Evolutionary Constraints and Potential of the Influenza A Virus RNA-dependent RNA Polymerase

by

Yuan Li

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Microbiology and Immunology)
in the University of Michigan
2024

Doctoral Committee:

Professor Adam Lauring, Chair
Professor Janet Smith
Associate Professor Evan Snitkin
Associate Professor Andrew Tai

Yuan Li

yuanrli@umich.edu

ORCID iD: 0000-0002-5771-064X

Dedication

To my beloved husband, John

Acknowledgements

I would like to express my deep gratitude to my advisor, Dr. Adam Lauring, who teaches by how he works. He has been exceptionally encouraging and supportive during difficult times and prepared me with critical thinking, creativity, and communication skills that are essential to both scientific study and my individual development plan.

I am grateful to the members of my committee: Dr. Andrew Tai, Dr. Evan Snitkin, and Dr. Janet Smith. Thank you for providing me with invaluable feedback and suggestions throughout the project. I was unfamiliar with some aspects of my project when I started, but with you, I always feel that I am well supported in exploring something new.

I would like to thank everyone in the Lauring lab. I feel included in a non-superficial way and enjoyed working and talking and having fun with you. Special thanks to Sarah Arcos for being a great captain of Team Polymerase and discussing with me the good and bad results. Having you bounce thoughts back and forth generated some of the best ideas in this project. I must also thank Will Fitzsimmons for being the best lab manager and keeping the lab a lively place. Thanks also to Emily Bendall and Jules Gilbert for their support in coding, to Chris Blair and Leigh Papalambros for their support with sequencing results, to lab alumni, Andrew Valesano and Danny Lyons, for setting up great role models as scientists, and everyone else in the lab. I will miss you all.

Finally, I want to thank people who have offered help selflessly in my project. I thank Nicole Koropatkin, Gideon Bradburd and Aaron King at the Department of Ecology and

Evolutionary Biology, Nicholas Wu and Ruipeng Lei at the University of Illinois Urbana-Champaign, and UMICH Microbiome Core for their expertise. I thank my cohort in Microbiology and Immunology for their friendship.

Table of Contents

List of Tables

List of Figures

List of Appendices

Abstract

Due to rapid evolution and adaptation, influenza viruses remain a major health concern despite co-existing with human beings for centuries. The influenza virus polymerase is a major driver of influenza virus evolution. Mutations within the viral polymerase can change replication efficiency, affecting the replicative fitness of the virus. The polymerase also controls the rate at which influenza virus acquires mutations, opening up possibilities for new phenotypes such as host range expansion, drug resistance, and antigenic drift. Despite its importance to viral evolution, our understanding of the mutational effects on the influenza virus polymerase is relatively limited. The influenza virus's segmented genome and the multi-unit structure of its polymerase add further complexity to the polymerase's evolutionary constraints and potential.

My dissertation focuses on characterizing the mutational effects of the core subunit of the influenza virus polymerase complex, the RNA-dependent RNA polymerase (RdRp) subunit, and reveals key constraints on the RdRp that shape influenza virus evolution in nature. The second chapter of my thesis evaluated the fitness effects of mutations and the mutational tolerance of influenza virus RdRp. I performed deep mutational scanning of the influenza A virus PB1 protein and measured the replicative fitness of nearly all variants with single amino acid substitutions. Deep mutational scanning measured replicative fitness with high accuracy and precision and revealed purifying selection against mutations with more dramatic changes. While most missense and nonsense mutations were highly detrimental, some near-neutral and beneficial mutations did exist. I calculated mutational tolerance as the Shannon entropy of the enrichment

of all amino acid variants at a site. The mutational tolerance of residues on the influenza virus RdRp was highly constrained by specific functions and site interactions and was not well characterized by the global protein structure. Many beneficial mutations revealed by deep mutational scanning were seen in the natural evolution history of PB1 or shown important to adaptation experimentally. Accessibility by single nucleotide mutations was a crucial factor in determining whether a beneficial mutation would arise in nature. My third chapter established a foundation to study the key mutations that would influence the virus's replicative fidelity using the variant library created by deep mutational scanning. I examined the growth of the influenza A virus under different concentrations of five mutagenic drugs in different cell lines and determined the proper drug concentrations to induce a moderate selective pressure. 5-Azacytidine and molnupiravir exhibited similar inhibition curves when the infections happened in MDCK or A549 cells, while the inhibition curves of ribavirin, favipiravir, and 5-fluorouracil were vastly different in different cells. These results highlight the complexity of the mechanisms by which mutagenic drugs inhibit influenza virus replication and the varying cell responses to mutagens. Overall, my dissertation provided a comprehensive map of mutational effects on a viral RdRp and revealed the evolutionary constraints and potential of influenza virus polymerase, which would be a valuable resource for future studies on influenza and RNA virus evolution.

Chapter 1 Introduction

Influenza has co-existed with human beings for centuries and caused four pandemics in the last one hundred years. The causative agent, influenza virus, belongs to the Orthomyxoviridae family and can be categorized into four types: A, B, C, and D. Among the four types, only influenza A and B viruses cause substantial morbidity and mortality in humans, while influenza C virus generally causes mild illness (1). Influenza D virus mainly infects cattle with common spillover to other species but is not known to infect humans (2). Since influenza A virus is the only type with pandemic potential (3), it draws most research interest and is the object of this study. Despite interacting with human beings for a long time, influenza viruses still cause significant morbidity and mortality in all age groups around the world. This is because influenza virus evolves quickly to become more "fit", which can be a manifestation of faster replication, an increased ability to survive and spread to new hosts, and/or being able to escape host immune clearance and anti-viral treatments.

The RNA-dependent RNA polymerase (RdRp) is central to influenza virus evolution. It performs genome replication and transcription functions and determines replication speed and fidelity. After decades of research, we now have a relatively clear understanding of its constitution as well as the mechanisms of each step of the replication and transcription processes. However, our knowledge regarding the mutational effects and evolutionary constraints on the RdRp is limited to isolated residues and mutations. The project in this dissertation is a systematic assessment of the mutational effects on the influenza virus RdRp, including the fitness effects of nearly all mutations and the mutational tolerance of RdRp residues. By linking mutational effects

and constraints to protein structure and functions, this project generated a map linking RdRp genotype to viral phenotype and provides a valuable resource for surveillance and future studies on influenza virus evolution. Below, I review the principal forces that have shaped influenza virus evolution, the structural basis of evolutionary constraints on influenza RdRp, existing studies of RdRp mutations, and novel technologies used in this project to thoroughly examine the functional constraints and evolutionary potential of influenza RdRp.

1.1 Influenza Virus Evolution

1.1.1 Source for Genome Diversity

Influenza virus is a negative-sense, single-stranded RNA virus with a segmented genome. The evolutionary rates for RNA viruses range from 10^{-4} to 10^{-3} substitutions per nucleotide site per year (s/n/y) (4). The evolutionary rate of influenza viruses can vary significantly in different hosts. For example, the evolutionary rate for avian influenza viruses (H6) is approximately 2.5×10^{-3} (s/n/y) in their natural reservoir wild birds but can be as high as 4×10^{-3} (s/n/y) when infecting poultry (5). This phenomenon could be due to the need for quick adaptation to new hosts or different biochemical environments for replication within different hosts. Additionally, different genes evolve at different rates in the same virus (6).

The evolutionary rates of RNA viruses are positively correlated with their mutation rates, and influenza viruses have extremely high mutation rates. Studies show that the mutation rates of RNA viruses can reach as high as 10^{-6} to 10^{-4} substitutions per nucleotide site per cell infection (s/n/c), 100 times higher than that of DNA viruses (4). Consistent with that result, the mutation rate for influenza A virus is 1.8×10^{-4} to 2.5×10^{-4} per nucleotide per strand copied (s/n/r), depending on the strains, or equal to 2~3 mutations every time it replicates its genome (7). Influenza viruses accumulate both silent and amino acid-changing substitutions at approximately

constant rates (8, 9), following the molecular clock. With an extensive collection of patient samples and deep sequencing, researchers can perform phylogenetic analysis to determine the transmission route in an epidemic from the changes in the influenza virus genome (10). Frequent mutation constantly creates new influenza A strains. Based on the sequences of two proteins on the surface, hemagglutinin (HA) and neuraminidase (NA), influenza A viruses are further divided into subtypes. In this study, I focused on subtype H1N1 because it is the subtype that caused two severe pandemics in 1918 and 2009.

The evolution of influenza virus is shaped by antigenic drift. New influenza strains escape the host's immune clearance and add a significant healthcare burden – adults on average get flu twice a decade, while children on average are infected every other year (11). However, the frequent mutation does not dramatically change the total number of circulating stains: new strains commonly replace the old ones in circulating populations, which is reflected in a "ladder shaped phylogenetic tree" (12). This phenomenon indicates evolutionary forces very different from those that shape other RNA viruses. Norovirus, for example, had as many as 21 distinct strains co-circulating in the north of England from 1998 to 2001 alone (13).

Apart from mutation, there are two important mechanisms driving virus evolution, recombination and reassortment. In the case of influenza virus, homologous recombination through template-switching during the replication of the RNA genome is very rare (14). During replication, the newly synthesized viral genomes are rapidly encapsidated into viral ribonucleoproteins (vRNPs), largely reducing their chance to reach a second replicating polymerase complex. Recombination is also rare in all negative-sense RNA viruses (15). On the other hand, reassortment happens in influenza A viruses frequently. Reassortment is the shuffling of genome sequences when different strains of viruses co-infect a host cell (16). It has played an

important role in the evolution of the influenza virus. Segmental reassortment can cause antigenic shift, a new combination of influenza virus genomes with major sequence change in one or more gene segments (17). Antigenic shift often generates mismatches in host immune recognition, antibody response, and vaccination, and is therefore associated with more severe cases and widespread transmission (18, 19, 20). The 2009 pandemic H1N1 strain is a result of the reassortment between viruses of the American triple-reassortant (TR) and Eurasian avian-like (EA) swine influenza lineages (21). The two gene segments, neuraminidase (NA) and matrix (M), from the EA swine lineages significantly increased the transmission of viral particles via respiratory droplets (22). Reassortment also contributes to the evolution of seasonal influenza viruses. It can lead to new antigenic clades (21), altered replication efficiency (23), and the emergence and spread of drug resistance (24). Reassortment can be deleterious, due to the breakage of segments that have been co-evolved for a long time (25, 26). Phipps *et. al.* found that seasonal influenza H3N2 strains often reassort with the 2009 pandemic H1N1, but the reassortants are attenuated compared to parental strains (27). Influenza A virus is the only type of influenza virus that experiences significant reassortment and antigenic shift, while the other three types are confined to one species and only accumulate mutations gradually through antigenic drift.

1.1.2 Viral Fitness

Viral fitness can be viewed at different levels. For an individual viral particle, viral fitness means its capacity to produce infectious progeny (28). This definition focuses on the virus's intrinsic capacity to replicate in a given environment, therefore it is also referred to as replicative fitness. On a populational level, viral fitness typically refers to the prevalence of viral genetic material in the field over time and therefore is influenced by the virus's ability to evade

immune clearance, survive outside the hosts, and transmit to new hosts (29). The capacity of a virus type (variant, serotype, or subtype) to become dominant is referred to as epidemiological fitness (30). In this study, I will mainly examine the virus's replicative fitness.

