM in 20 mM potassium phosphate at pH 7 and 25 °C. Urea titrations were performed under the same conditions using samples that were equilibrated for 30 minutes. Urea concentrations were determined based on their refractive index. CD ellipticity at 222 nm was used to follow urea induced unfolding and the resulting data was fit to a two-state unfolding model as previously described [313].

EMPIRIC analyses of point-mutants. The effect of point mutants on yeast growth was analyzed as previously described [326]. Time points in dextrose were selected for analysis where control cells lacking a rescue Hsp90 began to stall in growth in order to observe the rapid decrease in relative abundance of deleterious mutants (e.g. premature stop codons). The growth rate of cells harboring the WT coding sequence in bulk competitions was estimated from monoculture growth of WT constructs performed in parallel to the bulk competitions. For the GPD, TEF and TEFΔter constructs we analyzed time points in dextrose of 12, 16, 20, 24, 32, 40, and 48 hours (Table 2.5). For the CYC, ADH, CYCΔter, and ADHΔter constructs where the relative decrease of deleterious

mutants was less severe (due to slower growth rate of fit mutants) we analyzed time points in dextrose of 16, 20, 24, 32, 40, and 48 hours. To process these time point samples, yeast pellets were lysed with zymolyase and total DNA was extracted and purified through a silica column. The DNA encoding amino acids 582-590 was PCR amplified, and prepared for 36 base single-read Illumina sequencing. 3.4×10^7 high quality reads (>99% confidence across all 36 bases) were obtained and analyzed. The relative abundance of each point mutant at each time point for each promoter was tabulated. Effective selection coefficients for yeast growth were determined by linear fits to the change in mutant abundance relative to wild type for each possible codon substitution. To account for the rapid depletion of null-like mutants to noise levels, only the first three timepoints in selection were used to determine effective selection coefficients for stop codons and all other mutants with effective selection coefficients within two standard deviations of stop codons (corresponding to s=-0.28 for GPD, s=-0.37 for TEF, s=-0.4 for TEFΔter, s=-.0.35 for CYC, s=-0.46 for ADH, s=0.44 for CYCΔter, and s=-0.43 for ADHΔter). Because these null and near-null mutants rapidly deplete from the culture it is challenging to precisely measure their relative growth effects and they were binned as “null-like” (Table 2.2).Potential noise was analyzed by calculating normalized residuals (residuals/time points fit). Codon substitutions with residuals per time point greater than 0.25 or low initial mutant abundance (mutant/wt less than 0.004) were omitted (~7% of codons). For mutants that persist in the bulk competition (s>-0.1) synonymous codons exhibit a narrow distribution (Figure 2.17) indicating that the amino acid sequence is a dominant determinant of fitness. The

Figure 2.17: Effects of synonymous substitutions

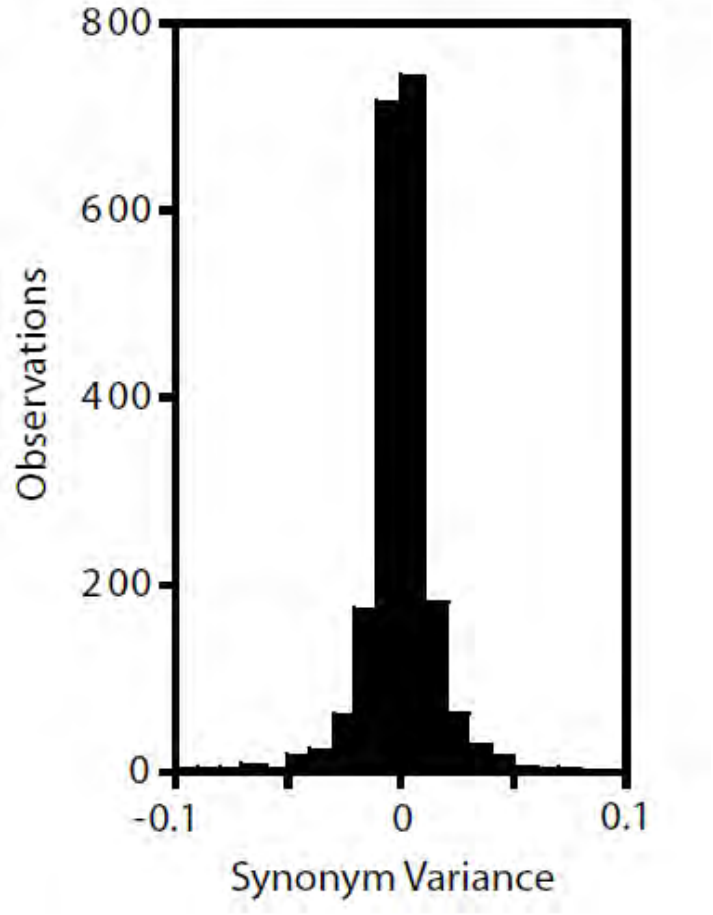


Figure 2.17 Effects of synonymous substitutions. For all mutations that persisted in the bulk competitions (s>-0.1), we calculated synonym variance as the difference between the effective selection coefficient for each codon and the average of all synonyms encoding the same amino acid.

effective selection coefficient for each amino acid substitution was estimated as the average of the effective selection coefficients of all synonymous codons. Epistasis between expression strength and amino acid substitutions was calculated as the difference in effective selection coefficient at reduced expression strengths relative to endogenous strength. For the epistasis calculations, null-like mutants were considered as true nulls. Thus, a mutant with wild type fitness at endogenous expression strength, and null-like fitness at the reduced expression strength would have an epistasis of -1.

Estimations of mutant effects on function. Function per molecule was calculated based on observed selection coefficients, the elasticity function, and the expression level for each different promoter construct using the following equations.

$$G=G_{max}/(1+E_m/EF) \quad (1)$$

$$G_{mut}/G_{WT} = W_{mut} = 1+s \quad (2)$$

Where G is growth rate, G_{max} is the maximal growth rate, E_m is the relative expression level that results in half maximal growth, E is the expression level relative to the endogenous level, F is the per molecule functionof a mutant relative to WT, W_{mut} is the growth rate of a mutant relative to WT, and *s* is the effective selection coefficient. Equation 1 is an extension of the elasticity equation (Figure 2.4), where the expression of functional molecules or net function (EF) is explicitly modeled. With the WT coding sequence (F=1 by definition), equation 1 simplifies to the elasticity function in Figure 2. These equations can be combined and rearranged to define F as follows.

$$F=E_m W_{mut}/(E+E_m-EW_{mut}) \quad [1]$$

Equation 3 was used to estimate mutant effects on function (Table 2.3) using the observed selection coefficients (Table S2.2), E_m=0.014 (Figure 2.4), E for each promoter construct based on experimental measurements (E_{GPD}=1,E_{TEF}=0.32,$E_{TEF\Delta ter}$=0.094), or estimated from the observed growth rate and the elasticity function for weak promoter constructs where experimental measures of expression were noisier (E_{CYC}=0.028,$E_{CYC\Delta ter}$=0.015,E_{ADH}=0.014,$E_{ADH\Delta ter}$=0.010). Where growth rates prohibited accurate estimation of fitness (null-like mutants, or absolute growth rates within 5% of G_{max}), bounds on relative per molecule function were calculated (Table 2.4). For each amino acid substitution, a final per molecule function estimate was generated by averaging across all promoter constructs that yielded a numerical estimate (and not a bound). For all pair-wise numerical function estimates (e.g. at two different expression strengths), we compared function effects between all constructs with adjacent expression levels (Figure 2.14). To facilitate biophysical comparisons, we used the Blosum62 matrix [327] to calculate the amino acid similarity to wild type for each possible point mutation, and Rosetta [321] to simulate effects on thermodynamic folding stability.

Model assumptions. We make the simplifying assumption that expression level is independent of mutations to the coding sequence. Steady state expression level is determined by the rates of both synthesis and degradation. Because degradation occurs after protein synthesis, it should depend primarily on the protein sequence such that synonymous substitutions minimally impact degradation rates. Across our data set we

noted that synonymous substitutions did not have dramatic impacts on fitness, suggesting that synthesis rates were relatively independent of mutation. Protein degradation rates vary depending on protein sequence, but all of the mutants that we analyze are single amino acid substitution, and hence minimally differ in overall sequence. In the event that a point mutant impacts degradation rate, it should be consistent across each promoter construct. Thus, mutant impacts on degradation should be rare (see Figure 2.16), but would be incorporated into our estimates of function.

In analyzing the effect of mutations relative to wild type, we make the simplifying assumption that function is independent of expression level. We examined the validity of this assumption by analyzing the standard deviation in function for each amino acid substitution determined at different expression levels. The average standard deviation was 0.1, indicating that this assumption is valid on a rough scale (on the order of 0.1) and is appropriate for interpreting the main features of the distribution of mutant effects on function. Of note, the mutations that we observe to improve function at reduced Hsp90 expression levels (Figure 2.13, Table 2.3) may be an artifact of this assumption.

The elasticity function does not include a cost of expression and as such has a maximum fitness at infinite expression level. Thus, we assume that expression cost is negligible relative to expression benefit over the range of our analyses. As the expression cost of native proteins is below experimental detection in yeast [328], this assumption appears reasonable.

We infer differences in cellular growth rates from measurements of DNA abundance. This inference is valid if DNA and cellular abundance are coupled. In previous work, we demonstrated that EMPIRIC measurements of fitness based on measures of plasmid abundance correlate strongly with cellular growth rates for a large set of mutants [312], indicating that plasmid abundance and cellular abundance are coupled. In addition, the copy number of the CEN plasmids utilized in this study is regulated, as cells maintaining multiple CEN plasmids grow slowly [329]. In addition, the low copy number of CEN plasmids is dominant to the addition of high copy genetic elements [330] and genetic alterations that increase CEN abundance are rare [331]. Nonetheless, CEN plasmids are not as stable as chromosomally encoded DNA, which may lead to a small amount of noise in our measurements.

Acknowledgments

We are appreciative of helpful discussions with B. Roscoe, R. Gilmore, N. Rhind, and O. Rando, to C.R. Matthews for the generous use of his CD instrument, and to R. Konz and the University of Massachusetts Medical School Core Flow Cytometry Lab for assistance and guidance with flow cytometry experiments.

Chapter III - Saturation mutagenesis of the HIV-1 Envelope CD4 binding loop reveals residues controlling distinct trimer conformations

This work has been submitted previously to *Nature Structural & Molecular Biology as Duenas-Decamp M*, Jiang L*, Bolon DN[#] and Clapham PR[#]. Saturation mutagenesis of the HIV-1 Envelope CD4 binding loop reveals residues controlling distinct trimer conformations.(* equal contribution; [#] co-corresponding author)*

The work presented in this chapter was a collaborative effort. I performed the plasmid library construction, cloning of individual mutations, reverse transcription, DNA sequencing library preparation and sequencing, sequence and statistical analysis. Dr. Maria Duenas-Decamp performed viral library recovery and titration, bulk competition experiment, viral RNA isolation and antibody neutralization assays and IC50 determination Dr. Daniel N.A. Bolon and Dr. Paul R. Clapham supervised the research. I, Dr. Maria Duenas-Decamp, Dr. Daniel N.A. Bolon and Dr. Paul R. Clapham analyzed the data and wrote the manuscript.

Abstract

The conformation of HIV-1 envelope (Env) glycoprotein trimers is key in ensuring protection against waves of neutralizing antibodies generated during infection, while maintaining sufficient exposure of the CD4 binding site (CD4bs) for viral entry. The CD4 binding loop on Env is an early contact site for CD4 while penetration of a proximal cavity by CD4 triggers Env conformational changes for entry.

The role of residues in the CD4 binding loop in regulating the conformation of the trimer and trimer association domain [168] was investigated using a novel saturation mutagenesis approach. Single mutations identified, resulted in distinct trimer conformations affecting CD4bs exposure, the glycan shield and the TAD across diverse HIV-1 clades. Different trimer conformations will affect the specificity and breadth of neutralizing antibodies elicited *in vivo* and are important to consider in design of Env immunogens for vaccines.

Introduction

The HIV-1 envelope glycoprotein (Env) comprises a surface gp120 and a transmembrane gp41 non-covalently associated on heterodimeric trimers. When gp120 on the Env trimer binds CD4 at the cell surface, conformational changes are triggered that open the trimer to expose a site for binding to a coreceptor, usually CCR5. Trimer opening involves the disengagement of the trimer association domain [168] at the trimer apex enabling (1) movement of the V1V2 loops to expose the V3 loop and (2) full exposure of determinants on the V1V2 stem recruited by CD4 to assemble the bridging sheet. The V3 loop and sections of the bridging sheet form the coreceptor binding site [332].

The CD4 binding loop on Env is an early contact site for CD4[161], while penetration of a proximal cavity by the hydrophobic side chain of CD4's Phe-43 triggers Env conformational changes and trimer opening [176, 333, 334] (Figure 3.1).

HIV-1 Envs in brain tissue use CCR5 as a coreceptor and are highly macrophage-tropic. These Env variants interact efficiently with low CD4 levels on macrophages for infection [335, 336]. Determinants that modulate mac-tropism of R5 Envs lie within or proximal to the CD4bs [177, 337] as well as in V1V2 and V3 loops of the TAD [177, 338, 339]. They include residues within the variable N-terminal flank of the CD4 binding loop that together with V3 loop amino acids modulated mac-tropism in a highly mac-tropic brain Env from a subject with neurological complications [177].

Figure 3.1: The HIV-1 envelope trimer structure

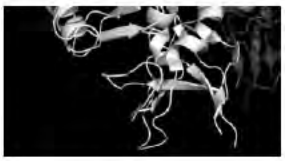


Figure 3.1 The HIV-1 envelope trimer structure. Structures shown were derived from PDB 4NCO [159]. (a) Side view of trimer depicted as a cartoon showing the location of the target sequences of two EMPIRIC libraries that span the CD4 binding loop and flanks and comprise residues 361-370 (yellow spheres) and 371-380 (green spheres). (b,c,d,e) Residues 361-370 (b,d) and 371-380 (c,e) are shown as spheres. (b-c) Red amino acids indicate CD4 contact residues [161]. (d-e) Blue and red residues show amino acids that surround the Phe-43 cavity, with red spheres directly contacting Phe-43 [159]. (f) CD4 binding loop residues 360-380 for clade B Envs, LN40, LN8 and clade C Env Z1792M. Stars denote CD4 contact residues, circles denote residues that line the Phe-43 cavity.

Here, a novel saturation mutagenesis approach; EMPIRIC (Exceedingly Meticulous and Parallel Investigation of Randomized Individual Codons) was exploited to investigate individual residues in a 20 amino acid region encompassing the CD4 binding loop, for effects on replication and Env conformation. This 20 residue region includes conserved residues that contact CD4 and/or form part of the Phe-43 cavity (Figure 3.1).

EMPIRIC involves the generation of libraries of mutations encoding all possible individual amino acid substitutions across important regions of genes [51, 54, 56, 326, 340, 341]. Libraries are subject to selection or competition before analyzing by deep sequencing to quantify the frequency change of each mutation. Using EMPIRIC, substitutions in the CD4 binding loop and flanks were identified that conferred enhanced or *wt* levels of replication in peripheral blood mononuclear cells (PBMCs). Several substitutions modulated the Env trimer with different mutations imparting distinct conformations that enhanced the exposure of the CD4bs and had varying effects on the TAD including the V3 loop and the glycan shield. One mutation enhanced the presentation of the trimer specific V2q, PG9/PGT145 epitope in V1V2 of the TAD, consistent with a modified but closed trimer conformation. The effects of the different mutations were transferable to diverse clade B and C Envs. These observations confirm the capacity of EMPIRIC to identify single Env residues in the CD4 binding loop region that induce different conformational states in the TAD and trimer. This data is relevant for design of trimeric Env immunogens in vaccines that aim to protect against diverse HIV-1.

Results

The primary LN40 Env and saturation libraries

Saturation mutant libraries were introduced into the primary LN40 Env [177, 182, 342]. LN40 env was PCR amplified and cloned from the lymph node of an AIDS patient with neurological complications. The LN40 R5 Env is not mac-tropic and is typical of Envs from immune tissue throughout disease [335, 343-345]. Most transmitted, founder R5 Envs are also not mac-tropic [346-349]. LN40 and other non-mac-tropic R5 Envs may form tightly closed trimers that protect against neutralizing antibodies (nabs). Determinants of LN40 non-mac-tropism were previously mapped to residues on the N-terminal flank of the CD4 binding loop in addition to residues within V3. Presumably, these residues reduce access to CD4 (as well as nabs) and restrict replication to T-cells expressing high CD4 levels [168, 177, 182, 342]. It was predicted that mutations in the CD4 binding loop, its flanks and Phe-43 cavity would have strong potential to increase viral fitness by enhancing efficiency of Env/CD4 interactions.

Two plasmid libraries were made containing all possible point mutations for Env amino acids 361-380 (361-370 and 371-380 in each library) of LN40 env in full length, replication competent, pNL4.3 [343]. The vast majority of the mutants were present in the plasmid libraries and virions (P0) produced by transfection of 293T cells (Figure 3.2a, b). The frequency of most mutants in plasmid and P0 libraries was well above the background from all processing steps including RT, estimated by sequencing virus recovered from a plasmid with *wt* env (Figure 3.2d). The frequency of mutants in the P0

Figure 3.2: Frequency of mutations in plasmid library, P0 and P1 viruses

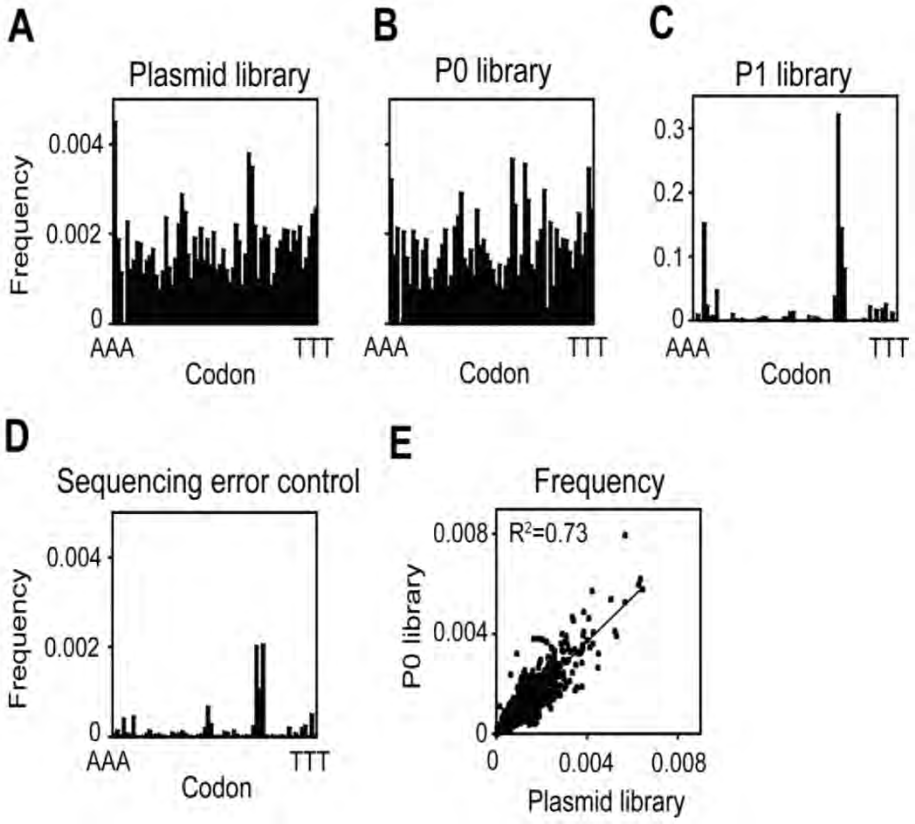


Figure 3.2 Frequency of mutations in plasmid library, P0 and P1 viruses. The frequency of mutations at position 377 in plasmid library (a), P0 library (b), P1 library (c) and *wt* plasmid as noise level estimate (d). (e) The frequency of mutants were strongly correlated in P0 library and in plasmid library. The data from library-371-380 is shown.

library was highly correlated with that in plasmid library, indicating that P0 library recovery by transfection achieved a good sampling of mutants in the plasmid library (Figure 3.2e).

Bulk competition of mutant libraries in PHA/IL-2 stimulated PBMCs

P0 virus of each library was competed in bulk for amplification in PBMCs (Figure 3.3a). The abundance of each mutant was measured before and after amplification using Illumina deep sequencing and fitness estimated (see Methods). Stop codons were consistently depleted in both libraries (Figure 3.3b). All *wt*-synonyms in library-371-380 displayed *wt*-like fitness effects, although slightly more variation in fitness effects of *wt* -synonyms in library-361-370 was noted.

Following eight days of infection, a strong correlation between enrichment and depletion of mutants in replicates was observed in library-371-380, but a weaker correlation between replicates in library-361-370 (Figure 3.3c, d). One explanation for the difference in reproducibility is that library-361-370 had less infectivity, and fewer virions mediated infection of PBMC. It was therefore more prone to insufficient sampling of P0 library, perhaps leading to stochastic enrichment or depletion of mutants.

Despite slightly higher variation in library-361-370, the results indicate that the majority of selection was reproducible and caused by introduced mutations. Data from the two replicates were pooled to obtain more precise measurements for analysis

Figure 3.3: EMPIRIC protocol, depletion of stop codons and reproducibility between assays

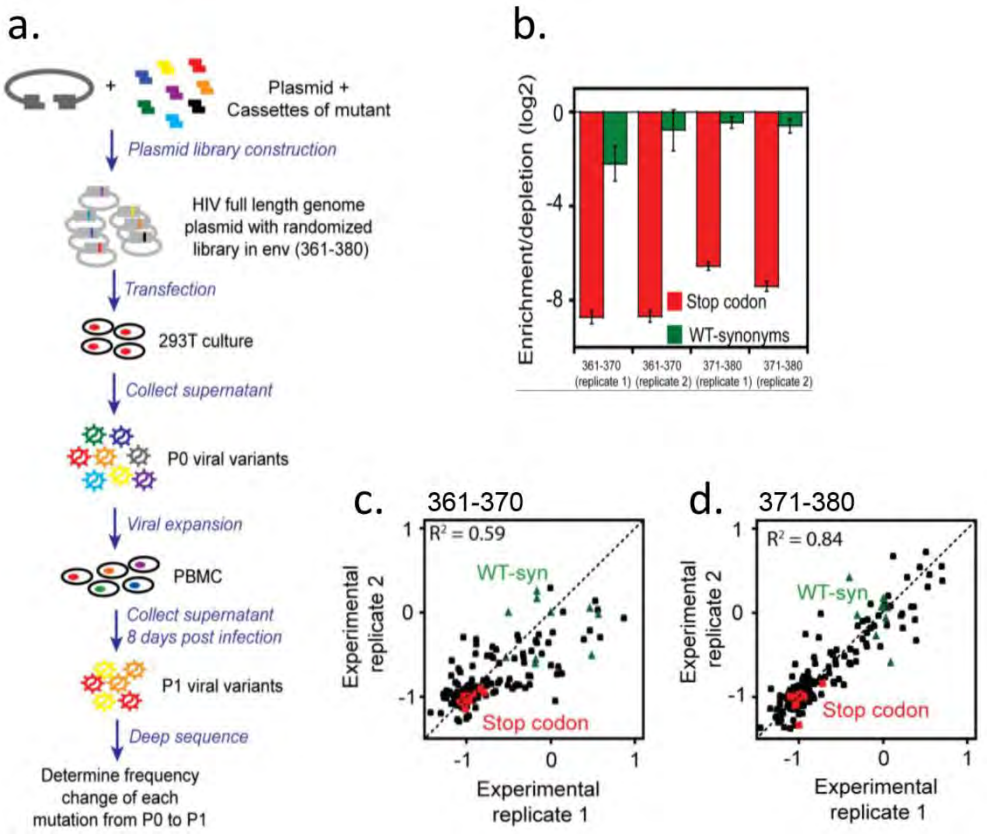


Figure 3.3 EMPIRIC protocol, depletion of stop codons and reproducibility between assays. *(*a) A cassette ligation strategy was used to introduce all 64 possible codons into each position of the CD4 binding loop region to form two libraries encompassing residues 361-370 and 371-380. (b) Depletion of stop codons and enrichment of *wt* synonyms (change in log2 frequency) in two experimental replicates of each library. Stop codons are shown in red and *wt* synonyms are shown in green. (c and d) Reproducibility of EMPIRIC measurements in HIV. Correlation in relative abundance of ~600 point mutants at amino acid positions 361-370 (c) and 371-380 (d) in the Env gene following infection of PBMCs. Green spots in panels b and c represent codons synonymous with *wt* codons.

(Table 3.1). The fitness effect of each amino acid was compared to that of repeatedly resampled *wt* -synonyms and an empirical p value of each amino acid (significantly lower fitness than *wt*, synonyms), or statistically *wt*-like (Tables 3.1 and 3.2). The majority of mutations in both libraries were strongly deleterious, indicating the 2 regions are extremely sensitive to missense mutations.

Both libraries had a limited number of mutants with *wt* -like fitness or above (Tables 3.1 and 3.2; Figure 3.4a), with library-361-370 containing more fit mutations (19%) than library-371-380 (16%), consistent with an increased variability of amino acids N-terminal to the CD4 contact residues. No other amino acids except *wt* residues were fully functional in CD4 contact residues ($GGD_{368}\_E_{370}$) (Figure 3.4b, c). In contrast, substitutions of proline at position P369 to cysteine, alanine, glutamine and aspartic acid conferred *wt* -like fitness. Positions 361-365 exhibited relatively higher tolerance of mutations, especially positions 362 and 363, where the *wt* amino acids are asparagine and glutamine respectively. Charged amino acids (except for histidine) were well tolerated at these two positions (Figure 3.4b, c). At position 365, the *wt* is serine, whereas valine and alanine exhibited slightly increased fitness.

In the 371-380 library, positions 373, 375 and 377 were tolerant of mutations. Even residues carrying side chains with different structures were at least *wt* -like in fitness (Figure 3.4b, c). For example, at position 373, glutamic acid was strongly beneficial over *wt* arginine. At position 375, where the *wt* amino acid is serine, all amino

Table 3.2: Beneficial mutations

Mutant	Fitness effect
R373E	0.36
R373K	0.34
R373N	0.24
R373Q	0.43
S375F	0.63
S375H	0.63
S375T	0.44
S375W	0.36
S375Y	0.64
N377L	0.16
N377T	0.18
N377V	0.49
G380P	0.27

Figure 3.4: EMPIRIC analysis of the CD4 binding loop and flanks of LN40 Env and catergorization of mutations as beneficial, wt or deleterious

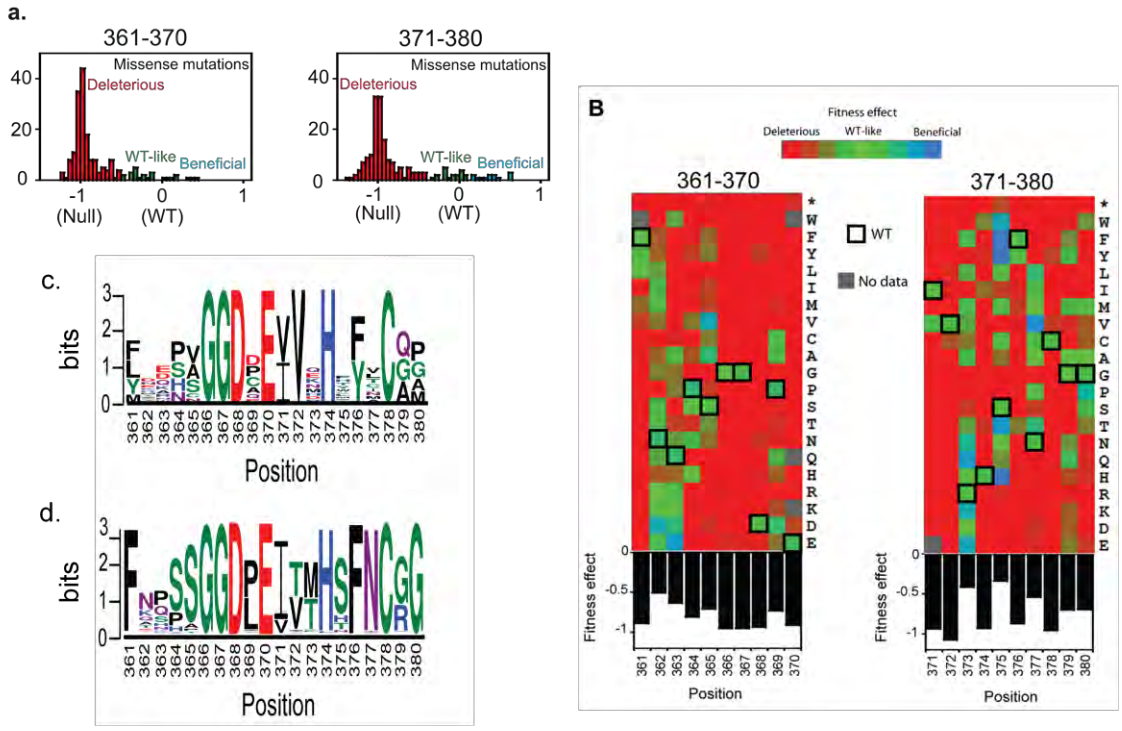


Figure 3.4 EMPIRIC analysis of the CD4 binding loop and flanks of LN40 Env and catergorization of mutations as beneficial, wt or deleterious. (a) Most mutations were deleterious. However, subsets of mutations that imparted wt or higher levels of replication were identified. (b) Mutations that affect CD4 contact residues were strongly deleterious. However, twelve mutations in regions downstream of the CD4 contact residues dramatically increased the efficiency of viral replication. These residues are proximal to the Phe-43 cavity. (c) The most common amino acids in positions 361-380 following PBMC passage of the two libraries. (d) The most common amino acids in positions 361-380 in all subtypes in the patient sequence database (http://www.hiv.lanl.gov/content/sequence/HIV/mainpage.html).

acids with aromatic rings as well as threonine and histidine were strongly beneficial. Glycine is the *wt* residue at position 380, and this residue is relatively conserved in natural isolates (Figure 3.4d). Proline was slightly beneficial at 380. Positions 371, 372, 374, 376 and 378 were resistant to change, although substitutions to amino acids with very similar structures (e.g. Ile/Val and Phe/Tyr) were tolerated at some positions e.g. 371 and 376.

Overall, 13 beneficial mutations were identified, all in library-371-380 (Table 3.2). Although a few mutations in library-361-370 exhibited small positive fitness effects, they were not statistically significant. This could be partly attributable to higher variation of fitness measurements in this library (leading to reduced statistical power) or to the optimal adaptation of this region to its function. All beneficial mutations in library-371-380 have a greater than 15% increase in fitness, which is very large compared to EMPIRIC studies in other systems [51].

Fitness benefit and *wt* mutations identified by EMPIRIC result in changes in LN40 Env conformation and function

35 mutations that conferred increased or wt levels of replication in PBMCs were investigated to elucidate effects on Env conformation. Each mutation was introduced into the pSVIIIenv Env expression construct carrying LN40 env, before producing Env+ pseudoviruses (Methods). Changes in Env conformation were evaluated by testing sensitivity LN40 Env+ pseudovirions to inhibition by soluble CD4 (sCD4) and Env

monoclonal antibodies (mabs) including CD4bs mabs, b6, b12, V3 loop mab, 447-52D, the glycan specific 2G12, V3 specific PGT128 and CD4i mab, 17b (Table 3.3).

Increases in sCD4 sensitivity

LN40 is relatively resistant to sCD4 inhibition due either to steric restrictions to binding and/or to resistance to CD4-induced conformational changes [168]. Sharp increases in sCD4 sensitivity for mutant Envs are consistent with changes in Env conformation that enhance CD4 interactions. Substitutions that increased sCD4 sensitivity included several at residue 375 *e.g.* S375W. Residue 375 is proximal to the Phe-43 cavity and S375W was reported to increase Env sampling of the CD4-bound form in gp120 monomers [176]. Substitutions at residues 373 (R373E), 377(N377V) and 380 (G380P) also conferred increases in sCD4 sensitivity. Other substitutions had less or no effect on sCD4 sensitivity (Table 3.3).

Increases in sensitivity to CD4bs mabs

LN40 is resistant to the CD4bs mab b12. The b12 epitope is present on monomeric LN40 gp120 but occluded on the trimer [168]. Five substitutions at residue 373 (373E, 373M, 373N, 373Q and 373K) increased LN40 sensitivity to b12. This observation is not surprising since it was previously reported that the side chain of R373 (together with the glycan at N386) sterically restricted W100 of b12 from accessing a

Table 3.3: The effect of LN40 mutations identified by EMPIRIC on Env structure and function.

LN40 Env *wt* and mutants		sCD4	447-52D *-V3 crown*	b6 *-CD4bs*	b12 *-CD4bs*	17b *-CD4i*	PGT128	2G12 *-glycans*
		IC50s (µg/ml)						
LN40 wt		>50	40	>50	34.9	>50	0.008	5.2
361	F361I	49.1	32.8	>50	>50	nt	0.007	3.1
	F361L	>50	34.6	>50	40.7	nt	0.007	3.5
	F361Y	35.8	30	>50	>50	nt	0.008	1.9
362	N362D	39.9	33.8	>50	29.6	nt	0.008	9.3
	N362E	>50	29.2	>50	>50	nt	0.008	3.5
	N362K	>50	21.6	>50	>50	nt	0.007	1.1
	N362S	>50	>50	>50	>50	nt	0.006	1.9
	N362T	>50	20	>50	>50	nt	0.008	2.0
	N362A	>50	>50	>50	>50	nt	0.007	1.9
363	Q363D	36.8	17.8	>50	4.5	nt	0.008	11.1
	Q363E	34.7	18.1	>50	2.5	nt	0.007	10.7
	Q363G	>50	29.4	>50	>50	nt	0.007	6.3
	Q363H	27.2	14.5	>50	>50	nt	0.008	6.1
365	S365A	32.1	21.7	>50	>50	nt	0.009	5.0
	S365V	44.6	28.5	>50	>50	>50	0.008	5.2
369	P369A	>50	>50	>50	>50	nt	nt	9.0
	P369C	>50	>50	>50	>50	nt	nt	5.1
	P369D	>50	26	>50	>50	nt	nt	14
	P369E	>50	29	>50	>50	nt	nt	10
371	R371V	>50	>50	>50	34.5	nt	0.01	5.0

373	R373K	>50	35.2	>50	2.8	nt	0.009	8.2
	R373M	>50	>50	>50	1.1	nt	0.009	7.5
	R373Q	44.2	25.2	>50	1.5	nt	0.012	14.9
	R373E	23.9	2.4	>50	2.0	>50	0.012	26.4
	R373N	43.9	12.9	>50	3.4	>50	0.01	9.9
375	S375H	15.3	>50	>50	<50	nt	0.014	15.9
	S375T	38.2	>50	>50	34.0	nt	0.009	8.4
	S375F	7.6	>50	>50	>50	nt	0.013	21.3
	S375W	8.7	>50	>50	42.8	>50	0.014	32.4
	S375Y	5.9	>50	>50	45.4	>50	0.01	30.1
377	N377V	19.0	1.3	>50	41.2	>50	0.01	12.3
380	G380A	32.0	<0.2	33.5	26.9	>50	0.01	11.4
	G380P	21.7	<0.2	15.3	25.2	>50	0.02	26.4

nt; not tested

green, >10<25; yellow, >1<10, red, <1.

pocket on gp120 for binding [342]. All substitutions at 373 with a shorter side chain than arginine might be expected to expose the W100 pocket for b12 [342]. Substitutions Q363E and Q363D also conferred sensitivity to b12. These negatively charged residues may alter the structure of the b12 W100 pocket (*e.g.* by moving the glycan at N386) so that it is now open. Decreased 2G12 sensitivity supports a shift in the glycan shield including the glycan at N386, a 2G12 target [182]. The same 363 and 373 substitutions also impacted the TAD as detected by increases in 447-52D sensitivity indicating a more exposed V3 loop.

Mab b6 also targets the CD4bs. However, its epitope is occluded on primary Envs [336, 350]. Here, b6 failed to neutralize any mutant, except for G380A and G380P, which conferred weak sensitivity. This indicates that other substitutions tested did not sufficiently open the trimer to enable b6 binding.

Increased sensitivity to 447-52D and V3 loop exposure

Mab 447-52D recognizes a GPGR motif on the V3 crown which is occluded within the trimers of most primary HIV-1 strains, including LN40.

Substitutions that increased sensitivity to 447-52D, indicating a more exposed V3 loop, included Q363D, Q363E, Q363H, R373E and R373N. Substitutions at 377 (N377V) and 380 (G380A and G380P) also conferred enhanced sensitivity to 447-52D as well as to sCD4, consistent with a larger impact on the TAD. G380A and G380P mutants

were exquisitely sensitive to 447-52D implying a more dramatic shift in the TAD and V3 loop exposure. In contrast, substitutions at residue 375 (*e.g.* S375W, S375Y and S375F) that conferred increased in sCD4 sensitivity, remained resistant to 447-52D, indicating that the V3 loop was not exposed and that these changes induced a distinct trimer conformation.

Decreased sensitivity to 2G12 indicates a shift in the glycan shield

All the substitutions that increased sensitivity to sCD4 and/or CD4bs and V3 loop mabs also had modestly increased resistance to the glycan specific mab, 2G12, consistent with a shift in the orientation of the glycan shield.

Env mutants do not carry an exposed CD4i epitope

We focused on substitutions at 373, 375, 377 and 380, which imparted the most significant shifts in Env conformation and tested their sensitivity to the CD4i mab, 17b [161, 351]. All were resistant, indicating that changes in trimer conformation for these mutant Envs did not expose the CD4i epitope or coreceptor binding site.

Several different residues substituted at P369 impart wt LN40 replication

CD4 contact residues, $GGD_{368}\_E_{370}$, in the CD4 binding loop were not readily substituted. However, mutations that substituted P369 with A, C, D and E residues imparted wt-like replication. P369 is relatively conserved in clade B Envs and its replacement with other amino acids might be expected to confer properties selected against in vivo. However, 369 mutants carrying A, C, D E, did not exhibit differences in sensitivity to sCD4, 447-52D, b6 or b12, while only D and E slightly decreased sensitivity to 2G12. These results indicated that this position accepts several different amino acids without detectable changes in Env conformation or replication fitness.

Summary of LN40 mutants

The use of multiple neutralizing mabs to investigate each substitution allowed an assessment of the changes distal to the residue in question and in overall Env conformation. The data presented show that different residues within or closely associated with the Phe-43 cavity have distinct effects on the conformation of the TAD and trimer.

The effect of introducing an N160 glycan into LN40 Env

V2q mabs, PG9, PG16 and PGT145 bind the V2 glycan N160 and preferentially recognize the TAD at the apex of trimers rather than monomers. These mabs can be used as probes to assess whether this site has been disrupted consistent with trimer opening

[168, 352]. Unfortunately, LN40 Env does not carry the N160 glycan, critical for binding V2q mabs, although a glycan at N156 is also targeted by V2q mabs. The N160 potential N-linked glycosylation signal was restored to LN40 Env by mutagenesis in the hope of reconstituting V2q mab binding. However, LN40 Y160N remained resistant to V2q mabs PG9, PG16 and PGT145 (not shown) and presumably doesn't bind these mabs.

Nevertheless, the introduction of N160 conferred increased sensitivity to sCD4 (Figure 3.5a, Table 3.4), consistent with a TAD conformation that enhances access to the CD4bs. Several substitutions identified by EMPIRIC conferred further enhancement in sensitivity to sCD4, the V3 mab 447-52D and CD4bs mab b6 when introduced together with Y160N (Figure 3.5a-c, Table 3.4). In particular, LN40 160N mutants, 377V and 380P, were more sensitive to sCD4 compared to Env mutants without Y160N. The presence of N160 also enhanced sensitivity of 377V to V3-specific 447-52D and more modestly to b6 (Figure 3.5b-c, Table 3.4). In contrast, the LN40 160N 375W mutant was only modestly more sensitive to sCD4 and remained relatively resistant to both 447-52D and b6.

These data indicate that the N160 glycan on LN40 conferred a trimer conformation where the CD4bs site is more exposed enabling enhanced interactions with CD4 and antibodies targeting V3 and the CD4bs.

Mutations that change LN40 Env conformation conferred similar effects on another clade B Env and a clade C transmitter, founder Env

Several mutations were introduced into another clade B Env, LN8, and into a clade C Env, Z1792M. LN8 is an R5 envelope derived by PCR cloning from the lymph node of

Figure 3.5: The effect of N160 on LN40 wt and mutant Envs on sensitivity to sCD4

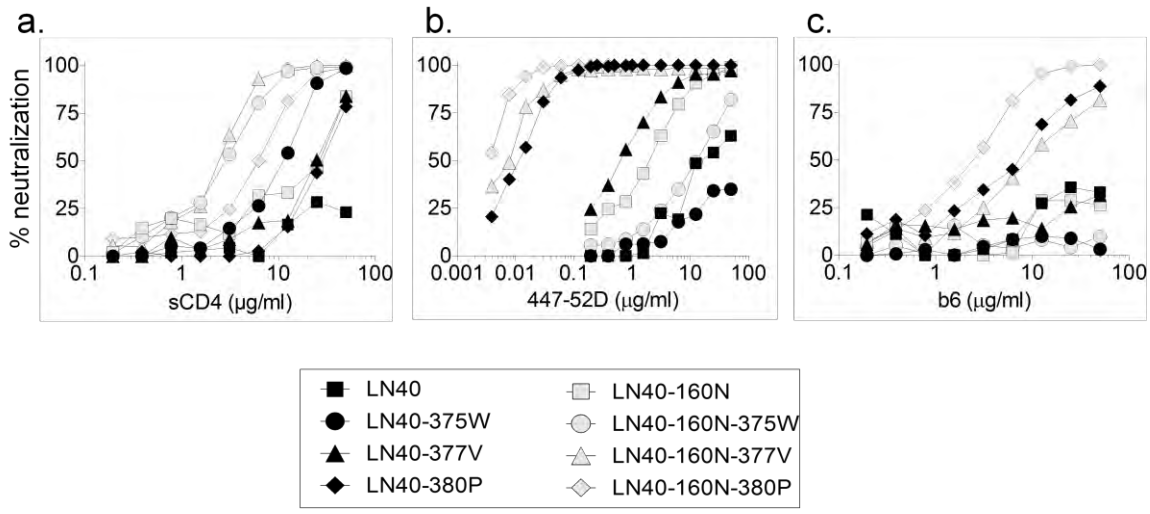


Figure 3.5 The effect of N160 on LN40 wt and mutant Envs on sensitivity to sCD4 *(a), V3 loop mab, 447-52D (b) and CD4bs mab, b6 (c).*

Table 3.4: N160 enhances the sensitivity of LN40 wt and mutant Envs to sCD4 and mab neutralization.

Mab, inhibitor	Env	IC50s +/- N160	
		-	+
sCD4	*wt*	>50	21.2
	R373E	27.0	8.4
	S375W	8.7	4.6
	S377V	19.0	2.1
	G380P	22.2	6.1
447-52D	*wt*	32.4	4.7
	R373E	2.4	<0.2
	S375W	>50	24.2
	N377V	1.1	0.008
	G380P	0.014	0.004
b6	*wt*	>50	>50
	R373E	>50	19.5
	S375W	>50	>50
	N377V	>50	18.7
	G380P	15.3	4.3
b12	*wt*	28.9	45.9
	R373E	2.0	1.4
	S375W	>50	>50
	N377V	>50	>50
	G380P	>50	>50

green, >10<25; yellow, >1<10, red, <1.

subject NA20, who died of AIDS with neurological disease [335, 343], while Z1792M is a transmitter, founder, clade C R5 Env from Zambia [353].

Substitutions at residue S375

Substitutions S375H and S375W both enhanced sCD4 sensitivity of LN8 and clade C, Z1792M, as they did for LN40 (Figure 3.6a; Tables 3.5 and 3.6). These S375 substitutions had little effect on LN8 sensitivity to b12, b6 and 447-52D, yet conferred more resistance to 2G12. This is consistent with a shift in the orientation of one or more glycans resulting in a more exposed CD4bs and increased sCD4 sensitivity. The changes implicated in LN8 Env conformation closely follow those described above for LN40 Env. Z1792M does not carry the b12, 447-52D or 2G12 epitopes. However, S375 substitutions had no effect on b6 resistance.

The N377V substitution

N377V conferred increased sensitivity to sCD4 for LN8 and Z1792M and modestly increased sensitivity of LN8 to V3 mab, 447-52D, following observations made for the equivalent LN40 mutant (Figure 3.6a, b).

Figure 3.6: Substitutions at residues 375, 377 and 380 confer similar effects on clade B LN40, LN8 and clade C Z1792M Envs

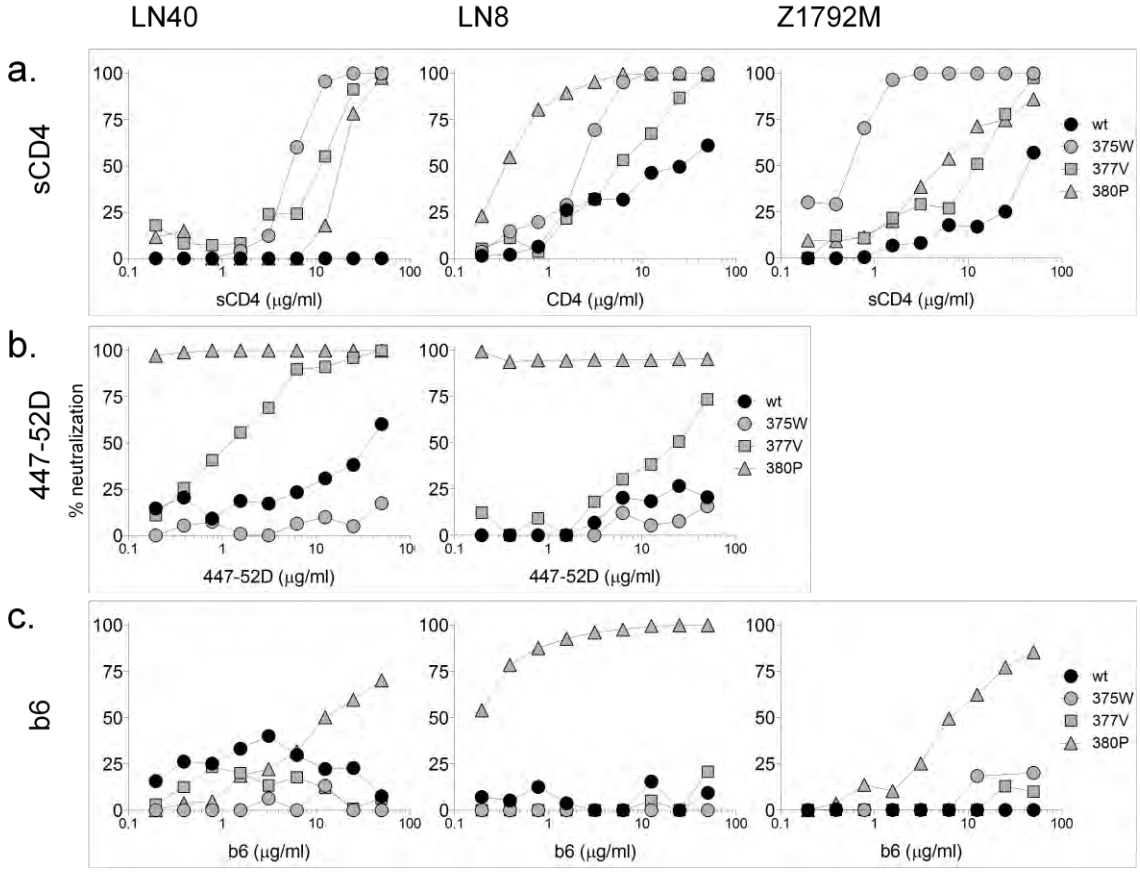


Figure 3.6 Substitutions at residues 375, 377 and 380 confer similar effects on clade B LN40, LN8 and clade C Z1792M Envs. Neutralization profiles for sCD4 (a), 447-52D (b) and b6 (c) are shown. Each substitution had similar effects on the three different Envs: consistently increasing sensitivity to sCD4. However, only 380P conferred sensitivity to mab b6, while 377V and 380P exposed the V3 loop crown (447-52D epitope), while 375W didn't. Of note, Z1792M does not carry the 447-52D epitope.

Table 3.5: The effect of mutations identified by EMPIRIC on LN8 Env structure and function.

LN8 Env *wt* and mutants		sCD4	447-52D *-V3 crown*	b6 *-CD4bs*	b12 *-CD4bs*	2G12 *-glycans*	PGT128 *-V3 glycans*	PG9 *-V2 N160*	PGT145 *-V2 N160*
		IC50s (µg/ml)							
LN8 *wt*		40.7	>50	>50	28.7	6.16	0.019	0.17	0.0006
362	N362D	31.4	>50	>50	2.87	>50	0.033	0.19	0.0007
363	Q363D	34.5	>50	>50	14.1	21.8	0.027	0.16	0.0004
365	S365A	25.4	>50	>50	48.5	6.59	0.028	0.13	0.0006
	S365V	8.0*	>50	>50	26.9	3.61	0.015	0.05	0.0003
373	M373E	37.5	37.4	>50	>50	6.98	0.011	0.11	0.0004
	M373N	34.6	>50	>50	>50	6.87	0.012	0.22	0.0005
375	S375H	12.3	>50	>50	>50	30.8	0.027	1.0	0.0009
	S375W	3.4	>50	>50	>50	35.3	0.046	5.3	0.0011
377	N377V	9.8	36.7	>50	30.1	6.71	0.016	0.22	0.0007
380	G380A	24.1	3.6	>50	2.1	7.36	0.02	0.16	0.0007
	G380P	0.76	<0.2	0.2	<0.2	1.85	0.019	4.3	0.0016

sCD4, 447-52D, b6, b12, 2G12: green, >10<25; yellow, >1<10, red, <1.

PGT128: red, <0.1. PG9: yellow, >1; red, <1. PGT145: yellow, >0.001; red, <0.001.

*50% neutralization at 8.0 µg/ml. However, neutralization curves were erratic and neutralization incomplete, reaching a maximum of 75-85% at 50 µg/ml.

Table 3.6: The effect of mutations identified by EMPIRIC on Z1792M Env structure and function

Z1792M Env *wt* and mutants		sCD4	b6 *-CD4bs*	PG9 *-V2 N160*	PGT145 *-V2 N160*
		IC50s (µg/ml)			
Z1792M wt		44.2	>50	0.20	0.29
362	E362D	>50	>50	0.14	0.3
363	E363D	>50	>50	0.15	0.14
364	H364S	>50	>50	0.18	0.15
365	S365A	>50	>50	0.14	0.32
	S365V	46.3	>50	0.03	0.005
369	L369P	12.1	>50	0.41	0.79
373	T373E	43.2	>50	0.08	0.03
	T373M	19.1	>50	0.32	0.6
	T373K	41.1	>50	0.12	0.54
	T373Q	43.1	>50	0.20	0.3
375	S375W	0.57	>50	2.6	>4.0
	S375Y	1.2	>50	1.5	1.89
	S375F	2.0	>50	0.46	0.55
	S375H	2.5	>50	1.2	1.28
	S375T	26.0	>50	0.22	0.89
377	N377V	13.5	>50	0.20	1.9
	N377L	15.0	>50	0.20	1.32
	N377T	45.4	>50	0.16	0.68
380	G380A	35.2	>50	0.12	0.64
	G380P	7.2	9.5	0.32	>4.0

sCD4, b6, b12: green, >10<50; yellow, >1<10, red, <1.
PG9, PGT145: yellow, >1; beige, >0.1<1.0; red, <0.1.

Substitutions at G380

G380P enhanced sCD4, b6 and b12 sensitivity for LN8 and conferred exquisite sensitivity to the V3 mab, 447-52D. G380P also enhanced sensitivity to sCD4 and b6 for Z1792M (Figure 3.6, Tables 3.5 and 3.6). These results indicate that G380P conferred more dramatic changes to the trimer conformation of LN40, LN8 and Z1792M Envs resulting in improved CD4 interactions and enhanced exposure of the V3 loop and b6 epitope, consistent with a more open conformation of the trimer.

Together, these findings indicate that substitutions at 375, 377 and 380 enhance Env/CD4 interactions and trimer function across diverse HIV-1 clades.

L369P substitution in clade C Z1792M Env

P369 is conserved on the clade B database, while L369 is dominant for clade C. A L369P substitution was made in clade C, Z1792M Env. The presence of P369, rendered Z1792M more sensitive to sCD4 (Table 3.6), perhaps indicating an Env trimer more open to nabs and possibly explaining the selective advantage of L over P at 369 for clade C Envs in patients.

The effects of Env substitutions on the V2q epitope of LN8 and Z1792M

Unlike LN40, LN8 and Z1792M Envs carry the glycan at N160 in V2 and are sensitive to the trimer specific V2q mabs PG9 and PGT145 [354]. Measuring sensitivity to V2q mabs could help monitor whether trimers are closed for each mutant. Surprisingly, all LN8 and Z1792M substitutions remained sensitive to PG9 and PGT145, although Envs carrying each of S375W, N377V and G380P substitutions in LN8 and Z1792M were less sensitive to either PG9, PGT145 or both (Figure 3.7a; Tables 3.5 and 3.6). The modest impact of these substitutions on sensitivity to V2q mabs is curious. How can the LN8 G380P Env trimer be open sufficiently to bind b6 and 447-52D, yet remain substantially sensitive to V2q mabs, PG9 and PGT145 that recognize closed trimers? One possibility is that PG9 or PGT145 bind open trimers by retaining an interaction with monomeric gp120. However, using ELISAs, it was confirmed that these mabs do not bind LN8 or Z1792M gp120 monomers (not shown). Perhaps, the G380P trimers transition to and from closed and open forms [355], so that b6 captures the open state, while V2q mabs capture the closed state (Figure 3.8B). Alternatively, the native trimer may take up a conformation where the PG9, PGT145 epitope in the TAD is largely maintained while still exposing the V3 loop crown and CD4bs including the b6 epitope (Figure 3.8A).

An S365V substitution in a CD4 contact residue enhanced the V2q epitope

While substitutions downstream of the GGD-E (366-370) CD4 contact residues, resulted in changes consistent with TAD and trimer opening, mutations in upstream residues had little effect on sensitivity to sCD4 and other mabs

Figure 3.7: Env substitutions in the CD4 binding loop modulate the trimer specific, V2q epitope

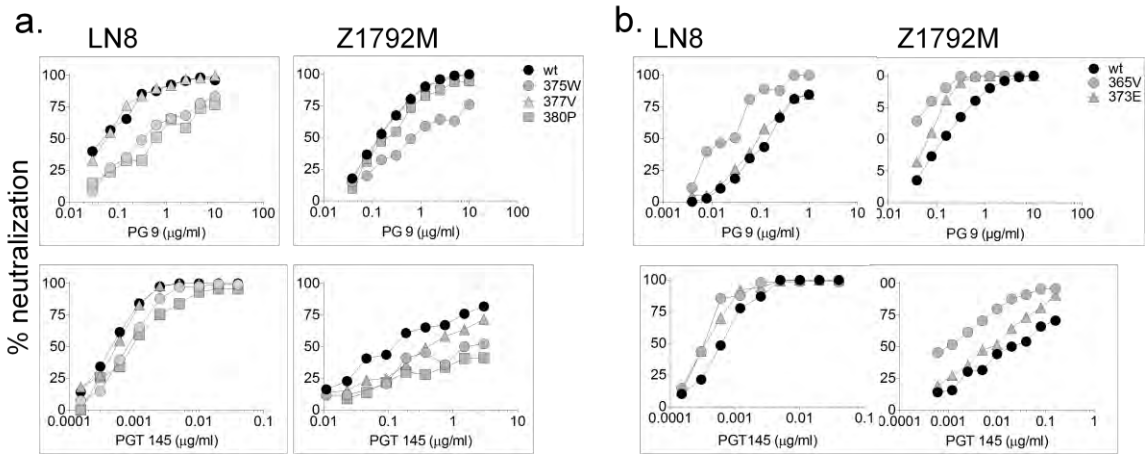


Figure 3.7 Env substitutions in the CD4 binding loop modulate the trimer specific, V2q epitope. (a) Env substitutions proximal to the Phe-43 cavity, reduced sensitivity to V2q, mabs, PG9 and PGT145, consistent with a more open TAD and trimer. (b) Substitution S365V on the variable, N-terminal flank of the CD4 binding loop enhanced sensitivity to V2q mab PGT145, consistent with a modified (but not more open TAD conformation).

Figure 3.8: Alternative models to explain how Env trimers may expose the CD4bs while retaining a closed TAD at the trimer apex

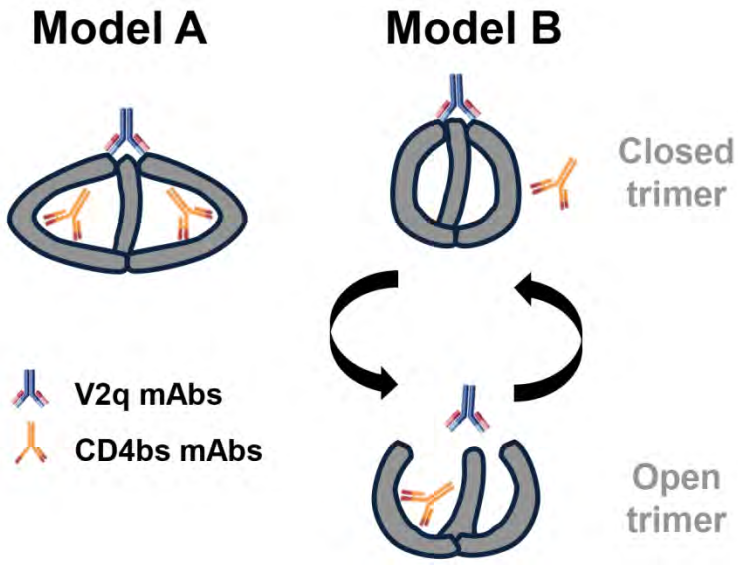


Figure 3.8 Alternative models to explain how Env trimers may expose the CD4bs while retaining a closed TAD at the trimer apex. (a) The Env trimer moves from a closed to open form and back again. V2q mabs capture the closed form, CD4bs mabs capture the open form. (b) Localized conformational changes expose CD4bs epitopes without impacting the TAD and V2q epitopes.

(Table 3.3, 3.5 and 3.6). Exceptions include N362D, Q363D and Q363E, which enhanced b12 sensitivity for either LN8 or LN40.

An S365V substitution modestly increased sensitivity to both PG9 and PGT145, particularly for Z1792M. Effects on LN8 were more marginal, although S365V was still the most sensitive LN8 mutant for PG9 and PGT145 (Figure 3.7b, Tables 3.5 and 3.6). The 373E substitution also enhanced (rather than reduced) the V2q epitope, although less than for S365V.

Discussion

An HIV-1 vaccine that induces potent and broad nabs will require detailed knowledge of the residues that control the configuration of the Env trimer. This information will help design Env immunogens that optimally present the best epitopes for eliciting the most rigorous vaccine response.

A novel saturation mutagenesis approach, EMPIRIC, was applied to investigate the role of residues in a 20 amino acid stretch of gp120 in regulating the conformation of the TAD and trimer. This region encompasses the CD4 binding loop and flanking regions. Several residues were identified that modulated the TAD and trimer (Figure 3.9). However, different residues enhanced Env: CD4 interactions by imparting distinct TAD conformations with differential exposure of the V3 loop and other epitopes usually occluded within the trimer. The majority of mutations that modified LN40 TAD and

Figure 3.9: CD4 binding loop residues identified that control trimer conformation

Figure 3.9 CD4 binding loop residues identified that control trimer conformation. Residues are depicted as spheres in different colors. The V1 (yellow), V2 (green) and V3 (orange) loops that form the trimer association domain [168] are shown at the trimer apex. Light yellow spheres show the NAG (N-acetyl glucosamine) of N160 (Cyan) on the top of the trimer. CD4 contact sites [2], are red in the cartoon structure. Structure is based on a side view of trimer (PDB 4NCO [36]).

trimer had similar effects on LN8 and clade C, Z1792M, indicating that residues identified, control Env conformation across clades.

Mutations likely to affect trimer conformation were identified in competition assays by replication in PBMCs. This approach selects for mutations that support viral replication in the absence of nabs and enabled the identification of residues infrequent on the HIV sequence database. For example, residues 375W, 377V and 380P increased CD4 binding and led to more open trimers where epitopes (usually occluded) are exposed. Two types of substitution were observed. (1) Substitutions that modified the local structure around the CD4bs e.g. N362D in LN8, where the b12 epitope was exposed without affecting other sites. (2) Substitutions that modified distal sites including V3 and the TAD e.g. S375W, N377V ad G380P. One concern with the competition assay was whether spontaneous mutations contributed to enhanced replication. This was not investigated. However, many mutations mediated significant effects on the trimer when investigated individually using pseudoviruses carrying mutant Envs. These observations confirm that EMPIRIC identifies physiologically relevant amino acid substitutions and is a formidable mutagenesis tool for analyses of Env conformation, fitness or other properties, relevant for vaccines.

Two libraries were prepared covering residues 361-370, and 371-380 (Figure 3.1). Library-361-370 contained the variable N-terminal flank [168] and CD4 contact residues GGD_{368}-E_{370}. Few mutants from this library were identified that increased replicative capacity over *wt* virus or conferred a more open Env conformation. Nevertheless, many

variant residues on the N-terminal flank conferred wt-like replication, contrasting with mutations in conserved CD4 contact residues where *wt* residues predominated during PBMCs replication (Figure 3.4b, c). These observations indicate that these latter amino acids have been highly selected during evolution as would be predicted for sites contacting a major receptor. The lack of beneficial mutants in the variable 361-365 region was curious, since a motif on this variable flank region was previously identified that influenced the non-mac-tropic phenotype of LN40 Env [177]. However, this motif required the presence of further determinants in V3 to mediate maximal effects on LN40 Env properties [177].

A small number of substitutions on the N-terminal flank did affect Env properties. LN40 Q363D and Q363E mutants were more sensitive to the CD4bs mab, b12. However, minimal effects on sCD4, 447-52D and 2G12, suggested a localized effect on b12 binding, rather than significant trimer opening. The 365V substitution conferred a small (not significant) increase in replication for LN40 Env+ virus in PBMCs. When introduced into other Envs, an increase in sensitivity to PG9 and PGT145 was detected for Z1789M (and more marginally for LN8) consistent with effects on the conformation of the TAD but without detectable opening of the trimer.

Several residues in library-371-380 could be replaced by residues that enhanced replication in PBMCs. These residues had major effects on the TAD and trimer conformation as indicated by significant increases in sensitivity to sCD4, CD4bs mabs b6 and b12, and to V3 crown mab 447-52D depending on substitution. The positions in 371-

380 are each closely associated with the Phe-43 cavity and likely play a role in controlling Env conformational changes in response to CD4 binding [333, 334].

Residues that opened the TAD and trimer had different effects on the exposure of mab epitopes. Several residues in the 371-380 region conferred enhanced sCD4 sensitivity, yet affected exposure of the 447-52D V3 epitope differently. Thus, G380P increased sensitivity to sCD4 in LN40, LN8 and Z1792M, but also imparted greatly enhanced sensitivity to 447-52D for both LN40 and LN8. In contrast, S375W, S375Y and S375F mediated increases in sCD4 sensitivity but had no effect on 447-52D resistance. Thus different residues associated with the Phe-43 cavity, regulate distinct conformations of the TAD at the trimer apex and highlight a potential role for this region in regulating immune protection.

For several mutants that had modified TAD and/or trimers, it was noted that sensitivity to 2G12 was reduced. Mab 2G12 recognizes several glycans on gp120 and its ability to bind and neutralize virions depends on their orientation. Reduced sensitivity to 2G12 suggests that modified, more open Envs have shifted the glycans forming this mab's epitope as previously reported [182].

The glycan N160 in V2 is a target for V2q antibodies and its lack explains why LN40 Env is insensitive to PG9 and PGT145 mabs. This was unfortunate since V2q mabs can be used to investigate changes in TAD configuration. Introduction of the potential glycan site at N160 in LN40 Env failed to restore sensitivity to V2q mabs. However, the presence of the glycan site at N160 increased sensitivity to sCD4 and V3 mab, 447-52D.

These observations indicate that in the context of LN40, N160 modified the trimer conformation, exposing the V3 loop and increasing access for CD4. N160 also enhanced the effects of other substitutions on Env conformation. For example, 373E, 377V and 380P were all more sensitive to V3 loop mab, 447-52D in the presence of N160. N160 on LN40 may loosen the TAD and facilitate some substitutions to confer more extensive shifts in conformation. It is worth noting that the effects of the S375W substitution were not enhanced by N160. These observations add further support to the conclusion that S375 substitutions induce a trimer conformation distinct from those imposed by other substitutions e.g at 377, 380.

It was surprising that several mutant LN8 and Z1792M Env trimers were sufficiently open to bind b6 and (for LN8 mutants) 447-52D, yet remained sensitive to V2q mabs (that predominantly recognize closed trimers), albeit with modestly increased IC50s. One explanation is that Envs transition from closed to open states and back, so that mabs b6 and 447-52D capture the open form, while the V2q mabs capture the closed form. However, it is also possible that some substitutions can expose the b6 epitope in the CD4bs region as well as the 447-52D epitope without significantly opening the TAD (Figure 3.8).

In summary, mutations in the CD4 binding loop and flanks that affect trimer conformation were identified in replication competition assays. Different substitutions identified were associated with distinct conformations that impacted on the exposure of the V3 loop and TAD, the CD4bs and the efficiency of CD4 interactions in Envs from

different clades. The information presented contributes to the establishment of universal, cross clade rules for regulating trimer conformation and will be invaluable in the design of next generation Env immunogens.

Methods

Construction of plasmid-encoded libraries

Env saturation mutant libraries were generated using a previously described approach [51, 326]. Briefly, the env gene was cloned into pRNDM to generate a plasmid without BsaI restriction sites. Inverted BsaI sites were then introduced to allow for a cassette ligation strategy with each single codon randomized as NNN to efficiently generate libraries of all possible codon substitutions; a separate cassette was used to mutate each codon to all 63 non-parental ones. Libraries of single codon mutants at 10 consecutive codons were combined and the resulting pool transferred from pRNDM to replication competent pNL4.3 with LN40 env, using sequence and ligation independent cloning (SLIC) [307]. A SLIC destination vector was generated that encoded the HIV genome with the majority of the env gene removed and a unique BmtI restriction site at this location. The destination vector was digested with BmtI and resected with T4 DNA polymerase as described previously [307] to leave approximately 30 base pairs of single stranded DNA at each end. Linear fragments of the env libraries from pRNDM with single stranded regions matching the prepared destination vector were generated by PCR (using Pfusion high fidelity polymerase and eight cycles of amplification to minimize amplification errors) and treatment with T4 DNA polymerase. The prepared library and destination DNA were mixed at equal molar amounts, annealed for 30 minutes at 37^{o}C, and transformed into bacteria to generate the plasmid libraries encoding full-length viral genomes.

Viral library recovery and competition experiments

2.5 μg of DNA encoding full length NL4.3 carrying LN40 or 361-370 and 371-380 mutant library envelopes were transfected into 293T cells using calcium phosphate. Supernatants carrying full length NL4.3-LN40env or libraries (P0) were harvested 48 h after transfection, clarified (1,000*g* for 10 min), aliquoted, and stored at 152°C. HeLa TZM-bl cells were used to titrate the P0 stock libraries using the LTR-controlled β-galactosidase reporter gene to identify infected cells as described previously [356].

$20x10^6$ PHA treated peripheral blood mononuclear cells (PBMCs) were recovered from a frozen stock and infected with 2 ml wild type (wt) LN40 virus, or with each library virus stock (P0) in duplicate. After 3 hours, infected PBMCs were centrifuged at 1200 rpm for 10 min. Supernatant was harvested and frozen as day 0 (D0). Cells were washed with 5ml of RPMI/10% fetal calf serum twice before adding 10 ml RPMI/10% fetal calf serum with IL-2 (Roche Inc.). Medium was changed after 4 days and supernatants collected on day 8. 200 μl samples were treated with recombinant DNase I for 2h at room temperature to eliminate any carry over of plasmid DNA before extracting RNA using the High pure viral RNA kit (Roche Inc.).

Sequence analyses and estimation of fitness

HIV genomic RNA was extracted from supernatants containing virions using High Pure Viral RNA kit (Roche Inc.). Viral RNA was reverse transcribed into cDNA using primers binding downstream of randomized libraries and SuperScript III (Life

Technologies Inc.). Subsequent processing steps were as described previously for analyzing mutant frequency [326]. Briefly, samples were barcoded to distinguish replicates as well as plasmid, P0, and P1 samples and submitted for Illumina 36bp single read sequencing on a Genome Analyzer II. Reads with a phred score of 20 or above (>99% confidence across all 36 bases) were analyzed (Table 3.7). The relative abundance (*A*) of each point mutant of plasmid, P0 and P1 library was estimated from read abundances (*R*) as indicated below in equation (1).

$$A = \log_2\left(\frac{R_{mut}}{R_{WT}}\right) \qquad \text{(Eq. 1)}$$

The frequency change (*F*) of a mutation from P0 to P1 (equation 2) was used as an estimate of the enrichment or depletion during viral expansion. Two replicates of P1 were determined separately.

$$F = A_{P1} - A_{P0} \qquad \text{(Eq. 2)}$$

Selection coefficients (*s*) representing the experimental effects of each mutation were calculated by normalizing the median of stop codon to -1 (representing null fitness) and wild-type synonyms to 0 (representing no fitness effect), as indicated in equation 3.

$$s = \frac{F_{mut} - F_{\overline{WTsyn}}}{F_{\overline{WTsyn}} - F_{\overline{stop}}} \qquad \text{(Eq. 3)}$$

The above analyses yielded estimates of fitness effects for each codon with frequency > 0.015% in the P0 library (Table 3.8). Mutations below this frequency in P0 were likely subject to highly stochastic sampling in the pool of viruses used to start P1

passages. These mutations (4% of the data) also had very low frequency in the plasmid library. The high correlation between frequency of mutants in P0 and the plasmid library suggested that viral recovery by transfection provided sufficient sampling of mutants in the plasmid library (Figure 3.2E), so mutants with low frequency in P0 library was due to their inherent low frequency in the plasmid library and not due to a bottleneck effect or selection in viral recovery.

As estimates of selection coefficient (s) had some noise, in particular in the 361-370 library (Figure 3.4b,c), the median of s of synonymous codons encoding the same amino acid was used to represent s of that amino acid, to minimize impact from outliers. RMSD was determined between the two replicates to estimate variation in s of amino acids. The two replicates in P1 were then pooled to estimate the selection coefficient of amino acids to further improve reproducibility. Specifically, the median was computed for s of all synonymous codons encoding the same amino acid in both replicates as s for that amino acid (Table 3.1).

To determine whether a mutant is statistically beneficial, or *wt* -like, or deleterious, the s of each amino acid was compared to the median of s of resampled *wt* -synonyms, and defined mutants with s significantly greater than that of *wt* -synonyms as beneficial; mutants with s significantly less than that of *wt*-synonyms as deleterious and the rest as *wt* -like. All *wt*-synonyms of both replicates were pooled for each library (38 for Library-361-370 library and 50 for Library-371-380 library). For each amino acid, *wt*-synonyms were resampled as twice the number of synonymous codons encoding that

amino acid and compared the median of s of resampled *wt*-synonyms with s of that amino acid. This process was repeated 10,000 times for each amino acid and computed the proportion (f) when s of amino acid is greater than median s of resampled *wt* -synonyms. 1-f (if f>0.5) or f (if f<=0.5) is the empirical p value of this amino acid having a fitness effect greater than or less than *wt* -synonyms. Before multiple test corrections, mutants with p value <0.025 have a significantly different s with *wt* -synonyms (Table 3.1). A two-sided 5% False Discovery Rate (FDR) was then applied as multiple test correction. After that, amino acid mutants with a sufficiently small p value were classified as statistically beneficial or deleterious and the rest as statistically *wt* -like. False negative rates were not estimated, so that a small number of mutants that were classified as *wt*-like might be beneficial or deleterious. Amino acids with more synonymous codons were treated as having more replicates so that they would have stronger statistical power in classification.

Cloning of individual mutants

A panel of individual gp120 mutations were cloned into the pSVIIIenv vector that carried clade B LN40, LN8 or clade C Z1792M env genes and analyzed in isolation. A cassette with a single mutant was ligated into BsaI digested pRNDM, as described above, and then subcloned into pSVIIIenv. For LN40 env, one KpnI site in pSVIIIenv vector outside env coding region, was eliminated with quick change mutagenesis. pRNDM with mutants and pSVIIIenv were both digested by KpnI and SpeI, and the mutant fragments from

pRNDM ligated into digested pSVIIIenv. For LN8 env, two sets of primers were utilized to generate two fragments of env by PCR. The 3' region of one fragment shared a 27-nucleotide homologous region with the 5' region of the other fragment, and the 3 nucleotides in the middle of the homologous region were mutated to desired mutations by PCR. Primers are described in Table 3.9. pSVIIIenv was digested by KpnI and SpeI, followed by T4 polymerase trimming to generate matched ends ready for homologous recombination. The digested vector and the two PCR generated fragments were then assembled back into the full length pSVIIIenv vector through Gibson assembly to generate a set of mutants (New England Biolabs, Ipswich, MA).

HIV Env clones, sCD4 and monoclonal antibodies

EMPIRIC libraries were cloned into pNL4.3 carrying the LN40 env gene. A version of LN40 env was used that was chimeric with LN40 gp120 and gp41 sequences derived from the B33, a brain env derived from the same subject as LN40. This chimeric Env is non-mac-tropic and carries determinants and properties of non-mac-tropism in gp120 as reported previously [177, 335, 336, 343]. Soluble CD4 was from Progenics Inc. Monoclonal antibodies (mabs) PG9, PG16 (V2q), b12 (CD4bs), 447-52D PGT145 (V2q) and VRC0 (V3 loop) and 2G12 (glycans) were from Polymun Scientific

Table 3.9: PCR primers used to introduce individual mutations into Env expression vectors.