Replicative fitness can be assessed with multiple methods, each with its own strengths and weaknesses. At the most detailed level, polymerase activity assays can measure the amount of RNA replication product, viral RNA (vRNA), with or without an infection. More vRNA accumulation within a certain period of time indicates higher replicative fitness. This method focuses on the functions of the polymerase and is helpful in revealing important structural factors contributing to replication and transcription. Some assays are even able to measure the efficiency of individual steps of the replication or transcription process (31, 32). On the other hand, polymerase activity is not the sole factor contributing to viral fitness (33). Cases in which replicative fitness during infection differs from polymerase activity are not uncommon.

Replicative fitness can also be assessed by mixed infections of two or more viral variants, where the difference in genome frequency between two variants becomes bigger over the passages (34). The fitness of the mutated virus relative to the wildtype is then determined by the copy number of its genome compared to that of the wildtype by RT-PCR. These competition assays are more sensitive to minor differences in replication efficiency but can only examine the fitness of one variant at a time. In this study, I measured viral fitness by letting all variants compete against each other during serial passaging and calculating the frequency change of every variant before and after passages. I will demonstrate the precision and accuracy of this method in later chapters.

1.1.3 Mutational Effects on Viral Fitness

Mutation has mixed effects on viral fitness. Mutation is the source of new genetic material, providing possibilities for more beneficial phenotypes. However, as a result of

prolonged adaptation, the replicative fitness of the virus is usually at a high level, if not at an optimum in its current host environment. The vast majority of mutations are detrimental to fitness for viruses (35, 36, 37). In a site-directed mutagenesis study to measure the genome-wide distribution of mutational fitness effects of an H1N1 influenza virus, 40% of 95 randomly generated mutations were lethal or highly detrimental, and 50% were mildly detrimental to neutral (38). Even when facing new selective pressure, most adaptation happens with a replicative fitness cost. Under host immunity, variants capable of escaping neutralizing antibodies or cytotoxic lymphocytes have an enormous fitness advantage, but the escape mutations often come with a significant reduction in replicative fitness, needing a compensatory mutation to restore the virus's overall fitness (39, 40).

Since most mutations are deleterious, the high mutation rate of influenza viruses adds great fitness burden. While it seems counterintuitive that influenza and other RNA viruses evolved to maintain this disadvantageous trait of high mutation rate, research on a poliovirus anti-mutator suggests that the high mutation rate of RNA polymerase may be a result of selection for elevated replication speed, and that high mutation rate is an unfavorable byproduct of selection (41). In fact, the mutation rate of RNA viruses is close to the upper limit for the viruses to carry the fitness cost from the accumulation of deleterious mutations (42). Chemical mutagens or mutagenic nucleoside analogs can reduce viral fitness or be used as anti-viral drugs by elevating the mutation rate over the upper limit (43, 44), and lethal mutagenesis has been used as a therapeutic strategy to treat influenza viruses with low traditional drug resistance (45). When treating with mutagenesis drugs, variants with high replicative fidelity are selected, but they frequently have lower replication efficiency compared to wild-type viruses without drug treatment (46, 47).

For organisms with high mutation rates like influenza viruses, mutational robustness becomes essential for the longevity of the viral population. High mutational robustness means the fitness of a virus is less altered by mutations. Influenza variants with high mutational robustness can tolerate more mutations without significantly lowering their fitness and wait for complementary or beneficial mutations to appear. In contrast, even if a new mutation is highly beneficial, it will quickly get lost due to the high mutation rate before it rises to a frequency that can pass the transmission bottleneck. Therefore, the competition among influenza viruses is the competition of who can tolerate most mistakes. If you visualize the mutational effects in a 3D space, the results can also be framed as “survival of the flattest” (48).

When focusing on a protein, mutational robustness is often called mutational tolerance and is different for every amino acid residue. Mutation tolerance is related to the residue’s functions. Residues that possess crucial enzymatic functions are usually less tolerant of mutations, while antigenic proteins are generally more tolerant. Visher *et al.* found that the mean fitness of mutations in the influenza virus surface protein hemagglutinin and neuraminidase was 0.88, while that for other proteins was only 0.78 (49). The difference in average mutational effects can reflect the history of positive selection on surface proteins related to immune escape and host entry (50). Another discovery supporting this theory is that the globular head of influenza surface protein hemagglutinin exhibits a much higher inherent tolerance for mutations than the stalk (51).

Synonymous mutations can also affect viral fitness or evolutionary fate. In a mutagenesis experiment to measure the mutational landscape of poliovirus polymerase, Acevedo *et al.* found that approximately 10% of synonymous mutations are lethal, indicating strong incompatibility at the codon level (52). Translational efficiency and the constraints on RNA structures have been

proposed as the major explanations for the fitness differences among codons encoding the same amino acids in RNA viruses (53, 54). Additionally, synonymous mutations may alter the virus's mutational robustness by opening up different possibilities for subsequent mutations (55). For example, although both AGG and CGG code for arginine, among their next possible mutations, AGG has 78% non-synonymous mutations, while CGG has only 56%. The accessible evolutionary pathways are particularly important for the adaptability of RNA viruses. Synonymous populations with mutated codons are shown to have lower fitness and attenuated virulence both *in vitro* and *in vivo* (55, 56, 57).

Interactions between mutations introduce further complexity to the mutational effects. The genetic interaction between multiple mutations in a genome is called epistasis (58). Because of epistasis, the same mutation can have varying effects depending on the presence of other mutations, and the nucleotide sequence of all other related genes that may interact with the mutation of interest is called the genetic background (59). Epistasis greatly expands the potential of fitness change brought by mutations and makes viral evolution highly path-dependent. The mutation that appears first opens or closes possibilities of the rise of new strains. Therefore, the evolution pathways we observe in nature are only a small fraction of all paths possible in the mutational landscape. Examining the mutations that did not happen and the paths untraveled by natural evolution history can help us better understand the selective forces that have shaped influenza viruses and recognize the mutations of epidemiologic importance.

1.1.4 Selective Forces that Shape Influenza Virus Evolution

Positive selection, negative selection, and genetic drift create a complex web of evolutionary forces that shape influenza evolution. Influenza virus produces a large population size within hosts and mutates frequently, yet patient samples show low genetic diversity (60).

The evolutionary force for influenza virus population within hosts is dominated by strong purifying selection (61). Between hosts, stochastic processes are the major determinants of which influenza genome sequences get transmitted and preserved, a process referred to as genetic drift (62). From time to time, beneficial mutations arise. However, the successful spread of a novel mutation largely depends on the size of the transmission bottleneck. Influenza virus has a very narrow transmission bottleneck of 1~2 genomes, therefore, a beneficial mutation needs to rise to a genome frequency above 2% to be passed to a new host (60). A typical influenza infection is acute, and the viruses are cleared by host immunity within a week or two. The limited replication time is usually not long enough for a novel mutation to accumulate to the transmission threshold; therefore, the evolution of local or seasonal influenza viruses is also driven by genetic drift (61, 63). On the other hand, positive selection is observed both in lab cultures and in patients with persistent infections (64, 65). The immune-driven antigenic variants concentrated by rare positive selections often cause larger epidemics and drive global influenza virus evolution.

1.2 Evolutionary Constraints on Influenza Virus Polymerase

1.2.1 General Constraints on Protein Evolution

Protein evolution is under various functional and structural constraints. Solvent accessibility, molecular flexibility, maintenance of intermolecular interactions, and the preservation of key secondary structures are major constraints that apply to all proteins (66). Solvent accessibility describes how much a molecule is buried or exposed. Ramsey *et al.* found a linear relationship between the relative solvent accessibility of residues in a protein and their evolutionary rate (67). Among the buried areas, solvent-inaccessible polar side chains provide the strongest structural and functional constraints. Mutations of residues containing these side chains could destabilize the protein structure. In addition, buried residues form more contacts

than exposed residues on the surface. Therefore, the buried residues are usually less tolerant to mutation and evolve at a lower rate (68). Amino acids can pose their side chains in different ways, and the position of one amino acid blocks or allows the position of other amino acids near it. The intrinsic flexibility of amino acid ensembles within a protein is important to the protein's conformational change and function and shows a strong correlation with residues' evolutionary rates (69). The level of solvent exposure and molecular flexibility are interconnected because buried sites face restrictions from more nearby molecules; but there are additional features that can be reflected by molecular flexibility but not solvent accessibility, such as intrinsically disordered regions. Interacting proteins co-evolve to preserve the functionality of the connections, and therefore evolve at a lower rate (70). Notably, the evolutionary rate of a residue is not dependent on the number of protein-protein interactions the residue participates in, but the group of residues with the largest number of interactions evolves slower than the rest of the protein (71). Some interactions, such as interfaces of subunits that comprise larger complexes, are permanent and obligate (72), while others, such as enzyme-substrate, are transient encounters that are only needed for certain processes (73). In co-evolution, the frequency of an obligate interaction requiring compensatory mutations is significantly higher than that of a transient interaction. Therefore, while interface residues are in general less tolerant to mutation, residues that are involved in obligate interactions tend to be more conserved than those in transient interactions (74). Finally, a mutation cannot disrupt the key secondary structures of the protein. The key secondary structures include α-helix, β-sheet, and coil, which all need proper torsion angles for maintenance. Some amino acids have special features on their side chains, and therefore support certain torsion angles and that are structurally necessary. For instance, glycine

has no side chain and allows a positive φ torsion angle and is abundant in helices; amino acids with side chains, on the other hand, support coils and turns (75).

1.2.2 Specific Constraints for Influenza Virus Polymerase

Compared to polymerases of other RNA viruses, influenza virus RdRp is under additional constraints, because it goes through complicated interactions with RNAs, other viral proteins, and host proteins during replication and transcription.

Influenza A virus has eight RNA segments, namely polymerase basic 2 (PB2), polymerase basic 1 (PB1), polymerase acidic (PA), hemagglutinin (HA), nucleoprotein (NP), neuraminidase (NA), matrix (M), and non-structural proteins (NS) (Figure 1.1) (76). The eight RNA segments encode for eleven proteins, and three proteins – PB1, PB2, and PA – together form the heterotrimeric polymerase complex. The RdRp domain, PB1 needs to cooperate with the cap-binding domain (PB2) and the endonuclease domain (PA) to perform replication and transcription functions.

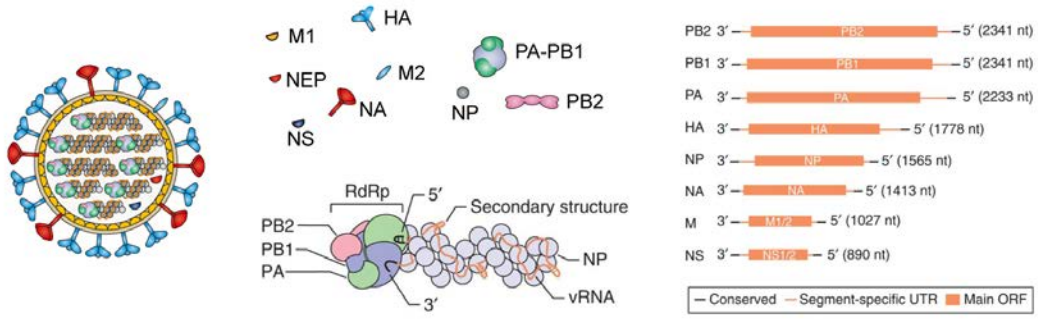


Figure 1.1 The scheme of influenza A virus viral particle and genome. Adapted from te Velthuis et al., 2016 (87) and Fodor & te Velthuis, 2019 (95). M1: matrix protein 1; NEP: nuclear export protein; NS: non-structural protein; HA: hemagglutinin; NA: neuraminidase; M2: matrix protein 2; NP: nucleoprotein; PA: polymerase acidic; PB1: polymerase basic 1; PB2: polymerase basic 2. In viral particles, genome segments bind to nucleoproteins and fold over to bind with a polymerase complex.