Primer Set	Primer nomenclature	Primer
Set 1	ln40_upKpnF	GGGTCACAGTCTATTATGGG
	ln40_downKpnR	GTAAGTCATTGGTCTTAAAG
Set 2	LN8_373E_F	CCAGAAATTGTAgagCACAGTTTTAAT
	LN8_373E_R	ATTAAAACTGTGctcTACAATTTCTGG
	LN8_373N_F	CCAGAAATTGTAaatCACAGTTTTAAT
	LN8_373N_R	ATTAAAACTGTGattTACAATTTCTGG
	LN8_375H_F	ATTGTAATGCACcacTTTAATTGTGGA
	LN8_375H_R	TCCACAATTAAAgtgGTGCATTACAAT
	LN8_375W_F	ATTGTAATGCACtggTTTAATTGTGGA
	LN8_375W_R	TCCACAATTAAAccaGTGCATTACAAT
	LN8_377V_F	ATGCACAGTTTTgtcTGTGGAGGGGAA
	LN8_377V_R	TTCCCCTCCACAgacAAAACTGTGCAT
	LN8_380P_F	TTTAATTGTGGAccgGAATTTTTCTAC
	LN8_380P_R	GTAGAAAAATTCcggTCCACAATTAAA
	LN8_380A_F	TTTAATTGTGGAgctGAATTTTTCTAC
	LN8_380A_R	GTAGAAAAATTCagcTCCACAATTAAA

Inc. (Austria). (CD4bs) were prepared in house. Mab b6 (CD4bs) was provided by Dennis Burton (Scripps Research Institute), mab 17b (CD4i) by George Lewis (Institute of Human Virology) and mab PGT128 (V3, glycans) was provided by IAVI.

Antibody neutralization assays and IC50 determination

Env+ pSVIIIenv constructs carrying different single mutations were cotransfected into 293T cells with *env*-minus pNL43. Env+ pseudovirions were harvested after 48 hours, clarified by low speed centrifugation and frozen as aliquots at -152°C. Env+ pseudovirions were titrated on HeLa TZM-bl cells cells, which carry β-galactosidase and luciferase reporter genes controlled by HIV LTR promoters [356]. Infected cells were visualized 48 hours after infection as focus forming units (FFU) following staining for β-galactosidase activity. Since Env+ pseudovirions are only capable of a single round of replication, individual cells or small groups of divided cells were counted as foci.

Neutralization and inhibition assays were performed as described previously using 200 FFU of Env+ pseudovirus and evaluating residual infectivity on HeLa TZM-bl cells via a luminescence readout [168, 336].

Acknowledgements

We thank Dr. Cynthia Derdeyn (Emory University School of Medicine) for providing HIV-1 envelope clone for clade C Z1792M and Dr. George Lewis (Institute of Human Virology, University of Maryland School of Medicine) for mab, 17b. We also thank Ms. Briana Quitadamo for undertaking ELISA assays. This work was supported by NIH R01 grants NS084910, AI089334 and GM112844.

Chapter IV - A balance between inhibitor binding and substrate processing confers influenza drug resistance

This work has been published previously as *Jiang L, Liu P, Bank C, Renzette N, Prachanronarong K, Yilmaz LS, et al. A balance between inhibitor binding and substrate processing confers influenza drug resistance. J Mol Biol. 2016;428[1]:538-53.*

The work presented in this chapter was a collaborative effort. I performed the plasmid library construction, cloning of individual mutations, viral library recovery and bulk competition experiment, plaque assay, viral RNA isolation and reverse transcription, DNA sequencing library preparation and sequencing, sequence and statistical analysis, measurements of *in vitro* enzymatic activity. Ping Liu optimized viral library recovery and bulk competition experiment and plaque assay. Dr. Claudia Bank developed the statistical approach to estimate variation of fitness estimates. Dr. Nicholas Renzette optimized viral RNA isolation. Kristina Prachanronarong, Dr. Lutfu S. Yilmaz, Dr. Daniel R. Caffrey, Dr. Konstantin B. Zeldovich, Dr. Celia A. Schiffer, Dr. Timothy F. Kowalik, Dr. Jeffrey Dr. Jensen and Dr. Robert W. Finberg contributed to design of experiments. Dr. Jennifer P. Wang and Dr. Daniel N.A. Bolon supervised the work. I, Dr. Jennifer P. Wang and Dr. Daniel N.A. Bolon wrote the manuscript with inputs from other authors.

Abstract

The therapeutic benefits of the neuraminidase (NA) inhibitor oseltamivir are dampened by the emergence of drug resistance mutations in influenza A virus (IAV). To investigate the mechanistic features that underlie resistance, we developed an approach to quantify the effects of all possible single nucleotide substitutions introduced into important regions of NA. We determined the experimental fitness effects of 450 nucleotide mutations encoding positions both surrounding the active site and at more distant sites in an N1 strain of IAV in the presence and absence of oseltamivir. NA mutations previously known to confer oseltamivir resistance in N1 strains, including H275Y and N295S, were adaptive in the presence of drug, indicating that our experimental system captured salient features of real-world selection pressures acting on NA. We identified mutations, including several at position 223, that reduce the apparent affinity for oseltamivir *in vitro*. Position 223 of NA is located adjacent to a hydrophobic portion of oseltamivir that is chemically distinct from the substrate, making it a hotspot for substitutions that preferentially impact drug binding relative to substrate processing. Furthermore, two NA mutations, K221N and Y276F, each reduce susceptibility to oseltamivir by increasing NA activity without altering drug binding. These results indicate that competitive expansion of IAV in the face of drug pressure is mediated by a balance between inhibitor binding and substrate processing.

Introduction

Influenza A virus (IAV) causes a highly contagious acute respiratory illness responsible for significant morbidity and mortality in humans. IAV has two surface glycoproteins, hemagglutinin (HA) and neuraminidase (NA) that are used to distinguish subtypes. The most common IAV subtypes that infect humans are H1N1 and H3N2. H1N1 IAV has caused several influenza pandemics, including the 1918 Spanish flu and the 2009 swine flu [183, 193]. HA binds to sialic acid that is part of glycoproteins located on the surface of host cells and is critical for initial attachment and infection. NA cleaves sialic acid from host cell glycoproteins during the release of newly formed viral progeny, thus reducing viral affinity for previously infected cells [209]. The NA competitive inhibitor oseltamivir is widely used for treatment of influenza [239]. Oseltamivir is a successful example of structure-based drug design, in which electrostatic interactions have been optimized between the drug and the protein [357]. As NA is an enzyme, the active site is more conserved than the rest of the protein surface to preserve the necessary activity for viral release [358]. Yet antiviral resistance is a persistent problem with IAV; the use of oseltamivir to prevent morbidity and mortality has been disappointing due to widespread drug resistance [359-361]. Improved approaches to combat influenza infection and an increased understanding of drug resistance mechanisms are in great demand.

Clinical reports have shown the emergence of a handful of different oseltamivir-resistance mutations in H1N1 IAV following the clinical use of oseltamivir [361]. Thus far, mutations that have been associated with oseltamivir resistance occur at only a few

positions that neighbor the NA active site [202, 264, 362, 363]. The most prevalent resistance mutation in H1N1 IAV encodes the H275Y substitution (N1 numbering system used throughout), which spread globally in 2008 [360]. Most oseltamivir resistance mutations that have been studied cause defects in viral expansion in the absence of drug pressure [252, 265]. For example, the H275Y mutation that is commonly observed in H1N1 isolates with oseltamivir resistance, caused a reduced titer in the absence of drug in the WSN strain [262]. In the case of H275Y, secondary mutations including R222Q and V234M can restore fitness and the combined H275Y/R222Q/V234M genotype became predominant in circulating H1N1 in 2008 [364]. The analyses of individual IAV clones indicate that costs of adaptation can mediate the mutations that emerge in response to drug pressure. Because only a limited number of mutations have been studied, often in different strain backgrounds, the extent to which fitness costs mediate the emergence of drug resistance mutations in IAV is unclear.

To effectively probe drug resistance mechanisms, precise measurements of the effects of individual mutations are critical because small differences can distinguish mutants that will have different evolutionary outcomes. This remains a technical challenge despite exciting recent findings from gene-wide analyses of mutant effects in IAV [50, 71, 72, 75, 77, 365] and genome-wide analyses in poliovirus [366]. Gene-wide studies of mutations in IAV provide useful estimates of the average impact of mutations at each amino acid position and effectively delineate the strength of selection acting at each position in the NA, NS, NP and HA genes of IAV. However, the effects of specific

amino acid substitutions from gene-wide analyses of IAV [72, 77, 365] are only modestly reproducible (R^2 ranging from 0.34 to 0.62).

To measure fitness effects in IAV with high precision, we adapted the EMPIRIC (Exceedingly Meticulous and Parallel Investigation of Randomized Individual Codons) approach that we previously developed to investigate fitness landscapes in yeast [51, 53, 54, 56, 326, 340, 367]. The EMPIRIC approach, along with related approaches developed by others [47, 57, 58, 368], utilizes bulk competitions of engineered mutational libraries and next generation sequencing to estimate the frequency of each mutation before and after selection. A similar strategy used by Sun and colleagues was successful in identifying resistance mutations in the NS5A gene of hepatitis C virus [76]. Here, we adapted the EMPIRIC approach to systematically quantify the fitness effects of mutations in the NA gene of influenza A/WSN/33, which is an H1N1 strain. The results of this study provide improved precision relative to other high throughput studies of IAV mutants and enabled a robust assessment of drug resistance and fitness costs in the absence of drug pressure.

Our results indicate that a balance between mutant effects on binding to drug and processing of substrates mediates drug resistance mutations. In the absence of drug pressure, most mutations exhibited fitness defects. The presence of oseltamivir changed the fitness effects of many mutations including a handful that became adaptive. The strongest drug adaptive mutations (H275Y, N295S, and I223M) have previously been associated with drug resistance in clinical N1 isolates [255, 280, 360]. These drug adaptive mutations had similar fitness defects (30-33%) compared to the parental strain in

the absence of oseltamivir. In contrast, mutations associated with drug resistance in clinical N2 isolates caused severe fitness defects when introduced into the N1 strain used in our experiments (60-100% defects relative to the parental N1 strain in the absence of drug) that hinder them from being adaptive in N1. These observations suggest that fitness costs govern the resistance mutations that emerge in different IAV subtypes. We observed the vast majority of mutants at position 223 became adaptive to oseltamivir and had decreased drug binding. We also identified two drug adaptive mutations, K221N and Y276F that did not decrease drug binding, but increased the efficiency of substrate processing. These observations demonstrate that resistance to oseltamivir can occur by two distinct mechanisms: decreased binding to drug or increased efficiency of substrate processing.

Results

We analyzed the fitness effects of all single nucleotide mutations in five specific 30 base regions of NA, focusing on mutations encoding regions of NA immediately adjacent to the active site as well as a control region on the surface of NA far from the active site (Table 4.1 and Figure 4.1A). Regions around the active site were selected to include amino acid positions previously associated with oseltamivir resistance in human isolates, as well as positions that could encompass potentially novel mutations. Mutations were site-specifically engineered into plasmid-encoded viral genomes using a previously described plasmid system in which each influenza gene segment of the A/WSN/33 H1N1 strain is encoded on an individual plasmid [369]. Libraries of NA plasmids containing all possible single nucleotide mutations at 30 consecutive bases were generated. The NA plasmid libraries were combined with plasmids containing the remaining seven segments and transfected into co-cultured 293T and Madin-Darby canine kidney (MDCK) cells to recover an initial viral pool containing engineered mutations (P0) (Fig. 4.1B). This pool was used to infect MDCK cells in the absence or presence of oseltamivir, a competitive inhibitor of NA. P1 viruses were isolated from these infected MDCK cells.

Systematic approach to quantify the fitness effect of NA mutants with high precision

We analyzed a total of 50 amino acid positions in NA. The frequency of each mutation in the plasmid, P0, and P1 samples provided a direct estimate of the fitness effects of all 450 single nucleotide mutations in these regions without drug pressure

Figure 4.1: A high throughput approach was optimized to precisely analyze the fitness effects of NA mutations in influenza A virus

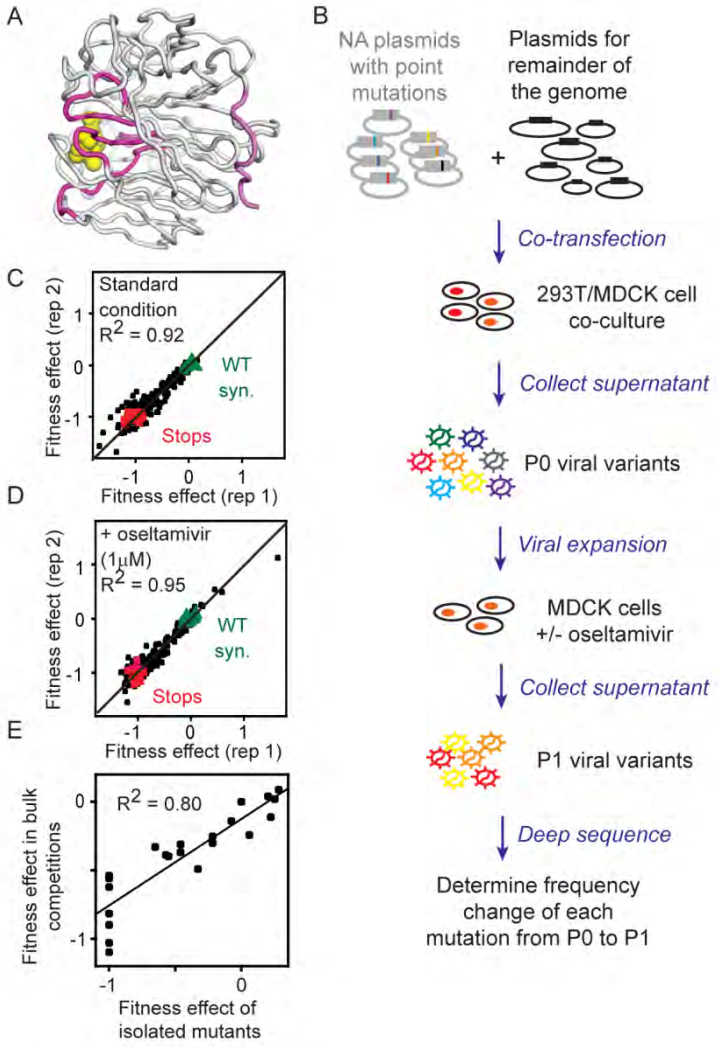


Figure 4.1 A high throughput approach was optimized to precisely analyze the fitness effects of NA mutations in influenza A virus. (A) Molecular image based on the **PDB ID: 3B7E** structure [197] of NA showing a monomer with the regions near the active site as well as a control region far from the active site that were chosen for mutational analysis highlighted in magenta. A competitive inhibitor bound in the active site of NA is shown as yellow spheres. (B) Site-directed mutagenesis was used to introduce point mutations into the plasmid encoding the NA gene from the influenza A/WSN/33 mixed with plasmids encoding wild-type versions of the other seven gene segments from this strain. Plasmids were co-transfected into 293T/MDCK cells to generate an initial P0 library of viral particles that were subsequently expanded through infection of MDCK cells to generate a P1 library. Focused deep sequencing was utilized to quantify the enrichment or depletion of each mutation during viral expansion. (C) Reproducibility of high throughput estimates of fitness effects from experimental replicates of viral expansion.

Silent substitutions (wild-type synonyms) in this plot are green and nonsense mutations (stop codons) are red. (D) Reproducibility of estimates of fitness effects in replicates of viral expansion with 1 μM oseltamivir. (E) Correlation between fitness effects estimated for a panel of individual mutations analyzed in isolation with estimates from bulk competitions.

Table 4.1: Amino acid regions of NA analyzed

Region	**Amino acid positions**
Active-site proximal 1	112-121
Active-site proximal 2	220-229
Active-site proximal 3	271-280
Active-site proximal 4	292-301
Surface loop distant from active site	83-92

(Table 4.2 and 4.3). Stop codons were universally depleted in P0 samples relative to plasmid. In contrast, synonymous mutations were relatively unchanged in frequency between plasmid, P0, and P1 pools (Figure 4.2). For mutations whose frequencies could be assessed in the plasmid, P0, and P1 pools, we observed consistent changes in frequency during viral recovery (plasmid to P0) and viral propagation (P0 to P1) (Figure 4.3). These observations indicate that similar selection occurred during P0 viral recovery and P1 viral propagation. To compare the effects of mutations in different experiments with potential region-specific bias, we normalized fitness effects, setting the average stop codon for each bulk competition to -1 to represent null NA function and average synonymous substitutions in each competition to 0 to represent WT-like fitness. We estimated fitness effects of mutations independently from changes in frequency from plasmid to P0, and from P0 to P1.

Fitness estimates for mutations based on P1 analyses in the absence and presence of drug are shown in Fig. 4.1C and Fig. 4.1D, respectively, and results were highly reproducible with $R^2 > 0.9$ for experimental replicates of viral expansion. Comparing bulk analyses with mutations analyzed in isolation is an important control that probes both stochastic noise and potential systematic differences between these independent approaches. We performed fitness analyses on a panel of 22 clones of individual mutations and observed strong correlation with fitness estimates from our bulk analyses (Figure 4.1E and Figure 4.4). These observations indicate that our bulk studies are accurate estimates of fitness effects of isolated clones.

Figure 4.2: Observed effects of nonsense mutations encoding stop codons and silent mutations

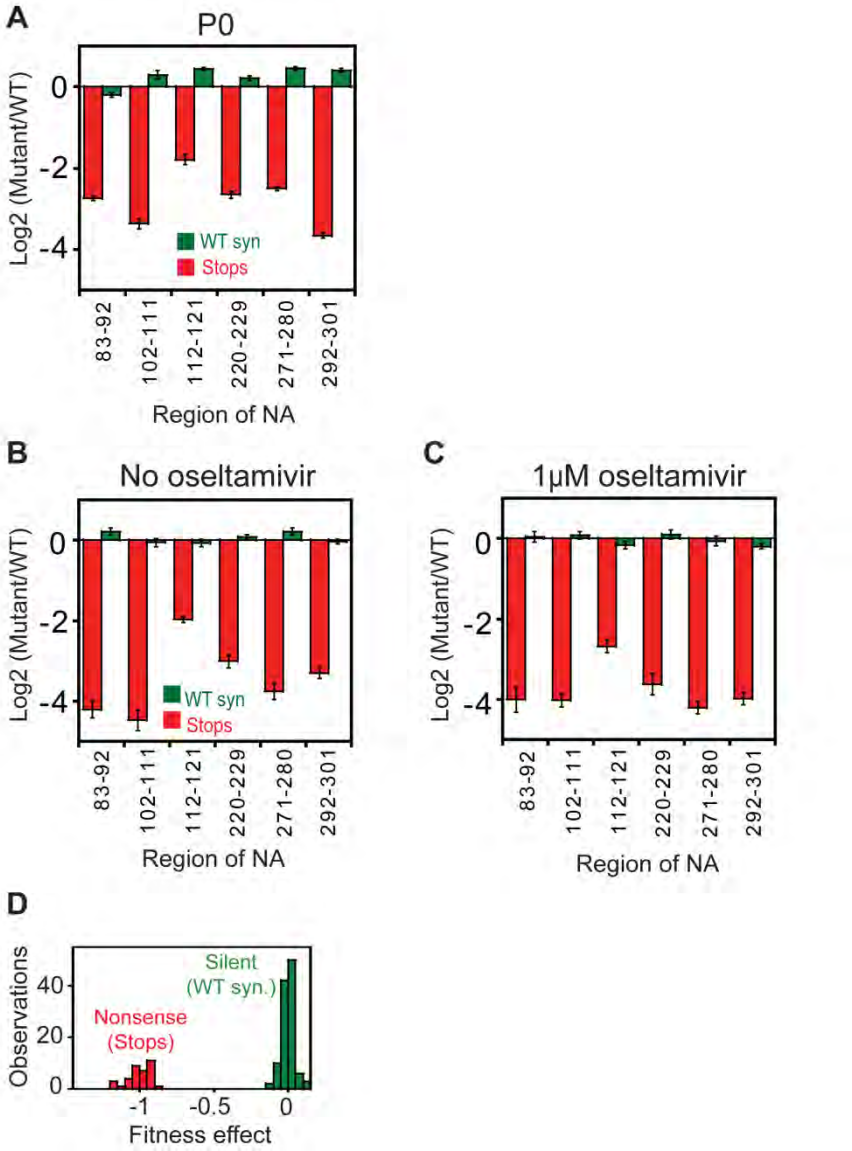


Figure 4.2 Observed effects of nonsense mutations encoding stop codons [38] and silent mutations encoding wild-type synonyms (green) in P0 library relative to plasmid library (A), without drug (B) and with 1 μM oseltamivir (C). Panels B and C depict the average and standard deviation for each class of mutations without normalization. Data in (B) and (C) are represented as mean ± SEM. (D) Distributions of observed effects for stop codons and wild-type synonyms across all regions analyzed with normalization within each region of stop codons to -1 and wild-type synonyms to 0.

Figure 4.3: Mutant frequency changes between plasmid and P0 correlate with changes between P0 and P1

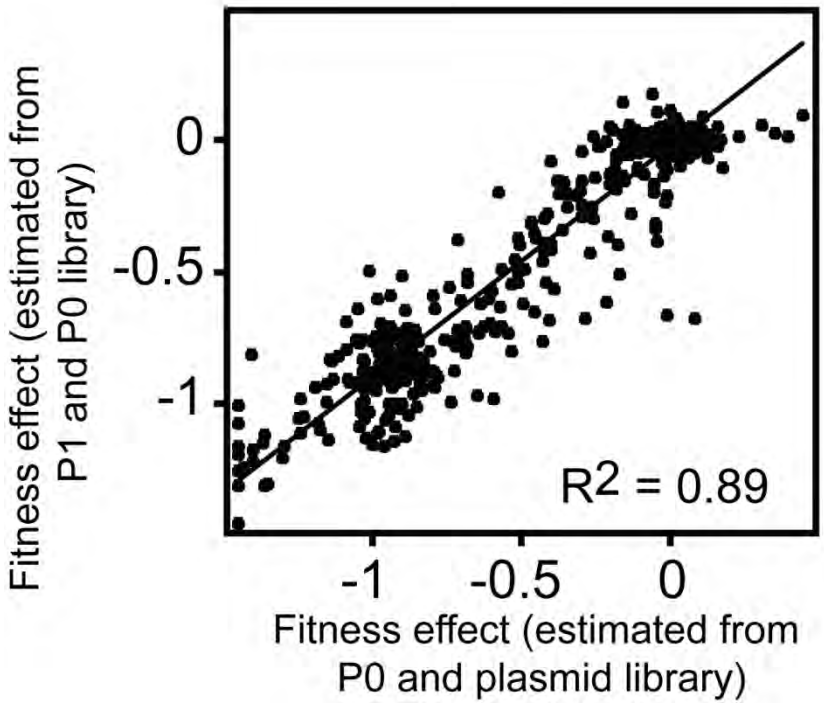


Figure 4.3 Mutant frequency changes between plasmid and P0 correlate with changes between P0 and P1. Fitness effects under standard conditions for all regions were determined based on enrichment or depletion estimates from sequencing. For each region, stop codons were normalized to s = -1, and wild-type synonyms to s = 0.

Figure 4.4: Studies of individual mutations analyzed in isolation

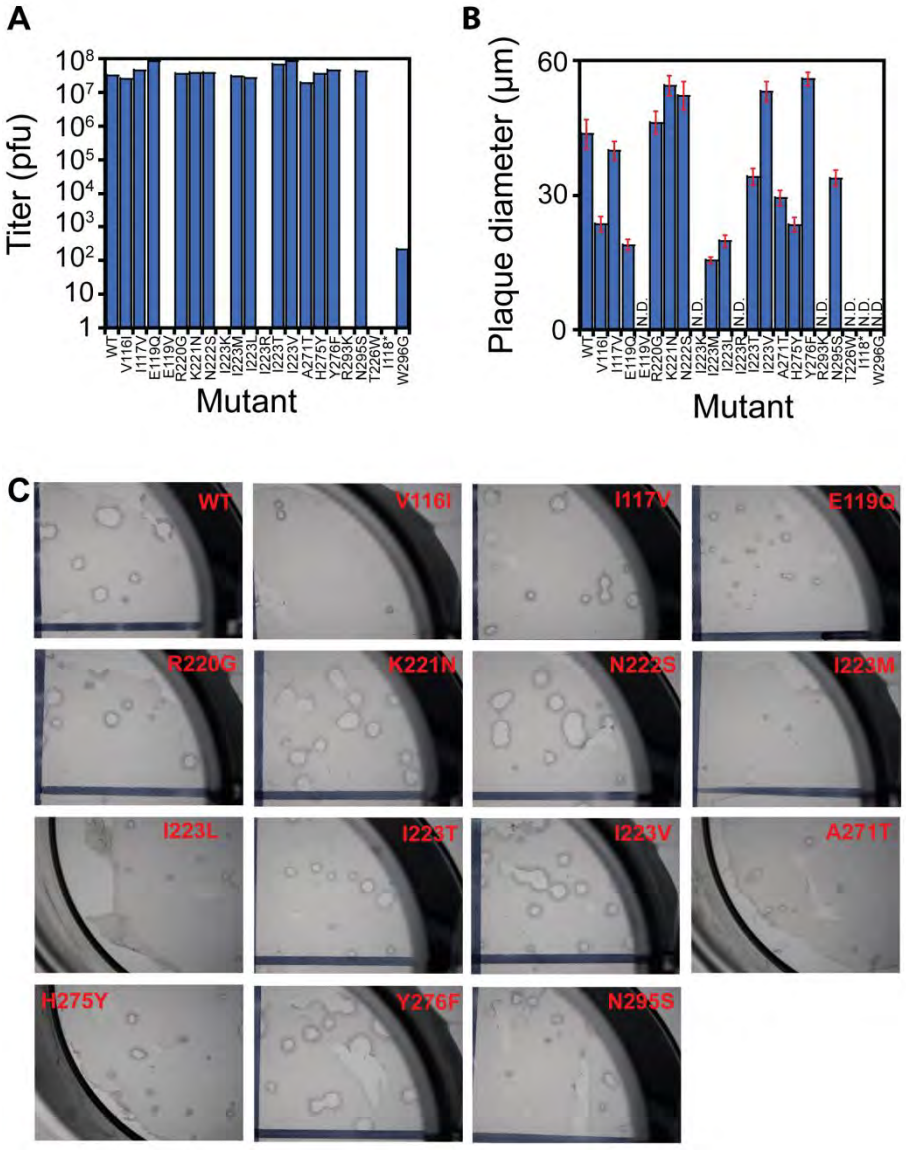


Figure 4.4 Studies of individual mutations analyzed in isolation. (A) Titers of individual mutants after P0 rescue determined by plaque assay. Error bars indicate the standard deviation of three independent titer determinations. (B) Average outer diameter of plaque size of individual mutants with titers greater than 10^4. Error bars indicate the standard deviation of measurements of 20 randomly chosen plaques. N.D. indicates radius of plaques not determined due to very low titer. Data in (A) and (B) are represented as mean ± SEM. (C) Images of plaques for all mutants with P0 titer greater than 10^4.

Fitness effects of mutations in NA without drug pressure

We examined how missense mutations that change the amino acid sequence impact fitness in the absence of drug pressure (Figure 4.5A). The vast majority of amino acid changes in regions of NA close to the active site caused severe fitness defects. In contrast, many amino acid changes in the control region located far from the active site were compatible with robust fitness (Figure 4.5B). These observations are consistent with the intuition that enzyme catalysis is sensitive to the physical composition of the amino acids surrounding the active site. To estimate the sensitivity of each amino acid position to mutation, we calculated the average fitness effect observed for missense mutations (Figure 4.6A). This metric of mutational sensitivity correlated ($R^2 = 0.46$) with solvent exposure (Figure 4.6B), indicating that solvent exposure contributes to the sensitivity of a position to mutation.

We examined the relevance of these experimental fitness analyses of the WSN strain under laboratory conditions to natural evolution. The WSN strain was chosen for these experiments because it can be efficiently recovered from plasmids, which helped facilitate the recovery of diverse viral libraries necessary for the competition experiments. WSN was cloned from a strain originally isolated in 1933 that had subsequently been passaged under laboratory conditions. The NA gene of WSN contains an unusual deletion in the stalk region that has been observed to partially impair viral expansion in chicken eggs, but not in mammalian cells [370]. The sequence of amino acids that comprise the active site of NA are similar to circulating viruses. To investigate how lab adaptation of WSN may influence the interpretation of our results, we compared the experimental

Figure 4.5: Fitness effects of mutants in WSN correspond to the frequency of amino acids observed in sequenced H1N1 isolates

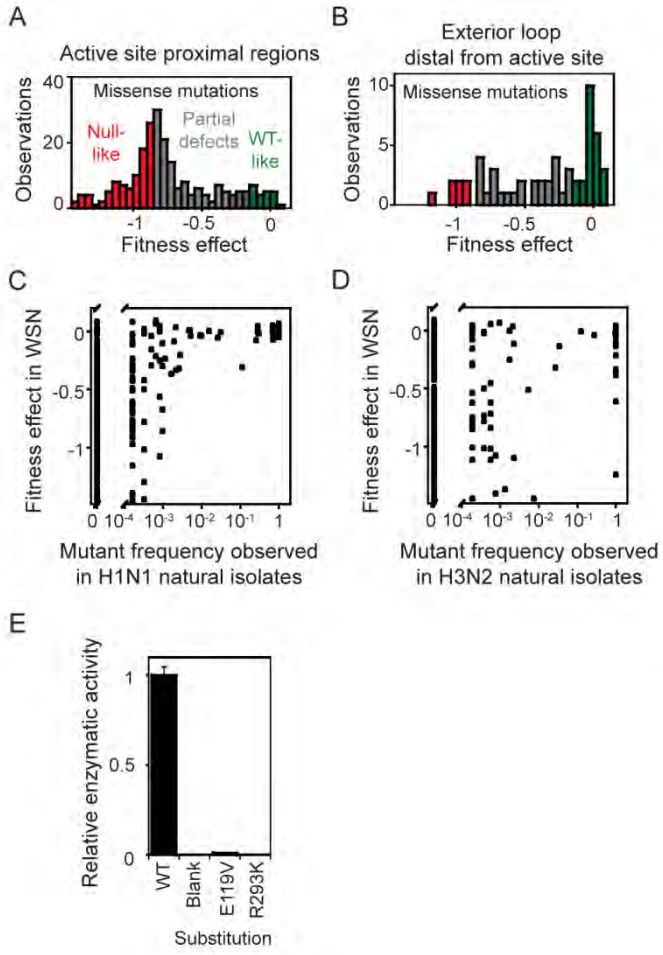


Figure 4.5. Fitness effects of mutants in WSN correspond to the frequency of amino acids observed in sequenced H1N1 isolates. The distribution of fitness effects observed for mutations in four regions proximal to the active site (A) and one region distant from the active site (B). Null-like and WT-like fitness effects are colored red and green and were based on observations of stop codons and WT synonyms. Fitness effects of amino acid changes were compared to the frequency of amino acids observed in 6205 sequenced H1N1 isolates (C) and 5279 sequenced H3N2 isolates (D) available from the Influenza Research Database [371]. (E) Relative MUNANA activity of NA variants of WSN expressed in 293T cells monitored using a fluorescent substrate. Cells lacking the NA gene were included as a blank control. Error bars indicate the standard deviation of 3 biological replicates.

Figure 4.6: Sensitivity to missense mutations of each amino acid position

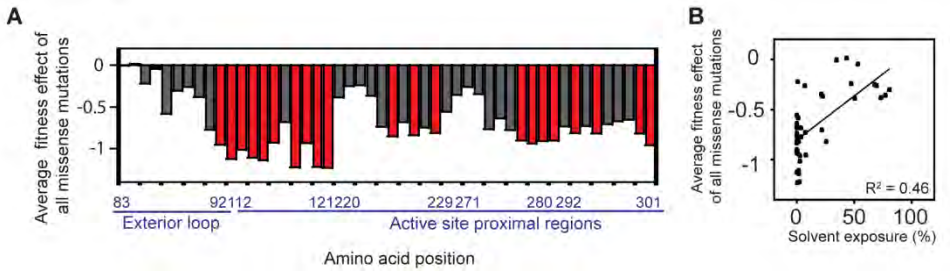


Figure 4.6 Sensitivity to missense mutations of each amino acid position. (A) The sensitivity to mutation of each amino acid position. In graphs, bars are colored red if the average fitness effect of a missense mutation was null-like, green if it was wild-type-like, and grey for all others. (B) Correlation between average experimental fitness effect and fraction of solvent accessible surface area of each amino acid position.

fitness effects of amino acid changes that we observed in WSN with the frequency of the same amino acids in 6,205 sequenced H1N1 isolates available from the influenza research database [371] (Figure 4.5C).

The amino acids most frequently sampled in the sequenced H1N1 isolates, including the most frequent amino acid at each position that we experimentally analyzed, all supported robust viral expansion when generated in the WSN genetic background (Figure 4.5C). In contrast, the vast majority of mutations that caused severe experimental fitness defects in WSN were either never or rarely observed in sequenced H1N1 isolates. Multiple factors could contribute to the observation that some mutations with severe fitness defects in WSN were rarely observed in sequenced isolates including a strong dependence on genetic background for these mutations as well as potential rare errors in the sequencing process of these isolates. Overall, the correspondence we observe between WSN fitness effects and the frequency of amino acids observed in sequenced H1N1 isolates suggests that our analyses of NA mutations in WSN capture many salient features of natural selection acting on the NA protein of circulating H1N1 IAV.

To investigate the extent to which our fitness experiments in WSN could extend to other IAV strains, we compared them to the frequency of amino acids observed in 5,279 sequenced H3N2 isolates available from the influenza research database [371] (Figure 4.5D). Many of the most common amino acids identified in H3N2 sequenced isolates were moderately to strongly deleterious when generated in WSN. This observation indicates that strongly epistatic or context-dependent amino acid substitutions have accrued during the divergence of the N1 and N2 proteins.

N1 and N2 subtype specific mutant effects

To further examine how divergence of N1 and N2 proteins contribute to fitness effects of mutants in different genetic backgrounds, we analyzed the mutations E119V and R293K, which have been specifically associated with oseltamivir resistance in N2 isolates. Both E119V and R293K had strong fitness defects when generated in the N1 of WSN: E119V was lethal and R293K exhibited a 62% defect relative to the parental WSN strain without drug pressure. We attempted to recover viral stocks from individually cloned plasmids encoding E119V and R293K in WSN, but neither yielded measureable titers (data not shown). The failure to recover viral stocks of these variants is consistent with the severe fitness defects observed for these mutations in our bulk competitions. A previous study demonstrates that oseltamivir binding is decreased by the E119V mutation in the WSN strain and showed that E119V dramatically reduces the susceptibility of WSN NA to oseltamivir inhibition compared to both the parental wild type NA as well as the H275Y variant that is most commonly associated with oseltamivir resistance in the N1 strain[263]. In the same study, the enzymatic activity of the R293K mutation was too low to be accurately measured by the NA inhibition experiment [263]. When expressed in 293T cells, both the E119V and R293K variants of WSN NA had enzyme activity in the absence of oseltamivir that was less than 5% that observed for wild type NA in the absence of oseltamivir (Figure 4.5E). These observations suggest that severe fitness costs in the absence of oseltamivir prevent E119V and R293K from contributing to drug adaptation in WSN. In future research beyond the scope of this work, it will be of interest

to investigate the process by which N1 and N2 proteins have accumulated context-dependent amino acid substitutions.

Fitness effects of mutations in NA with oseltamivir

To examine the impact of drug pressure on NA mutations, we compared fitness effects observed with and without oseltamivir (Figure 4.7A). To provide a sensitive readout, we utilized a concentration of drug (1 μM) that partially reduced expansion of the parental WSN strain. Using a Student's *t*-test with multiple test correction, we identified 24 drug responsive mutations whose fitness effect was significantly increased in the presence of this concentration of oseltamivir (Table 4.4**;** plotted in orange in Figure 4.7A). Most of these drug responsive mutations exhibited fitness defects without drug, indicating that they reduce substrate processing as well as drug binding. The fitness defects of many of the drug responsive mutations hindered their ability to outcompete the parental strain under the conditions of these experiments. Using a similar statistical approach, we determined that only five mutations (K221N, I223M, H275Y, Y276F, and N295S, N1 numbering system) were significantly more fit than the parental strain in the presence of 1 μM oseltamivir, and we refer to these mutations as drug adaptive.

Three of these five drug adaptive mutations from our experiments (H275Y, N295S, and I223M) have previously been associated with oseltamivir resistance (Table 4.5) [280, 361, 372]. In our experiments, H275Y was the strongest drug resistance mutation, with greater than two-fold increased fitness in the presence of oseltamivir compared to

Table 4.5 Experimental fitness effects of previously reported oseltamivir resistance mutants that were covered in the libraries analyzed in the study [255, 274, 277, 279-281, 283, 361, 372, 373]

Mutation	Subtype	No oseltamivir	1 μM oseltamivir	Increase in fitness effect
I117V	H5N1	-0.14	0.12	0.26
E119V	H3N2, H7N9	-1.10	-1.00	0.10
I223K	H1N1, H7N9	-0.54	-0.15	0.39
I223R	H1N1, H7N9	-0.56	-0.13	0.43
I223M	H5N1	-0.33	0.23	0.56
I223T	H1N1, H5N1	-0.25	-0.01	0.24
I223L	H5N1, H3N2	-0.40	-0.02	0.38
H275Y	H1N1, H5N1	-0.31	1.12	1.43
R293K	H3N2, H7N9	-0.62	-0.38	0.24
N295S	H1N1, H5N1, H3N2	-0.30	0.49	0.79

wild-type (Figure 4.7A). H275Y is the predominant oseltamivir resistance mutant that has appeared in seasonal and pandemic H1N1 influenza virus [361]. The N295S mutation, which exhibited a 50% fitness improvement in the presence of oseltamivir compared to wild-type in our experiments, has also been associated with oseltamivir resistance in different serotypes of IAV, although less frequently than H275Y [255, 274, 372]. The I223M mutation has been associated with oseltamivir resistance in H5N1 IAV [280]. These three drug adaptive mutations (H275Y, N295S, and I223M) all had similar fitness defects ranging from 30-33% relative to the parental strain in the absence of drug. These observations indicate that the influenza virus can tolerate large fitness costs in the face of a novel selection pressure. In principle, fitness defects of this magnitude impose subsequent selection for compensatory mutations that can restore substrate processing and fitness. This scenario has been observed with influenza viruses carrying the H275Y mutation that accumulated the compensatory R222Q/V234M mutations and became the dominant circulating strain in 2008 [364].