PB1 is the center of the polymerase complex (Figure 1.2). It has 15 N-terminal residues that interact with PA (77) and 80 C-terminal residues that interact with PB2 (78). The central

region of PB1 (residues 21-699) has a right-handed fold that is conserved in all RdRp, comprising fingers, fingertips, palm, and thumb subdomains. This shape helps position replication or transcription substrates and metal ions for catalysis, at the same time preserving enough freedom for dynamic changes needed during RNA synthesis. The influenza virus PB1 contains conserved residues within the subdomains, called motifs. The six motifs (pre-A, also known as F, and A to E motifs) have sequences shared by many polynucleotide polymerases with important functions (79, 80). One aspartic acid residue in motif A helps RdRp to preferably incorporate nucleoside triphosphates (NTPs) rather than deoxyribonucleoside triphosphates (dNTPs). Another Asp-Asp sequence in motif C forms a cluster with the Asp in motif A in a three-dimensional structure. The three-Asp cluster catalyzes RNA synthesis by coordinating two Mg^{2+} ions, with one Mg^{2+} promoting the deprotonation of the nascent strand 3′-OH group, and the other helping to stabilize the transitioning NTP and release its pyrophosphate group (81). Motif B has a methionine-rich loop involved in stabilizing the interaction between the template and the incoming NTP (82). Motif D residues Lys 480 and Lys 481 are involved in NTP binding, and Arg 486 is necessary for extracting the proton from the 3′-OH of the primer and transferring it from the active site to pyrophosphate (83). The β-hairpin in motif E helps stabilize the substrate and priming nucleotide during replication (82). Motif pre-A (motif F) forms an NTP entry tunnel and only presents in RdRp, because RdRp has a closed structure where the active site is buried, while other polynucleotide polymerases resemble an open hand (84).

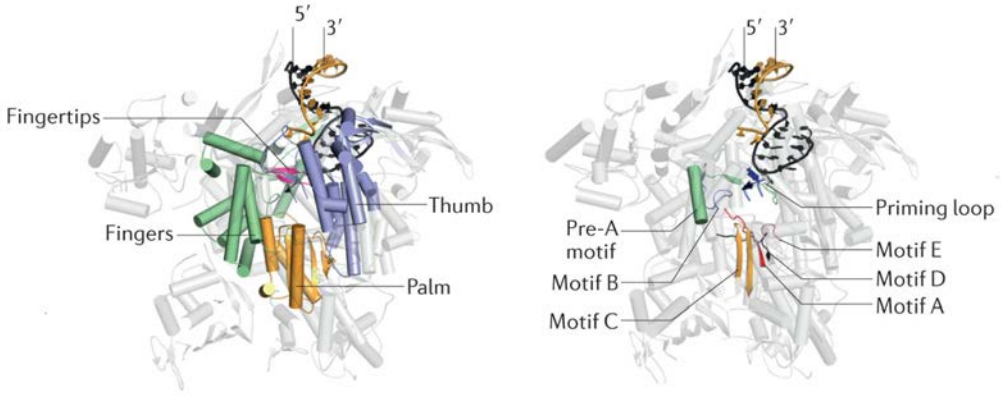


Figure 1.2 PB1 subunit in the influenza A virus polymerase complex. Adapted from te Velthuis et al., 2016 (87). Cartoon models of the influenza A virus polymerase complex (Protein Data Bank (PDB) entry: 4WSB), including PB1, PB2, and PA subunits, as well as the negative-sense, template viral RNA. Left: the right-handed arrangement of the PB1 fingers, palm and thumb subdomains and the fingertips; Right: the conserved sequences in most influenza viruses as well as their relative positions to viral RNA and the priming loop. The 5′ and 3′ termini of the viral RNA were marked in dark grey and yellow, respectively.

The PA subunit consists of two subdomains facing each other, wrapping around the external side of PB1 fingers and palm subdomains (Figure 1.3). On the N-terminal side, the endonuclease subdomain (residues 1-195) is exposed to solvent and interacts with the PB2 cap-binding subdomain. The C-terminal subdomain (residues 258-714) binds to the N-terminus of PB1. Two subdomains are linked by the residues 196-257. These linker residues also interact with the PB1 surface at multiple points, forming hydrophobic and polar contacts.

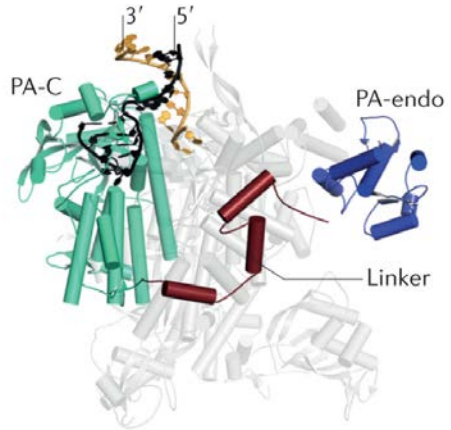


Figure 1.3 PA subunit in the influenza A virus polymerase complex. Adapted from te Velthuis et al., 2016 (87). Cartoon models of the influenza A virus polymerase complex (PDB entry: 4WSB), including PB1, PB2, and PA subunits, as well as the negative-sense, template viral RNA. From left to right, the colored structures showed the C-terminal domain, the linker, and the endonuclease domain of the PA subunit.

The PB2 subunit attaches to PB1 at the opposite site of PA and can also be divided into N-terminal and C-terminal parts (Figure 1.4). The N-terminal part includes the N terminus, N-linkers, and the lid subdomains. It is positioned above, located opposite to the PA linker and interacts with the PB1 C-terminal extension and thumb subdomain. Helix α4 interacts with the template, helping it enter the RdRp active site, while residues 55-103, containing β-sheet 1, 2, 3 and helix α5, 6, often form parallel structures to PB1 helices, to sustain the PB1 thumb subdomain from the back. The lower part of the N-terminal features more β-ribbons. They make hydrophobic contacts with the PA C-terminal and PB1 thumb and palm subdomains. Below the entire PB1, PA, and PB2 N-terminal group, the PB2 C-terminal lies horizontally with the cap-binding subdomain, cap-627 linker, 627-subdomain, and the nuclear-localization signal (NLS) subdomain. The mid subdomain connects the N- and C-terminal parts of PB2. The loop containing residue 627 is of special importance because it specifies the host range of the virus (85). Residue 627 is usually lysine for human influenza viruses, glutamate for avian influenza viruses, and serine for bat influenza viruses.

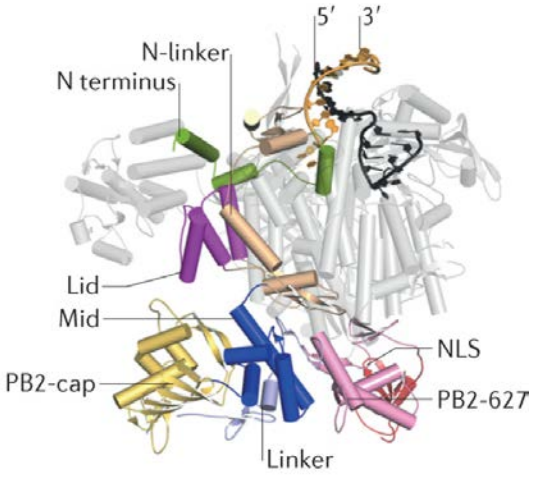


Figure 1.4 PB2 subunit in the influenza A virus polymerase complex. Adapted from te Velthuis et al., 2016 (87). Cartoon models of the influenza A virus polymerase complex (PDB entry: 4WSB), including PB1, PB2, and PA subunits, as well as the negative-sense, template viral RNA. The structure was turned 180° from the structures in Figure 1.2 and 1.3 with z-axis fixed. The colored structures marked the N terminus, N-linker, lid, mid, cap-binding, cap-627 linker, 627-, and nuclear-localization signal (NLS) domains of the PB2 subunit.

In viral particles, all influenza gene segments (vRNA) are bound to multiple copies of nucleoprotein, and fold over to bind both ends of a single, trimeric polymerase complex (86). This vRNA-NP-polymerase assembly is called viral ribonucleoprotein (vRNP). The influenza virus infection cycle begins with the binding of virus particles to the receptors on host cell surface mediated by HA. The viral particle is internalized through endocytosis. The low pH in late endosome triggers the fusion of the viral and endosomal membranes, causing the release of vRNPs into cytoplasm. The vRNPs will then be transported into the nucleus. In the nucleus, the polymerase complex bound to vRNA transcribes the vRNA to viral mRNAs for the production of viral proteins and replicates vRNA to generate progeny vRNPs. Newly synthesized genomic RNAs will then be transported out of the nucleus and assembled with other viral proteins to form progeny viral particles (87).

Influenza virus replication is a two-step, primer-independent process. First, the polymerase complex replicates the negative-sense vRNA into positive-sense complementary RNAs (cRNAs). The cRNAs will bind to nucleoproteins and a second polymerase complex and form complementary RNPs (cRNPs). Then the cRNA will act as the intermediate template to produce progeny genome vRNAs. The two rounds of replication require a series of conformational changes of the three subunits but have some differences in mechanism. When copying vRNA into cRNA, influenza virus RdRp initiates replication with a β-hairpin protruding from the PB1 thumb subdomain to the active site called priming loop (81). This initiation happens at positions 1 and 2 of the 3′ end of the vRNA template and, therefore is called terminal initiation. When the RdRp copies the cRNA back to vRNA, the polymerase initiates at positions 4 and 5 of the 3′ end of the template and is called internal initiation. Internal initiation does not require support from the priming loop (31). Following the initiation, replication starts by

synthesizing the dinucleotide pppApG and continues by pulling the template vRNA or cRNA through the active site. The replication from cRNA to vRNA requires a second polymerase (88), which is speculated to induce the necessary conformational change of the primary polymerase to activate the replication process (Figure 1.5). Host protein ANP32 is essential for the recruitment of the secondary polymerase and is therefore necessary for an efficient vRNA synthesis (89, 90). ANP32 directly interacts with the PB2 627 domain (91), and both ANP32 and the amino acid at the residue 627 of PB2 are species specific (92, 93). The compatibility between ANP32 and PB2 residue 627 is a major restriction factor for influenza virus host range (91-94). Human isoforms of ANP32, ANP32A and ANP32B, lack a 33-amino-acids sequence compared to avian ANP32, and therefore are unable to support the activity of influenza polymerase with PB2 627E. The mutation E627K allows the avian influenza polymerase to replicate efficiently in human cells (94, 95). Finally, at the termination stage, the 5′ end of the vRNA or cRNA template will be released from the binding pocket and replicated.

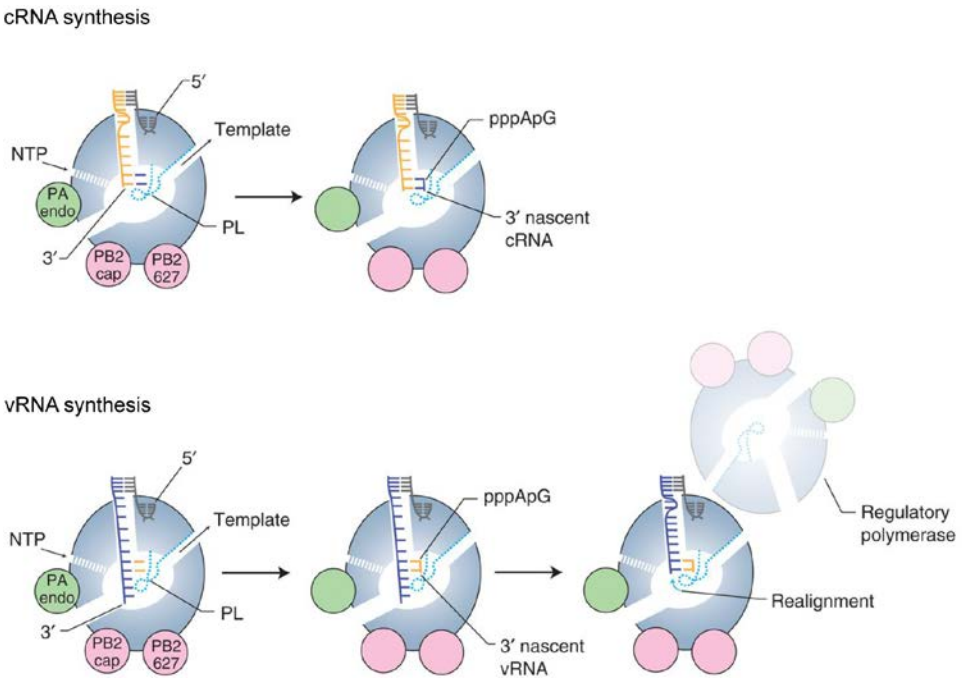


Figure 1.5 Differences in the initiation mechanism between the two steps of influenza virus replication. Adapted from Fodor & te Velthuis, 2019 (95). The synthesis of cRNA begins with terminal initiation, which happens at the positions 1 and 2 of the 3′ end of the template RNA; the synthesis of vRNA begins with internal initiation, requiring the backtracking and realignment of the template RNA with the aid of a regulatory polymerase.

The transcription of influenza vRNA, however, is a primer-dependent process. The primer used for this process is a capped pre-mRNA obtained from the host mRNAs via a process called ‘cap-snatching’ (96). The cap-snatching is realized by PB2 and PA, during which PB2 captures the 5′ cap of the nascent host capped mRNA, and PA cleaves it (85). The primer forms 1~3 base pairs with the U at the 3′ end of the vRNA and is directed towards the polymerase active site to initiate RNA synthesis. The priming loop, though useful for RNA replication, blocks the entry of the 3′ end vRNA template and the template-primer binding in the transcription. The transition from initiation to elongation requires the priming loop to become disordered and extrude out of the active cavity (97). In the meantime, the PB1 thumb subdomain rotates to accommodate for the priming loop extrusion. At the elongation stage, the PB2 lid subdomain enforces strand separation, directing the template into the template exit channel and

mRNA into the product exit channel (Figure 1.6). When the U-stretch reaches the active site, the polymerase stutters, creating a poly(A) tail.

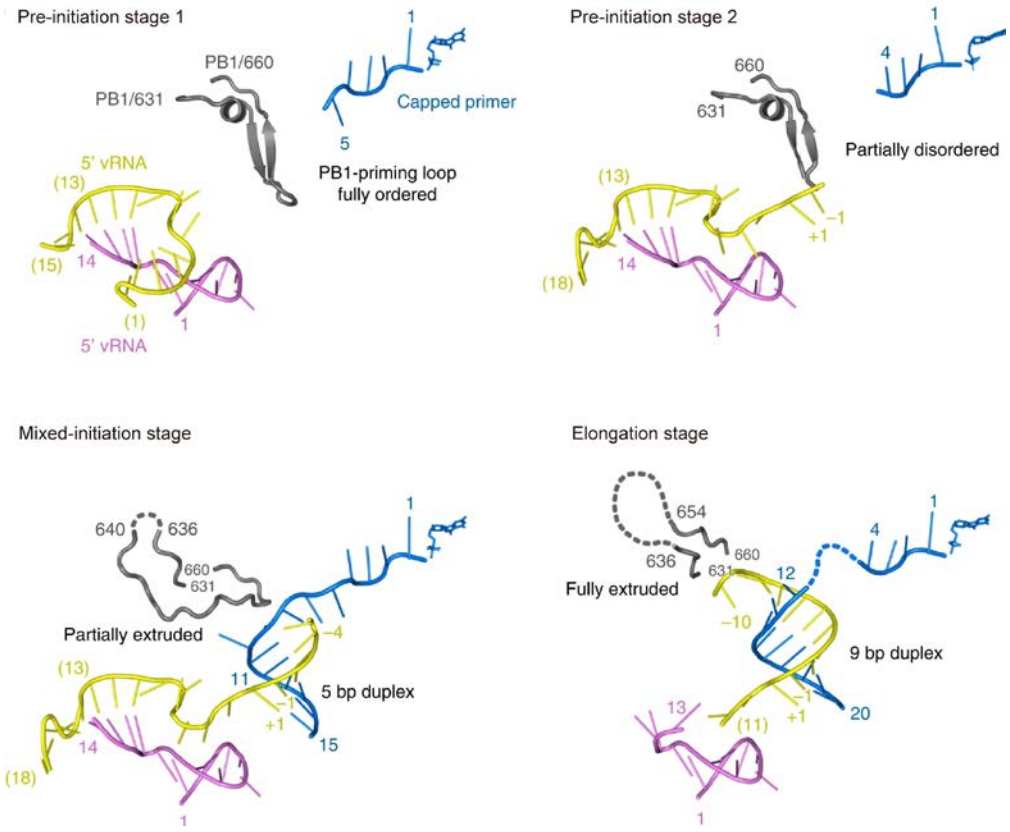


Figure 1.6 Influenza A virus transcription initiation. Adapted from Kouba et al., 2019 (97). During the initiation of transcription, the priming loop (gray) goes from fully ordered to fully disordered and extrudes out of the active site as the template (yellow) and the primer (blue) enter the active site.

While mutations and selection happen frequently, adaptation often requires the accumulation of multiple (98, 99) or complementary mutations (47). Compared to antigenic proteins HA and NA, polymerase proteins accumulate adaptative substitutions at much lower rates (100). Maintaining the aforementioned interactions and frequent conformational changes is speculated to be a substantial evolutionary constraint. On the other hand, each of the numerous interactions is carried out by a few key residues and amino acids. Sometimes, a single mutation can have a significant mutational effect on viral fitness.

1.3 Existing Research regarding Mutational Effects on Influenza Virus Polymerase

The mutational effects on the influenza RdRp have been a constant research interest, especially after the 2009 H1N1 outbreak. By comparing the 2009 H1N1 pandemic virus to previous strains, Santos *et al.* found that three mutations, L298I, R386K, and I517V in PB1, improve overall viral fitness and putatively support viral adaptation to the human host (98). Mutation of conserved amino acids usually ends up in attenuation. As an example, PB1 L319Q was found to reduce both pathogenicity and transmission in animal models. The mutation also made the virus more sensitive to higher or lower temperature (101).

Some mutations increase viral fitness in new hosts. Influenza H3N8 viruses typically infect wild birds, with sporadic spill-over to mammals. The strain with high transmissibility among mammals was found to have a unique PB1 mutation, S524G. Minigenome polymerase assays showed that PB1 S524G enhances polymerase activity, therefore rescuing the poor replication of avian flu in mammalian cells (102). A/duck/Hunan/S4020/2008 (DK/08) and A/chicken/Guangxi/S2039/2009 (CK/09) are both H5N1 viruses, but CK/09 is much more virulent in mammals. Feng *et al.* created and tested over ten CK/09 and DK/08 chimeric viruses with mutations and demonstrated that the glycine at position 622 of PB1 contributed to the high virulence of CK/09 in mice by more efficient vRNA binding (103). Other examples include PB1 473V and 598P, which were also reported to increase H5N1 polymerase activity in mammalian cells (104).

Some mutations affect viral fitness by influencing post-translational modifications. Residue 612 of influenza A virus PB1 is a SUMOylation site and is occupied by a conserved lysine residue. SUMOylation at site 612 has no effect on PB1 protein stability or localization in host cells but is essential for PB1 to bind viral RNA, therefore vital to replication. Mutant viruses

with K612R were found to be less efficient in binding vRNA and have reduced replication and transmission in ferrets (105).

Some mutations alter viral fitness by changing replication fidelity. Depending on the environment, higher or lower replication fidelity can both be beneficial to the virus. When a strain first expands to a new host species, such as the H1N1 pandemic strain in 2009, low replication fidelity is beneficial for the virus to generate new mutations quickly and adapt to humans. H1N1 influenza viruses with the S216G mutation on PB1 exhibited greater mutation potential and were less virulent in mammalian infections (106). Both features are beneficial to viral propagation. On the other hand, as humans start to treat RNA viruses with mutagenic antiviral drugs, forcing viruses to accumulate detrimental mutations, high replication fidelity can bring significant fitness advantages. PB1 K229R increases polymerase replication fidelity and confers favipiravir resistance (47). This type of mutation usually lowers viral fitness by itself, due to the “speed-fidelity trade-off”. Therefore, the fixation of an “anti-mutator” mutation is often accompanied by additional mutations rescuing for reduced polymerase activity. In the case of K229R, a compensatory mutation, P653L, in the PA subunit can restore the polymerase activity, and the double mutant supports effective replication both with and without favipiravir (47).

Some mutations influence viral fitness through alternative functions of PB1. PB1-F2 is a small protein expressed by some influenza A viruses through an alternate open reading frame. It was first identified in 2001 and found to be immunogenic and to antagonize interferon activities (107, 108). A single amino acid change, N66S, in the PB1-F2 protein increases virulence in both H5N1 and H1N1 backgrounds (109). However, this finding may not apply to all influenza strains. The 2009 H1N1 pandemic strain does not code for full-length PB1-F2, and the

expression of PB1-F2 either with 66N or 66S resulted only in different proinflammatory cytokine levels but not symptoms in mice or ferrets (110).

However, most studies only examined isolated PB1 mutations, either naturally occurring or created based on previous hypotheses of their effect. One of the only two-large scale research studies examined 16 mutations on A/WSN/33 (H1N1) PB1, among which 6 were lethal, and all viable mutants had lower-than-wildtype fitness except for two synonymous mutants (49). The other study focused on the conserved motifs on PB1 (79). Using mini-replicon assays, Chu *et al.* assessed the transcription and replication efficiency of 25 mutations on any of the four PB1 motifs found in natural isolates. While most mutations showed significantly reduced polymerase activity, one mutation, K480R, supported more efficient vRNA, cRNA, and mRNA synthesis. However, higher polymerase activity did not bring higher viral fitness: viruses with the PB1 K480R mutation exhibited similar growing kinetics as the wild type in human airway epithelial cells.

1.4 Novel Technologies to Study Influenza Virus Polymerase

Targeted mutagenesis is a powerful tool to assess the functional consequences of amino acid changes on proteins but limited by its scale (111). Random mutagenesis by error-prone PCR is a simplistic technique that generates massive mutations based on a single parent gene, but the number, position, or type of the mutations on each individual variant cannot be controlled (112, 113, 114). To systematically study the functional constraints and evolutionary potential of the influenza virus polymerase, we need a method to generate single amino acid variants accurately on a large scale and a sensitive readout for the fitness of each variant.