To examine if mutations with more extreme fitness costs could outcompete the parental strain at different drug concentrations, we measured the fitness effects of mutations at positions 292-301 under a range of oseltamivir concentrations (Table 4.6). This region includes two drug responsive mutations with different fitness costs relative to the WSN virus without oseltamivir: N295S (30% fitness cost) and R293K (70% fitness cost). These mutations both exhibited fitness effects that increased with oseltamivir concentration (Figure 4.7B), supporting the conclusion that these mutations reduce susceptibility to oseltamivir. While we observed fitness effects responding to oseltamivir

Figure 4.7: Oseltamivir adaptive mutations were identified by comparing fitness effects of NA mutations in the presence or absence of oseltamivir

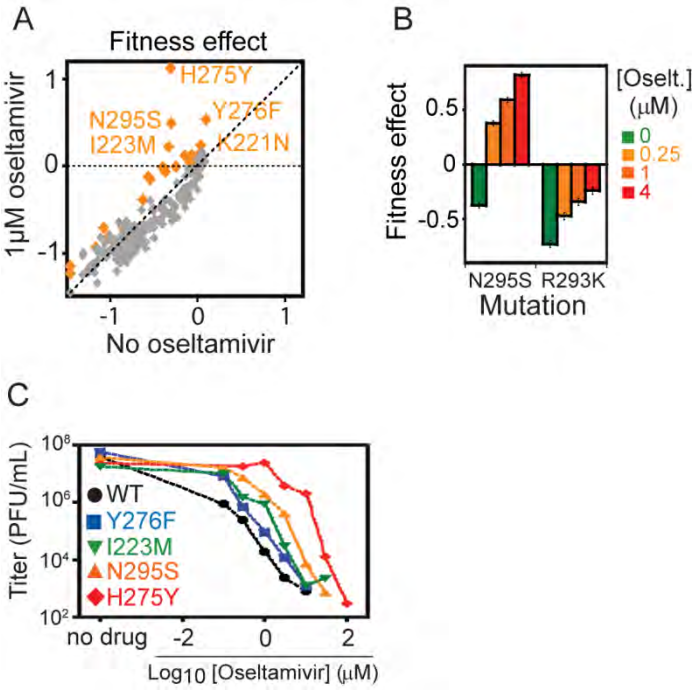


Figure 4.7 Oseltamivir adaptive mutations were identified by comparing fitness effects of NA mutations in the presence or absence of oseltamivir. (A) Comparison of the fitness effects of NA mutations with and without oseltamivir. Mutations with a statistically significant (detail in Materials and Methods) increase in fitness effect in the presence of oseltamivir compared to without drug (above and to the left of the diagonal) are colored orange. Mutations that were significantly more fit than WT in the presence of drug (above the horizontal dashed line) are labeled. (B) Fitness effects of two mutations associated with oseltamivir resistance in H1N1 (N295S) or H3N2 (R293K) from bulk competitions performed over a range of oseltamivir concentrations. The mutations exhibited fitness effects that increased with escalating amounts of oseltamivir. (C) Expansion of individually isolated mutations as a function of oseltamivir concentration.

concentration, the changes in fitness effects were small (2-fold) relative to the changes in oseltamivir concentration (16-fold). Over this range of oseltamivir concentrations, the R293K mutation is always less fit than the parental strain. These observations suggest that the strong fitness cost of the R293K mutation hinders its ability to outgrow the parental strain and cause drug resistance.

We investigated how the fitness effects of mutations based on bulk competitions in the presence and absence of oseltamivir compared with analyses of individual clones of five variants (Figure 4.7C). This comparison is complicated because different references are utilized in each experiment: in bulk competitions, the fitness of mutations are determined directly relative to WT, and in the analyses of individual clones the effect of drug is relative to the expansion rate of each viral variant in the absence of drug. The response of an isolated clone to oseltamivir when not in competition with other viruses should be a function of fitness effects measured in bulk competitions both with oseltamivir and in the absence of drug. Consistent with this principle, the order of the sensitivity of four isolated clones to oseltamivir (Figure 4.7C) was the same as the order of mutations based on the difference between fitness effects with and without oseltamivir (Y276F < I223M < N295S < H275Y).

Position 223 is a hotspot for mutations that decrease binding to oseltamivir

Hotspots have been associated with drug resistance in many systems [374, 375]. To scan for potential hotspots that impact oseltamivir binding relative to substrate processing, we calculated the average drug responsiveness (defined as change in average

fitness effect of mutations with oseltamivir compared to without drug) at each analyzed position (Fig. 4.8A). Mutations at position 223 and 275 exhibited the strongest averaged oseltamivir responsiveness. Analyzing the individual mutations at these positions indicated that the responsiveness of position 275 was almost completely due to one mutation (H275Y) of very large effect (Fig. 4.8B). In contrast, the responsiveness of position 223 was due to multiple mutations with intermediate drug responsive effects.

To investigate the drug responsive effects of mutations at position 223 in further detail, we generated individual clones and examined their responsiveness to oseltamivir *in vitro* (Figure 4.8C). We observed that six different mutations at position 223 (Arg, Lys, Met, Leu, Thr, and Val) were less sensitive to oseltamivir than WT. All of the mutations at position 223 that we analyzed were more sensitive to oseltamivir than the H275Y mutation. Together, our results indicate that a very specific physical change at position 275 is required to disrupt drug binding. This is consistent with structural analyses of H275Y, which indicate that subtle conformational rearrangements mediated by the side chain of E277 that is located between position 275 and oseltamivir are responsible for disrupting drug binding [202]. While position 275 does not directly contact oseltamivir or substrate, position 223 is located at the substrate binding site and can directly contact oseltamivir (Figure 4.8D). Position 223 contacts hydrophobic atoms in oseltamivir that are outside the substrate envelope and that differ in physical properties from the polar substrate. Distinctions between substrate and inhibitors have been associated with hotspots for drug resistance in HIV protease [376]. The differences between oseltamivir and substrate where these contact position 223 likely cause this position to be a hotspot

Figure 4.8: Multiple mutations at position 223 reduced the apparent binding affinity of NA to oseltamivir

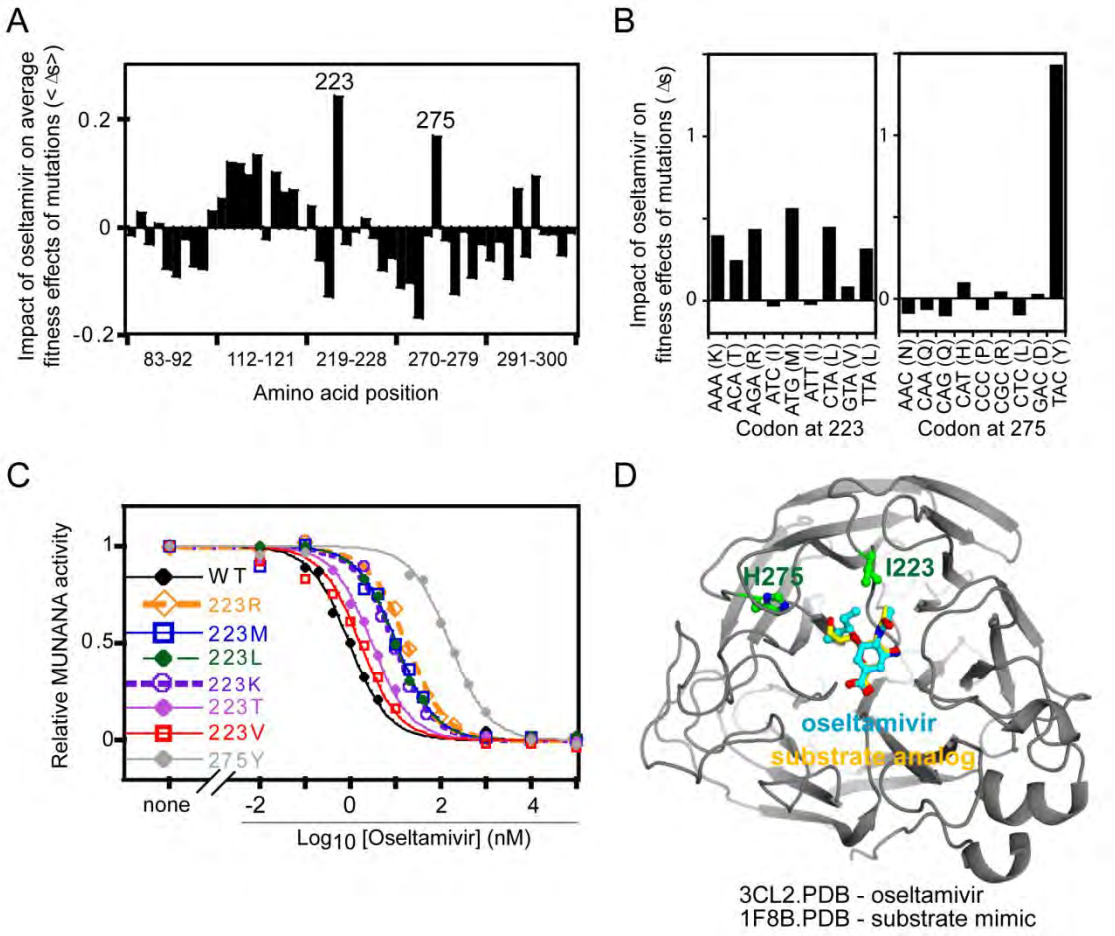


Figure 4.8 Multiple mutations at position 223 reduced the apparent binding affinity of NA to oseltamivir. (A) To identify hotspots for mutations with adaptive potential to oseltamivir, we examined the average effect of mutations at each position on responsiveness to oseltamivir in the bulk competitions. (B) Oseltamivir responsiveness of single nucleotide mutations at positions 223 and 275. (C) Enzyme activity of individually isolated viral variants as a function of oseltamivir concentration. The activity of I223K and I223R was estimated in 293T cells while the activity of other mutants and WT was estimated using virus. (D) Structural image of an NA monomer indicating the location of positions 223 and 275 relative to oseltamivir and a substrate analog. The image was generated from **PDB ID: 3CL2** [202] and **PDB ID: 1F8B** [377].

for mutations that disrupt drug binding in NA. Of note, the I223K, I223R, and I223T mutations have been observed in 2009 H1N1 isolates with reduced sensitivity to oseltamivir [279, 281], although other mutations at position 223 have not yet been associated with adaptation to oseltamivir in the H1N1 subtype to the best of our knowledge. Structural analyses of neuraminidase with I223R demonstrate that this mutation physically disrupts binding to oseltamivir [277].

Drug adaptive mechanism of Y276F and K221N

The Y276F and K221N mutations had similar or slightly improved fitness compared to the parental strain in the absence of drug pressure (Figure 4.9A). In bulk competition experiments, K221N exhibited smaller differences in fitness effects with and without drug compared to Y276F. The small fitness differences observed for K221N in bulk competitions would be difficult to discern with experiments on an individual clone. For these reasons, conclusions regarding K221N must be considered cautiously. Y276F exhibited a larger fitness increase and in isolation we observed that it did require elevated concentrations of oseltamivir to slow expansion relative to WT (Figure 4.7C). Taken together, our observations indicate that Y276F improves the fitness of WSN both in the presence and absence of oseltamivir. To the best of our knowledge, the Y276F mutation has not previously been associated with oseltamivir resistance.

To investigate potential drug adaptive mechanisms of K221N and Y276F, we closely examined structural and biochemical properties of each mutation. The Cα-Cβ bond of amino acids at positions 221 and 276 are oriented away from the active site

Figure 4.9: K221N and Y276F are adaptive in the presence of oseltamivir because they increase NA activity without altering drug binding

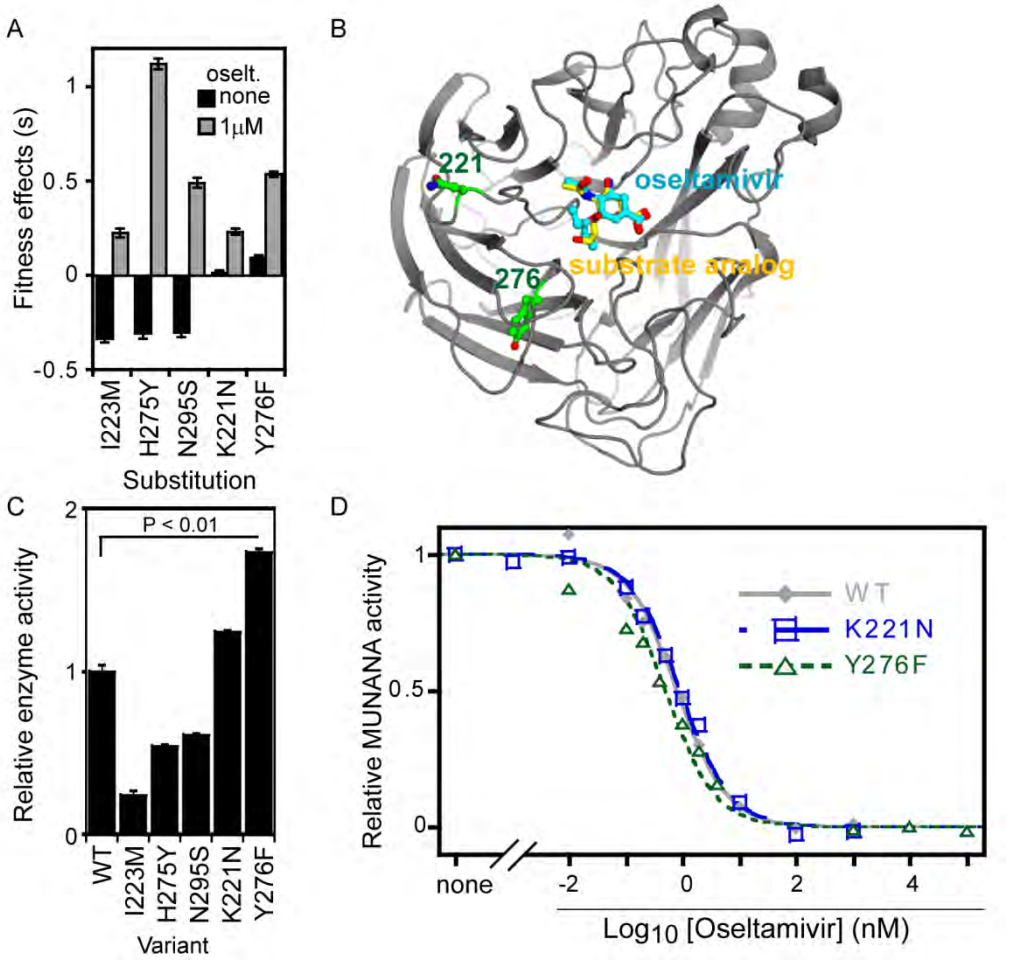


Figure 4.9 K221N and Y276F are adaptive in the presence of oseltamivir because they increase NA activity without altering drug binding. (A) Fitness effects of mutations that were statistically more fit than WT during bulk competitions with oseltamivir. Zero represents WT fitness, positive values represent increased fitness relative to WT, and negative values represent decreased fitness relative to WT. Error bar indicates standard error of the mean with N =3. (B) Structural image of an NA monomer indicating the location of positions 221 and 276 relative to oseltamivir (blue) and a synthetic substrate used in enzymatic assays (yellow). The image was generated from **PDB ID: 3CL2**[202] and **PDB ID: 1F8B** [377]. (C) Relative activity of isolates of individual viral variants determined using the fluorescent substrate MUNANA. Enzyme activities were normalized to viral titer to estimate the relative enzyme activity per infectious unit. Error bars indicate the standard deviation of 5 biological replicates. The standard deviations are as follows: WT: 0.0499, Blank: 0.0002, E119V: 0.0004, R292K: 0.0001 (D) Enzyme activity of individually isolated viral variants as a function of oseltamivir concentration. The experimentally determined IC_{50}'s are as follows: WT: 0.89 nM, K221N: 1.04 nM, and Y276F: 0.55 nM.

(Figure 4.9B) such that the side chains are unlikely to directly contact either the inhibitor oseltamivir or the sialic acid moiety of the substrate. This observation suggests that the fitness effects of the K221N and Y276F mutations are likely due to subtle alterations to the conformation or dynamics of nearby positions that do contact substrate and inhibitor.

We analyzed the enzymatic activity of K221N and Y276F as well as the other three identified drug adaptive mutations (I223M, H275Y, and N295S) *in vitro* using a fluorescent substrate (Fig. 4.9C). The three drug-adaptive mutations with fitness costs in the absence of drug (I223M, H275Y, and N295S) all had reduced enzymatic activity relative to WT. These results suggest that reduced substrate turnover caused by these mutations is responsible for decreased fitness in the absence of drug pressure. In contrast, K221N and Y276F both exhibited increased enzymatic activity relative to WT. In principle, mutations that increase enzymatic activity without impacting drug binding can provide an adaptive advantage in the face of drug pressure. For example, a mutation that increases the efficiency of substrate processing two-fold without impacting drug binding will increase WT enzymatic activity in the presence of drug concentrations that reduce the WT enzyme efficiency 2-fold. To examine if this mechanism is relevant to the K221N and Y276F mutations, we analyzed the effects of these mutations on inhibition by oseltamivir *in vitro* (Figure 4.9D). Full experimental titration experiments with WT indicate that the results of this inhibition assay are highly reproducible. Both K221N and Y276F had inhibition profiles and 50% enzyme-inhibitory concentration (IC_{50}) values that were similar or slightly more sensitive to oseltamivir than WT (Figure 4.9D). These

results indicate that the adaptive advantage of K221N and Y276F are due to increased efficiency of substrate processing rather than decreased binding to drug.

Discussion

Many pathogens, including IAV, accumulate mutations that make them resistant to currently available drugs. While some mutations that cause influenza to become resistant to oseltamivir have been identified, the impact of most mutations has not been fully resolved. In particular, the effect of most mutations in the absence of drug pressure has not been experimentally characterized. Here, we systematically analyzed all possible single nucleotide mutations in regions of the active site of the viral NA gene using the EMPIRIC method. Our approach provides rapid and highly reproducible fitness measurements of IAV mutants. We comprehensively examined fitness effects of all possible single nucleotide mutations in defined regions of NA in the absence and presence of oseltamivir, identifying mutations that confer resistance to oseltamivir and validating mutations with individual clones.

Our results support the conclusion that a balance between substrate processing and drug binding determines the potential of an NA mutation to adapt to oseltamivir (Figure 4.10A). Mutations with either decreased drug binding (Figure 4.10B, purple line) or increased substrate processing (Figure 4.10B, green line) should expand more efficiently compared to WT in the presence of drug (Figure 4.10B, black line). In principle, mutations that decrease drug binding without causing a defect in substrate processing provide the ability for viruses to expand most efficiently over the broadest

Figure 4.10: Mechanisms of adaptation to drug pressure

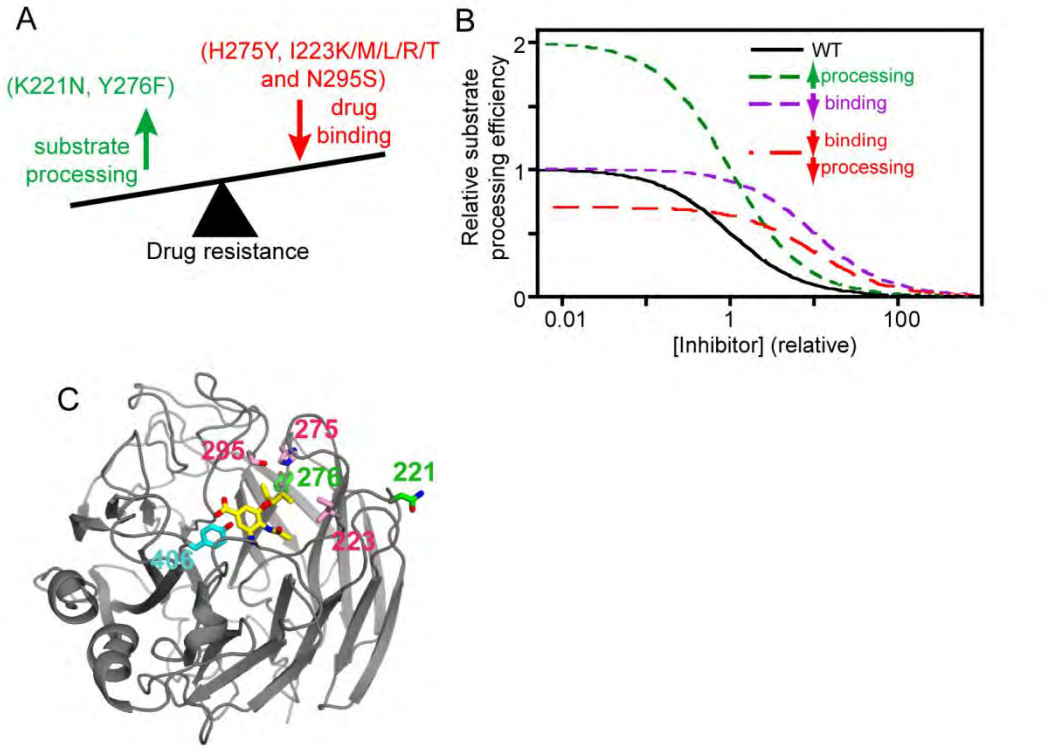


Figure 4.10 Mechanisms of adaptation to drug pressure. (A) Drug resistance is a balance between the effects of mutations on drug binding and substrate processing. Mutations that increase substrate processing or reduce drug binding favor drug resistance. (B) Theoretical model of the efficiency of substrate processing as a function of inhibitor concentration for mutations with different impacts. (C) Structural representation based on 3CL2.PDB of a NA subunit illustrating the location of mutations that either increased substrate processing (colored pink), or decrease binding to oseltamivir (colored green). Oseltamivir is shown in yellow and the catalytic tyrosine at position 406 is shown in cyan.

concentrations of drug (Figure 4.10B). Multiple lines of evidence indicate that this type of mutation is extremely rare: none were identified in our mutational scan, and drug resistance mutations identified in clinical isolates exhibit fitness costs without drug [252, 265]. Instead, most mutations associated with oseltamivir resistance in clinical isolates had clear experimental fitness defects (~30%) in our experiments (Figure 4.10B, red line). The fitness defects of these drug-resistant mutations should impose selection for compensatory mutations that increase the efficiency of substrate processing, which has been observed both experimentally [378] and in the majority of H1N1 clinical isolates from 2008 [364].

Our results indicate that NA mutations can also provide an adaptive advantage in the presence of oseltamivir by increasing the efficiency of substrate processing without decreasing sensitivity to drug. The Y276F mutation that exhibited the largest increase in substrate processing efficiency and fitness effect with oseltamivir has been observed at low frequency (0.1%, only observed in 2009-2010) in sequenced H1N1 isolates. The low frequency of Y276F in sequenced H1N1 isolates suggests that the effects of Y276F may be distinct in different N1 sequence backgrounds. Nonetheless, our observations of the effects of Y276F in WSN highlight the general mechanistic conclusion that increased efficiency of substrate processing by NA without reduced sensitivity to oseltamivir binding can lead to drug adaptation. While this type of mutation does not provide resistance to high levels of oseltamivir, it may act as an enabling mutation that provides greater sampling of secondary mutations that could lead to strong drug resistance. For example, a mutation that increases the efficiency of substrate processing may offset the

fitness costs of a mutation that reduces binding to drug and processing of substrate such that the double mutant disrupts binding to drug while maintaining efficient substrate processing.

Our results, as well as other studies [364, 379], indicate that multiple mutations in NA are required to reduce oseltamivir binding while maintaining efficient substrate processing. The evolution of drug resistance in this scenario will be a complex combination of mutational probabilities as well as the effects of individual and combined mutations on binding to drug and processing of substrate. Additional studies of combinations of mutations will be critical to appreciate the full spectrum of potential oseltamivir resistance NA variants. This should also serve as a caution that widespread use of oseltamivir at sub-neutralizing concentrations will likely lead to an increased frequency of multiple NA mutations with adaptive benefits.

Materials and Methods

Construction of plasmid-encoded libraries

NA point mutant libraries were generated using a previously described approach [326, 380]. Plasmids encoding the parental NA gene as well as the other seven gene segments encoding the H1N1 A/WSN/33 strain in the pHW2000 vector were kindly provided by R. Webster (St. Jude Children's Research Hospital, Memphis, TN). The NA gene was cloned into pRNDM to generate a plasmid without any BsaI restriction sites. Inverted BsaI sites were then introduced to enable a cassette ligation strategy to efficiently generate libraries of single nucleotide mutations; a separate cassette was used to mutate each base to all three non-parental bases. Libraries of single nucleotide mutants at 30 consecutive bases were combined and blended with parental (wild-type) plasmid as well as a panel of three stop codons at about 4-fold elevated frequency as negative controls for NA function. The resulting pool of NA libraries was transferred from pRNDM to pHW2000 [369] using sequence and ligation independent cloning (SLIC) [325, 380]. The pHW2000 construct contains a CMV promoter to drive expression of mRNA and a polII promoter in the opposite orientation to generate genomic negative strand RNA.

Cell culture

293T and MDCK cell lines were obtained from the American Type Culture Collection (Manassas, VA). The 293T cell line was maintained in 293T cell culture media consisting of Opti-MEM I reduced serum media (Gibco, Grand Island, NY)

supplemented with 5% fetal bovine serum (Hyclone, Logan, UT), 100 U/mL penicillin, and 100 μg/mL of streptomycin at 37°C and 5% carbon dioxide. The MDCK cell line was maintained in Eagle's minimal essential medium (MEM) supplemented with 10% fetal bovine serum, 2 mM L-glutamine, 10 mM sodium pyruvate, 1X non-essential amino acid, 100 U/mL penicillin, and 100 μg/mL of streptomycin at 37°C and 5% carbon dioxide. All cell culture reagents were from Corning (Manassas, VA) unless otherwise indicated.

Viral library recovery and selection experiments

Viral libraries were recovered from plasmids as previously described [369]. Briefly, equal numbers of 293T and MDCK cells were mixed in 293T cell culture media, and seeded in 6-well plates at a density of 2–6 × 10^5 cells/well. 293T-MDCK co-cultures were transfected with 1 μg of NA plasmid library and 1 μg of each plasmid encoding the other seven gene segments (8 μg total plasmid) using TransIT-LT1 Reagent (Mirus, Madison, WI). The ratio of DNA (μg) to TransIT-LT1 (μL) was 1:2. At 6 hours post-transfection, cell growth media was replaced with fresh Opti-MEM I reduced serum media. At 30 hours post-transfection, TPCK-trypsin (Sigma-Aldrich, St. Louis, MO) was added to cell growth media to a final concentration of 0.5 μg/mL. At 72 hours post-transfection, supernatant containing viral particles was harvested and centrifuged at 300x*g* for 15 minutes to remove cell debris. Supernatants were stored at -80℃ . These recovered pools of viral variants are referred to as P0 libraries. Plaque assays were

performed to determine the titer (plaque forming units or PFU/mL) of each P0 sample as previously described [68].

MDCK cells were used for additional viral competition experiments. 10^6 MDCK cells were seeded in individual wells of a 6-well plate and grown for one day. Cells were washed twice with 1X PBS and once with cDMEM/BSA (DMEM, 100 U/mL penicillin, 100 μg/mL streptomycin, and 7.5% BSA) before infection. P0 viral libraries were diluted in influenza virus growth medium (IVGM: cDMEM/BSA with 1 μg/mL TPCK trypsin). MDCK cells were infected at a multiplicity of infection [178] of 0.001. For each experimental dataset, three independent infections were conducted in the presence or absence of 1 μM oseltamivir. Replicate datasets were obtained by performing three additional independent P1 infections. Oseltamivir carboxylate was a kind gift from Hoffmann-La Roche Pharmaceuticals (Basel, Switzerland). Viral binding was conducted at 37°C with 10% carbon dioxide for one hour, followed by two washes with PBS. After washing, 2 mL of fresh IVGM was added to each well of MDCK cells, which were maintained at 37°C with 10% carbon dioxide. The supernatant containing viruses was collected when 50%-90% CPE was observed or at 120 hours post-infection. Supernatants were centrifuged at 300x*g* for 15 minutes and stored at -80°C. Samples recovered from MDCK expanded viral pools were referred to as P1.

Analyses of individual mutants

A panel of individual NA variants were cloned into plasmids and analyzed in isolation. Viral samples were recovered from 293T-MDCK cells as described for library

samples. Titers of P0 samples were determined for each variant using plaque assays. For variants that produced P0 samples with titers >10,000 PFU/mL, we also analyzed plaque size. Plaque size was measured using a Nikon SMZ1500 microscope for 20 randomly selected and well separated plaques for each variant analyzed (Figure 4.4).

For a subset of variants, infectivity was analyzed as a function of oseltamivir concentration as previously described [381]. Briefly, confluent MDCK cells in 24-well plates were infected at an MOI of 0.01. The infection was conducted in a range of oseltamivir concentrations (0, 0.1, 0.3, 1, 3, 10, 30 and 100 μM), and incubated at 37°C with 10% carbon dioxide. Supernatants were collected 3 days post-infection and virus titer was determined by plaque assay.

Analyses of enzyme activity *in vitro*

Enzyme activity of NA was determined using fluorogenic 2'-(4-methylumbelliferyl)-alpha-D-N-acetylneuraminic acid (MUNANA) substrate according to the manufacturer's instructions (Life Technologies, Carlsbad, CA). Briefly, recombinant viruses were incubated with MUNANA substrate at a final concentration of 0.1 mM for one hour at 37℃ with shaking. After this incubation, fluorescence was measured using a Victor X5 plate-reader (PerkinElmer, Waltham, MA) with a 355 +/- 10 nm excitation filter and a 460 +/-20 nm emission filter. RFU was normalized to the titer of viruses determined by plaque assay to obtain estimates of relative NA activity between different variants. To estimate sensitivity to oseltamivir, MUNANA assays were performed in the presence of a range of oseltamivir concentrations. Recombinant viruses

were incubated with various concentrations of oseltamivir for 45 minutes at 37°C with shaking and then reacted with MUNANA as described. The fluorescent signal at each concentration of oseltamivir was normalized to the signal in the absence of oseltamivir and the resulting data fit to a standard binding equation in order to estimate IC_{50} values. Enzymatic activity of E119V and R293K was estimated by transiently expressing NA on the surface of 293T cells according to a previously published protocol with modifications [364]. Briefly, an equal amount of plasmid harboring WT or mutant NA was transfected into an equal number of 293T cells using TransIT-LT1 Reagent (Mirus, Madison, WI). 293T cells were harvested 24 hours post transfection and resuspended in non-lysis buffer before subject to MUNANA assays as described for recombinant viruses. Sensitivity of I223R and I223K to oseltamivir was also estimated by expressing NA in 293T cells and then incubated with oseltamivir, followed by reaction with MUNANA as described. IC_{50} values were estimated by fitting normalized signal to a standard binding equation.

Sequence analyses

Influenza genomic RNA was extracted from supernatants containing virions using the QIAamp Viral RNA Mini Kit (Qiagen, Germantown, MD). Viral RNA was reverse transcribed into cDNA using primers binding upstream of randomized libraries and SuperScriptIII (Life Technologies, Beverly, MA). Subsequent processing steps were as described previously for analyzing mutant frequency [326]. Briefly, samples were barcoded to distinguish replicates as well as plasmid, P0, and P1 samples and submitted for Illumina 36bp single read sequencing on a Genome Analyzer II. 2.05×10^7 high

quality reads (>99.5% confidence across all 36 bases) were obtained and analyzed. Raw sequencing data has been submitted to the NIH Short Read Archive under accession number: Bioproject PRJNA272490 or SRA SRX839403. Read abundance (*R*) is the count of each mutant. The relative abundance (*A*) of each point mutant in plasmid or P0 or P1 library was estimated from logarithmic frequency of mutant normalized to WT, as indicated below in Equation 1.

$$A = \log_2\left(\frac{R_{mut}}{R_{WT}}\right) \qquad \text{(Eq. 1)}$$

The frequency change (*F*) of a mutation from P0 to the mean of three P1 replicates (Equation 2) was used as an estimate of the enrichment or depletion during viral expansion. Estimates of selection during virus recovery were made by comparing frequency changes between plasmid and P0 using the same equations.

$$F = A_{\overline{P1}} - A_{P0} \qquad \text{(Eq. 2)}$$

Selection coefficients (*s*) representing the experimental effects of each mutation on viral replication were calculated by normalizing the average stop codon to -1 (representing null fitness) and wild-type synonyms to 0 (representing no fitness effect), as indicated in Equation 3. Mutants with *s* less than 0 had a fitness defect, whereas mutants with *s* greater than 0 had a fitness benefit, relative to the parental sequence.

$$s = \frac{F_{mut} - F_{\overline{WTsyn}}}{F_{\overline{WTsyn}} - F_{\overline{stop}}} \qquad \text{(Eq. 3)}$$

The above analyses yielded experimentally reproducible estimates of fitness effects for mutations with frequency greater than 0.2% (Figure 4.11). Mutations below this frequency in P0 were likely subject to highly stochastic sampling in the pool of

Figure 4.11: Plots showing pooled error estimate and the resulting residuals for all data sets

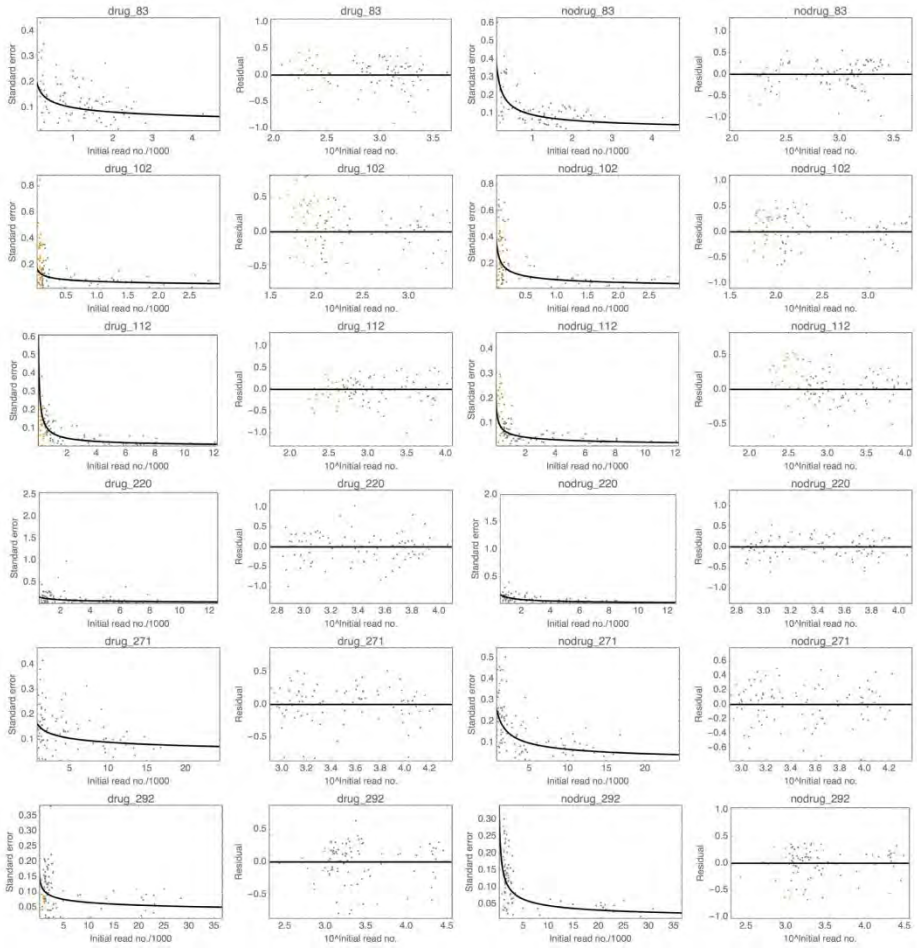


Figure 4.11 Plots showing pooled error estimate and the resulting residuals for all data sets. The header for each plot indicates the condition (labeled drug for 1 μM oseltamivir conditions, and no drug for standard conditions) and the first amino acid position of the region analyzed. The first and third graphs in each row show the actual standard errors as dots, and the estimated pooled error curve as solid black lines. Excluded data points, based on a 0.2% cutoff as explained in the Materials and methods, are shown in orange. The second and fourth graphs in each row show the residuals that are obtained from the pooled error regression. Residuals are evenly distributed around 0 throughout the whole range of initial read numbers, supporting the hypothesis that initial read number is the major source of noise.

viruses used to start P1 passages. The majority of mutations with low frequency in P0 were prevalent at far greater abundance in the plasmid-encoded libraries. These observations indicated that selection during viral recovery depleted viruses with unfit NA mutations, and is consistent with expected viral expansion in MDCK cells during the co-culture viral recovery to generate P0 samples. We observed that the frequency change of mutations from plasmid to P0 correlated (R^2=0.89) with frequency changes from P0 to P1 (Figure 4.3). From these observations we infer that selection pressure on NA mutations during plasmid to P0 rescue mimics selection during P0 to P1 expansion and that for mutations at high relative abundance in P0, frequency changes from P0 to P1 may be more accurate estimates of fitness effects than frequency changes from plasmid to P1. For these reasons, we calculated fitness effects without drug pressure for mutations with low frequency (<0.2%; 14% of our dataset) based on frequency changes from plasmid to P0, while for mutations that were abundant in P0 (>0.2%) fitness effects were calculated solely based on frequency changes from P0 to P1. The severe fitness defects for these mutations with low frequency in the absence of drug make it unlikely that they would contribute to drug resistance.