Deep mutational scanning (DMS) is developed to couple genotype-phenotype relationships (115). It consists of three parts: a library of all or nearly all mutations of a protein,

an assay where the phenotype of interest can be read out, and deep sequencing to identify mutants' genotypes with high accuracy and precision. DMS has been applied to human immunodeficiency virus (HIV-1) envelope protein (116, 117), Zika virus envelope protein (118), and SARS-CoV-2 receptor binding domain (119, 120) to evaluate mutational effects on viral growth, sensitivity to neutralizing antibodies, proper expression and folding, and receptor binding. The application of DMS is not limited to viral proteins. The technology has also been used to reveal the amino acid sequences that affect enzyme solubility and catalytic activity (121, 122), the key sites for RNA binding affinity in yeast RNA recognition motif (123), and the structural basis for α-synuclein activity (124). For influenza virus, DMS has been performed on hemagglutinin (51, 125, 126), neuraminidase (127, 128), matrix protein M1 (129), nucleoprotein (130), and PB2 protein (131) to study amino acid preferences, replicative fitness, mutational tolerance, drug resistance, sensitivity to host inhibitory proteins, and adaptation to a new host. A DMS on the core subunit of the RdRp complex, PB1, will uncover the key amino acids and residues for polymerase functions and help complete the genotype-phenotype map for influenza virus.

This project used an overlapping polymerase chain reaction (PCR)-based, site-directed mutagenesis approach to construct an exhaustive library of PB1 single amino acid substitutions. The mutagenesis was achieved by two rounds of PCR, in which target codon mutations were coded within two pairs of overlapping mutagenesis primers (131, 132). In the first round, the mutagenesis primers created two segments that covered the genome before and after the target codon mutation, with the end of the first segment overlapping with the beginning of the second. The second round of PCR used the products of the first round as the template and created a full genome with target codon mutation. The genome segments with single codon mutations were

then ligated into plasmid vectors and transformed into competent cells for amplification. Finally, the variant virus libraries were generated by transfecting the variant plasmid library along with the plasmids encoding the rest of the influenza genes (133). The variant virus libraries would ideally have every possible mutation of PB1, therefore, were highly diverse. In the meantime, each virus in the libraries would only have a single mutation. When comparing all codons or amino acids at the same residue, the frequency of a certain mutation was very low. Due to the unusual characteristics of the variant virus libraries, deep sequencing and the fitness measurement of variant viruses also required customized library preparation.

In this project, the replicative fitness of each variant was measured by their frequency change before and after serial passaging in cells. High-throughput sequencing (HTS) technology allows the sequencing of millions of DNA fragments simultaneously and makes it possible to quantify the abundance of a mutation in a population and the frequency change of a mutation before and after any given selection (134). However, the average per-base error rate of HTS platforms such as Illumina is around 1% (135), which is higher than the frequency of most codon mutations in the library. When the frequency of the target variant is lower than the error rate of variant calling, the accuracy of detecting the variant drops dramatically (136). To solve this problem, I adopted a sequencing library preparation strategy called barcoded-subamplicon sequencing to better identify the rare mutations (51, 137, 138). In barcoded-subamplicon sequencing, unique barcodes were not only attached to each library, but appended to each sequence. Sequencing reads with the same barcode were considered to originate from the same molecule. True mutations would occur in all reads with the same barcode, while sequencing errors would only appear in some of the reads and be identified. Barcoded-subamplicon sequencing also replaced the random fragmentation of long genes in conventional next-

generation sequencing (139) with designed subamplicons, which prevented the double counting of the same mutation in different fragments.

1.5 References

1. CDC. 2019. Types of Influenza Viruses. Centers for Disease Control and Prevention. https://www.cdc.gov/flu/about/viruses/types.htm#:~:text=There%20are%20four%20types%20of.

2. Liu R, Sheng Z, Huang C, Wang D, Li F. 2020. Influenza D virus. Current Opinion in Virology 44:154–161.

3. Zambon MC. 2001. The pathogenesis of influenza in humans. Reviews in Medical Virology 11:227–241.

4. Peck KM, Lauring AS. 2018. Complexities of Viral Mutation Rates. Journal of Virology 92.

5. Fourment M, Holmes EC. 2015. Avian influenza virus exhibits distinct evolutionary dynamics in wild birds and poultry. BMC Evolutionary Biology 15.

6. Bhatt S, Holmes EC, Pybus OG. 2011. The Genomic Rate of Molecular Adaptation of the Human Influenza A Virus. Molecular Biology and Evolution 28:2443–2451.

7. McCrone JT, Woods RJ, Monto AS, Martin ET, Lauring AS. 2020. The effective population size and mutation rate of influenza A virus in acutely infected individuals. Preprint. https://www.biorxiv.org/content/10.1101/2020.10.24.353748v1.full.

8. Gojobori T, Moriyama EN, Kimura M. 1990. Molecular clock of viral evolution, and the neutral theory. Proceedings of the National Academy of Sciences 87:10015–10018.

9. Buonagurio DA, Nakada S, Parvin JD, Krystal M, Palese P, Fitch WM. 1986. Evolution of Human Influenza A Viruses Over 50 Years: Rapid, Uniform Rate of Change in NS Gene. Science 232:980–982.

10. Alpert T, Brito AF, Lasek-Nesselquist E, Rothman J, Valesano AL, MacKay MJ, Petrone ME, Breban MI, Watkins AE, Vogels CBF, Kalinich CC, Dellicour S, Russell A, Kelly JP, Shudt M, Plitnick J, Schneider E, Fitzsimmons WJ, Khullar G, Metti J. 2021. Early introductions and transmission of SARS-CoV-2 variant B.1.1.7 in the United States. Cell 184:2595-2604.e13.

11. Kucharski AJ, Lessler J, Read JM, Zhu H, Jiang CQ, Guan Y, Cummings DAT, Riley S. 2015. Estimating the Life Course of Influenza A(H3N2) Antibody Responses from Cross-Sectional Data. PLOS Biology 13:e1002082.

12. Lee J, Neher R, Bedford T. 2023. Real-time tracking of influenza A/H3N2 evolution. Nextstrain. https://nextstrain.org/flu/seasonal/h3n2/ha/2y. Retrieved 10 September 2023.

13. Gallimore C, Green J, Lewis DA, Richards A, Lopman BA, Hale A, Eglin R, Gray J, Brown D. 2004. Diversity of Noroviruses Cocirculating in the North of England from 1998 to 2001. Journal of Clinical Microbiology 42:1396–1401.

14. Boni MF, Zhou Y, Taubenberger JK, Holmes EC. 2008. Homologous Recombination Is Very Rare or Absent in Human Influenza A Virus. Journal of Virology 82:4807–4811.

15. Chare ER, Gould EA, Holmes EC. 2003. Phylogenetic analysis reveals a low rate of homologous recombination in negative-sense RNA viruses. Journal of General Virology 84:2691–2703.

16. Vijaykrishna D, Mukerji R, Smith GJD. 2015. RNA Virus Reassortment: An Evolutionary Mechanism for Host Jumps and Immune Evasion. PLOS Pathogens 11:e1004902.

17. Nelson MI, Viboud C, Simonsen L, Bennett RT, Griesemer SB, St. George K, Taylor J, Spiro DJ, Sengamalay NA, Ghedin E, Taubenberger JK, Holmes EC. 2008. Multiple Reassortment Events in the Evolutionary History of H1N1 Influenza A Virus Since 1918. PLoS Pathogens 4:e1000012.

18. Goldstein EJ, Harvey WT, Wilkie GS, Shepherd SJ, MacLean AR, Murcia PR, Gunson RN. 2018. Integrating patient and whole-genome sequencing data to provide insights into the epidemiology of seasonal influenza A(H3N2) viruses. Microbial Genomics 4.

19. Saunders-Hastings P, Krewski D. 2016. Reviewing the History of Pandemic Influenza: Understanding Patterns of Emergence and Transmission. Pathogens 5:66.

20. Vijaykrishna D, Poon LLM, Zhu HC, Ma SK, Li OTW, Cheung CL, Smith GJD, Peiris JSM, Guan Y. 2010. Reassortment of Pandemic H1N1/2009 Influenza A Virus in Swine. Science 328:1529–1529.

21. Garten RJ, Davis CT, Russell CA, Shu B, Lindstrom S, Balish A, Sessions WM, Xu X, Skepner E, Deyde V, Okomo-Adhiambo M, Gubareva L, Barnes J, Smith CB, Emery SL, Hillman MJ, Rivailler P, Smagala J, de Graaf M, Burke DF. 2009. Antigenic and Genetic Characteristics of Swine-Origin 2009 A(H1N1) Influenza Viruses Circulating in Humans. Science 325:197–201.

22. Lakdawala SS, Lamirande EW, Suguitan AL, Wang W, Santos CP, Vogel L, Matsuoka Y, Lindsley WG, Jin H, Subbarao K. 2011. Eurasian-Origin Gene Segments Contribute to the

Transmissibility, Aerosol Release, and Morphology of the 2009 Pandemic H1N1 Influenza Virus. PLoS Pathogens 7:e1002443.

23. Hsieh E-F, Lin S, Chee Keng Mok, Chen G-W, Huang C-H, Wang Y-C, Chen T, Chen C-J, Ojcius DM, Shih S. 2011. Altered Pathogenicity for Seasonal Influenza Virus by Single Reassortment of the RNP Genes Derived From the 2009 Pandemic Influenza Virus. The Journal of Infectious Diseases 204:864–872.

24. Yang J-R, Lin Y-C, Huang Y-P, Su C-H, Lo J, Ho Y-L, Yao C-Y, Hsu L-C, Wu H-S, Liu M-T. 2011. Reassortment and Mutations Associated with Emergence and Spread of Oseltamivir-Resistant Seasonal Influenza A/H1N1 Viruses in 2005–2009. PLoS ONE 6:e18177.

25. Villa M, Lässig M. 2017. Fitness cost of reassortment in human influenza. PLOS Pathogens 13:e1006685.

26. Lowen AC. 2018. It's in the mix: Reassortment of segmented viral genomes. PLoS Pathogens 14.

27. Phipps KL, Marshall N, Tao H, Danzy S, Onuoha N, Steel J, Lowen AC. 2017. Seasonal H3N2 and 2009 Pandemic H1N1 Influenza A Viruses Reassort Efficiently but Produce Attenuated Progeny. Journal of Virology 91.

28. Domingo E, Holland JJ. 1997. RNA Virus Mutations and Fitness for Survival. Annual Review of Microbiology 51:151–178.

29. Wargo AR, Kurath G. 2012. Viral fitness: definitions, measurement, and current insights. Current Opinion in Virology 2:538–545.

30. Domingo E, de Ávila AI, Gallego I, Sheldon J, Perales C. 2019. Viral fitness: history and relevance for viral pathogenesis and antiviral interventions. Pathogens and Disease 77.

31. Deng T, Vreede FT, Brownlee GG. 2006. Different De Novo Initiation Strategies Are Used by Influenza Virus RNA Polymerase on Its cRNA and Viral RNA Promoters during Viral RNA Replication. Journal of Virology 80:2337–2348.

32. te W, Long JS, Barclay WS. 2018. Assays to Measure the Activity of Influenza Virus Polymerase. Springer eBooks 1836:343–374.

33. Fan H, Walker AP, Carrique L, Keown JR, Serna Martin I, Karia D, Sharps J, Hengrung N, Pardon E, Steyaert J, Grimes JM, Fodor E. 2019. Structures of influenza A virus RNA polymerase offer insight into viral genome replication. Nature 573:287–290.

34. Lanxon-Cookson EC, Swain JV, Manocheewa S, Smith RA, Maust B, Kim M, Westfall D, Rolland M, Mullins JI. 2013. Factors affecting relative fitness measurements in pairwise competition assays of human immunodeficiency viruses. Journal of Virological Methods 194:7–13.

35. Domingo-Calap P, Cuevas JM, Sanjuán R. 2009. The Fitness Effects of Random Mutations in Single-Stranded DNA and RNA Bacteriophages. PLoS Genetics 5:e1000742.

36. Sanjuán R, Moya A, Elena SF. 2004. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. Proceedings of the National Academy of Sciences 101:8396–8401.

37. CarrascoP, de la Iglesia F, Elena SF. 2007. Distribution of Fitness and Virulence Effects Caused by Single-Nucleotide Substitutions in Tobacco Etch Virus. Journal of Virology 81:12979–12984.

38. Lyons DM, Lauring AS. 2018. Mutation and Epistasis in Influenza Virus Evolution. Viruses 10.

39. Peut V, Kent S. 2006. Fitness Constraints on Immune Escape from HIV: Implications of Envelope as a Target for Both HIV-Specific T Cells and Antibody. Current HIV Research 4:191–197.

40. Song H, Pavlicek JW, Cai F, Bhattacharya T, Li H, Iyer SS, Bar KJ, Decker JM, Goonetilleke N, Liu MK, Berg A, Hora B, Drinker MS, Eudailey J, Pickeral J, Moody MA, Ferrari G, McMichael A, Perelson AS, Shaw GM. 2012. Impact of immune escape mutations on HIV-1 fitness in the context of the cognate transmitted/founder genome. Retrovirology 9.

41. Fitzsimmons WJ, Woods RJ, McCrone JT, Woodman A, Arnold JJ, Yennawar M, Evans R, Cameron CE, Lauring AS. 2018. A speed–fidelity trade-off determines the mutation rate and virulence of an RNA virus. PLOS Biology 16:e2006459.

42. Lauring AS, Andino R. 2010. Quasispecies Theory and the Behavior of RNA Viruses. PLoS Pathogens 6:e1001005.

43. Lee CH, Gilbertson DL, Novella IS, Huerta R, Domingo E, Holland JJ. 1997. Negative effects of chemical mutagenesis on the adaptive behavior of vesicular stomatitis virus. Journal of virology 71:3636–3640.

44. Crotty S, Maag D, Arnold JJ, Zhong W, Lau JYN, Hong Z, Andino R, Cameron CE. 2000. The broad-spectrum antiviral ribonucleoside ribavirin is an RNA virus mutagen. Nature Medicine 6:1375–1379.

45. Hadj Hassine I, Ben M'hadheb M, Menéndez-Arias L. 2022. Lethal Mutagenesis of RNA Viruses and Approved Drugs with Antiviral Mutagenic Activity. Viruses 14:841.

46. Lloyd SB, Kent SJ, Winnall WR. 2014. The High Cost of Fidelity. AIDS Research and Human Retroviruses 30:8–16.

47. Goldhill DH, te Velthuis AJW, Fletcher RA, Langat P, Zambon M, Lackenby A, Barclay WS. 2018. The mechanism of resistance to favipiravir in influenza. Proceedings of the National Academy of Sciences 115:11613–11618.

48. Lauring AS, Frydman J, Andino R. 2013. The role of mutational robustness in RNA virus evolution. Nature Reviews Microbiology 11:327–336.

49. Visher E, Whitefield SE, McCrone JT, Fitzsimmons W, Lauring AS. 2016. The Mutational Robustness of Influenza A Virus. PLoS Pathogens 12.

50. Plotkin JB, Dushoff J. 2003. Codon bias and frequency-dependent selection on the hemagglutinin epitopes of influenza A virus. Proceedings of the National Academy of Sciences 100:7152–7157.

51. Doud M, Bloom J. 2016. Accurate Measurement of the Effects of All Amino-Acid Mutations on Influenza Hemagglutinin. Viruses 8:155.

52. Acevedo A, Brodsky L, Andino R. 2013. Mutational and fitness landscapes of an RNA virus revealed through population sequencing. Nature 505:686–690.

53. Plotkin JB, Kudla G. 2010. Synonymous but not the same: the causes and consequences of codon bias. Nature Reviews Genetics 12:32–42.

54. Goodfellow I, Chaudhry Y, Richardson AJ, Meredith J, Almond JW, Barclay WS, Evans DJ. 2000. Identification of a cis -Acting Replication Element within the Poliovirus Coding Region. Journal of Virology 74:4590–4600.

55. Lauring Adam S, Acevedo A, Cooper Samantha B, Andino R. 2012. Codon Usage Determines the Mutational Robustness, Evolutionary Capacity, and Virulence of an RNA Virus. Cell Host & Microbe 12:623–632.

56. Burns CC, Shaw J, Campagnoli R, Jaume Jorba, Vincent A, Quay J, Kew OM. 2006. Modulation of Poliovirus Replicative Fitness in HeLa Cells by Deoptimization of Synonymous Codon Usage in the Capsid Region. Journal of Virology 80:3259–3272.

57. Agashe D, Martinez-Gomez NC, Drummond DA, Marx CJ. 2012. Good Codons, Bad Transcript: Large Reductions in Gene Expression and Fitness Arising from Synonymous Mutations in a Key Enzyme. Molecular Biology and Evolution 30:549–560.

58. Lyons D. 2022. Finding General Patterns in Fitness Landscapes. deepbluelibumichedu. https://deepblue.lib.umich.edu/handle/2027.42/172512. Retrieved 10 September 2023.

59. Atsushi Yoshiki, Moriwaki K. 2006. Mouse Phenome Research: Implications of Genetic Background. ILAR Journal 47:94–102.

60. McCrone JT, Woods RJ, Martin ET, Malosh RE, Monto AS, Lauring AS. 2018. Stochastic processes constrain the within and between host evolution of influenza virus. eLife 7:e35962.

61. Xue KS, Moncla LH, Bedford T, Bloom JD. 2018. Within-Host Evolution of Human Influenza Virus. Trends in Microbiology 26:781–793.

62. Nelson MI, Simonsen L, Viboud C, Miller MA, Taylor J, George KSt, Griesemer SB, Ghedin E, Sengamalay NA, Spiro DJ, Volkov I, Grenfell BT, Lipman DJ, Taubenberger JK, Holmes EC. 2006. Stochastic Processes Are Key Determinants of Short-Term Evolution in Influenza A Virus. PLoS Pathogens 2:e125.

63. Debbink K, McCrone JT, Petrie JG, Truscon R, Johnson E, Mantlo EK, Monto AS, Lauring AS. 2017. Vaccination has minimal impact on the intrahost diversity of H3N2 influenza viruses. PLOS Pathogens 13:e1006194.

64. Gatherer D. 2010. Passage in egg culture is a major cause of apparent positive selection in influenza B hemagglutinin. Journal of Medical Virology 82:123–127.

65. Hamada N, Hara K, Kashiwagi T, Imamura Y, Nakazono Y, Watanabe H, Imamura Y, Chijiwa K. 2012. Intrahost emergent dynamics of oseltamivir-resistant virus of pandemic influenza A (H1N1) 2009 in a fatally immunocompromised patient. Journal of Infection and Chemotherapy 18:865–871.

66. Franzosa EA, Xia Y. 2009. Structural Determinants of Protein Evolution Are Context-Sensitive at the Residue Level. Molecular Biology and Evolution 26:2387–2395.

67. Ramsey DC, Scherrer MP, Zhou T, Wilke CO. 2011. The Relationship Between Relative Solvent Accessibility and Evolutionary Rate in Protein Evolution. Genetics 188:479–488.

68. Tokuriki N, Stricher F, Schymkowitz J, Serrano L, Tawfik DS. 2007. The Stability Effects of Protein Mutations Appear to be Universally Distributed. Journal of Molecular Biology 369:1318–1332.

69. Marsh JA, Teichmann SA. 2013. Parallel dynamics and evolution: Protein conformational fluctuations and assembly reflect evolutionary changes in sequence and structure. BioEssays 36:209–218.

70. Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW. 2002. Evolutionary Rate in the Protein Interaction Network. Science 296:750–752.

71. Jordan IK, Wolf YI, Koonin EV. 2003. No simple dependence between protein evolution rate and the number of protein-protein interactions: only the most prolific interactors tend to evolve slowly. BMC Evolutionary Biology 3:1.

72. Jones S, Thornton JM. 1996. Principles of protein-protein interactions. Proceedings of the National Academy of Sciences 93:13–20.

73. Nooren IMA, Thornton JM. 2003. Structural Characterisation and Functional Significance of Transient Protein–Protein Interactions. Journal of Molecular Biology 325:991–1018.

74. Mintseris J, Weng Z. 2005. Structure, function, and evolution of transient and obligate protein-protein interactions. Proceedings of the National Academy of Sciences 102:10930–10935.

75. Worth CL, Gong S, Blundell TL. 2009. Structural and functional constraints in the evolution of protein families. Nature Reviews Molecular Cell Biology 10:709–720.

76. Chou Y, Vafabakhsh R, Doğanay S, Gao Q, Ha T, Palese P. 2012. One influenza virus particle packages eight unique viral RNAs as shown by FISH analysis. Proceedings of the National Academy of Sciences 109:9101–9106.

77. He X, Zhou J, Bartlam M, Zhang R, Ma J, Lou Z, Li X, Li J, Joachimiak A, Zeng Z, Ge R, Rao Z, Liu Y. 2008. Crystal structure of the polymerase PAC–PB1N complex from an avian influenza H5N1 virus. Nature 454:1123–1126.

78. Sugiyama K, Obayashi E, Kawaguchi A, Suzuki Y, Tame JRH, Nagata K, Park S-Y. 2009. Structural insight into the essential PB1–PB2 subunit contact of the influenza virus RNA polymerase. The EMBO Journal 28:1803–1811.

79. Chu C, Fan S, Li C, Macken C, Kim JH, Hatta M, Neumann G, Kawaoka Y. 2012. Functional Analysis of Conserved Motifs in Influenza Virus PB1 Protein. PLoS ONE 7:e36113.

80. te Velthuis AJW. 2014. Common and unique features of viral RNA-dependent polymerases. Cellular and Molecular Life Sciences 71:4403–4420.

81. te JW, Robb NC, Kapanidis AN, Fodor E. 2016. The role of the priming loop in influenza A virus RNA synthesis. Nature microbiology 1.

82. Pflug A, Guilligay D, Reich S, Cusack S. 2014. Structure of influenza A polymerase bound to the viral RNA promoter. Nature 516:355–360.

83. Peramachi Palanivelu. 2023. An insight into the active sites of the catalytic basic protein subunit PB1 of the RNA polymerase of human influenza viruses. World Journal of Advanced Research and Reviews 17:625–656.

84. S Jane Flint, Enquist LW, Racaniello VR, Rall GF, Skalka AM. 2015. Principles of virology. Volume 1, Molecular biology. Asm Press, Washington, D.C.

85. Reich S, Guilligay D, Pflug A, Malet H, Berger I, Crépin T, Hart D, Lunardi T, Nanao M, Ruigrok RWH, Cusack S. 2014. Structural insight into cap-snatching and RNA synthesis by influenza polymerase. Nature 516:361–366.