Estimate of variation of experimental fitness measurements

We assessed potential sampling limitations (*e.g.*, bottlenecking) in our viral recovery and passaging experiments by analyzing stop codons and synonymous codons. Stop codons had consistent and strong depletion in P0 samples relative to plasmid, whereas synonymous mutations had very limited changes in frequency between these

pools (Figure 4.2). This indicates that selection on NA occurs during P0 viral recovery (consistent with infection of co-cultured MDCK cells during this step) and that sampling of plasmid variants during viral recovery is sufficient to reproducibly sample the number of mutations in our libraries. If sampling were insufficient, stochastic depletion of some synonymous mutations due to random under-sampling should have been observed. In addition, four mutants with strong depletion in P0 samples relative to plasmid were studied in isolation and had strongly reduced or even no infectivity in an independent experiment, suggesting that mutants became depleted in P0 samples mainly because they had detrimental effects on viral infectivity.

The variation in fitness effects between EMPIRIC experimental repeats (Figure 4.11) suggests that sampling during initial viral entry was the main source of variance between replicates. This sampling was sufficient to analyze all single nucleotide variants, but would hinder the investigation of more complex libraries (*e.g.*, libraries of all possible amino acid substitutions). To calculate confidence intervals on estimates of fitness effects from mutants that passed the P0 cutoff, we used a pooled error function for each data set. This approach provides a more robust estimate of standard errors for each mutant than calculations of the standard error of the mean based on three measurements. Pooled error functions were obtained by a log-linear regression of the individual standard deviations as a function of mutant frequency in P0 (Figure 4.11). An evaluation of the resulting residuals supported the validity of the pooled error approach. Mutations whose standard error appeared as an outlier of the estimated error distribution, which indicates potentially different sources of error for these mutations, are identified in Table 4.3. To estimate

noise from sample processing and sequencing, viral samples generated solely from wild-type NA were processed by identical procedures and sequenced as control (Table 4.2).

Statistical analysis to determine oseltamivir responsive mutants

t-tests using 2×(N-1) degrees of freedom (where N represents independent fitness measurements for each amino acid substitution in both drug and no-drug conditions) were used to compare fitness effects of amino acid substitutions with and without oseltamivir together with a multiple-test correction using a 5% false discovery rate.

Analysis of mutant frequency in natural isolates

6205 H1N1 NA protein sequences and 5279 H3N2 NA protein sequences isolated from humans were downloaded from the Influenza Research Database [371]. These sequences were from viruses collected from 1933 to 2013, although the majority of viruses (76%) were from the 2004-2013. The dataset was curated to exclude partial sequences and duplicate entries. Multiple sequence alignment was conducted using Multiple Sequence Comparison by Log- Expectation, MUSCLE [382]. Positional amino acid frequencies were tabulated and compared to experimental fitness measurements.

Acknowledgments

We acknowledge the contributions of all members of the ALiVE (Algorithms to Limit Viral Epidemics) working group. We thank Melanie Trombly and Nese Kurt-Yilmaz for assistance with the preparation of the manuscript. This work was supported by the Prophecy Program, Defense

Advanced Research Projects Agency, Defense Sciences Office [DSO], contract HR0011-11-C-0095.

Chapter V – General Discussion

Summary

The work presented in this dissertation is mainly focused on applying high throughput systematic mutagenesis approaches to understand sequence-function-fitness relationship and explore sequence space for molecular adaptation in diverse proteins and organisms. In Chapter II, I first applied the EMPIRIC approach developed by Hietpas *et al* to investigate a fundamental protein sequence-function relationship question about effect of interplay between protein expression and sequence on organismal fitness. We discovered a non-linear relationship between yeast growth rate and Hsp90 expression: nearly 80% reduction in Hsp90 expression does not have strong effect on yeast growth, while stronger reduction in Hsp90 expression significantly impairs yeast growth. Many mutations that exhibit WT-like fitness effect at endogenous level of Hsp90 expression show fitness defect at reduced expression level, indicating that high expression of essential proteins, *e.g.* Hsp90, mask the fitness defect of many mutations and buffer the detrimental effect of many stochastic mutations on organismal fitness. In Chapter III, I optimized the EMPIRIC approach to map the functional constraint on the CD4 binding loop and flanking regions of the trimeric env complex of HIV. We show that the vast majority of mutations at these positions cause intermediate to strong fitness defect, indicating the functional significance of these residues. Only wild-type amino acids are functional at residues in the CD4 binding loop (GGDPE), whereas several residues in the flanking regions are tolerant of amino acid substitution even with distinct physical-chemical properties. Thirteen mutations are beneficial compared to the WT sequence

under the experimental condition that mainly selects for higher infectivity. These mutations all exhibit stronger affinity to sCD4 than WT, suggesting higher CD4 affinity as a common pathway to enhanced infectivity. Neutralization profiles of these mutations reveal that many of these beneficial mutations also impose distinct and complicated impact on the conformation of the trimeric env complex. In Chapter IV, I optimized the EMPIRIC approach to quantify the fitness of mutations in the active site and proximal regions of NA in influenza virus in the presence or absence of a NA competitive inhibitor (oseltamivir) and screen for beneficial mutations that adapt to drug selection. We identify previously known drug resistance mutations, which validate our approach. We also identify novel drug resistance mutations that have not been clinically associated with reduced sensitivity to oseltamivir. The majority of these drug adaptive mutations occur in the active site, in particular at the drug resistance hotspot residue 223. Biochemical analyses on drug adaptive mutations uncover two distinct biochemical pathways towards drug resistance: reduced binding to drug and increased substrate processing. In summary, I start with a basic protein sequence-function relationship question about collective effect of protein expression and protein sequence in a well-characterized model system, *Saccharomyces cerevisiae*. Then I map functional constraints on important regions of the receptor of a fast evolving virus that causes pandemics, HIV. Lastly, I approach a more clinical relevant problem about drug resistance by systematically screening for mutations with reduced drug sensitivity in influenza virus and using biochemical assays to dissect underlying mechanisms.

Distinct distribution of fitness effect in yeast and RNA viruses

Although mutations can be broadly divided into three categories: deleterious, neutral and beneficial based on their effect on organismal fitness (fitness effect), the fitness effect is actually a continuous parameter, *i.e.* deleterious mutations have distinct detrimental effect, while beneficial mutations have distinct advantageous effect. The relative frequency of mutations with different fitness effect can be described by the distribution of fitness effect (DFE). DFE is essential for mapping the architecture of sequence space and genetic variation for any gene and provides valuable information for frequency of neutral and beneficial mutations that mediate neutral and adaptive evolution [383]. In the presented thesis, I investigated the DFE of Hsp90 for yeast, Env of HIV and NA of IAV (Figure 5.1). These proteins are all essential proteins that determine organismal fitness, and these regions under investigation are involved in protein-protein interaction or substrate processing.

Under standard laboratory growth condition, the yeast Hsp90 exhibits a bimodal distributed DFE with one major peak at the WT-like fitness and the other minor one at the null-like fitness, indicating that a large fraction of mutations have little effect while several mutations have strong deleterious effect (Figure 2.2). Only a small fraction of mutations exhibit intermediate fitness defect. RNA viruses (HIV and IAV), on the other hand, has a quite distinct unimodal DFE: the majority of mutations exhibit strong fitness defects while very few mutations exhibit WT-like fitness (Figure 3.4 and 4.5). Notably, no mutations are strongly beneficial under standard laboratory condition in yeast or IAV,

Figure 5.1 Relationship between fitness and total protein function as a predominant determinant of the distribution of fitness for different proteins

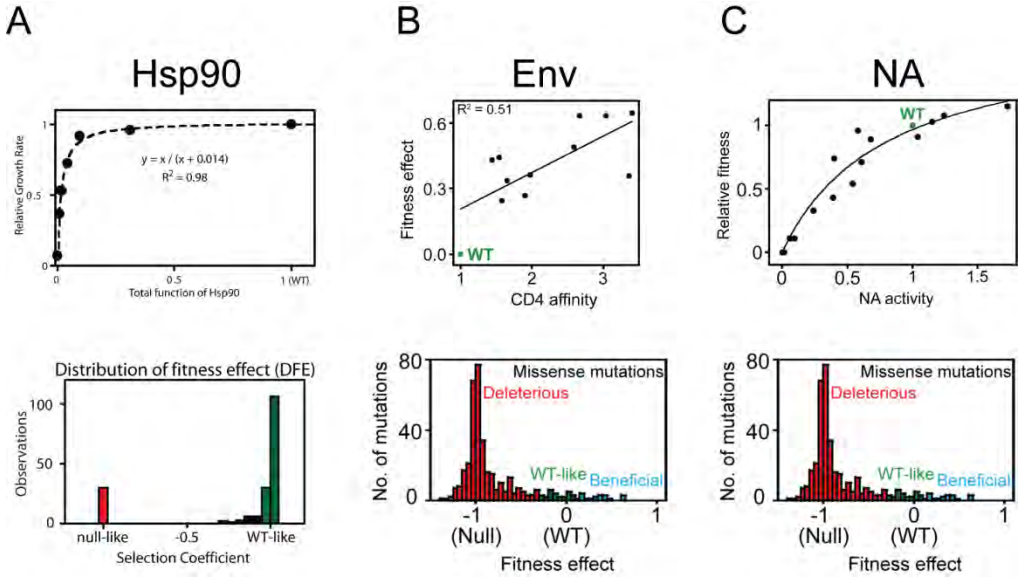


Figure 5.1 Relationship between fitness and total protein function as a predominant determinant of the distribution of fitness for different proteins. The relationship between organismal fitness and protein total function (upper panel) and the distribution of fitness effect (DFE, lower panel) is shown for a substrate/receptor binding site for yeast Hsp90 (A), HIV Env (B) and influenza NA (C).

indicating that the sequences of these proteins have been highly optimized after repeated culturing and adaptation. Another shared feature is that few mutations in either yeast or RNA viruses have intermediate deleterious fitness, indicating that missense mutations either maintain or completely abolish protein functions. The distinct DFE between yeast and IAV indicate that yeast is much more robust to mutations than IAV. Most mutations in other cellular organisms (e.g. *E. coli* and *C. elegans*) also have little effect on growth [384, 385], while many mutations in RNA or single strand DNA (ssDNA) viruses generally exhibit detrimental effect on viral infectivity [386].

Genome complexity appears to correlate with sensitivity to mutations [387]. Signaling network with abundant redundancy and alternative metabolic pathway in cellular organisms are able to buffer the detrimental effect of mutations, so cellular organisms are more tolerant of mutational effects. However, absence of these robustness mechanisms from the simple and compact genomes of RNA or ssDNA viruses potentiates the effects of selection pressure, sensitizing RNA viruses to mutational effects. In fact, this differential sensitivity to mutations suggests two distinct evolutionary strategies. Cellular organisms accumulate many cryptic mutations that appear neutral. These mutations may provide growth advantage in a shifted environmental condition. This prediction is consistent with buffering role of Hsp90 in masking phenotypes of cryptic mutations. Only upon occasional environmental stresses that tax the buffering capacity of Hsp90 and impose more selection pressures, the effect of mutations becomes

revealed and beneficial mutations are selected and fixed. RNA viruses are constantly subject to strong purifying selection, so most deleterious mutations will be purged. However, lack of buffering mechanisms also accelerate fixation of adaptive mutations. This hypothesis is consistent with the notion that mutation rates should maximize adaptation by modulating the balance between sampling adaptive mutations and tolerating deleterious mutations. The low mutation rate of cellular organisms ensures relatively low mutation load, which can be suppressed by the buffer system; high mutation rate of RNA viruses enables efficient purification of deleterious mutations and fixation of adaptive mutations. The intermediate genome complexity, tolerance of mutational effect and mutation rate of double strand DNA viruses also support this explanation [388].

Differential sensitivity to mutations leads to distinct strengths of beneficial mutations. In yeast, mutations confer growth advantage up to 7% under elevated salinity selection [56]. In contrast, adaptive mutations in IAV confer up to ~110% benefit under drug selection pressure (Figure 4.5), consistent with potentiated mutational sensitivity of IAV. The DFE of HIV presents an interesting case. The *env* gene was cloned from a patient sample, so it is mal-adapted under the standard laboratory condition. Adapting the patient derived env to the laboratory condition leads to beneficial mutations of growth advantage up to ~63%, 9 fold higher than that of yeast adaptive mutations (Figure 3.4 and Table 3.3). Mutations of large benefit have also been reported in ssDNA phages [389,

390]. Beneficial mutations of large effect rapidly spread, mediating adaptation of RNA viruses to changing selection pressures, such as immune response and cross-host barrier.

Different regions of the genome appear to have varied shape of DFE, depending on their functional importance. In an exterior loop of NA, more than half of the mutations have little defect, while few mutations have strong fitness defect [Figure 4.5B]. This indicates that regions of greater functional or stability significance (catalysis, substrate binding or scaffolding) tend to be more sensitive to mutations, whereas regions of less functional or structural significance appear to be more tolerant of mutations. DFE from randomly selected mutations across the genome for several RNA and ssDNA viruses assumes a bimodal distribution with one peak at WT-like fitness and the other one at null-like fitness, though the relative height of two peaks vary among different viruses [292-295]. The combined DFE for the active site and proximal regions and exterior loop of NA exhibits a bimodal distribution similar to that observed for other RNA and ssDNA viruses, indicating that DFE derived from fitness effects of randomly generated mutations across the genome represents a convolution of DFE from functionally important and unimportant regions. Indeed, different genes of ssDNA virus f1 exhibited varied sensitivity to missense mutations [295]. In summary, genome wide DFE may mask local variation in sensitivity to mutations, so determining region specific DFE may reveal functional significance of different regions of a genome.

Protein activity per molecule as a predominant determinant of fitness

The bimodal DFE is consistent with a model that proposes protein stability as a dominant determinant of fitness [297]. This model is based on two assumptions: positions that are critical for protein function are rare and native protein folding is required for protein function. The elasticity-like non-linear relationship between protein stability and function suggests that most mutant proteins are either completely folded or unfolded. There is only a narrow range corresponding to partially folded mutant proteins [297, 391]. This predicts a bimodal distribution of protein function and hence a bimodal DFE. Simulation using protein stability as the sole parameter to model fitness successfully produces bimodal DFE in *in silico* viruses [296]. In the stability model, mutations at critical residues that abolish protein function are rare and therefore make little contribution to DFE.

Protein function per molecule (F) derived from the fitness measurements and the elasticity function between fitness and protein function indicates that F without apparent contribution from stability effect also contributes to determine organismal fitness. The estimated F exhibits a unimodal distribution, indicating that most mutations reduce F. This is inconsistent with a bimodal distribution of F derived from stability effect alone, suggesting stability independent effect that also contribute to protein function per molecule, *e.g.* catalysis and substrate binding. Most mutations exhibiting intermediate deleterious effects also indicate that more mutations directly affect protein function than predicted by a stability model. Notably, our definition of F combines all properties

including stability, binding affinity and catalysis efficiency, providing a more comprehensive measurement of mutational effect on protein function. Fitness measurements of Hsp90 mutations display very weak correlation with their stability effect estimated by Rosetta simulation (Figure 2.13C) [104, 321]. Together with results from biochemical assays (circular dichroism spectra and urea-induced unfolding curve) on a panel of six mutations [Figure 2.16 B and C], they verify the minor contribution of stability effect to protein function and yeast growth for mutations in this region of yeast Hsp90. Of note, this region is not involved in ATP binding and hydrolysis, but implicated in substrate binding. The unimodal distribution of protein function per molecule is consistent with its potential biological function. The frequency of regions of biological function may vary among different proteins and will be examined in future experiments that systematically map rate-limiting regions.

Distinct conformation of the Env mutants with enhanced binding affinity to CD4

All Env mutants with increased infectivity exhibited enhanced binding affinity to CD4 (Table 3.3), indicating that stronger binding of the trimer Env complex for CD4 serves as a common evolutionary pathway to higher replication fitness in a standard laboratory condition. The parental strain (LN40) shows relatively weak binding to CD4 due to steric restrictions and/or refractory to conformational changes. Thus, the increased CD4 binding of these beneficial mutations suggest enhanced exposure of the Env to CD4. Of note, S375W drives the conformation of the Env into a more CD4-bound and open

state, increasing the exposure of the trimeric env complex to CD4 [176]. The bulk competition was conducted in a standard laboratory condition without selection pressure from immune response *in vivo*, so it should select for mutations with more open Env conformation, elevated affinity for CD4 and hence higher infectivity. However, Env with open conformation is often more vulnerable to antibody neutralization, so the complex *in vivo* condition should select for mutations that are best able to keep a delicate balance between escaping neutralizing antibody and maximizing infectivity. The differences in environmental conditions select for distinct adapting gp120 mutations, indicating the enormous potential of HIV to adapt to varied environmental conditions even with the simplest evolutionary step – single amino acid substitutions.

We were able to examine the sensitivity of these beneficial mutations to b6, b12, 447-52D and 17b. This combination of antibodies enables interrogation of conformational changes at regions that are critical for CD4 and/or coreceptor binding. Both b12 and b6 targets the CD4 binding site, but they are only exposed in the monomeric gp120. Increased sensitivity to these antibodies indicate opening of the trimer and exposure of their epitopes, which are excluded in the trimer [166, 350]. In particular, b6 requires extensive opening of gp120 and fails to neutralize almost all primary isolates [350]. 447-52D binds to the V3 apex, which is only exposed after sufficient opening of the env trimer, so it only neutralizes lab-adapted strains [392]. 17b is a CD4 binding induced antibody (CD4i) and neutralizes gp120 that completes CD4 induced

conformational change [161, 351]. This set of antibodies allows us to gauge the openness of the trimeric env complex.

The trimeric Env spike exhibits varied conformational state and sensitivity to different neutralizing antibodies. The parental strain (LN40) is generally resistant to antibodies that target epitopes exposed only in Env monomer. These epitopes are excluded in trimeric Env complex, but become exposed after trimer dissociation. Many mutations at residue 373, 375, 377 and 380 were more sensitive to b12, indicating that mutations at these positions facilitate exposure of the b12 epitope and probably opening of the trimer. These residues are within or very close to the Phe43 cavity, so local perturbations are likely to affect gp120:CD4 interaction. This is supported by increased sensitivity of these mutations to sCD4 neutralization. Only G380A and G380P exhibited measurable sensitivity to both b6 and 447-12D, indicating that mutations at this residue open up the trimeric Env complex more than mutations at other positions. However, none of these mutations were sensitive to 17b, indicating that they are not open enough to expose the 17b epitope. Interestingly, most beneficial mutations developed intermediate resistance to glycan-targeted 2G12 compared to the parental strain, indicating that these mutations confer a shift in glycan distribution, presumably at residue 386 [182]. However, these mutations and the WT were still highly sensitive to PGT128, which targets a different conserved glycan at V3 [393], indicating that the glycan shift is restricted locally.

These beneficial mutations in LN40, a clade B HIV, impose similar effect on the conformation states of the trimeric Env complex from another clade B (LN8) and a clade C (Z1792M) primary isolate. For example, mutations at residue 373, 375, 377 and 380 were more sensitive to neutralization by sCD4, indicating that they enhance the binding interaction between CD4 and gp120. Moreover, pseudovirus with G380P became extremely sensitive to b6, b12 and 447-52D, indicating that this mutation opens up the trimer complex and unmasks many epitopes not available in the closed trimer. Thus, despite large sequence variation among different HIV strains and/or clades, many mutations have similar effect on conformation states of the trimer Env complex. LN8 and Z1792M both carry a glycan at N160, which forms the epitope recognized by the trimer specific V2q mab PG9 and PGT145 [354, 394]. In fact, they are both very sensitive to PG9 and PGT145 (Table 3.5 and 3.6). Thus, sensitivity to PG9 and PGT145 reflects the epitope integrity, which is part of the trimer association domain [168], so disruption of the TAD should lead to resistance to PG9 and PGT145 neutralization. As shown before, many mutations at residue 373, 375, 377 and 380 (in particular G380P) appear to drive the trimeric env complex into a more open and exposed conformation, which should disrupt the TAD. Surprisingly, these mutations are still sensitive to PG9 and PGT145 as compared to the WT (Table 3.5 and 3.6), indicating an intact TAD and closed conformation of the trimer. One explanation to this apparent contradiction is that these mutations promote more frequent transition into CD4-bound state of gp120 without triggering a trimer-opening event, so that these mutants are sensitive to both trimer-

specific and CD4BS antibodies. This is consistent to a recent finding that applied FRET to determine the conformation dynamics of Env: despite a ground closed conformation state, neutralization sensitive strain NL4.3 frequently transits into the open state, while neutralization resistant strain JR-FL mostly stays in the closed state [355]. Another possible explanation is that some mutations are able to expose the CD4 binding site and simultaneously maintain the TAD.

Permissive/compensatory mutations of oseltamivir resistance mutations

The majority of oseltamivir resistance mutations identified in the EMPIRIC screening exhibit fitness defect (mostly 20-50% defect) in the absence of drug pressure [Table 4.5]. For example, I222M, H275Y and N295S show ~30% defect. The observed fitness defect in the bulk competition is consistent with reduced infectivity measured on individual generated clones. For example, H275Y and N295S in NA all show negative effect on the replication and/or transmission capability of IAV [252, 262, 263, 265]. Adaptive mutations to one particular selection pressure frequently results in fitness defect under other environmental conditions, which is known as cost of adaptation. The tradeoff in fitness during adaptation to distinct selection pressures is common for many microorganisms in response to shifting environments [395-399]. For example, numerous yeast mutations that confer up to 7% benefit under high osmotic pressure exhibit fitness defect under standard growth condition [56]. Another example is that mutations in HIV protease or reverse transcriptase that reduce binding to HLA molecule tend to have

impaired *in vitro* replicative fitness [400]. Many of these studies also focused on negative pleiotropic effects of mutations that conferred antibiotic resistance on other cellular functions [396-399]. This is consistent with predictions from the Fisher's geometric model (FGM) [56]. In a phenotypic space, each single phenotypic optimum corresponds to a different selection pressure. When the wild-type sequence is highly optimized for adaptation to one selection pressure, it is located very close to the optimum. Under a different selection pressure, mutations that reduce the distance to the new optimum are considered adaptive mutations, but these mutations typically deviate away from the initial optimum. Thus they are deleterious mutations under the original selection pressure.

Most oseltamivir resistance mutations occur at either functional or framework residues in the active site of NA. They appear to interfere with the binding of NA for oseltamivir as well as the natural substrate [252, 281, 401]. For instance, H275Y leads to greater than 400 fold higher Ki, but also approximately 2 fold higher Km. H275Y also leads to reduced surface NA enzymatic activity when NA is expressed in mammalian cells, mainly due to reduced surface expression of NA [364, 401]. Destabilizing mutations may cause reduced surface NA expression, while mutations that impair catalysis or substrate binding may decrease enzymatic activity per NA molecule (data not shown).

Permissive or compensatory mutations restore the fitness cost of oseltamivir resistance mutations. Secondary permissive or compensatory mutations are able to compensate for the loss of the total enzymatic activity and restore replication efficiency. Pioneering work by Bloom *et al* showed that R194G, R222Q and V234M are able to raise the reduced surface expression of H275Y mutant NA in the seasonal and pandemic H1N1 strains and rescue their replication fitness defect [364]. These permissive mutations had very high frequency in H1N1 in the influenza season of 2007 and 2008 [207, 401]. Fixation of permissive mutations provides a plausible explanation for competent growth and transmission fitness of H275Y and its global circulation in the influenza season of 2007-2008 even in the absence of massive use of oseltamivir. Experiments in a ferret model showed that V241I and N369K rescued the replication and transmission defect of a pandemic H1N1 strain with H275Y [379]. I222V partially restored the enzymatic activity (Vmax) of NA in a seasonal H3N2 strain and rescued the fitness cost of E119V [402].

Y276F is a hyperactive mutation that increases the enzymatic activity of NA and has the potential to rescue the fitness defect of oseltamivir resistance mutation through buffering loss of NA activity. Y276F exhibits big increase in fitness in the presence of oseltamivir, indicating that it is an adaptive mutation to oseltamivir. However, it shows similar affinity to oseltamivir as compared to the WT, indicating that other features initiate its adaptation to oseltamivir. Y276F turns out to be a hyperactive mutation that enhances the enzymatic function (per infectious unit) of NA by approximately 70%

(Figure 4.9). It creates a buffer zone in NA activity, which compensates for the inhibition of NA activity by oseltamivir. The residual NA activity is able to support growth of IAV and confers superior replication fitness compared to the WT. I hypothesize that Y276F might compensate for the loss of NA activity from oseltamivir resistance mutations and rescue their growth. Notably, Y276F is in close proximity of residue 275, so it might also be able to change the local conformation and offset the fitness defect from H275Y (data not shown).

Merits and limitations of EMPIRIC in studying viral evolution

The EMPIRIC approach has been proved useful in exploring the evolvability of RNA viruses (*e.g.* IAV) to therapeutic selection pressures (Chapter IV). It has also been shown instrumental in mapping biophysical constraints on important regions of RNA viruses (*e.g.* HIV) under strong purifying selection (Chapter III). A major advantage of utilizing EMPIRIC to investigate viral evolution is its high reproducibility. The R-squared value (R^2) of the correlation between fitness measurements from two replicates of IAV expansion is greater than 0.9, which is higher than whole-gene mutational scanning approaches (0.34-0.62), indicating highly reproducible measurements of IAV replication fitness. For HIV, the R^2 is approximately 0.6-0.8, which is still higher than R^2 of most other approaches of higher throughput. The improved precision in fitness measurements ensures accurate characterization of individual mutations regarding their survival probability under purifying selection and adaptation potential under positive

selection, *e.g.* drug pressure. This is important in screening for rare adaptive mutations where small differences in fitness may lead to distinct evolutionary outcome. In addition, accurate assessment of fitness for mutations in the absence of therapeutic selection pressure enables robust analysis of likelihood of fixation of mutations in the treatment-naïve population. Precise gauging of cost and benefit of adaptation is critical for thorough characterization of drug resistance mutations.

Other advantages of EMPIRIC include systematic mutagenesis and relatively simple experimental setup. EMPIRIC uses cassette ligation to generate every possible single codon or nucleotide substitution. The former enables fitness characterization for all possible amino acid mutations, leading to detailed delineation of positional amino acid preference and adapting potential. The latter provides fitness measurements for mutations that are only one nucleotide step away from the WT, revealing evolutionary potential for the most likely mutational step in nature. Several other approaches that employ error-prone PCR based random mutagenesis may not cover all possible single substitutions. Gene-wide mutational scanning encompasses large number of mutations (usually > 10000), so it is inherently prone to bottleneck effect during transfer of libraries (*e.g.* transfection or infection). To achieve higher reproducibility, whole-gene mutational scanning requires huge number of cells and virions in many replicates, increasing the difficulty to manage such large scale experiments. In contrast, EMPIRIC focuses on specific regions of interest and uses relatively smaller scale of reagents.

The major limitation of EMPIRIC is its intermediate throughput. EMPIRIC has been used for systematic mutational scanning of several ten-amino-acid regions in several parallel experiments, while other approaches allow for whole-gene (hundreds of amino acids) mutational scanning in a single experiment. EMPIRIC can be scaled up to investigate an entire gene, but it requires considerable experimental effort. There are two rate-limiting steps: library construction and bulk competition. Cassette based randomized mutagenesis is more labor-intensive than PCR based one used in most whole-gene mutational scanning approaches. Because each library only contains ten amino acids, several libraries are analyzed in parallel. As the number of libraries grows, the workload involved with bulk competition and sequencing sample preparation grows correspondingly, leading to more effort than studying an entire gene in a single pooled library. Another limitation is that EMPIRIC requires prior information such as structure, chemical properties, function, and conservation of proteins to select which part(s) of the protein to analyze. Conversely, the whole-gene mutational scanning follows a shotgun concept and relies less on other information. This may prove useful for investigating function of proteins with limited prior information. However, the intermediate throughput of EMPIRIC significantly contributes to the improved reproducibility of fitness measurements. In fact, a recent whole-gene mutational scanning study on PA of IAV used a “small library” approach similar to EMPIRIC: they divided the gene into nine sections, generated nine independent libraries and analyzed them individually to get higher experimental reproducibility [74]. Although pursuing higher throughput proves rewarding in many ways, it is critical to retain the balance between throughput and

reproducibility of results. This balance is particularly important for mutational scanning on viral proteins, which are especially sensitive to bottleneck effect during library transfer.

EMPIRIC can be further optimized from several perspectives. In the yeast system, population management is robust to bottleneck effect since yeast can be grown to high density. A barcode can be engineered into the plasmid outside the open reading frame and associated with individual mutations by pair-end Illumina sequencing. For example, an 18 consecutive nucleotide randomized barcode encompasses nearly 69 billion unique combinations, so each individual mutation is most likely to be associated with a unique barcode that is several nucleotides different from any other barcode. Although additional experimental and computational steps are required, this modification enables combination of multiple ten amino acid libraries into a single pool, which can be subject to diverse environmental conditions for rapid probe of genotype-environment relationship. In addition, the barcode strategy allows sequencing error correction by taking consensus of multiple sequencing reads of the same barcode. Secondly, the Illumina read length was 36bp when EMPIRIC was first developed, which was sufficient to sequence a 10 amino acid region (30 nucleotides). It has now grown to be 300bp, allowing for more residues in a single library. A library with more residues reduces the workload of bulk competition and sequencing sample preparation, because fewer libraries are processed for the same number of amino acids analyzed. Thirdly, EMPIRIC can be combined with other higher throughput whole-gene mutational scanning approaches. For example, whole-gene

scanning identifies critical residues that are sensitive to mutations; EMPIRIC on these residues then generates reproducible fitness measurements on individual mutations under various selection pressures. In summary, EMPIRIC can be further improved regarding throughput and efficiency by taking advantage of longer sequencing reads and novel barcoding strategies.

Future directions

EMPIRIC can be used to map resistance profile for other competitive inhibitors of NA. There are four competitive NA inhibitors (NAI): oseltamivir, zanamivir, peramivir and laninamivir. The first three have been approved by FDA, while Laninamivir is still in clinical trial in the US. All four NAI are derived from the natural substrate (sialic acid) using structural based drug design. Thus, it is of utmost clinical, pharmacological and basic science interest to compare the resistance profiles of these four closely related NAIs. From a clinical standpoint, knowing resistance mutations to each NAI and whether mutations develop cross-resistance to two or more NAIs assists clinical monitor of drug resistance mutations and rationale administration of NAI to avoid or delay drug resistance. From a pharmacological standpoint, associating differences in resistance profiles to divergences in chemical structures of NAIs provides valuable information to design novel ones with minimized resistance. From a basic science standpoint, biophysical and biochemical measurements of drug resistance mutations are likely to provide training set for machine learning algorithm that predicts drug resistance mutations. In addition,

fitness landscape of NA mutations in response to different but closely related drug selection pressure will illuminate evolutionary and biochemical pathways of molecular adaptation of IAV. EMPIRIC can be used to determine the fitness landscape of NA with different NAIs and answer the above questions.

EMPIRIC may be incorporated into development of new inhibitors of RNA viruses. During preclinical development, EMPIRIC analysis may identify mutations with reduced sensitivity to the inhibitor. The biochemical and structural features of drug resistance mutations may aid refinements of the initial inhibitor to target isolated drug resistance mutations. Current EMPIRIC approach and many mutational scanning approaches use lab adapted strains because they generally have robust growth and efficient reverse genetic systems, which lead to higher reproducibility. In order to be incorporated into drug discovery pipeline, EMPIRIC needs to be optimized for generating highly reproducible measurements on primary isolates. This has been tested in a patient derived HIV env (Chapter III), but has yet to be implemented in other pathogenic RNA viruses.

EMPIRIC can be utilized to isolate escaping mutations from broad neutralizing antibodies of HIV. There has been no successful HIV vaccine to date. High mutation rate of HIV enables rapid diversification of HIV, so strain-specific antibodies are incapable of neutralizing a heterogeneous population of different HIV strains. Several antibodies that

were isolated from patients with chronic HIV infection exhibit much broader spectrum of neutralization. These broad neutralizing antibodies (bNab) are able to neutralize 50-95% of primary isolates, providing a novel avenue for HIV vaccine research. EMPIRIC can be used to analyze the adaptive potential of the epitopes of HIV Env under the selection of bNabs. The structural information of escaping mutations should assist in optimization of naturally evolved bNab to achieve broader breadth and even higher potency.

EMPIRIC can also be adapted to analyze the functional constraints and evolutionary potential of other fast growing organisms such as HCV and cancer cells. There have been numerous reports of drug resistance in other pathogenic RNA viruses such as HCV and dengue viruses [403, 404]. EMPIRIC can be optimized to investigate drug adaptive potential of these viruses with robust reverse genetics system. Of note, there has been one study that used an EMPIRIC-like approach to identify drug resistance mutations in NS5A of HCV in response to Daclatasivr [76], corroborating the potential of applying EMPIRIC to other RNA viruses. Cancer cells quickly evolve resistance to traditional chemotherapy or recently developed targeted therapy, leading to poor prognosis of many tumors [405]. EMPIRIC has the potential to elucidate the adaptive fitness landscape of mutations in tumors under the selection pressure of chemo or targeted therapy, revealing drug resistance mutations and associated molecular mechanism. For example, EMPIRIC has been successfully applied to study secondary mutations in BRAF-V600E melanoma that confer resistance to a targeted drug -

PLX4720. A novel mutation with reduced sensitivity to both BRAF and MEK inhibitors was discovered, which was later isolated from a BRAF mutant melanoma patient treated with PLX4720 [341, 406]. This indicates that EMPIRIC can be modified to pinpoint single substitutions that cause resistance to targeted cancer therapy.

Broader impact

The sequence-function relationship is a fundamental question at the intersection of different disciplines of biomedical studies. In this dissertation, I utilize a novel mutagenesis approach – high throughput mutational scanning to map the sequence-function relationship in diverse proteins under various selection pressure. I analyze the interplay between protein function and sequence on organismal fitness, unifying the combined effect of two mechanisms for protein evolution. I explore the adaptive landscape of pathogenic RNA viruses, illuminating the collective constraints of positive and negative selection on governing the evolutionary pathway of RNA viruses. In summary, the work presented in this dissertation explores the role of sequence-function relationship in basic protein evolution and clinical relevant adaptation of RNA viruses to immune/drug selection pressure, opening up many new avenues to demarcate the sequence space and underlying molecular mechanisms available for continual adaptation to the ever-changing environment.

Appendix – Systematic identification of zanamivir resistance mutations

Introduction

Zanamivir is a competitive inhibitor of neuraminidase (NAI) of IAV. Zanamivir is administered through inhalation and has been used for IAV treatment since 1999. Compared to the other NAI, oseltamivir, zanamivir has showed much less drug resistance. Some explanations include its close resemblance to the natural substrate and less clinical usage. NA mutations (E119V and R293K, N1 numbering) displaying reduced sensitivity to zanamivir were identified in cell culture prior to clinical use of zanamivir, but these mutations caused severe fitness defect and were not observed in patients under treatment of zanamivir[407-409]. Many oseltamivir resistance mutations, including the predominant H275Y, impair the binding of pentyloxyl group at C-6 of oseltamivir. Since zanamivir has a glycerol group at C-6, these mutations still allow zanamivir binding and remain sensitive to zanamivir [202]. Several mutations, including I223R and N295S (N1 numbering used throughout), do impair binding of NA for zanamivir and hence confer zanamivir resistance. They generally dampen interactions between contact residues of NA and functional groups of NAI, so they cause cross resistance to both oseltamivir and zanamivir [202, 249, 282, 283]. However, the majority of these mutations were only isolated sporadically from patient samples or only characterized as possessing reduced sensitivity to zanamivir *in vitro*. These mutations also affect the binding of NA for the natural substrate and severely impair the replication fitness of IAV, so they are unlikely to reach a high frequency in human populations. To systematically map the adaptive

potential of NA under selection of zanamivir, we utilized EMPIRIC, which was optimized to determine the mutational fitness landscape in IAV, to quantify the fitness effects of NA mutations with or without zanamivir.

Result and discussion

Sialic acid (Neu5Ac), zanamivir and oseltamivir share similar chemical backbone, but had different chemical groups at critical positions (Figure 6.1). The major difference between Neu5Ac and zanamivir is that a hydroxyl group at C-4 of Neu5Ac was replaced with a guanidinyl group in zanamivir. Introducing a larger and more basic group at C-4 improves the affinity between NA and zanamivir through interactions between Glu119/228 of NA and the guanidinyl moiety of zanamivir. The major difference between Neu5Ac and oseltamivir is that a polar glycerol group at C-6 is substituted with a large and hydrophobic pentyloxyl group. The hydrophobic group at C-6 orients E277 of NA away from the active site to create a hydrophobic pocket, while glycerol group at C-6 forms hydrogen bonds with E277. The distinction in chemical structure of zanamivir and oseltamivir appears to elicit different drug adaptive mutations

We utilized EMPIRIC to determine the fitness effects of all possible single nucleotide substitutions in focused regions of NA in the presence or absence of zanamivir (Figure 6.2A). Plasmid libraries of mutations at four regions in the active site and proximal regions, one in the tetramerization interface and the other one in an exterior loop were examined (Table 6.1). These regions were chosen because they encompassed the zanamivir binding site and were most likely to encompass zanamivir resistance

Figure 6.1: The natural substrate and competitive inhibitors of NA

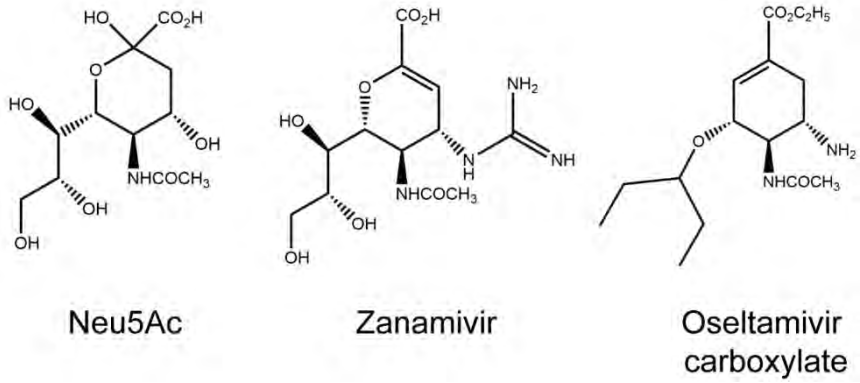


Figure 6.1 The natural substrate and competitive inhibitors of NA. Chemical structure of sialic acid (Neu5Ac), zanamivir and oseltamivir carboxylate.

Figure 6.2: A high throughput approach to determine the fitness effect of mutations in NA with or without zanamivir

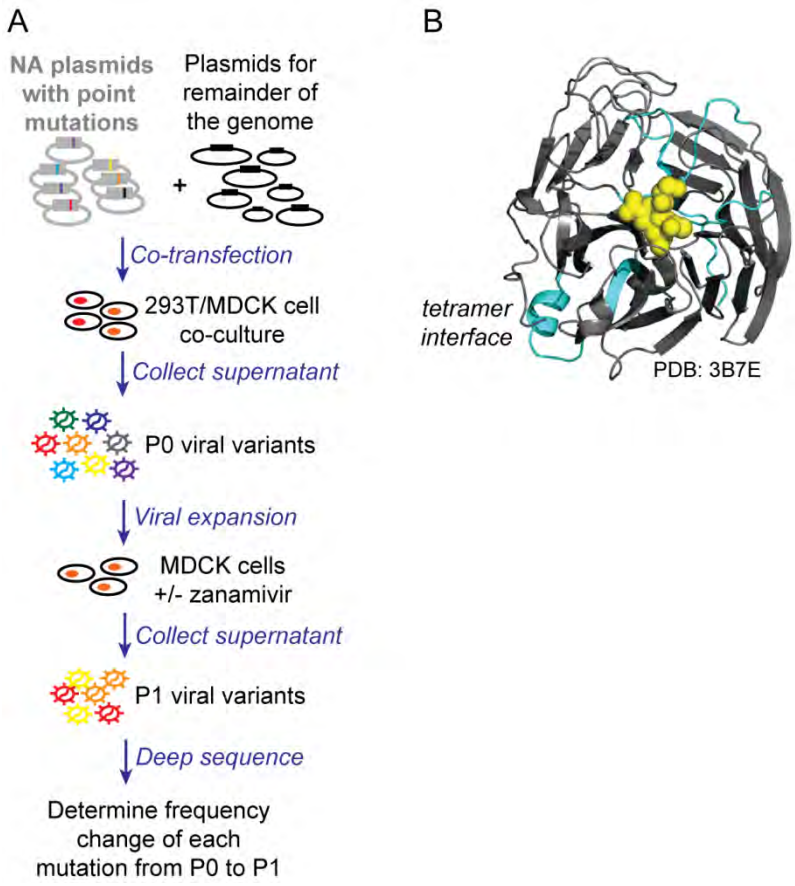


Figure 6.2 A high throughput approach to determine the fitness effect of mutations in NA with or without zanamivir. (A) An overview of the experimental setting. (B) The molecular structure image of NA (PDB ID: 3B7E). The NAI, zanamivir, is represented as yellow spheres in the active site. The regions with engineered all possible single nucleotide substitutions are in light blue.

Table 6.1 Analyzed Amino acid regions of NA

Region	Amino acid positions
Active-site proximal 1	112-121
Active-site proximal 2	220-229
Active-site proximal 3	271-280
Active-site proximal 4	292-301
Tetramer interface	102-111
Surface loop distant from active site	83-92

mutations (Figure 6.2B). This approach was previously developed for measuring fitness effect of mutations with or without oseltamivir (Chapter IV). NA plasmid harboring single nucleotide substitutions and plasmids encoding the other 7 IAV gene segments were co-transfected into co-culture of 293T and MDCK cells to recover viral library of mutants (P0). P0 library was used for infection of fresh MDCK cells with or without zanamivir. The supernatant containing viral progenies (P1) were harvested. Both P0 and P1 viruses were sequenced to determine log frequency change as a proxy of fitness effect. The NA mutation libraries had two internal controls: stop codons (nonsense mutations) and WT-synonyms (silent mutations). Nonsense mutations produce truncated NA, which should abolish NA function and impair IAV infection; silent mutations do not change the primary sequence of NA, which should support robust IAV infection. Indeed, nonsense mutations exhibited strong depletion while silent mutations behaved similarly to the WT (Figure 6.3). Thus, the log frequency change from P0 to P1 was normalized to that of nonsense mutations and silent mutations to so that fitness effect of -1 corresponded to null-like fitness and 0 indicated WT-like fitness.

Few mutations were adaptive or responsive to zanamivir

We compared the fitness effects of all mutations with or without zanamivir to identify mutations that were either adaptive or responsive to zanamivir selection (Figure 6.4A). We used 1μM of zanamivir for two reasons. Firstly, it was about 10 fold higher than EC50 (data not shown), allowing for amplification of small differences in fitness

Figure 6.3: Stop codons were strongly depleted, while WT-synonyms did not have much change in frequency

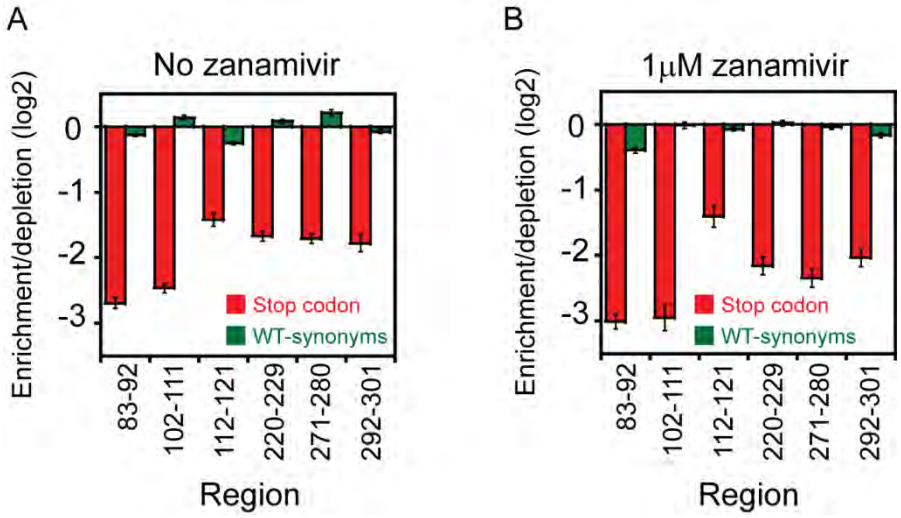


Figure 6.3 Stop codons were strongly depleted, while WT-synonyms did not have much change in frequency. The average log2 frequency change of stop codons (nonsense mutations, colored in red) and WT-synonyms (silent mutations, colored in green) were measured without zanamivir (A) or with zanamivir (B). The error bar is standard error of the mean with N equal to number of stop codons or WT-synonyms in each library.

Figure 6.4: Comparison of fitness effects of all mutations +/- zanamivir revealed few mutations that were adaptive or responsive to zanamivir

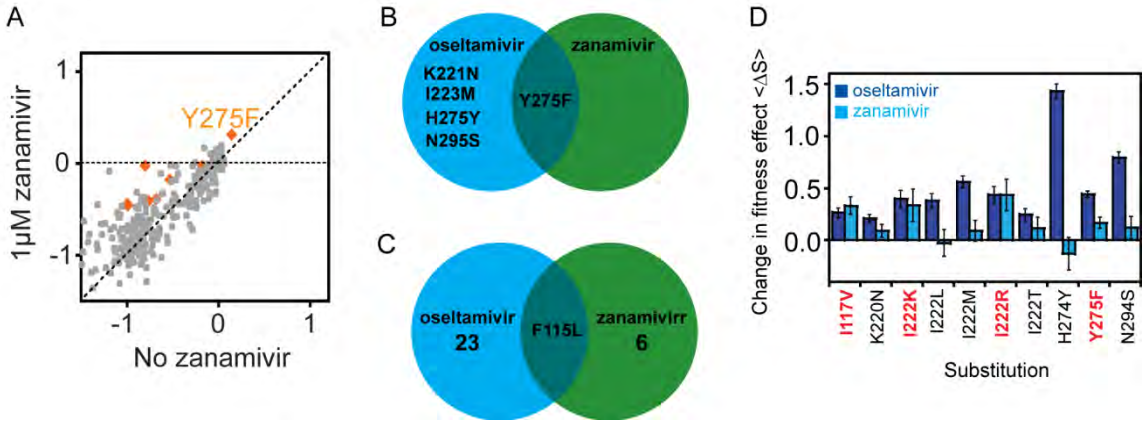


Figure 6.4 Comparison of fitness effects of all mutations +/- zanamivir revealed few mutations that were adaptive or responsive to zanamivir. (A) Comparison of fitness effect of all mutations with or without zanamivir. 0 indicates WT-like fitness while -1 corresponds to null-like fitness. Zanamivir responsive mutations (mutations with increased fitness effect with zanamivir than without) are colored in orange. Zanamivir adaptive mutations (mutations that are more fit than WT in the presence of zanamivir) are labeled by their positions and amino acid changes. (B) Venn diagram to show mutations that were adaptive to oseltamivir or zanamivir or both. (C) Venn diagram to show the number of mutations that were responsive to oseltamivir or zanamivir. The position and amino acid change of the mutation that was responsive to both oseltamivir and zanamivir was shown. (D) Bar graph to show changes in fitness effect (ΔS) +/- zanamivir or oseltamivir for a panel of oseltamivir responsive mutations. ΔS for zanamivir is shown in light blue, while ΔS for oseltamivir is shown in dark blue. Error bars are standard deviation with N = 3.

effect with zanamivir. Secondly, it was the same concentration used for oseltamivir bulk competition experiment, enabling comparison of fitness landscape of mutations with oseltamivir or zanamivir. As shown in figure 6.4A, most mutations had similar fitness in the presence or absence of zanamivir, indicating that zanamivir only affected the fitness effect of a small fraction of mutations. Thus the inhibitory effect of zanamivir on most mutations appeared to be similar to the WT. To identify mutations that had statistically significant differences in fitness effect +/- zanamivir, we used student t-test with multiple test correction to compare the fitness effect of mutations with or without zanamivir. We identified 7 mutations that had statistically significant higher fitness with zanamivir compared to without zanamivir (Table 6.2). These mutations were coined zanamivir responsive mutations. As shown in Chapter IV, responsive mutations to NAI frequently have reduced sensitivity to NAI. Notably, none of these mutations had been clinically associated with zanamivir resistance, demonstrating the power of EMPIRIC to identify novel mutations with reduced sensitivity to zanamivir. Several other mutations exhibited large increase in fitness with zanamivir compared to without zanamivir, but they were not statistically significant. Using a similar statistical approach, we identified only Y276F that was significantly more fit than the WT in the presence of zanamivir, indicating that Y276F was the only drug adaptive mutation under this experimental condition. In summary, we only identified very few mutations that exhibited adaptation or responsiveness to zanamivir, indicating high barrier for NA to evolve resistance to zanamivir.

Table 6.2: The fitness effect of zanamivir responsive mutations (+/- zanamivir)

Substitution	Fitness effect (no zanamivir)	Fitness effect (zanamivr)
K102N	-1.00	-0.44
I106L	-0.18	-0.03
R107S	-0.99	-0.46
F115L	-0.76	-0.41
E119A	-0.80	-0.03
H275Q	-0.53	-0.18
E277D	-0.68	-0.39

We sought to compare the resistance profile of zanamivir and oseltamivir (Figure 6.4B and C). As stated above, Y276F was the solely identified adaptive mutation to zanamivir. It was also adaptive to oseltamivir (Figure 6.4B), indicating the potential of this mutation to confer cross resistance to both oseltamivir and zanamivir. Four other mutations (K221N, I223M, H275Y and N295S) were adaptive to oseltamivir, but not zanamivir, so zanamivir might be administered as a substitute of oseltamivir to treat patients carrying these oseltamivir specific adaptive mutations. Of note, the predominant H275Y remained sensitive to zanamivir. The structural explanation is that the binding of zanamivir for NA does not rely on conformational change of E277 to create a hydrophobic pocket, which is disrupted by H275Y [202]. Moreover, only one mutation (F115L) exhibited responsiveness to both zanamivir and oseltamivir. The limited overlap of resistance profiles of oseltamivir and zanamivir suggests low probability of mutations simultaneously developing resistance to both zanamivir and oseltamivir. Notably, 24 mutations were responsive to oseltamivir while only seven mutations were responsive to zanamivir, indicating that NA is less likely to develop resistance to zanamivir than oseltamivir.

Most Oseltamivir responsive mutations were sensitive to zanamivir. As shown in Figure 6.4C, only F115L appeared to be responsive to both zanamivir and oseltamivir, indicating that the other 23 identified oseltamivir responsive mutations were still sensitive to zanamivir. Several of these mutations were isolated from patient samples and

shown to have reduced sensitivity to oseltamivir by *in vitro* biochemical assays (Figure 4.8C and Table 4.5). We compared their change in fitness effect in the presence and absence of oseltamivir or zanamivir (Figure 6.4D). Six out of the ten selected mutations showed little changes in fitness effect with or without zanamivir, confirming their sensitivity to zanamivir. Four of the ten mutations appeared to have increased fitness effect with zanamivir, though not statistically significant. It indicates that these four mutations (I117V, I223K, I223R and Y276F) exhibited a general trend to become refractory to zanamivir, but the trend was not as strong as that for oseltamivir. In summary, the majority of mutations that were responsive to oseltamivir remained sensitive to zanamivir.

We analyzed the structural basis underlying the higher resistance barrier for zanamivir. As competitive inhibitors of sialic acid, both zanamivir and oseltamivir resemble the chemical structure of sialic acid. Although they are built upon similar backbones, zanamivir entails minimal modification to the native structure of sialic acid. The large and basic group at C-4 of zanamivir strengthens interactions with conserved residues of NA without additional perturbations, while the hydrophobic group at C-6 of oseltamivir forces local conformational changes in NA. The binding of oseltamivir for NA requires reorientation of E277, which is not necessary for the binding of sialic acid, so it increases the possibility of mutations with reduced binding to oseltamivir, consistent with a substrate envelope hypothesis [410]. In summary, compared to oseltamivir,

zanamivir better resembled the natural substrate and selected for fewer mutations with adaptive potential.

Y276F was a drug adaptive mutation to both zanamivir and oseltamivir

We identified only Y276F that was adaptive to both zanamivir and oseltamivir (Figure 6.5). In the absence of any NAI, Y276F exhibited significantly positive fitness effect, indicating that it was more fit than the WT even without NAI. In fact, it showed the highest fitness effect among the pool of examined mutations (Figure 6.4A). In the presence of oseltamivir or zanamivir, Y276F became more fit and exhibited 30-50% higher fitness than the WT (Figure 6.5A), indicating that it was highly adapted to selection pressure from oseltamivir or zanamivir. Of note, Y276F has not been clinically associated with resistance to oseltamivir or zanamivir, so it is important to examine its adaptive potential to NAI in a clinical relevant IAV strain. Y276F has been isolated from patient samples sporadically in 2009 and 2010, indicating a high probability that it may have similar effect on clinical isolates.

We sought to explore the molecular mechanism of higher fitness of Y276F with oseltamivir or zanamivir. We measured the enzymatic activity of Y276F or WT to cleave a synthetic substrate (MUNANA) in increasing concentrations of oseltamivir by a fluorescent readout. Y276F appeared to have similar binding affinity to oseltamivir

Figure 6.5: Y275F was the sole identified zanamivir adaptive mutation

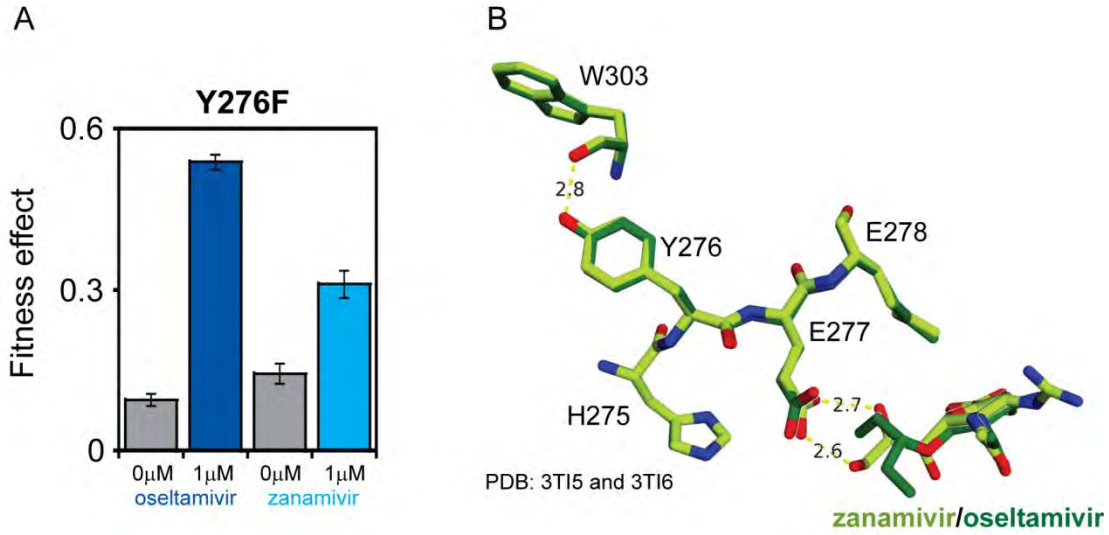


Figure 6.5 Y275F was the sole identified zanamivir adaptive mutation. (A) The fitness effect of Y275F in the presence or absence of oseltamivir or zanamivir. Two no drug controls from two independent viral propagation experiments were shown. Error bars are standard deviation with N =3. (B) Molecular structure image of residues 275-278 of NA bound to oseltamivir or zanamivir. The structure image of NA bound to oseltamivir (PDB ID: 3TI6, dark green) is superimposed onto the structure image of NA bound to zanamivir (PDB ID: 3TI5, light green). The distance between carboxyl group of E277 and glycerol group of zanamivir indicates formation of hydrogen bonds.

compared to the WT (Figure 4.9D). Thus, we predicted that its affinity to zanamivir might also be similar to the WT based on structure similarity of oseltamivir and zanamivir, though additional experiments are necessary for validation. Then we used the same MUNANA approach to measure the enzymatic activity of Y276F and WT normalized to the viral titer (infectious unit). Y276F had approximately 70% higher activity than WT, consistent with its higher fitness (Figure 4.4B and 4.9A). It also suggests that enhanced enzymatic activity of Y276F may contribute to its further increased fitness in the presence of either oseltamivir or zanamivir. One explanation is that it may create a buffer zone in NA activity, which compensates for the inhibition of NA activity by oseltamivir or zanamivir. The residual NA activity is able to support growth of IAV, leading to higher fitness of Y276F compared to WT.

To investigate the biophysical underpinnings of increased enzymatic activity of Y276F and adaptation of Y276F to oseltamivir and zanamivir, we examined the conformation of Tyr276 in the crystal structure of NA (Figure 6.5B). The Cα-Cβ of Tyr276 appeared to orient away from the binding substrate without forming any strong interaction with the substrate. In addition, the only difference between tyrosine and phenylalanine was that tyrosine had an extra hydroxyl group, but the hydroxul group pointed away from the substrate. Y276 is located between a framework residue H275 and two functional residues E277 and E278, so perturbations to Y276 may affect the conformation of these residues and hence binding interaction of sialic acid or NAI for NA.

Molecular dynamics analysis on NA with the WT sequence showed that Y276 formed a stable hydrogen bond with W303 through its hydroxyl group, so Y276F should break the hydrogen bond (personal communication, Prachanronarong). However, the effect of loss of the hydrogen bond has not been experimentally tested. One hypothesis is that disruption of the hydrogen bond makes residues 275-278 more flexible, enhancing the activity of the catalytic residue E278. Moreover, the increased flexibility may allow the region to restore the integrity of the hydrophobic pocket and better accommodate H275Y. Indeed, Y276F restored the enzymatic activity of H275Y and rescued its fitness defect, consistent with this hypothesis (data not shown).

Responsive mutations to zanamivir exhibited strong fitness defect without drug

We sought to explore common features shared by identified zanamivir responsive mutations. Six of the seven identified zanamivir responsive mutations showed strong fitness defect (>50% defect) in the absence of zanamivir (Figure 6.6A and table 6.3), indicating a strong fitness cost associated with adaptation to zanamivir. The severe cost of adaptation provides another rationale explanation for the difficulty of NA to evolve resistance to zanamivir: most mutations that adapt to selection pressure from zanamivir strongly impair the fitness of IAV in the absence of zanamivir, hindering their propagation and spread. Of note, the fitness of E119A was almost null-like without zanamivir, but became comparable to the WT with zanamivir, indicating that zanamivir exerted strong selection pressure on the replication of IAV. We mapped these zanamivir responsive mutations to the crystal structure of NA bound to zanamivir (Figure 6.6B).

Figure 6.6: Fitness effects of zanamivir responsive mutations

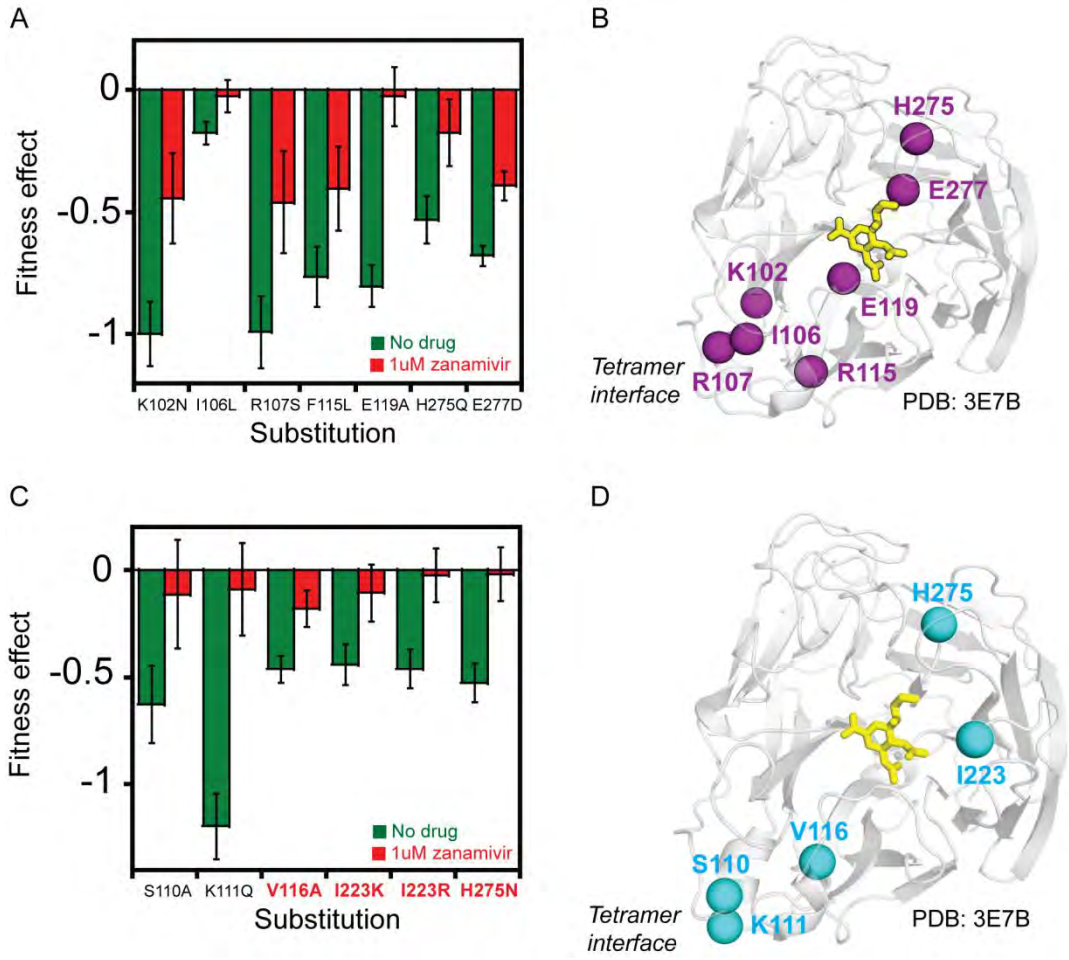


Figure 6.6 Fitness effects of zanamivir responsive mutations. (A) A bar graph of fitness effects of zanamivir responsive mutations with (red) or without (green) zanamivir. Error bars are standard deviation with N=3. (B) Molecular structure image of NA (PDB ID: 3B7E). Residues with zanamivir responsive are colored purple. (C) A bar graph of fitness effects of mutations with only small fitness defect in the presence of zanamivir. The fitness effect of mutations with zanamivir is colored in green, without zanamivir colored in red. Error bars are standard deviation with N=3. (D) Structure image of NA (PDB ID: 3B7E). Residues that harbor mutations with small fitness in the presence of zanamivir are colored light blue.

Table 6.3: The fitness effect of mutations with WT-like fitness in the presence of zanamivir

Substitution	Fitness effect (no zanamivir)	Fitness effect (zanamivr)
S110A	-0.63	-0.11
K111Q	-1.20	-0.09
V116A	-0.46	-0.18
I222K	-0.44	-0.11
I222R	-0.46	-0.03
H275N	-0.53	-0.02

Surprisingly, four of the seven identified zanamivir responsive mutations were within or close to the tetramizaiton interface. Three of them were almost lethal without zanamivir, indicating that tetramization of NA was required for the enzymatic function of NA. Mutations to the tetramization interface exhibited reduced sensitivity to zanamivir, indicating that tetramization of NA was also required for efficient binding of zanamivir for NA. The other three mutations (E119A, H275Q and E277D) occurred at functional or framework residues in the active site. They caused strong fitness defect, indicating that these mutations severely impaired NA function, presumably by reducing their binding affinity to NA; however they should also hinder binding of zanamivir to NA.

We also identified several mutations with little fitness defect (less than 20%) in the presence of zanamivir, though they were not more fit than WT. Most of them had approximately 45-80% fitness defect without zanamivir, except for K111Q, which appeared to be null-like (Figure 6.6C and Table 6.3). Although their increase in fitness effect with zanamivir was not statistically significant, collectively they displayed a general trend to be responsive to zanamivir. Moreover, three of the six mutations exhibited reduced sensitivity to zanamivir in IAV samples isolated from patients [249, 282, 283]. We mapped these mutations to the crystal structure of NA and discovered a similar spatial distribution to that of zanamivir responsive mutations (Figure 6.6D). These mutations were also clustered at the tetramization interface and the active site, confirming important roles of these regions in mediating adaptation to zanamivir. In particular, two

mutations were identified at residue 223 (I223K and I223R) and 275 (H275N and H275Q) respectively. Position 223 had been identified as a hotspot for oseltamivir responsive mutations (Figure 4.8). Amino acids with positive charges appeared to reduce binding of zanamivir for NA presumably by impairing interactions between C-5 of zanamivir and R225 of NA. Position 275 harbored the predominant N1 oseltamivir resistance mutation H275Y. Polar amino acids, instead of the parental positively charged histidine with an aromatic ring, appeared to affect the binding of zanamivir for NA. In summary, we identified several mutations that were responsive to zanamivir at the tetrameriation interface and active site of NA, but strong cost of adaptation likely hindered their fixation in human population.

To sum up, we identified a limited number mutations that exhibit adaptive potential to zanamivir. These mutations were clustered in the tetramizaiton interface or active site of NA and incurred strong fitness defect in the absence of zanamivir selection. Our results demonstrated the higher resistance barrier to zanamivir and revealed cost of adaptation as a predominant determinant of limited resistance to zanamivir.

Material and methods

The plasmid library construction, viral library recovery, bulk competition, sequencing sample preparation, sequencing and statistical analysis are described in the material and methods section of Chapter IV.

Bibliography

[1] Loeb DD, Swanstrom R, Everitt L, Manchester M, Stamper SE, Hutchison CA, 3rd. Complete mutagenesis of the HIV-1 protease. Nature. 1989;340:397-400.
[2] Stenson PD, Mort M, Ball EV, Howells K, Phillips AD, Thomas NS, et al. The Human Gene Mutation Database: 2008 update. Genome Med. 2009;1:13.
[3] Andre I, Bradley P, Wang C, Baker D. Prediction of the structure of symmetrical protein assemblies. Proc Natl Acad Sci U S A. 2007;104:17656-61.
[4] Koga N, Tatsumi-Koga R, Liu G, Xiao R, Acton TB, Montelione GT, et al. Principles for designing ideal protein structures. Nature. 2012;491:222-7.
[5] Crick F. Central dogma of molecular biology. Nature. 1970;227:561-3.
[6] Botstein D, Shortle D. Strategies and applications of in vitro mutagenesis. Science. 1985;229:1193-201.
[7] Muller HJ. Artificial Transmutation of the Gene. Science. 1927;66:84-7.
[8] Auerbach C, Robson JM. Tests of chemical substances for mutagenic action. Proc R Soc Edinb Biol. 1947;62:284-91.
[9] Beadle GW, Ephrussi B. The Differentiation of Eye Pigments in Drosophila as Studied by Transplantation. Genetics. 1936;21:225-47.
[10] Beadle GW, Tatum EL. Genetic Control of Biochemical Reactions in Neurospora. Proc Natl Acad Sci U S A. 1941;27:499-506.
[11] Ruvkun GB, Ausubel FM. A general method for site-directed mutagenesis in prokaryotes. Nature. 1981;289:85-8.
[12] Berg P, Mertz JE. Personal reflections on the origins and emergence of recombinant DNA technology. Genetics. 2010;184:9-17.
[13] Muller W, Weber H, Meyer F, Weissmann C. Site-directed mutagenesis in DNA: generation of point mutations in cloned beta globin complementary dna at the positions corresponding to amino acids 121 to 123. J Mol Biol. 1978;124:343-58.
[14] Shortle D, Grisafi P, Benkovic SJ, Botstein D. Gap misrepair mutagenesis: efficient site-directed induction of transition, transversion, and frameshift mutations in vitro. Proc Natl Acad Sci U S A. 1982;79:1588-92.
[15] Bartlett JM, Stirling D. A short history of the polymerase chain reaction. Methods Mol Biol. 2003;226:3-6.
[16] Carter P, Bedouelle H, Winter G. Improved oligonucleotide site-directed mutagenesis using M13 vectors. Nucleic Acids Res. 1985;13:4431-43.
[17] Braman J, Papworth C, Greener A. Site-directed mutagenesis using double-stranded plasmid DNA templates. Methods Mol Biol. 1996;57:31-44.
[18] Lo MC, Ha S, Pelczer I, Pal S, Walker S. The solution structure of the DNA-binding domain of Skn-1. Proc Natl Acad Sci U S A. 1998;95:8455-60.
[19] Cadwell RC, Joyce GF. Mutagenic PCR. PCR Methods Appl. 1994;3:S136-40.
[20] Ibarra-Molero B, Zitzewitz JA, Matthews CR. Salt-bridges can stabilize but do not accelerate the folding of the homodimeric coiled-coil peptide GCN4-p1. J Mol Biol. 2004;336:989-96.
[21] Carr J, Ives J, Kelly L, Lambkin R, Oxford J, Mendel D, et al. Influenza virus carrying neuraminidase with reduced sensitivity to oseltamivir carboxylate has altered properties in vitro and is compromised for infectivity and replicative ability in vivo. Antiviral Res. 2002;54:79-88.
[22] Chillakuri CR, Sheppard D, Lea SM, Handford PA. Notch receptor-ligand binding and activation: insights from molecular studies. Semin Cell Dev Biol. 2012;23:421-8.

[23] Oh Y, Chang KJ, Orlean P, Wloka C, Deshaies R, Bi E. Mitotic exit kinase Dbf2 directly phosphorylates chitin synthase Chs2 to regulate cytokinesis in budding yeast. Mol Biol Cell. 2012;23:2445-56.
[24] Mattheakis LC, Bhatt RR, Dower WJ. An in vitro polysome display system for identifying ligands from very large peptide libraries. Proc Natl Acad Sci U S A. 1994;91:9022-6.
[25] Smith MM, Shi L, Navre M. Rapid identification of highly active and selective substrates for stromelysin and matrilysin using bacteriophage peptide display libraries. J Biol Chem. 1995;270:6440-9.
[26] Francisco JA, Earhart CF, Georgiou G. Transport and anchoring of beta-lactamase to the external surface of Escherichia coli. Proc Natl Acad Sci U S A. 1992;89:2713-7.
[27] Feldhaus MJ, Siegel RW, Opresko LK, Coleman JR, Feldhaus JM, Yeung YA, et al. Flow-cytometric isolation of human antibodies from a nonimmune Saccharomyces cerevisiae surface display library. Nat Biotechnol. 2003;21:163-70.
[28] Bloom JD, Nayak JS, Baltimore D. A computational-experimental approach identifies mutations that enhance surface expression of an oseltamivir-resistant influenza neuraminidase. PLoS One. 2011;6:e22201.
[29] Levin AM, Weiss GA. Optimizing the affinity and specificity of proteins with molecular display. Mol Biosyst. 2006;2:49-57.
[30] Cunningham BC, Wells JA. High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. Science. 1989;244:1081-5.
[31] Kouadio JL, Horn JR, Pal G, Kossiakoff AA. Shotgun alanine scanning shows that growth hormone can bind productively to its receptor through a drastically minimized interface. J Biol Chem. 2005;280:25524-32.
[32] Simonsen SM, Sando L, Rosengren KJ, Wang CK, Colgrave ML, Daly NL, et al. Alanine scanning mutagenesis of the prototypic cyclotide reveals a cluster of residues essential for bioactivity. J Biol Chem. 2008;283:9805-13.
[33] Akbulut N, Tuzlakoglu Ozturk M, Pijning T, Issever Ozturk S, Gumusel F. Improved activity and thermostability of Bacillus pumilus lipase by directed evolution. J Biotechnol. 2013;164:123-9.
[34] Ma J, Wu L, Guo F, Gu J, Tang X, Jiang L, et al. Enhanced enantioselectivity of a carboxyl esterase from Rhodobacter sphaeroides by directed evolution. Appl Microbiol Biotechnol. 2013;97:4897-906.
[35] Acharya P, Rajakumara E, Sankaranarayanan R, Rao NM. Structural basis of selection and thermostability of laboratory evolved Bacillus subtilis lipase. J Mol Biol. 2004;341:1271-81.
[36] Hoseki J, Okamoto A, Masui R, Shibata T, Inoue Y, Yokoyama S, et al. Crystal structure of a family 4 uracil-DNA glycosylase from Thermus thermophilus HB8. J Mol Biol. 2003;333:515-26.
[37] Buonpane RA, Moza B, Sundberg EJ, Kranz DM. Characterization of T cell receptors engineered for high affinity against toxic shock syndrome toxin-1. J Mol Biol. 2005;353:308-21.
[38] Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409:860-921.
[39] Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet. 2010;11:31-46.
[40] Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, et al. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science. 2012;337:64-9.
[41] Gnad F, Baucom A, Mukhyala K, Manning G, Zhang Z. Assessment of computational methods for predicting the effects of missense mutations in human cancers. BMC Genomics. 2013;14 Suppl 3:S7.