86. Eisfeld AJ, Neumann G, Kawaoka Y. 2014. At the centre: influenza A virus ribonucleoproteins. Nature Reviews Microbiology 13:28–41.

87. te Velthuis AJW, Fodor E. 2016. Influenza virus RNA polymerase: insights into the mechanisms of viral RNA synthesis. Nature Reviews Microbiology 14:479–493.

88. York A, Hengrung N, Vreede FT, Huiskonen JT, Fodor E. 2013. Isolation and characterization of the positive-sense replicative intermediate of a negative-strand RNA virus. Proceedings of the National Academy of Sciences of the United States of America 110:E4238-4245.

89. Sugiyama K, Kawaguchi A, Okuwaki M, Nagata K. 2015. pp32 and APRIL are host cell-derived regulators of influenza virus RNA synthesis from cRNA. eLife 4.

90. Staller E, Sheppard CM, Neasham PJ, Mistry B, Peacock TP, Goldhill DH, Long JS, Barclay WS. 2019. ANP32 Proteins Are Essential for Influenza Virus Replication in Human Cells. Journal of Virology 93.

91. Baker SK, Ledwith MP, Mehle A. 2018. Differential Splicing of ANP32A in Birds Alters Its Ability to Stimulate RNA Synthesis by Restricted Influenza Polymerase. Cell Reports 24:2581-2588.e4.

92. Domingues P, Hale BG. 2017. Functional Insights into ANP32A-Dependent Influenza A Virus Polymerase Host Restriction. Cell Reports 20:2538–2546.

93. Zhang H, Zhang Z, Wang Y, Wang M, Wang X, Zhang X, Ji S, Du C, Chen H, Wang X. 2019. Fundamental Contribution and Host Range Determination of ANP32A and ANP32B in Influenza A Virus Polymerase Activity. Journal of Virology 93.

94. Long JS, Giotis ES, Moncorgé O, Frise R, Mistry B, James J, Morisson M, Iqbal M, Vignal A, Skinner MA, Barclay WS. 2016. Species difference in ANP32A underlies influenza A virus polymerase host restriction. Nature 529:101–104.

95. Fodor E, te Velthuis AJW. 2019. Structure and Function of the Influenza Virus Transcription and Replication Machinery. Cold Spring Harbor Perspectives in Medicine 10:a038398.

96. Plotch SJ, Bouloy M, Ulmanen I, Krug RM. 1981. A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. Cell 23:847–858.

97. Kouba T, Drncová P, Cusack S. 2019. Structural snapshots of actively transcribing influenza polymerase. Nature Structural & Molecular Biology 26:460–470.

98. Santos LA, Almeida F, Gíria M, Trigueiro-Louro J, Rebelo-de-Andrade H. 2023. Adaptive evolution of PB1 from influenza A(H1N1)pdm09 virus towards an enhanced fitness. Virology 578:1–6.

99. Shih AC-C, Hsiao T-C, Ho M-S, Li W-H. 2007. Simultaneous amino acid substitutions at antigenic sites drive influenza A hemagglutinin evolution. Proceedings of the National Academy of Sciences 104:6283–6288.

100. Bhatt S, Holmes EC, Pybus OG. 2011. The Genomic Rate of Molecular Adaptation of the Human Influenza A Virus. Molecular Biology and Evolution 28:2443–2451.

101. Nogales A, Steel J, Liu W-C, Lowen AC, Rodriguez L, Chiem K, Cox A, García-Sastre A, Albrecht RA, Dewhurst S, Martínez-Sobrido L. 2022. Mutation L319Q in the PB1 Polymerase Subunit Improves Attenuation of a Candidate Live-Attenuated Influenza A Virus Vaccine. Microbiology Spectrum 10.

102. Zhang X, Li Y, Jin S, Zhang Y, Sun L, Hu X, Zhao M, Li F, Wang T, Sun W, Feng N, Wang H, He H, Zhao Y, Yang S, Xia X, Gao Y. 2021. PB1 S524G mutation of wild bird-origin H3N8 influenza A virus enhances virulence and fitness for transmission in mammals. Emerging Microbes & Infections 10:1038–1051.

103. Feng X, Wang Z, Shi J, Deng G, Kong H, Tao S, Li C, Liu L, Guan Y, Chen H. 2016. Glycine at Position 622 in PB1 Contributes to the Virulence of H5N1 Avian Influenza Virus in Mice. Journal of Virology 90:1872–1879.

104. Xu C, Hu W-B, Xu K, He Y-X, Wang T-Y, Chen Z, Li T-X, Liu J-H, Buchy P, Sun B. 2012. Amino acids 473V and 598P of PB1 from an avian-origin influenza A virus contribute to polymerase activity, especially in mammalian cells. Journal of General Virology 93:531–540.

105. Li J, Liang L, Jiang L, Wang Q, Wen X, Zhao Y, Cui P, Zhang Y, Wang G, Li Q, Deng G, Shi J, Tian G, Zeng X, Jiang Y, Liu L, Chen H, Li C. 2021. Viral RNA-binding ability conferred by SUMOylation at PB1 K612 of influenza A virus is essential for viral pathogenesis and transmission. PLOS Pathogens 17:e1009336.

106. Lin R-W, Chen G-W, Sung H-H, Lin R-J, Yen L-C, Tseng Y-L, Chang Y-K, Lien S-P, Shih S-R, Liao C-L. 2019. Naturally occurring mutations in PB1 affect influenza A virus replication fidelity, virulence, and adaptability. Journal of Biomedical Science 26:55.

107. Chen W, Calvo PA, Malide D, Gibbs J, Schubert U, Bacik I, Basta S, O'Neill R, Schickli J, Palese P, Henklein P, Bennink JR, Yewdell JW. 2001. A novel influenza A virus mitochondrial protein that induces cell death. Nature Medicine 7:1306–1312.

108. Varga ZT, Ramos I, Hai R, Schmolke M, García-Sastre A, Fernandez-Sesma A, Palese P. 2011. The Influenza Virus Protein PB1-F2 Inhibits the Induction of Type I Interferon at the Level of the MAVS Adaptor Protein. PLoS Pathogens 7:e1002067.

109. Conenello GM, Zamarin D, Perrone LA, Tumpey T, Palese P. 2007. A Single Mutation in the PB1-F2 of H5N1 (HK/97) and 1918 Influenza A Viruses Contributes to Increased Virulence. PLoS Pathogens 3:e141.

110. Hai R, Schmolke M, Varga ZT, Manicassamy B, Wang TT, Belser JA, Pearce MB, García-SastreA, Tumpey TM, Palese P. 2010. PB1-F2 Expression by the 2009 Pandemic H1N1

Influenza Virus Has Minimal Impact on Virulence in Animal Models. Journal of Virology 84:4442–4450.

111. Zhu H. 2001. Global Analysis of Protein Activities Using Proteome Chips. Science 293:2101–2105.

112. McCullum EO, Williams BAR, Zhang J, Chaput JC. 2010. Random Mutagenesis by Error-Prone PCR. Methods in Molecular Biology 634:103–109.

113. McCullum EO, Williams BAR, Zhang J, Chaput JC. 2010. Random Mutagenesis by Error-Prone PCR. Methods in Molecular Biology 634:103–109.

114. Fujii R, Kitaoka M, Hayashi K. 2004. One-step random mutagenesis by error-prone rolling circle amplification. Nucleic Acids Research 32:e145–e145.

115. Araya CL, Fowler DM. 2011. Deep mutational scanning: assessing protein function on a massive scale. Trends in Biotechnology 29:435–442.

116. Dingens AS, Haddox HK, Overbaugh J, Bloom JD. 2017. Comprehensive Mapping of HIV-1 Escape from a Broadly Neutralizing Antibody. Cell Host & Microbe 21:777-787.e4.

117. Haddox HK, Dingens AS, Bloom JD. 2016. Experimental Estimation of the Effects of All Amino-Acid Mutations to HIV's Envelope Protein on Viral Replication in Cell Culture. PLOS Pathogens 12:e1006114.

118. Sourisseau M, Lawrence DA, Schwarz MC, Storrs C, Veit EC, Bloom JD, Evans M. 2019. Deep Mutational Scanning Comprehensively Maps How Zika Envelope Protein Mutations Affect Viral Growth and Antibody Escape. Journal of Virology 93.

119. Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KHD, Dingens AS, Navarro MJ, Bowen JE, Tortorici MA, Walls AC, King NP, Veesler D, Bloom JD. 2020. Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. Cell 182:1295-1310.e20.

120. Starr TN, Greaney AJ, Dingens AS, Bloom JD. 2021. Complete map of SARS-CoV-2 RBD mutations that escape the monoclonal antibody LY-CoV555 and its cocktail with LY-CoV016. Cell Reports Medicine 2:100255.

121. Klesmith JR, Bacik J-P, Wrenbeck EE, Michalczyk R, Whitehead TA. 2017. Trade-offs between enzyme fitness and solubility illuminated by deep mutational scanning. Proceedings of the National Academy of Sciences 114:2265–2270.

122. Romero PA, Tran TM, Abate AR. 2015. Dissecting enzyme function with microfluidic-based deep mutational scanning. Proceedings of the National Academy of Sciences 112:7159–7164.

123. Melamed D, Young DL, Gamble CE, Miller CR, Fields S. 2013. Deep mutational scanning of an RRM domain of the Saccharomyces cerevisiae poly(A)-binding protein. RNA 19:1537–1551.

124. Newberry RW, Leong JT, Chow ED, Kampmann M, DeGrado WF. 2020. Deep mutational scanning reveals the structural basis for α-synuclein activity. Nature Chemical Biology 16:653–659.

125. Thyagarajan B, Bloom JD. 2014. The inherent mutational tolerance and antigenic evolvability of influenza hemagglutinin. eLife 3.

126. Lee JM, Huddleston J, Doud MB, Hooper KA, Wu NC, Bedford T, Bloom JD. 2018. Deep mutational scanning of hemagglutinin helps predict evolutionary fates of human H3N2 influenza variants. Proceedings of the National Academy of Sciences 115.

127. Lei R, Milena A, Tan TC, Qi Wen Teo, Wang Y-Q, Zhang X, Luo S, Nair SK, Peng J, Wu NC. 2023. Mutational fitness landscape of human influenza H3N2 neuraminidase. Cell Reports 42:111951–111951.

128. Wang S, Tian Hao Zhang, Hu M, Tang K, Sheng L, Hong M, Chen D, Chen L, Shi Y, Feng J, Qian J, Sun L, Ding K, Sun R, Du Y. 2023. Deep mutational scanning of influenza A virus neuraminidase facilitates the identification of drug resistance mutations in vivo. MSystems https://doi.org/10.1128/msystems.00670-23.

129. Hom N, Gentles L, Bloom JD, Lee KK. 2019. Deep Mutational Scan of the Highly Conserved Influenza A Virus M1 Matrix Protein Reveals Substantial Intrinsic Mutational Tolerance. Journal of Virology 93:e00161-19.

130. Orr Ashenberg, Jai P, Doud MB, Bloom JD. 2017. Deep mutational scanning identifies sites in influenza nucleoprotein that affect viral inhibition by MxA. PLOS Pathogens 13:e1006288–e1006288.

131. Soh YS, Moncla LH, Eguia R, Bedford T, Bloom JD. 2019. Comprehensive mapping of adaptation of the avian influenza polymerase protein PB2 to humans. eLife 8:e45079.