[42] Gonzalez-Perez A, Lopez-Bigas N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. Am J Hum Genet. 2011;88:440-9.
[43] Cooper GM, Goode DL, Ng SB, Sidow A, Bamshad MJ, Shendure J, et al. Single-nucleotide evolutionary constraint scores highlight disease-causing mutations. Nat Methods. 2010;7:250-1.
[44] Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7:248-9.
[45] Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003;31:3812-4.
[46] Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46:310-5.
[47] Fowler DM, Fields S. Deep mutational scanning: a new style of protein science. Nat Methods. 2014;11:801-7.
[48] Gray DM, Gray CW, Yoo BH, Lou TF. Antisense DNA parameters derived from next-nearest-neighbor analysis of experimental data. BMC Bioinformatics. 2013;11:252.
[49] Boucher JI, Cote P, Flynn J, Jiang L, Laban A, Mishra P, et al. Viewing protein fitness landscapes through a next-gen lens. Genetics. 2014;198:461-71.
[50] Wu NC, Young AP, Dandekar S, Wijersuriya H, Al-Mawsawi LQ, Wu TT, et al. Systematic identification of H274Y compensatory mutations in influenza A virus neuraminidase by high-throughput screening. J Virol. 2013;87:1193-9.
[51] Hietpas RT, Jensen JD, Bolon DN. Experimental illumination of a fitness landscape. Proc Natl Acad Sci U S A. 2011;108:7896-901.
[52] Romero PA, Tran TM, Abate AR. Dissecting enzyme function with microfluidic-based deep mutational scanning. Proc Natl Acad Sci U S A. 2015;112:7159-64.
[53] Roscoe BP, Bolon DN. Systematic exploration of ubiquitin sequence, E1 activation efficiency, and experimental fitness in yeast. J Mol Biol. 2014;426:2854-70.
[54] Roscoe BP, Thayer KM, Zeldovich KB, Fushman D, Bolon DN. Analyses of the effects of all ubiquitin point mutants on yeast growth rate. J Mol Biol. 2013;425:1363-77.
[55] Stiffler MA, Hekstra DR, Ranganathan R. Evolvability as a function of purifying selection in TEM-1 beta-lactamase. Cell. 2012;160:882-92.
[56] Hietpas RT, Bank C, Jensen JD, Bolon DN. Shifting fitness landscapes in response to altered environments. Evolution. 2013;67:3512-22.
[57] McLaughlin RN, Jr., Poelwijk FJ, Raman A, Gosal WS, Ranganathan R. The spatial architecture of protein function and adaptation. Nature. 2012;491:138-42.
[58] Fowler DM, Araya CL, Fleishman SJ, Kellogg EH, Stephany JJ, Baker D, et al. High-resolution mapping of protein sequence-function relationships. Nat Methods. 2010;7:741-6.
[59] Araya CL, Fowler DM, Chen W, Muniez I, Kelly JW, Fields S. A fundamental protein property, thermodynamic stability, revealed solely from large-scale measurements of protein function. Proc Natl Acad Sci U S A. 2012;109:16858-63.
[60] Starita LM, Pruneda JN, Lo RS, Fowler DM, Kim HJ, Hiatt JB, et al. Activity-enhancing mutations in an E3 ubiquitin ligase identified by high-throughput mutagenesis. Proc Natl Acad Sci U S A. 2013;110:E1263-72.
[61] Melamed D, Young DL, Gamble CE, Miller CR, Fields S. Deep mutational scanning of an RRM domain of the Saccharomyces cerevisiae poly(A)-binding protein. Rna. 2013;19:1537-51.
[62] Starita LM, Young DL, Islam M, Kitzman JO, Gullingsrud J, Hause RJ, et al. Massively Parallel Functional Analysis of BRCA1 RING Domain Variants. Genetics. 2015;200:413-22.
[63] Bank C, Hietpas RT, Jensen JD, Bolon DN. A systematic survey of an intragenic epistatic landscape. Mol Biol Evol. 2015;32:229-38.

[64] Drake JW, Holland JJ. Mutation rates among RNA viruses. Proc Natl Acad Sci U S A. 1999;96:13910-3.
[65] Domingo E, Holland JJ. RNA virus mutations and fitness for survival. Annu Rev Microbiol. 1997;51:151-78.
[66] Domingo E, Menendez-Arias L, Holland JJ. RNA virus fitness. Rev Med Virol. 1997;7:87-96.
[67] Liao HX, Lynch R, Zhou T, Gao F, Alam SM, Boyd SD, et al. Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. Nature. 2013;496:469-76.
[68] Renzette N, Caffrey DR, Zeldovich KB, Liu P, Gallagher GR, Aiello D, et al. Evolution of the influenza A virus genome during development of oseltamivir resistance in vitro. J Virol. 2014;88:272-81.
[69] Herfst S, Schrauwen EJ, Linster M, Chutinimitkul S, de Wit E, Munster VJ, et al. Airborne transmission of influenza A/H5N1 virus between ferrets. Science. 2012;336:1534-41.
[70] Russell CA, Fonville JM, Brown AE, Burke DF, Smith DL, James SL, et al. The potential for respiratory droplet-transmissible A/H5N1 influenza virus to evolve in a mammalian host. Science. 2012;336:1541-7.
[71] Heaton NS, Sachs D, Chen CJ, Hai R, Palese P. Genome-wide mutagenesis of influenza virus reveals unique plasticity of the hemagglutinin and NS1 proteins. Proc Natl Acad Sci U S A. 2013;110:20248-53.
[72] Bloom JD. An experimentally determined evolutionary model dramatically improves phylogenetic fit. Mol Biol Evol. 2014;31:1956-78.
[73] Thyagarajan B, Bloom JD. The inherent mutational tolerance and antigenic evolvability of influenza hemagglutinin. Elife. 2015;3.
[74] Wu NC, Olson CA, Du Y, Le S, Tran K, Remenyi R, et al. Functional Constraint Profiling of a Viral Protein Reveals Discordance of Evolutionary Conservation and Functionality. PLoS Genet. 2015;11:e1005310.
[75] Wu NC, Young AP, Al-Mawsawi LQ, Olson CA, Feng J, Qi H, et al. High-throughput identification of loss-of-function mutations for anti-interferon activity in the influenza A virus NS segment. J Virol. 2014;88:10157-64.
[76] Qi H, Olson CA, Wu NC, Ke R, Loverdo C, Chu V, et al. A quantitative high-resolution genetic profile rapidly identifies sequence determinants of hepatitis C viral fitness and drug sensitivity. PLoS Pathog. 2014;10:e1004064.
[77] Wu NC, Young AP, Al-Mawsawi LQ, Olson CA, Feng J, Qi H, et al. High-throughput profiling of influenza A virus hemagglutinin gene at single-nucleotide resolution. Sci Rep. 2014;4:4942.
[78] Soskine M, Tawfik DS. Mutational effects and the evolution of new protein functions. Nat Rev Genet. 2010;11:572-82.
[79] Mumberg D, Muller R, Funk M. Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. Gene. 1995;156:119-22.
[80] Zaret KS, Sherman F. Mutationally altered 3' ends of yeast CYC1 mRNA affect transcript stability and translational efficiency. J Mol Biol. 1984;177:107-35.
[81] Liu J, Schmitz JC, Lin X, Tai N, Yan W, Farrell M, et al. Thymidylate synthase as a translational regulator of cellular gene expression. Biochim Biophys Acta. 2002;1587:174-82.
[82] Bergman Y, Cedar H. DNA methylation dynamics in health and disease. Nat Struct Mol Biol. 2013;20:274-81.
[83] Wagner A. The molecular origins of evolutionary innovations. Trends Genet. 2011;27:397-410.

[84] Manceau M, Domingues VS, Mallarino R, Hoekstra HE. The developmental role of Agouti in color pattern evolution. Science. 2011;331:1062-5.
[85] Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH. Why highly expressed proteins evolve slowly. Proc Natl Acad Sci U S A. 2005;102:14338-43.
[86] Dekel E, Alon U. Optimality and evolutionary tuning of the expression level of a protein. Nature. 2005;436:588-92.
[87] Ibarra RU, Edwards JS, Palsson BO. Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. Nature. 2002;420:186-9.
[88] Lenski RE, Ofria C, Pennock RT, Adami C. The evolutionary origin of complex features. Nature. 2003;423:139-44.
[89] Liebermeister W, Klipp E, Schuster S, Heinrich R. A theory of optimal differential gene expression. Biosystems. 2004;76:261-78.
[90] Bershtein S, Mu W, Serohijos AW, Zhou J, Shakhnovich EI. Protein quality control acts on folding intermediates to shape the effects of mutations on organismal fitness. Mol Cell. 2013;49:133-44.
[91] Borkovich KA, Farrelly FW, Finkelstein DB, Taulien J, Lindquist S. hsp82 is an essential protein that is required in higher concentrations for growth of cells at higher temperatures. Mol Cell Biol. 1989;9:3919-30.
[92] Lunzer M, Miller SP, Felsheim R, Dean AM. The biochemical architecture of an ancient adaptive landscape. Science. 2005;310:499-501.
[93] Rennell D, Bouvier SE, Hardy LW, Poteete AR. Systematic mutation of bacteriophage T4 lysozyme. J Mol Biol. 1991;222:67-88.
[94] Kacser H, Burns JA. The molecular basis of dominance. Genetics. 1981;97:639-66.
[95] Kacser H, Burns JA. The control of flux. Biochem Soc Trans. 1995;23:341-66.
[96] McKenzie SL, Henikoff S, Meselson M. Localization of RNA from heat-induced polysomes at puff sites in Drosophila melanogaster. Proc Natl Acad Sci U S A. 1975;72:1117-21.
[97] Taipale M, Jarosz DF, Lindquist S. Hsp90 at the hub of protein homeostasis: emerging mechanistic insights. Nat Rev Mol Cell Biol. 2010;11:515-28.
[98] Bardwell JC, Craig EA. Eukaryotic Mr 83,000 heat shock protein has a homologue in Escherichia coli. Proc Natl Acad Sci U S A. 1987;84:5177-81.
[99] Crevel G, Bates H, Huikeshoven H, Cotterill S. The Drosophila Dpit47 protein is a nuclear Hsp90 co-chaperone that interacts with DNA polymerase alpha. J Cell Sci. 2001;114:2015-25.
[100] Nadeau K, Das A, Walsh CT. Hsp90 chaperonins possess ATPase activity and bind heat shock transcription factors and peptidyl prolyl isomerases. J Biol Chem. 1993;268:1479-87.
[101] Ammirante M, Rosati A, Gentilella A, Festa M, Petrella A, Marzullo L, et al. The activity of hsp90 alpha promoter is regulated by NF-kappa B transcription factors. Oncogene. 2008;27:1175-8.
[102] Stephanou A, Amin V, Isenberg DA, Akira S, Kishimoto T, Latchman DS. Interleukin 6 activates heat-shock protein 90 beta gene expression. Biochem J. 1997;321 (Pt 1):103-6.
[103] Ripley BJ, Stephanou A, Isenberg DA, Latchman DS. Interleukin-10 activates heat-shock protein 90beta gene expression. Immunology. 1999;97:226-31.
[104] Ali MM, Roe SM, Vaughan CK, Meyer P, Panaretou B, Piper PW, et al. Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex. Nature. 2006;440:1013-7.
[105] Wayne N, Lai Y, Pullen L, Bolon DN. Modular control of cross-oligomerization: analysis of superstabilized Hsp90 homodimers in vivo. J Biol Chem. 2010;285:234-41.
[106] Prodromou C, Roe SM, O'Brien R, Ladbury JE, Piper PW, Pearl LH. Identification and structural characterization of the ATP/ADP-binding site in the Hsp90 molecular chaperone. Cell. 1997;90:65-75.

[107] Meyer P, Prodromou C, Hu B, Vaughan C, Roe SM, Panaretou B, et al. Structural and functional analysis of the middle segment of hsp90: implications for ATP hydrolysis and client protein and cochaperone interactions. Mol Cell. 2003;11:647-58.
[108] Hainzl O, Lapina MC, Buchner J, Richter K. The charged linker region is an important regulator of Hsp90 function. J Biol Chem. 2009;284:22559-67.
[109] Retzlaff M, Hagn F, Mitschke L, Hessling M, Gugel F, Kessler H, et al. Asymmetric activation of the hsp90 dimer by its cochaperone aha1. Mol Cell. 2010;37:344-54.
[110] Fontana J, Fulton D, Chen Y, Fairchild TA, McCabe TJ, Fujita N, et al. Domain mapping studies reveal that the M domain of hsp90 serves as a molecular scaffold to regulate Akt-dependent phosphorylation of endothelial nitric oxide synthase and NO release. Circ Res. 2002;90:866-73.
[111] Sato S, Fujita N, Tsuruo T. Modulation of Akt kinase activity by binding to Hsp90. Proc Natl Acad Sci U S A. 2000;97:10832-7.
[112] Pearl LH, Prodromou C. Structure and mechanism of the Hsp90 molecular chaperone machinery. Annu Rev Biochem. 2006;75:271-94.
[113] Young JC, Obermann WM, Hartl FU. Specific binding of tetratricopeptide repeat proteins to the C-terminal 12-kDa domain of hsp90. J Biol Chem. 1998;273:18007-10.
[114] Hessling M, Richter K, Buchner J. Dissection of the ATP-induced conformational cycle of the molecular chaperone Hsp90. Nat Struct Mol Biol. 2009;16:287-93.
[115] Cintron NS, Toft D. Defining the requirements for Hsp40 and Hsp70 in the Hsp90 chaperone pathway. J Biol Chem. 2006;281:26235-44.
[116] McLaughlin SH, Sobott F, Yao ZP, Zhang W, Nielsen PR, Grossmann JG, et al. The co-chaperone p23 arrests the Hsp90 ATPase cycle to trap client proteins. J Mol Biol. 2006;356:746-58.
[117] Meyer P, Prodromou C, Liao C, Hu B, Roe SM, Vaughan CK, et al. Structural basis for recruitment of the ATPase activator Aha1 to the Hsp90 chaperone machinery. Embo J. 2004;23:1402-10.
[118] Roe SM, Ali MM, Meyer P, Vaughan CK, Panaretou B, Piper PW, et al. The Mechanism of Hsp90 regulation by the protein kinase-specific cochaperone p50(cdc37). Cell. 2004;116:87-98.
[119] Silverstein AM, Grammatikakis N, Cochran BH, Chinkers M, Pratt WB. p50(cdc37) binds directly to the catalytic domain of Raf as well as to a site on hsp90 that is topologically adjacent to the tetratricopeptide repeat binding site. J Biol Chem. 1998;273:20090-5.
[120] Falsone SF, Leptihn S, Osterauer A, Haslbeck M, Buchner J. Oncogenic mutations reduce the stability of SRC kinase. J Mol Biol. 2004;344:281-91.
[121] McClellan AJ, Tam S, Kaganovich D, Frydman J. Protein quality control: chaperones culling corrupt conformations. Nat Cell Biol. 2005;7:736-41.
[122] Millson SH, Truman AW, King V, Prodromou C, Pearl LH, Piper PW. A two-hybrid screen of the yeast proteome for Hsp90 interactors uncovers a novel Hsp90 chaperone requirement in the activity of a stress-activated mitogen-activated protein kinase, Slt2p (Mpk1p). Eukaryot Cell. 2005;4:849-60.
[123] Zhao R, Davey M, Hsu YC, Kaplanek P, Tong A, Parsons AB, et al. Navigating the chaperone network: an integrative map of physical and genetic interactions mediated by the hsp90 chaperone. Cell. 2005;120:715-27.
[124] Jakob U, Lilie H, Meyer I, Buchner J. Transient interaction of Hsp90 with early unfolding intermediates of citrate synthase. Implications for heat shock in vivo. J Biol Chem. 1995;270:7288-94.
[125] Queitsch C, Sangster TA, Lindquist S. Hsp90 as a capacitor of phenotypic variation. Nature. 2002;417:618-24.

[126] Rutherford SL, Lindquist S. Hsp90 as a capacitor for morphological evolution. Nature. 1998;396:336-42.
[127] Cowen LE, Lindquist S. Hsp90 potentiates the rapid evolution of new traits: drug resistance in diverse fungi. Science. 2005;309:2185-9.
[128] Jarosz DF, Lindquist S. Hsp90 and environmental stress transform the adaptive value of natural genetic variation. Science. 2010;330:1820-4.
[129] Sangster TA, Lindquist S, Queitsch C. Under cover: causes, effects and implications of Hsp90-mediated genetic capacitance. Bioessays. 2004;26:348-62.
[130] Yeyati PL, Bancewicz RM, Maule J, van Heyningen V. Hsp90 selectively modulates phenotype in vertebrate development. PLoS Genet. 2007;3:e43.
[131] Faria NR, Rambaut A, Suchard MA, Baele G, Bedford T, Ward MJ, et al. HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. Science. 2014;346:56-61.
[132] Worobey M, Gemmel M, Teuwen DE, Haselkorn T, Kunstman K, Bunce M, et al. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. Nature. 2008;455:661-4.
[133] Gottlieb MS, Schroff R, Schanker HM, Weisman JD, Fan PT, Wolf RA, et al. Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. N Engl J Med. 1981;305:1425-31.
[134] Barre-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, et al. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science. 1983;220:868-71.
[135] Gallo RC, Sarin PS, Gelmann EP, Robert-Guroff M, Richardson E, Kalyanaraman VS, et al. Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). Science. 1983;220:865-7.
[136] Sanchez-Pescador R, Power MD, Barr PJ, Steimer KS, Stempien MM, Brown-Shimer SL, et al. Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2). Science. 1985;227:484-92.
[137] Wain-Hobson S, Sonigo P, Danos O, Cole S, Alizon M. Nucleotide sequence of the AIDS virus, LAV. Cell. 1985;40:9-17.
[138] Piatak M, Jr., Saag MS, Yang LC, Clark SJ, Kappes JC, Luk KC, et al. High levels of HIV-1 in plasma during all stages of infection determined by competitive PCR. Science. 1993;259:1749-54.
[139] Ribeiro RM, Bonhoeffer S. Production of resistant HIV mutants during antiretroviral therapy. Proc Natl Acad Sci U S A. 2000;97:7681-6.
[140] Collier AC, Coombs RW, Schoenfeld DA, Bassett RL, Timpone J, Baruch A, et al. Treatment of human immunodeficiency virus infection with saquinavir, zidovudine, and zalcitabine. AIDS Clinical Trials Group. N Engl J Med. 1996;334:1011-7.
[141] Gulick RM, Mellors JW, Havlir D, Eron JJ, Gonzalez C, McMahon D, et al. Treatment with indinavir, zidovudine, and lamivudine in adults with human immunodeficiency virus infection and prior antiretroviral therapy. N Engl J Med. 1997;337:734-9.
[142] Hammer SM, Squires KE, Hughes MD, Grimes JM, Demeter LM, Currier JS, et al. A controlled trial of two nucleoside analogues plus indinavir in persons with human immunodeficiency virus infection and CD4 cell counts of 200 per cubic millimeter or less. AIDS Clinical Trials Group 320 Study Team. N Engl J Med. 1997;337:725-33.
[143] Perelson AS, Essunger P, Cao Y, Vesanen M, Hurley A, Saksela K, et al. Decay characteristics of HIV-1-infected compartments during combination therapy. Nature. 1997;387:188-91.

[144] Deeks SG, Lewin SR, Havlir DV. The end of AIDS: HIV infection as a chronic disease. Lancet. 2013;382:1525-33.
[145] Fauci AS, Folkers GK. Toward an AIDS-free generation. Jama. 2012;308:343-4.
[146] Greenwood B. The contribution of vaccination to global health: past, present and future. Philos Trans R Soc Lond B Biol Sci. 2014;369:20130433.
[147] Lievano F, Galea SA, Thornton M, Wiedmann RT, Manoff SB, Tran TN, et al. Measles, mumps, and rubella virus vaccine (M-M-RII): a review of 32 years of clinical and postmarketing experience. Vaccine. 2012;30:6918-26.
[148] Voigt EA, Kennedy RB, Poland GA. Defending against smallpox: a focus on vaccines. Expert Rev Vaccines. 2016:1-15.
[149] Zachariah P, Stockwell MS. Measles vaccine: Past, present, and future. J Clin Pharmacol. 2016;56:133-40.
[150] Esparza J. A brief history of the global effort to develop a preventive HIV vaccine. Vaccine. 2013;31:3502-18.
[151] Corti D, Lanzavecchia A. Broadly neutralizing antiviral antibodies. Annu Rev Immunol. 2013;31:705-42.
[152] Kim JH, Excler JL, Michael NL. Lessons from the RV144 Thai phase III HIV-1 vaccine trial and the search for correlates of protection. Annu Rev Med. 2015;66:423-37.
[153] Wyatt R, Sodroski J. The HIV-1 envelope glycoproteins: fusogens, antigens, and immunogens. Science. 1998;280:1884-8.
[154] Dalgleish AG, Beverley PC, Clapham PR, Crawford DH, Greaves MF, Weiss RA. The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. Nature. 1984;312:763-7.
[155] Klatzmann D, Champagne E, Chamaret S, Gruest J, Guetard D, Hercend T, et al. T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. Nature. 1984;312:767-8.
[156] Feng Y, Broder CC, Kennedy PE, Berger EA. HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. Science. 1996;272:872-7.
[157] Moore JP. Coreceptors: implications for HIV pathogenesis and therapy. Science. 1997;276:51-2.
[158] Starcich BR, Hahn BH, Shaw GM, McNeely PD, Modrow S, Wolf H, et al. Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS. Cell. 1986;45:637-48.
[159] Julien JP, Cupo A, Sok D, Stanfield RL, Lyumkis D, Deller MC, et al. Crystal structure of a soluble cleaved HIV-1 envelope trimer. Science. 2013;342:1477-83.
[160] Burton DR, Mascola JR. Antibody responses to envelope glycoproteins in HIV-1 infection. Nat Immunol. 2015;16:571-6.
[161] Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature. 1998;393:648-59.
[162] Huang CC, Lam SN, Acharya P, Tang M, Xiang SH, Hussan SS, et al. Structures of the CCR5 N terminus and of a tyrosine-sulfated antibody with HIV-1 gp120 and CD4. Science. 2007;317:1930-4.
[163] Huang CC, Tang M, Zhang MY, Majeed S, Montabana E, Stanfield RL, et al. Structure of a V3-containing HIV-1 gp120 core. Science. 2005;310:1025-8.
[164] Kwon YD, Finzi A, Wu X, Dogo-Isonagie C, Lee LK, Moore LR, et al. Unliganded HIV-1 gp120 core structures assume the CD4-bound conformation with regulation by quaternary interactions and variable loops. Proc Natl Acad Sci U S A. 2012;109:5663-8.

[165] Pancera M, Zhou T, Druz A, Georgiev IS, Soto C, Gorman J, et al. Structure and immune recognition of trimeric pre-fusion HIV-1 Env. Nature. 2014;514:455-61.
[166] Zhou T, Xu L, Dey B, Hessell AJ, Van Ryk D, Xiang SH, et al. Structural definition of a conserved neutralization epitope on HIV-1 gp120. Nature. 2007;445:732-7.
[167] Myszka DG, Sweet RW, Hensley P, Brigham-Burke M, Kwong PD, Hendrickson WA, et al. Energetics of the HIV gp120-CD4 binding reaction. Proc Natl Acad Sci U S A. 2000;97:9026-31.
[168] O'Connell O, Repik A, Reeves JD, Gonzalez-Perez MP, Quitadamo B, Anton ED, et al. Efficiency of bridging-sheet recruitment explains HIV-1 R5 envelope glycoprotein sensitivity to soluble CD4 and macrophage tropism. J Virol. 2013;87:187-98.
[169] Mao Y, Wang L, Gu C, Herschhorn A, Xiang SH, Haim H, et al. Subunit organization of the membrane-bound HIV-1 envelope glycoprotein trimer. Nat Struct Mol Biol. 2012;19:893-9.
[170] Wei X, Decker JM, Wang S, Hui H, Kappes JC, Wu X, et al. Antibody neutralization and escape by HIV-1. Nature. 2003;422:307-12.
[171] Kwong PD, Doyle ML, Casper DJ, Cicala C, Leavitt SA, Majeed S, et al. HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. Nature. 2002;420:678-82.
[172] Bartesaghi A, Merk A, Borgnia MJ, Milne JL, Subramaniam S. Prefusion structure of trimeric HIV-1 envelope glycoprotein determined by cryo-electron microscopy. Nat Struct Mol Biol. 2013;20:1352-7.
[173] Sanders RW, Derking R, Cupo A, Julien JP, Yasmeen A, de Val N, et al. A next-generation cleaved, soluble HIV-1 Env trimer, BG505 SOSIP.664 gp140, expresses multiple epitopes for broadly neutralizing but not non-neutralizing antibodies. PLoS Pathog. 2013;9:e1003618.
[174] Pikora CA. Glycosylation of the ENV spike of primate immunodeficiency viruses and antibody neutralization. Curr HIV Res. 2004;2:243-54.
[175] Ratner L, Fisher A, Jagodzinski LL, Mitsuya H, Liou RS, Gallo RC, et al. Complete nucleotide sequences of functional clones of the AIDS virus. AIDS Res Hum Retroviruses. 1987;3:57-69.
[176] Xiang SH, Kwong PD, Gupta R, Rizzuto CD, Casper DJ, Wyatt R, et al. Mutagenic stabilization and/or disruption of a CD4-bound state reveals distinct conformations of the human immunodeficiency virus type 1 gp120 envelope glycoprotein. J Virol. 2002;76:9888-99.
[177] Duenas-Decamp MJ, Peters PJ, Burton D, Clapham PR. Determinants flanking the CD4 binding loop modulate macrophage tropism of human immunodeficiency virus type 1 R5 envelopes. J Virol. 2009;83:2575-83.
[178] Schader SM, Colby-Germinario SP, Quashie PK, Oliveira M, Ibanescu RI, Moisi D, et al. HIV gp120 H375 is unique to HIV-1 subtype CRF01_AE and confers strong resistance to the entry inhibitor BMS-599793, a candidate microbicide drug. Antimicrob Agents Chemother. 2012;56:4257-67.
[179] Finzi A, Pacheco B, Xiang SH, Pancera M, Herschhorn A, Wang L, et al. Lineage-specific differences between human and simian immunodeficiency virus regulation of gp120 trimer association and CD4 binding. J Virol. 2012;86:8974-86.
[180] Zhou T, Georgiev I, Wu X, Yang ZY, Dai K, Finzi A, et al. Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. Science. 2010;329:811-7.
[181] Zhou T, Lynch RM, Chen L, Acharya P, Wu X, Doria-Rose NA, et al. Structural Repertoire of HIV-1-Neutralizing Antibodies Targeting the CD4 Supersite in 14 Donors. Cell. 2015;161:1280-92.
[182] Duenas-Decamp MJ, Clapham PR. HIV-1 gp120 determinants proximal to the CD4 binding site shift protective glycans that are targeted by monoclonal antibody 2G12. J Virol. 2010;84:9608-12.

[183] Taubenberger JK, Kash JC. Influenza virus evolution, host adaptation, and pandemic formation. Cell Host Microbe. 2010;7:440-51.
[184] Webster RG, Bean WJ, Gorman OT, Chambers TM, Kawaoka Y. Evolution and ecology of influenza A viruses. Microbiol Rev. 1992;56:152-79.
[185] Molinari NA, Ortega-Sanchez IR, Messonnier ML, Thompson WW, Wortley PM, Weintraub E, et al. The annual impact of seasonal influenza in the US: measuring disease burden and costs. Vaccine. 2007;25:5086-96.
[186] Johnson NP, Mueller J. Updating the accounts: global mortality of the 1918-1920 "Spanish" influenza pandemic. Bull Hist Med. 2002;76:105-15.
[187] Taubenberger JK, Reid AH, Krafft AE, Bijwaard KE, Fanning TG. Initial genetic characterization of the 1918 "Spanish" influenza virus. Science. 1997;275:1793-6.
[188] Kash JC, Tumpey TM, Proll SC, Carter V, Perwitasari O, Thomas MJ, et al. Genomic analysis of increased host immune and cell death responses induced by 1918 influenza virus. Nature. 2006;443:578-81.
[189] Kobasa D, Jones SM, Shinya K, Kash JC, Copps J, Ebihara H, et al. Aberrant innate immune response in lethal infection of macaques with the 1918 influenza virus. Nature. 2007;445:319-23.
[190] Neumann G, Watanabe T, Ito H, Watanabe S, Goto H, Gao P, et al. Generation of influenza A viruses entirely from cloned cDNAs. Proc Natl Acad Sci U S A. 1999;96:9345-50.
[191] Kawaoka Y, Krauss S, Webster RG. Avian-to-human transmission of the PB1 gene of influenza A viruses in the 1957 and 1968 pandemics. J Virol. 1989;63:4603-8.
[192] Scholtissek C, Rohde W, Von Hoyningen V, Rott R. On the origin of the human influenza virus subtypes H2N2 and H3N2. Virology. 1978;87:13-20.
[193] Garten RJ, Davis CT, Russell CA, Shu B, Lindstrom S, Balish A, et al. Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. Science. 2009;325:197-201.
[194] Dawood FS, Iuliano AD, Reed C, Meltzer MI, Shay DK, Cheng PY, et al. Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: a modelling study. Lancet Infect Dis. 2012;12:687-95.
[195] Chen J, Lee KH, Steinhauer DA, Stevens DJ, Skehel JJ, Wiley DC. Structure of the hemagglutinin precursor cleavage site, a determinant of influenza pathogenicity and the origin of the labile conformation. Cell. 1998;95:409-17.
[196] Gamblin SJ, Haire LF, Russell RJ, Stevens DJ, Xiao B, Ha Y, et al. The structure and receptor binding properties of the 1918 influenza hemagglutinin. Science. 2004;303:1838-42.
[197] Xu X, Zhu X, Dwek RA, Stevens J, Wilson IA. Structural characterization of the 1918 influenza virus H1N1 neuraminidase. J Virol. 2008;82:10493-501.
[198] Kawaoka Y, Webster RG. Sequence requirements for cleavage activation of influenza virus hemagglutinin expressed in mammalian cells. Proc Natl Acad Sci U S A. 1988;85:324-8.
[199] Van Poucke SG, Nicholls JM, Nauwynck HJ, Van Reeth K. Replication of avian, human and swine influenza viruses in porcine respiratory explants and association with sialic acid distribution. Virol J. 2010;7:38.
[200] Shinya K, Ebina M, Yamada S, Ono M, Kasai N, Kawaoka Y. Avian flu: influenza virus receptors in the human airway. Nature. 2006;440:435-6.
[201] van Riel D, Munster VJ, de Wit E, Rimmelzwaan GF, Fouchier RA, Osterhaus AD, et al. H5N1 Virus Attachment to Lower Respiratory Tract. Science. 2006;312:399.
[202] Collins PJ, Haire LF, Lin YP, Liu J, Russell RJ, Walker PA, et al. Crystal structures of oseltamivir-resistant influenza virus neuraminidase mutants. Nature. 2008;453:1258-61.

[203] Russell RJ, Haire LF, Stevens DJ, Collins PJ, Lin YP, Blackburn GM, et al. The structure of H5N1 avian influenza neuraminidase suggests new opportunities for drug design. Nature. 2006;443:45-9.
[204] Varghese JN, Colman PM. Three-dimensional structure of the neuraminidase of influenza virus A/Tokyo/3/67 at 2.2 A resolution. J Mol Biol. 1991;221:473-86.
[205] Varghese JN, Laver WG, Colman PM. Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 A resolution. Nature. 1983;303:35-40.
[206] Taylor NR, von Itzstein M. Molecular modeling studies on ligand binding to sialidase from influenza virus and the mechanism of catalysis. J Med Chem. 1994;37:616-24.
[207] Abed Y, Pizzorno A, Bouhy X, Boivin G. Role of permissive neuraminidase mutations in influenza A/Brisbane/59/2007-like (H1N1) viruses. PLoS Pathog. 2011;7:e1002431.
[208] Chong AK, Pegg MS, Taylor NR, von Itzstein M. Evidence for a sialosyl cation transition-state complex in the reaction of sialidase from influenza virus. Eur J Biochem. 1992;207:335-43.
[209] Wagner R, Matrosovich M, Klenk HD. Functional balance between haemagglutinin and neuraminidase in influenza virus infections. Rev Med Virol. 2002;12:159-66.
[210] Kaverin NV, Matrosovich MN, Gambaryan AS, Rudneva IA, Shilov AA, Varich NL, et al. Intergenic HA-NA interactions in influenza A virus: postreassortment substitutions of charged amino acid in the hemagglutinin of different subtypes. Virus Res. 2000;66:123-9.
[211] Mitnaul LJ, Matrosovich MN, Castrucci MR, Tuzikov AB, Bovin NV, Kobasa D, et al. Balanced hemagglutinin and neuraminidase activities are critical for efficient replication of influenza A virus. J Virol. 2000;74:6015-20.
[212] Wagner R, Wolff T, Herwig A, Pleschka S, Klenk HD. Interdependence of hemagglutinin glycosylation and neuraminidase as regulators of influenza virus growth: a study by reverse genetics. J Virol. 2000;74:6316-23.
[213] Baigent SJ, McCauley JW. Glycosylation of haemagglutinin and stalk-length of neuraminidase combine to regulate the growth of avian influenza viruses in tissue culture. Virus Res. 2001;79:177-85.
[214] Xu R, Zhu X, McBride R, Nycholat CM, Yu W, Paulson JC, et al. Functional balance of the hemagglutinin and neuraminidase activities accompanies the emergence of the 2009 H1N1 influenza pandemic. J Virol. 2012;86:9221-32.
[215] Yen HL, Liang CH, Wu CY, Forrest HL, Ferguson A, Choy KT, et al. Hemagglutinin-neuraminidase balance confers respiratory-droplet transmissibility of the pandemic H1N1 influenza virus in ferrets. Proc Natl Acad Sci U S A. 2011;108:14264-9.
[216] Nobusawa E, Sato K. Comparison of the mutation rates of human influenza A and B viruses. J Virol. 2006;80:3675-8.
[217] Parvin JD, Moscona A, Pan WT, Leider JM, Palese P. Measurement of the mutation rates of animal viruses: influenza A virus and poliovirus type 1. J Virol. 1986;59:377-83.
[218] Ferguson NM, Galvani AP, Bush RM. Ecological and immunological determinants of influenza evolution. Nature. 2003;422:428-33.
[219] Couch RB, Atmar RL, Franco LM, Quarles JM, Wells J, Arden N, et al. Antibody correlates and predictors of immunity to naturally occurring influenza in humans and the importance of antibody to the neuraminidase. J Infect Dis. 2013;207:974-81.
[220] Nayak B, Kumar S, DiNapoli JM, Paldurai A, Perez DR, Collins PL, et al. Contributions of the avian influenza virus HA, NA, and M2 surface proteins to the induction of neutralizing antibodies and protective immunity. J Virol. 2010;84:2408-20.
[221] Koel BF, Burke DF, Bestebroer TM, van der Vliet S, Zondag GC, Vervaet G, et al. Substitutions near the receptor binding site determine major antigenic change during influenza virus evolution. Science. 2006;342:976-9.

[222] Smith DJ, Lapedes AS, de Jong JC, Bestebroer TM, Rimmelzwaan GF, Osterhaus AD, et al. Mapping the antigenic and genetic evolution of influenza virus. Science. 2004;305:371-6.
[223] Grenfell BT, Pybus OG, Gog JR, Wood JL, Daly JM, Mumford JA, et al. Unifying the epidemiological and evolutionary dynamics of pathogens. Science. 2004;303:327-32.
[224] Rambaut A, Pybus OG, Nelson MI, Viboud C, Taubenberger JK, Holmes EC. The genomic and epidemiological dynamics of human influenza A virus. Nature. 2008;453:615-9.
[225] Russell CA, Jones TC, Barr IG, Cox NJ, Garten RJ, Gregory V, et al. The global circulation of seasonal influenza A (H3N2) viruses. Science. 2008;320:340-6.
[226] Davies WL, Grunert RR, Haff RF, McGahen JW, Neumayer EM, Paulshock M, et al. Antiviral Activity of 1-Adamantanamine (Amantadine). Science. 1964;144:862-3.
[227] Wintermeyer SM, Nahata MC. Rimantadine: a clinical perspective. Ann Pharmacother. 1995;29:299-310.
[228] Douglas RG, Jr. Prophylaxis and treatment of influenza. N Engl J Med. 1990;322:443-50.
[229] Pinto LH, Holsinger LJ, Lamb RA. Influenza virus M2 protein has ion channel activity. Cell. 1992;69:517-28.
[230] Sansom MS, Kerr ID. Influenza virus M2 protein: a molecular modelling study of the ion channel. Protein Eng. 1993;6:65-74.
[231] Bright RA, Shay DK, Shu B, Cox NJ, Klimov AI. Adamantane resistance among influenza A viruses isolated early during the 2005-2006 influenza season in the United States. Jama. 2006;295:891-4.
[232] Bright RA, Medina MJ, Xu X, Perez-Oronoz G, Wallis TR, Davis XM, et al. Incidence of adamantane resistance among influenza A (H3N2) viruses isolated worldwide from 1994 to 2005: a cause for concern. Lancet. 2005;366:1175-81.
[233] De Clercq E. Antiviral agents active against influenza A viruses. Nat Rev Drug Discov. 2006;5:1015-25.
[234] Aoki FY, Macleod MD, Paggiaro P, Carewicz O, El Sawy A, Wat C, et al. Early administration of oral oseltamivir increases the benefits of influenza treatment. J Antimicrob Chemother. 2003;51:123-9.
[235] Kaiser L, Wat C, Mills T, Mahoney P, Ward P, Hayden F. Impact of oseltamivir treatment on influenza-related lower respiratory tract complications and hospitalizations. Arch Intern Med. 2003;163:1667-72.
[236] Welliver R, Monto AS, Carewicz O, Schatteman E, Hassman M, Hedrick J, et al. Effectiveness of oseltamivir in preventing influenza in household contacts: a randomized controlled trial. Jama. 2001;285:748-54.
[237] Abdel-Magid AF, Maryanoff CA, Mehrman SJ. Synthesis of influenza neuraminidase inhibitors. Curr Opin Drug Discov Devel. 2001;4:776-91.
[238] Varghese JN, McKimm-Breschkin JL, Caldwell JB, Kortt AA, Colman PM. The structure of the complex between influenza virus neuraminidase and sialic acid, the viral receptor. Proteins. 1992;14:327-32.
[239] von Itzstein M, Wu WY, Kok GB, Pegg MS, Dyason JC, Jin B, et al. Rational design of potent sialidase-based inhibitors of influenza virus replication. Nature. 1993;363:418-23.
[240] Goodford PJ. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. J Med Chem. 1985;28:849-57.
[241] von Itzstein M, Dyason JC, Oliver SW, White HF, Wu WY, Kok GB, et al. A study of the active site of influenza virus sialidase: an approach to the rational design of novel anti-influenza drugs. J Med Chem. 1996;39:388-91.
[242] Woods JM, Bethell RC, Coates JA, Healy N, Hiscox SA, Pearson BA, et al. 4-Guanidino-2,4-dideoxy-2,3-dehydro-N-acetylneuraminic acid is a highly effective inhibitor both of the

sialidase (neuraminidase) and of growth of a wide range of influenza A and B viruses in vitro. Antimicrob Agents Chemother. 1993;37:1473-9.
[243] Vavricka CJ, Li Q, Wu Y, Qi J, Wang M, Liu Y, et al. Structural and functional analysis of laninamivir and its octanoate prodrug reveals group specific mechanisms for influenza NA inhibition. PLoS Pathog. 2011;7:e1002249.
[244] Holzer CT, von Itzstein M, Jin B, Pegg MS, Stewart WP, Wu WY. Inhibition of sialidases from viral, bacterial and mammalian sources by analogues of 2-deoxy-2,3-didehydro-N-acetylneuraminic acid modified at the C-4 position. Glycoconj J. 1993;10:40-4.
[245] Li CY, Yu Q, Ye ZQ, Sun Y, He Q, Li XM, et al. A nonsynonymous SNP in human cytosolic sialidase in a small Asian population results in reduced enzyme activity: potential link with severe adverse reactions to oseltamivir. Cell Res. 2007;17:357-62.
[246] Kelso A, Hurt AC. The ongoing battle against influenza: Drug-resistant influenza viruses: why fitness matters. Nat Med. 2012;18:1470-1.
[247] Kim CU, Lew W, Williams MA, Liu H, Zhang L, Swaminathan S, et al. Influenza neuraminidase inhibitors possessing a novel hydrophobic interaction in the enzyme active site: design, synthesis, and structural analysis of carbocyclic sialic acid analogues with potent anti-influenza activity. J Am Chem Soc. 1997;119:681-90.
[248] Varghese JN, Smith PW, Sollis SL, Blick TJ, Sahasrabudhe A, McKimm-Breschkin JL, et al. Drug design against a shifting target: a structural basis for resistance to inhibitors in a variant of influenza virus neuraminidase. Structure. 1998;6:735-46.
[249] Nguyen HT, Sheu TG, Mishin VP, Klimov AI, Gubareva LV. Assessment of pandemic and seasonal influenza A (H1N1) virus susceptibility to neuraminidase inhibitors in three enzyme activity inhibition assays. Antimicrob Agents Chemother. 2012;54:3671-7.
[250] Treanor JJ, Hayden FG, Vrooman PS, Barbarash R, Bettis R, Riff D, et al. Efficacy and safety of the oral neuraminidase inhibitor oseltamivir in treating acute influenza: a randomized controlled trial. US Oral Neuraminidase Study Group. Jama. 2000;283:1016-24.
[251] Whitley RJ, Hayden FG, Reisinger KS, Young N, Dutkowski R, Ipe D, et al. Oral oseltamivir treatment of influenza in children. Pediatr Infect Dis J. 2001;20:127-33.
[252] Ives JA, Carr JA, Mendel DB, Tai CY, Lambkin R, Kelly L, et al. The H274Y mutation in the influenza A/H1N1 neuraminidase active site following oseltamivir phosphate treatment leave virus severely compromised both in vitro and in vivo. Antiviral Res. 2002;55:307-17.
[253] Abed Y, Bourgault AM, Fenton RJ, Morley PJ, Gower D, Owens IJ, et al. Characterization of 2 influenza A(H3N2) clinical isolates with reduced susceptibility to neuraminidase inhibitors due to mutations in the hemagglutinin gene. J Infect Dis. 2002;186:1074-80.
[254] Gubareva LV, Webster RG, Hayden FG. Detection of influenza virus resistance to neuraminidase inhibitors by an enzyme inhibition assay. Antiviral Res. 2002;53:47-61.
[255] Kiso M, Mitamura K, Sakai-Tagawa Y, Shiraishi K, Kawakami C, Kimura K, et al. Resistant influenza A viruses in children treated with oseltamivir: descriptive study. Lancet. 2004;364:759-65.
[256] Hurt AC, Ernest J, Deng YM, Iannello P, Besselaar TG, Birch C, et al. Emergence and spread of oseltamivir-resistant A(H1N1) influenza viruses in Oceania, South East Asia and South Africa. Antiviral Res. 2009;83:90-3.
[257] Meijer A, Lackenby A, Hungnes O, Lina B, van-der-Werf S, Schweiger B, et al. Oseltamivir-resistant influenza virus A (H1N1), Europe, 2007-08 season. Emerg Infect Dis. 2009;15:552-60.
[258] Hurt AC, Hardie K, Wilson NJ, Deng YM, Osbourn M, Gehrig N, et al. Community transmission of oseltamivir-resistant A(H1N1)pdm09 influenza. N Engl J Med. 2011;365:2541-2.

[259] Hurt AC, Hardie K, Wilson NJ, Deng YM, Osbourn M, Leang SK, et al. Characteristics of a widespread community cluster of H275Y oseltamivir-resistant A(H1N1)pdm09 influenza in Australia. J Infect Dis. 2012;206:148-57.
[260] Storms AD, Gubareva LV, Su S, Wheeling JT, Okomo-Adhiambo M, Pan CY, et al. Oseltamivir-resistant pandemic (H1N1) 2009 virus infections, United States, 2010-11. Emerg Infect Dis. 2012;18:308-11.
[261] Gubareva LV, Kaiser L, Matrosovich MN, Soo-Hoo Y, Hayden FG. Selection of influenza virus mutants in experimentally infected volunteers treated with oseltamivir. J Infect Dis. 2001;183:523-31.
[262] Abed Y, Goyette N, Boivin G. A reverse genetics study of resistance to neuraminidase inhibitors in an influenza A/H1N1 virus. Antivir Ther. 2004;9:577-81.
[263] Abed Y, Baz M, Boivin G. Impact of neuraminidase mutations conferring influenza resistance to neuraminidase inhibitors in the N1 and N2 genetic backgrounds. Antivir Ther. 2006;11:971-6.
[264] Collins PJ, Haire LF, Lin YP, Liu J, Russell RJ, Walker PA, et al. Structural basis for oseltamivir resistance of influenza viruses. Vaccine. 2009;27:6317-23.
[265] Herlocher ML, Truscon R, Elias S, Yen HL, Roberts NA, Ohmit SE, et al. Influenza viruses resistant to the antiviral drug oseltamivir: transmission studies in ferrets. J Infect Dis. 2004;190:1627-30.
[266] Baz M, Abed Y, Simon P, Hamelin ME, Boivin G. Effect of the neuraminidase mutation H274Y conferring resistance to oseltamivir on the replicative capacity and virulence of old and recent human influenza A(H1N1) viruses. J Infect Dis. 2010;201:740-5.
[267] Hamelin ME, Baz M, Abed Y, Couture C, Joubert P, Beaulieu E, et al. Oseltamivir-resistant pandemic A/H1N1 virus is as virulent as its wild-type counterpart in mice and ferrets. PLoS Pathog. 2010;6:e1001015.
[268] Rameix-Welti MA, Enouf V, Cuvelier F, Jeannin P, van der Werf S. Enzymatic properties of the neuraminidase of seasonal H1N1 influenza viruses provide insights for the emergence of natural resistance to oseltamivir. PLoS Pathog. 2008;4:e1000103.
[269] Seibert CW, Kaminski M, Philipp J, Rubbenstroth D, Albrecht RA, Schwalm F, et al. Oseltamivir-resistant variants of the 2009 pandemic H1N1 influenza A virus are not attenuated in the guinea pig and ferret transmission models. J Virol. 2010;84:11219-26.
[270] Wong DD, Choy KT, Chan RW, Sia SF, Chiu HP, Cheung PP, et al. Comparable fitness and transmissibility between oseltamivir-resistant pandemic 2009 and seasonal H1N1 influenza viruses with the H275Y neuraminidase mutation. J Virol. 2012;86:10558-70.
[271] Yen HL, Herlocher LM, Hoffmann E, Matrosovich MN, Monto AS, Webster RG, et al. Neuraminidase inhibitor-resistant influenza viruses may differ substantially in fitness and transmissibility. Antimicrob Agents Chemother. 2005;49:4075-84.
[272] Herlocher ML, Carr J, Ives J, Elias S, Truscon R, Roberts N, et al. Influenza virus carrying an R292K mutation in the neuraminidase gene is not transmitted in ferrets. Antiviral Res. 2002;54:99-111.
[273] Pizzorno A, Bouhy X, Abed Y, Boivin G. Generation and characterization of recombinant pandemic influenza A(H1N1) viruses resistant to neuraminidase inhibitors. J Infect Dis. 2011;203:25-31.
[274] Le QM, Kiso M, Someya K, Sakai YT, Nguyen TH, Nguyen KH, et al. Avian flu: isolation of drug-resistant H5N1 virus. Nature. 2005;437:1108.
[275] Morlighem JE, Aoki S, Kishima M, Hanami M, Ogawa C, Jalloh A, et al. Mutation analysis of 2009 pandemic influenza A(H1N1) viruses collected in Japan during the peak phase of the pandemic. PLoS One.6:e18956.

[276] Yen HL, Ilyushina NA, Salomon R, Hoffmann E, Webster RG, Govorkova EA. Neuraminidase inhibitor-resistant recombinant A/Vietnam/1203/04 (H5N1) influenza viruses retain their replication efficiency and pathogenicity in vitro and in vivo. J Virol. 2007;81:12418-26.
[277] van der Vries E, Collins PJ, Vachieri SG, Xiong X, Liu J, Walker PA, et al. H1N1 2009 pandemic influenza virus: resistance of the I223R neuraminidase mutant explained by kinetic and structural analysis. PLoS Pathog. 2012;8:e1002914.
[278] van der Vries E, Stelma FF, Boucher CA. Emergence of a multidrug-resistant pandemic influenza A (H1N1) virus. N Engl J Med. 2010;363:1381-2.
[279] van der Vries E, Veldhuis Kroeze EJ, Stittelaar KJ, Linster M, Van der Linden A, Schrauwen EJ, et al. Multidrug resistant 2009 A/H1N1 influenza clinical isolate with a neuraminidase I223R mutation retains its virulence and transmissibility in ferrets. PLoS Pathog. 2011;7:e1002276.
[280] McKimm-Breschkin JL, Barrett S, Pudjiatmoko, Azhar M, Wong FY, Selleck P, et al. I222 Neuraminidase mutations further reduce oseltamivir susceptibility of Indonesian Clade 2.1 highly pathogenic Avian Influenza A(H5N1) viruses. PLoS One. 2013;8:e66105.
[281] Huang L, Cao Y, Zhou J, Qin K, Zhu W, Zhu Y, et al. A conformational restriction in the influenza A virus neuraminidase binding site by R152 results in a combinational effect of I222T and H274Y on oseltamivir resistance. Antimicrob Agents Chemother. 2014;58:1639-45.
[282] McKimm-Breschkin JL. Influenza neuraminidase inhibitors: antiviral action and mechanisms of resistance. Influenza Other Respir Viruses. 2013;7 Suppl 1:25-36.
[283] Nguyen HT, Fry AM, Gubareva LV. Neuraminidase inhibitor resistance in influenza viruses and laboratory testing methods. Antivir Ther. 2012;17:159-73.
[284] Stemmer WP. Rapid evolution of a protein in vitro by DNA shuffling. Nature. 1994;370:389-91.
[285] Hughes AL, Nei M. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature. 1988;335:167-70.
[286] Yang Z, Bielawski JP. Statistical methods for detecting molecular adaptation. Trends Ecol Evol. 2000;15:496-503.
[287] King MC, Wilson AC. Evolution at two levels in humans and chimpanzees. Science. 1975;188:107-16.
[288] Gout JF, Kahn D, Duret L. The relationship among gene expression, the evolution of gene dosage, and the rate of protein evolution. PLoS Genet. 2010;6:e1000944.
[289] Pursell NW, Mishra P, Bolon DN. Solubility-promoting function of hsp90 contributes to client maturation and robust cell growth. Eukaryot Cell. 2012;11:1033-41.
[290] Wayne N, Mishra P, Bolon DN. Hsp90 and client protein maturation. Methods Mol Biol. 2011;787:33-44.
[291] Whitesell L, Lin NU. Hsp90 as a platform for the assembly of more effective cancer chemotherapy. Biochim Biophys Acta. 2012;1823:756-66.
[292] Sanjuan R, Moya A, Elena SF. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. Proc Natl Acad Sci U S A. 2004;101:8396-401.
[293] Carrasco P, de la Iglesia F, Elena SF. Distribution of fitness and virulence effects caused by single-nucleotide substitutions in Tobacco Etch virus. J Virol. 2007;81:12979-84.
[294] Domingo-Calap P, Cuevas JM, Sanjuan R. The fitness effects of random mutations in single-stranded DNA and RNA bacteriophages. PLoS Genet. 2009;5:e1000742.
[295] Peris JB, Davis P, Cuevas JM, Nebot MR, Sanjuan R. Distribution of fitness effects caused by single-nucleotide substitutions in bacteriophage f1. Genetics. 2010;185:603-9.

[296] Wylie CS, Shakhnovich EI. A biophysical protein folding model accounts for most mutational fitness effects in viruses. Proc Natl Acad Sci U S A. 2011;108:9916-21.
[297] Zeldovich KB, Chen P, Shakhnovich EI. Protein stability imposes limits on organism complexity and speed of molecular evolution. Proc Natl Acad Sci U S A. 2007;104:16152-7.
[298] Bershtein S, Segal M, Bekerman R, Tokuriki N, Tawfik DS. Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. Nature. 2006;444:929-32.
[299] Kacser H, Fell DA. The control of flux: 21 years on. Biochem Soc Trans. 1995;23:341-66.
[300] Lunzer M, Golding GB, Dean AM. Pervasive cryptic epistasis in molecular evolution. PLoS Genet. 2010;6:e1001162.
[301] Kacser H, Burns JA. The control of flux. Symp Soc Exp Biol. 1973;27:65-104.
[302] Hegreness M, Shoresh N, Hartl D, Kishony R. An equivalence principle for the incorporation of favorable mutations in asexual populations. Science. 2006;311:1615-7.
[303] Lynch M, Conery JS. The origins of genome complexity. Science. 2003;302:1401-4.
[304] Tsai IJ, Bensasson D, Burt A, Koufopanou V. Population genomics of the wild yeast Saccharomyces paradoxus: Quantifying the life cycle. Proc Natl Acad Sci U S A. 2008;105:4957-62.
[305] Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics. 2012;13:341.
[306] Adkar BV, Tripathi A, Sahoo A, Bajaj K, Goswami D, Chakrabarti P, et al. Protein model discrimination using mutational sensitivity derived from deep sequencing. Structure. 2012;20:371-81.
[307] Whitehead TA, Chevalier A, Song Y, Dreyfus C, Fleishman SJ, De Mattos C, et al. Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. Nat Biotechnol. 2012;30:543-8.
[308] Debartolo J, Dutta S, Reich L, Keating AE. Predictive bcl-2 family binding models rooted in experiment or structure. J Mol Biol. 2012;422:124-44.
[309] Pitt JN, Ferre-D'Amare AR. Rapid construction of empirical RNA fitness landscapes. Science. 2010;330:376-9.
[310] Mumberg D, Muller R, Funk M. Regulatable promoters of Saccharomyces cerevisiae: comparison of transcriptional activity and their use for heterologous expression. Nucleic Acids Res. 1994;22:5767-8.
[311] Nathan DF, Lindquist S. Mutational analysis of Hsp90 function: interactions with a steroid receptor and a protein kinase. Mol Cell Biol. 1995;15:3917-25.
[312] Roscoe BP, Thayer KM, Zeldovich KB, Fushman D, Bolon DN. Analyses of the Effects of All Ubiquitin Point Mutants on Yeast Growth Rate. J Mol Biol. in press.
[313] Wayne N, Bolon DN. Dimerization of Hsp90 is required for in vivo function. Design and analysis of monomers and dimers. J Biol Chem. 2007;282:35386-95.
[314] Picard D, Khursheed B, Garabedian MJ, Fortin MG, Lindquist S, Yamamoto KR. Reduced levels of hsp90 compromise steroid receptor action in vivo. Nature. 1990;348:166-8.
[315] Tokuriki N, Tawfik DS. Stability effects of mutations and protein evolvability. Curr Opin Struct Biol. 2009;19:596-604.
[316] Harris SF, Shiau AK, Agard DA. The crystal structure of the carboxy-terminal dimerization domain of htpG, the Escherichia coli Hsp90, reveals a potential substrate binding site. Structure. 2004;12:1087-97.
[317] Dill KA. Dominant forces in protein folding. Biochemistry. 1990;29:7133-55.
[318] King JL, Jukes TH. Non-Darwinian evolution. Science. 1969;164:788-98.

[319] Sauer RT, Milla ME, Waldburger CD, Brown BM, Schildbach JF. Sequence determinants of folding and stability for the P22 Arc repressor dimer. Faseb J. 1996;10:42-8.
[320] Toth-Petroczy A, Tawfik DS. Slow protein evolutionary rates are dictated by surface-core association. Proc Natl Acad Sci U S A. 2011;108:11151-6.
[321] Kellogg EH, Leaver-Fay A, Baker D. Role of conformational sampling in computing mutation-induced changes in protein structure and stability. Proteins. 2011;79:830-8.
[322] Dykhuizen DE, Dean AM. Enzyme activity and fitness: Evolution in solution. Trends Ecol Evol. 1990;5:257-62.
[323] Dykhuizen DE, Dean AM, Hartl DL. Metabolic flux and fitness. Genetics. 1987;115:25-31.
[324] Dykhuizen DE, Dean AM. Experimental Evolution from the Bottom Up. In: Theodor, editor. Experimental Evolution: Concepts, Methods, and Applications of Selection Experiments2009.
[325] Li MZ, Elledge SJ. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. Nat Methods. 2007;4:251-6.
[326] Hietpas R, Roscoe B, Jiang L, Bolon DN. Fitness analyses of all possible point mutations for regions of genes in yeast. Nat Protoc. 2012;7:1382-96.
[327] Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci U S A. 1992;89:10915-9.
[328] Geiler-Samerotte KA, Dion MF, Budnik BA, Wang SM, Hartl DL, Drummond DA. Misfolded proteins impose a dosage-dependent fitness cost and trigger a cytosolic unfolded protein response in yeast. Proc Natl Acad Sci U S A. 2010;108:680-5.
[329] Futcher B, Carbon J. Toxic effects of excess cloned centromeres. Mol Cell Biol. 1986;6:2213-22.
[330] Tschumper G, Carbon J. Copy number control by a yeast centromere. Gene. 1983;23:221-32.
[331] Tschumper G, Carbon J. Saccharomyces cerevisiae mutants that tolerate centromere plasmids at high copy number. Proc Natl Acad Sci U S A. 1987;84:7203-7.
[332] Wilen CB, Tilton JC, Doms RW. Molecular mechanisms of HIV entry. Adv Exp Med Biol. 2012;726:223-42.
[333] Haim H, Strack B, Kassa A, Madani N, Wang L, Courter JR, et al. Contribution of intrinsic reactivity of the HIV-1 envelope glycoproteins to CD4-independent infection and global inhibitor sensitivity. PLoS Pathog. 2012;7:e1002101.
[334] Madani N, Schon A, Princiotto AM, Lalonde JM, Courter JR, Soeta T, et al. Small-molecule CD4 mimics interact with a highly conserved pocket on HIV-1 gp120. Structure. 2008;16:1689-701.
[335] Peters PJ, Bhattacharya J, Hibbitts S, Dittmar MT, Simmons G, Bell J, et al. Biological analysis of human immunodeficiency virus type 1 R5 envelopes amplified from brain and lymph node tissues of AIDS patients with neuropathology reveals two distinct tropism phenotypes and identifies envelopes in the brain that confer an enhanced tropism and fusigenicity for macrophages. J Virol. 2004;78:6915-26.
[336] Peters PJ, Duenas-Decamp MJ, Sullivan WM, Brown R, Ankghuambom C, Luzuriaga K, et al. Variation in HIV-1 R5 macrophage-tropism correlates with sensitivity to reagents that block envelope: CD4 interactions but not with sensitivity to other entry inhibitors. Retrovirology. 2008;5:5.
[337] Dunfee RL, Thomas ER, Gorry PR, Wang J, Taylor J, Kunstman K, et al. The HIV Env variant N283 enhances macrophage tropism and is associated with brain infection and dementia. Proc Natl Acad Sci U S A. 2006;103:15160-5.

[338] Musich T, Peters PJ, Duenas-Decamp MJ, Gonzalez-Perez MP, Robinson J, Zolla-Pazner S, et al. A conserved determinant in the V1 loop of HIV-1 modulates the V3 loop to prime low CD4 use and macrophage infection. J Virol. 2011;85:2397-405.
[339] Walter BL, Wehrly K, Swanstrom R, Platt E, Kabat D, Chesebro B. Role of low CD4 levels in the influence of human immunodeficiency virus type 1 envelope V1 and V2 regions on entry and spread in macrophages. J Virol. 2005;79:4828-37.
[340] Jiang L, Mishra P, Hietpas RT, Zeldovich KB, Bolon DN. Latent effects of Hsp90 mutants revealed at reduced expression levels. PLoS Genet. 2013;9:e1003600.
[341] Wagenaar TR, Ma L, Roscoe B, Park SM, Bolon DN, Green MR. Resistance to vemurafenib resulting from a novel mutation in the BRAFV600E kinase domain. Pigment Cell Melanoma Res. 2014;27:124-33.
[342] Duenas-Decamp MJ, Peters P, Burton D, Clapham PR. Natural resistance of human immunodeficiency virus type 1 to the CD4bs antibody b12 conferred by a glycan and an arginine residue close to the CD4 binding loop. J Virol. 2008;82:5807-14.
[343] Peters PJ, Sullivan WM, Duenas-Decamp MJ, Bhattacharya J, Ankghuambom C, Brown R, et al. Non-macrophage-tropic human immunodeficiency virus type 1 R5 envelopes predominate in blood, lymph nodes, and semen: implications for transmission and pathogenesis. J Virol. 2006;80:6324-32.
[344] Schnell G, Joseph S, Spudich S, Price RW, Swanstrom R. HIV-1 replication in the central nervous system occurs in two distinct cell types. PLoS Pathog. 2011;7:e1002286.
[345] Sturdevant CB, Joseph SB, Schnell G, Price RW, Swanstrom R, Spudich S. Compartmentalized replication of R5 T cell-tropic HIV-1 in the central nervous system early in the course of infection. PLoS Pathog. 2015;11:e1004720.
[346] Isaacman-Beck J, Hermann EA, Yi Y, Ratcliffe SJ, Mulenga J, Allen S, et al. Heterosexual transmission of human immunodeficiency virus type 1 subtype C: Macrophage tropism, alternative coreceptor use, and the molecular anatomy of CCR5 utilization. J Virol. 2009;83:8208-20.
[347] Joseph SB, Swanstrom R, Kashuba AD, Cohen MS. Bottlenecks in HIV-1 transmission: insights from the study of founder viruses. Nat Rev Microbiol. 2015;13:414-25.
[348] Peters PJ, Gonzalez-Perez MP, Musich T, Moore Simas TA, Lin R, Morse AN, et al. Infection of ectocervical tissue and universal targeting of T-cells mediated by primary non-macrophage-tropic and highly macrophage-tropic HIV-1 R5 envelopes. Retrovirology. 2015;12:48.
[349] Salazar-Gonzalez JF, Salazar MG, Keele BF, Learn GH, Giorgi EE, Li H, et al. Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. J Exp Med. 2009;206:1273-89.
[350] Binley JM, Wrin T, Korber B, Zwick MB, Wang M, Chappey C, et al. Comprehensive cross-clade neutralization analysis of a panel of anti-human immunodeficiency virus type 1 monoclonal antibodies. J Virol. 2004;78:13232-52.
[351] Hoffman TL, LaBranche CC, Zhang W, Canziani G, Robinson J, Chaiken I, et al. Stable exposure of the coreceptor-binding site in a CD4-independent HIV-1 envelope protein. Proc Natl Acad Sci U S A. 1999;96:6359-64.
[352] Boyd DF, Peterson D, Haggarty BS, Jordan AP, Hogan MJ, Goo L, et al. Mutations in HIV-1 envelope that enhance entry with the macaque CD4 receptor alter antibody recognition by disrupting quaternary interactions within the trimer. J Virol. 2015;89:894-907.
[353] Murphy MK, Yue L, Pan R, Boliar S, Sethi A, Tian J, et al. Viral escape from neutralizing antibodies in early subtype A HIV-1 infection drives an increase in autologous neutralization breadth. PLoS Pathog. 2013;9:e1003173.

[354] Walker LM, Phogat SK, Chan-Hui PY, Wagner D, Phung P, Goss JL, et al. Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. Science. 2009;326:285-9.
[355] Munro JB, Gorman J, Ma X, Zhou Z, Arthos J, Burton DR, et al. Conformational dynamics of single HIV-1 envelope trimers on the surface of native virions. Science. 2014;346:759-63.
[356] Wei X, Decker JM, Liu H, Zhang Z, Arani RB, Kilby JM, et al. Emergence of resistant human immunodeficiency virus type 1 in patients receiving fusion inhibitor (T-20) monotherapy. Antimicrob Agents Chemother. 2002;46:1896-905.
[357] Lew W, Chen X, Kim CU. Discovery and development of GS 4104 (oseltamivir): an orally active influenza neuraminidase inhibitor. Curr Med Chem. 2000;7:663-72.
[358] Gubareva LV, Kaiser L, Hayden FG. Influenza virus neuraminidase inhibitors. Lancet. 2000;355:827-35.
[359] Dharan NJ, Gubareva LV, Meyer JJ, Okomo-Adhiambo M, McClinton RC, Marshall SA, et al. Infections with oseltamivir-resistant influenza A(H1N1) virus in the United States. Jama. 2009;301:1034-41.
[360] Moscona A. Global transmission of oseltamivir-resistant influenza. N Engl J Med. 2009;360:953-6.
[361] Samson M, Pizzorno A, Abed Y, Boivin G. Influenza virus resistance to neuraminidase inhibitors. Antiviral research. 2013;98:174-85.
[362] Oakley AJ, Barrett S, Peat TS, Newman J, Streltsov VA, Waddington L, et al. Structural and functional basis of resistance to neuraminidase inhibitors of influenza B viruses. J Med Chem. 2010;53:6421-31.
[363] Li Q, Qi J, Zhang W, Vavricka CJ, Shi Y, Wei J, et al. The 2009 pandemic H1N1 neuraminidase N1 lacks the 150-cavity in its active site. Nat Struct Mol Biol. 2010;17:1266-8.
[364] Bloom JD, Gong LI, Baltimore D. Permissive secondary mutations enable the evolution of influenza oseltamivir resistance. Science. 2010;328:1272-5.
[365] Thyagarajan B, Bloom JD. The inherent mutational tolerance and antigenic evolvability of influenza hemagglutinin. Elife. 2014;3.
[366] Acevedo A, Brodsky L, Andino R. Mutational and fitness landscapes of an RNA virus revealed through population sequencing. Nature. 2014;505:686-90.
[367] Bank C, Hietpas RT, Wong A, Bolon DN, Jensen JD. A bayesian MCMC approach to assess the complete distribution of fitness effects of new mutations: uncovering the potential for adaptive walks in challenging environments. Genetics. 2014;196:841-52.
[368] Fleishman SJ, Whitehead TA, Ekiert DC, Dreyfus C, Corn JE, Strauch EM, et al. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. Science. 2011;332:816-21.
[369] Hoffmann E, Neumann G, Kawaoka Y, Hobom G, Webster RG. A DNA transfection system for generation of influenza A virus from eight plasmids. Proc Natl Acad Sci U S A. 2000;97:6108-13.
[370] Castrucci MR, Kawaoka Y. Biologic importance of neuraminidase stalk length in influenza A virus. J Virol. 1993;67:759-64.
[371] Squires RB, Noronha J, Hunt V, Garcia-Sastre A, Macken C, Baumgarth N, et al. Influenza research database: an integrated bioinformatics resource for influenza research and surveillance. Influenza and other respiratory viruses. 2012;6:404-16.
[372] Morlighem JE, Aoki S, Kishima M, Hanami M, Ogawa C, Jalloh A, et al. Mutation analysis of 2009 pandemic influenza A(H1N1) viruses collected in Japan during the peak phase of the pandemic. PLoS One. 2011;6:e18956.

[373] Marjuki H, Mishin VP, Chesnokov AP, De La Cruz JA, Davis CT, Villanueva JM, et al. Neuraminidase Mutations Conferring Resistance to Oseltamivir in Influenza A(H7N9) Viruses. J Virol. 2015;89:5419-26.
[374] Zabriskie MS, Eide CA, Tantravahi SK, Vellore NA, Estrada J, Nicolini FE, et al. BCR-ABL1 compound mutations combining key kinase domain positions confer clinical resistance to ponatinib in Ph chromosome-positive leukemia. Cancer cell. 2014;26:428-42.
[375] Soumana DI, Ali A, Schiffer CA. Structural analysis of asunaprevir resistance in HCV NS3/4A protease. ACS chemical biology. 2014;9:2485-90.
[376] Ali A, Bandaranayake RM, Cai Y, King NM, Kolli M, Mittal S, et al. Molecular Basis for Drug Resistance in HIV-1 Protease. Viruses. 2010;2:2509-35.
[377] Smith BJ, Colman PM, Von Itzstein M, Danylec B, Varghese JN. Analysis of inhibitor binding in influenza virus neuraminidase. Protein science : a publication of the Protein Society. 2001;10:689-96.
[378] Abed Y, Pizzorno A, Bouhy X, Boivin G. Role of permissive neuraminidase mutations in influenza A/Brisbane/59/2007-like (H1N1) viruses. PLoS Pathog. 2012;7:e1002431.
[379] Butler J, Hooper KA, Petrie S, Lee R, Maurer-Stroh S, Reh L, et al. Estimating the fitness advantage conferred by permissive neuraminidase mutations in recent oseltamivir-resistant A(H1N1)pdm09 influenza viruses. PLoS Pathog. 2014;10:e1004065.
[380] Wagenaar TR, Ma L, Roscoe B, Park SM, Bolon DN, Green MR. Resistance to vemurafenib resulting from a novel mutation in the BRAFV600E kinase domain. Pigment Cell Melanoma Res. 2013;27:124-33.
[381] Hayden FG, Cote KM, Douglas RG, Jr. Plaque inhibition assay for drug susceptibility testing of influenza viruses. Antimicrob Agents Chemother. 1980;17:865-70.
[382] Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research. 2004;32:1792-7.
[383] Eyre-Walker A, Keightley PD. The distribution of fitness effects of new mutations. Nat Rev Genet. 2007;8:610-8.
[384] Elena SF, Ekunwe L, Hajela N, Oden SA, Lenski RE. Distribution of fitness effects caused by random insertion mutations in Escherichia coli. Genetica. 1998;102-103:349-58.
[385] Davies JE. Redundant genome sequencing? Science. 1996;273:1155.
[386] Elena SF, Carrasco P, Daros JA, Sanjuan R. Mechanisms of genetic robustness in RNA viruses. EMBO Rep. 2006;7:168-73.
[387] Sanjuan R. Mutational fitness effects in RNA and single-stranded DNA viruses: common patterns revealed by site-directed mutagenesis studies. Philos Trans R Soc Lond B Biol Sci. 2010;365:1975-82.
[388] Springman R, Keller T, Molineux IJ, Bull JJ. Evolution at a high imposed mutation rate: adaptation obscures the load in phage T7. Genetics. 2010;184:221-32.
[389] Bull JJ, Badgett MR, Wichman HA. Big-benefit mutations in a bacteriophage inhibited with heat. Mol Biol Evol. 2000;17:942-50.
[390] Rokyta DR, Joyce P, Caudle SB, Wichman HA. An empirical test of the mutational landscape model of adaptation using a single-stranded DNA virus. Nat Genet. 2005;37:441-4.
[391] Tokuriki N, Tawfik DS. Protein dynamism and evolvability. Science. 2009;324:203-7.
[392] Stanfield RL, Gorny MK, Williams C, Zolla-Pazner S, Wilson IA. Structural rationale for the broad neutralization of HIV-1 by human monoclonal antibody 447-52D. Structure. 2004;12:193-204.
[393] Walker LM, Huber M, Doores KJ, Falkowska E, Pejchal R, Julien JP, et al. Broad neutralization coverage of HIV by multiple highly potent antibodies. Nature. 2011;477:466-70.

[394] McLellan JS, Pancera M, Carrico C, Gorman J, Julien JP, Khayat R, et al. Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. Nature. 2012;480:336-43.
[395] Ayme V, Souche S, Caranta C, Jacquemond M, Chadoeuf J, Palloix A, et al. Different mutations in the genome-linked protein VPg of potato virus Y confer virulence on the pvr2(3) resistance in pepper. Mol Plant Microbe Interact. 2006;19:557-63.
[396] Charron C, Nicolai M, Gallois JL, Robaglia C, Moury B, Palloix A, et al. Natural variation and functional analyses provide evidence for co-evolution between plant eIF4E and potyviral VPg. Plant J. 2008;54:56-68.
[397] Kassen R, Bataillon T. Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria. Nat Genet. 2006;38:484-8.
[398] MacLean RC, Buckling A. The distribution of fitness effects of beneficial mutations in Pseudomonas aeruginosa. PLoS Genet. 2009;5:e1000406.
[399] Schenk MF, Szendro IG, Krug J, de Visser JA. Quantifying the adaptive potential of an antibiotic resistance enzyme. PLoS Genet. 2012;8:e1002783.
[400] Mostowy R, Kouyos RD, Hoof I, Hinkley T, Haddad M, Whitcomb JM, et al. Estimating the fitness cost of escape from HLA presentation in HIV-1 protease and reverse transcriptase. PLoS Comput Biol.8:e1002525.
[401] Duan S, Govorkova EA, Bahl J, Zaraket H, Baranovich T, Seiler P, et al. Epistatic interactions between neuraminidase mutations facilitated the emergence of the oseltamivir-resistant H1N1 influenza viruses. Nat Commun. 2014;5:5029.
[402] Simon P, Holder BP, Bouhy X, Abed Y, Beauchemin CA, Boivin G. The I222V neuraminidase mutation has a compensatory role in replication of an oseltamivir-resistant influenza virus A/H3N2 E119V mutant. J Clin Microbiol. 2011;49:715-7.
[403] Mateo R, Nagamine CM, Kirkegaard K. Suppression of Drug Resistance in Dengue Virus. MBio. 2015;6:e01960-15.
[404] Perales C, Quer J, Gregori J, Esteban JI, Domingo E. Resistance of Hepatitis C Virus to Inhibitors: Complexity and Clinical Implications. Viruses. 2015;7:5746-66.
[405] Holohan C, Van Schaeybroeck S, Longley DB, Johnston PG. Cancer drug resistance: an evolving paradigm. Nat Rev Cancer. 2013;13:714-26.
[406] Hoogstraat M, Gadellaa-van Hooijdonk CG, Ubink I, Besselink NJ, Pieterse M, Veldhuis W, et al. Detailed imaging and genetic analysis reveal a secondary BRAF(L505H) resistance mutation and extensive intrapatient heterogeneity in metastatic BRAF mutant melanoma patients treated with vemurafenib. Pigment Cell Melanoma Res. 2015;28:318-23.
[407] Blick TJ, Tiong T, Sahasrabudhe A, Varghese JN, Colman PM, Hart GJ, et al. Generation and characterization of an influenza virus neuraminidase variant with decreased sensitivity to the neuraminidase-specific inhibitor 4-guanidino-Neu5Ac2en. Virology. 1995;214:475-84.
[408] McKimm-Breschkin JL, McDonald M, Blick TJ, Colman PM. Mutation in the influenza virus neuraminidase gene resulting in decreased sensitivity to the neuraminidase inhibitor 4-guanidino-Neu5Ac2en leads to instability of the enzyme. Virology. 1996;225:240-2.
[409] McKimm-Breschkin JL, Sahasrabudhe A, Blick TJ, McDonald M, Colman PM, Hart GJ, et al. Mutations in a conserved residue in the influenza virus neuraminidase active site decreases sensitivity to Neu5Ac2en-derived inhibitors. J Virol. 1998;72:2456-62.
[410] Prabu-Jeyabalan M, Nalivaika EA, Romano K, Schiffer CA. Mechanism of substrate recognition by drug-resistant human immunodeficiency virus type 1 protease variants revealed by a novel structural intermediate. J Virol. 2006;80:3607-16.