132. Jain P, Varadarajan R. 2014. A rapid, efficient, and economical inverse polymerase chain reaction-based method for generating a site saturation mutant library. Analytical Biochemistry 449:90–98.

133. Hoffmann E, Neumann G, Kawaoka Y, Hobom G, Webster RG. 2000. A DNA transfection system for generation of influenza A virus from eight plasmids. Proceedings of the National Academy of Sciences of the United States of America 97:6108–6113.

134. Metzker ML. 2009. Sequencing technologies — the next generation. Nature Reviews Genetics 11:31–46.

135. Quail MA, Kozarewa I, Smith F, Scally A, Stephens PJ, Durbin R, Swerdlow H, Turner DJ. 2008. A large genome center's improvements to the Illumina sequencing system. Nature Methods 5:1005–1010.

136. McCrone J, Lauring AS. 2016. Measurements of Intrahost Viral Diversity Are Extremely Sensitive to Systematic Errors in Variant Calling. Journal of Virology 90:6884–6895.

137. Hiatt JB, Patwardhan RP, Turner EH, Lee C, Shendure J. 2010. Parallel, tag-directed assembly of locally derived short sequence reads. Nature Methods 7:119–122.

138. Kinde I, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B. 2011. Detection and quantification of rare mutations with massively parallel sequencing. Proceedings of the National Academy of Sciences 108:9530–9535.

139. Knierim E, Lucke B, Schwarz JM, Schuelke M, Seelow D. 2011. Systematic Comparison of Three Methods for Fragmentation of Long-Range PCR Products for Next Generation Sequencing. PLoS ONE 6:e28240.

Chapter 2 Deep Mutational Scanning Reveals the Functional Constraints and Evolutionary Potential of the Influenza A Virus PB1 Protein

Note: This chapter is a modified version of the published article:

Li Y, Arcos S, Sabsay KR, te Velthuis AJW, Lauring AS. 2023. Deep mutational scanning reveals the functional constraints and evolutionary potential of the influenza A virus PB1 protein. Journal of Virology https://doi.org/10.1128/jvi.01329-23.

2.1 Abstract

The influenza virus polymerase is central to influenza virus evolution. Adaptive mutations within the polymerase are often a prerequisite for efficient spread of novel animal-derived viruses in human populations. The polymerase also determines fidelity and, therefore, the rate at which the virus will acquire mutations that lead to host range expansion, drug resistance, or antigenic drift. Despite its importance to viral replication and evolution, our understanding of the mutational effects and associated constraints on the influenza RNA-dependent RNA polymerase (RdRp) is relatively limited. We performed deep mutational scanning of the A/WSN/1933 (H1N1) polymerase basic 1 (PB1), generating a library of 95.4% of amino acid substitutions at 757 sites. After accuracy filters, we were able to measure replicative fitness for 13,354 (84%) of all possible amino acid substitutions, and 13 were validated by results from pairwise competition assays. Functional and structural constraints were better revealed by individual sites involved in RNA or protein interactions than by major

subdomains defined by sequence conservation. Mutational tolerance, as defined by site entropy, was correlated with evolutionary potential, as captured by diversity in the available H1N1 sequences. Of the 29 beneficial sites, many have either been identified in the natural evolution of PB1 or shown experimentally to have important impacts on replication and adaptation. Accessibility of amino acid substitutions by single nucleotide mutation was a key factor in determining whether mutations appeared in natural PB1 evolution. Our work provides a comprehensive map of mutational effects on a viral RdRp and a valuable resource for subsequent studies of influenza replication and evolution.

2.2 Introduction

Viral RNA-dependent RNA polymerases (RdRp) are central to RNA virus replication and evolution. The RdRp replicates the genome and is a key determinant for replicative fitness and viral mutation rates. For negative-strand RNA viruses, the RdRp is also responsible for transcription, thereby regulating protein expression. The RdRp has been directly linked to virulence (1). Mutations within the RdRp influence host adaptation (2-4), replication fidelity (5-8), post-translational modifications (9), and host immune responses (10, 11).

The evolution of viral RdRp is functionally and structurally constrained. Functional constraints include requirements for interactions with RNAs and other proteins, adaptation to new replication environments (12), the deleterious impact of low fidelity (5), and viral codon abundance (13-15). Residues that are involved in obligatory interactions tend to be less tolerant to mutation and evolve at a slower rate (16-19). The primary structural constraints are solvent accessibility (20), maintenance of molecular flexibility (21-23), intermolecular interactions (24, 25), and key protein secondary structures (26). For example, the establishment of secondary structures requires certain biochemical characteristics conferred by a limited number of amino

acids (27), and mutations in buried residues often have a bigger fitness effect, as their change will impact nearby residues (27, 28).

The influenza virus RdRp is a heterotrimer that consists of three subunits: polymerase basic 1 (PB1), polymerase basic 2 (PB2), and polymerase acidic (PA), in which PB1 functions as the catalytic subunit. The PB1 subunit may have additional functional and structural constraints, because it cooperates with the two other polymerase subunits and viral nucleoproteins (NP) in transcription and genome replication. During transcription, PB1 guides the capped primer cleaved from a host pre-mRNA by PB2 and PA into the polymerase active site and stabilizes it on the 3′ end of the viral RNA (vRNA) template in the active site (29). The PB1 RdRp then extends the capped primer through the incorporation of nucleoside triphosphates, separates the template-product duplex downstream of the active site, and extrudes the viral mRNA through the product exit channel and the copied template through the template exit channel (29). The interactions among PB1, PB2, and PA shift at every stage of transcription (30). During replication, a vRNA is copied into a complementary RNA (cRNA). Next, the cRNA product serves as the template for negative-strand vRNA synthesis. The process of vRNA and cRNA synthesis not only requires the coordination of polymerase subunits but also interactions with an encapsidating RdRp and host protein ANP32 to form an RdRp dimer, a trans-activating RdRp to induce correct replication initiation, conformational changes to transfer the nascent vRNA or cRNA to the additional RdRp, and recruitment of viral nucleoprotein to encapsidate the nascent vRNA and cRNA molecules (31).

Given the importance of PB1 to influenza virus replication and evolution, defining the fitness effects of amino acid substitutions can elucidate the relevant functional and structural constraints. Deep mutational scanning (DMS) – saturation mutagenesis combined with deep

sequencing – is a massively parallel approach that has recently been used to explore the fitness landscapes of viral proteins (19, 32-35). Here, we applied deep mutational scanning to the influenza virus A/WSN/1933 (H1N1) (abbreviated WSN33) PB1 RdRp subunit, identifying constrained regions of the protein and relating beneficial mutations to those observed in natural evolution. Overall, our study provides a comprehensive resource for studies of influenza virus replication and evolution.

2.3 Materials and Methods

Cell lines and media

MDCK-SIAT1-TMPRSS2 and HEK293T-CMV-PB1 were provided by Dr. Jesse Bloom (Fred Hutchinson Cancer Research Center) and maintained in D10 media [Dulbecco's modified Eagle medium (DMEM), Invitrogen, 11995-065], with 10% heat-inactivated fetal bovine serum (FBS, Gibco, 26140-079), 1% L-Glutamine (100×, Gibco, 25030-081), and 1% Pen+Strep (10,000 U/mL P, 10,000 µg/mL S, Invitrogen, 15140-122). A549 cells were maintained in A549 growth media (DMEM, high glucose, with L-glutamine, without Na pyruvate (Invitrogen 11965-092), with 10% FBS, 1% Pen+Strep, 0.1875% bovine albumin fraction V (7.5%, Invitrogen 15260-037), and 2.5% HEPES (1M, Invitrogen, 15630-080). We used IGM+ media [Opti-MEM1 Reduced Serum Media (Gibco, 31985-070), with 0.5% heat-inactivated FBS, 1% Pen+Strep, 0.3% bovine albumin fraction V, and 500 µL of 100 mg/mL $CaCl_2$] for 24 hours following transfection. We used WNM media (Medium 199, Gibco, 11043-023, no phenol red, 0.5% heat-inactivated FBS, 1% Pen+Strep, 0.3% bovine albumin fraction V, 2.5% HEPES, and 500 µL of 100 mg/mL CaCl2) for TCID50 assays. We used A549 growth media for seeding cells for viral passages and A549 viral media [DMEM, high glucose, with L-glutamine, without Na

pyruvate (Invitrogen 11965-092), with 1% Pen+Strep, 0.1875% bovine albumin fraction V, 2.5% HEPES, and TPCK-trypsin at a final concentration of 4 μg/mL] for virus infections.

Construction of PB1 codon mutant plasmid libraries

PB1 codon mutant libraries were generated using an overlapping PCR strategy described in reference (32) with (36) as a reference. We used the code in reference (37) first described in reference (38) with the modifications from reference (39) to generate tiled primers for mutagenesis and a code from reference (40) to determine how library diversity would be impacted by restriction enzymes used in cloning. We performed 10 cycles of fragment PCR (round one) with 1.2 μg of plasmid (pHW2000) containing the wild-type (WT) PB1 sequence from WSN33 and 20 cycles of joining PCR (round two). The lengths of PCR products were checked by gel electrophoresis. In a pilot experiment in which we generated PB1 variants for 96 out of the 758 sites, we randomly picked PCR products from 24 clones for Sanger sequencing to evaluate the library mutation rate. Twenty out of 24 clones had only a single codon mutation at the target site, and four clones were wild type.

We pooled an equal volume from the 758 PCR reactions into 16 pools. Each pool was digested by restriction enzyme AarI and the 16 pools combined into one variant insert pool. We used T4 DNA ligase (NEB, #M0202L) to ligate the variant insert pool into BsmBI-digested pHW2000 plasmid and transformed Stellar Competent Cells (Takara, #636763) according to the manufacturer's instructions. We independently performed the ligation and transformation three times to create three libraries. We plated the transformed cells onto Nunc Square BioAssay Dishes (Thermo Scientific, #240845) and obtained 82,800-118,800 colonies for each library

replicate. Plasmid DNA was extracted directly from the pooled colonies using a QIAGEN Plasmid Maxi Kit (QIAGEN, #12162).

Transfection

We generated variant virus libraries by transfecting HEK293T-CMV-PB1 cells, which constitutively express the wild-type PB1 protein from WSN33. For each variant plasmid library, we seeded 36 wells of 6-well plates with 5×10^5 MDCK-SIAT1-TMPRSS2 cells and 5×10^5 HEK293T-CMV-PB1 cells. Seventeen hours later, we transfected each well with 1 µg in each of the seven plasmids containing the seven wild-type WSN33 genome segments and 1 µg of the PB1 variant library using TransIT-LT1 Transfection Reagent (MIR 2300). We used the same procedure to make the wild-type WSN33 viruses as control, only on a smaller scale (six wells) and using the wild-type WSN33 PB1 in place of the variant plasmid library. At 24 hours post-transfection, we replaced the transfection media with fresh IGM+ and then incubated for an additional 24 hours. At 48 hours post-transfection, we harvested viral supernatants by centrifuging at 200 × g for 5 minutes. Three virus variant libraries and the wild-type virus control were aliquoted and snap frozen in 0.5% glycerol prior to storage at -80°C.

Determination of virus titer

Viruses were titered by median Tissue Culture Infective Dose ($TCID_{50}$) on MDCK-SIAT1-TMPRSS2 cells. For each assay, we seeded 6×10^3 MDCK-SIAT1-TMPRSS2 cells in 100 µL of WNM media in each well of a 96-well plate. Seventeen hours later, we serially diluted the virus samples 1:10 with WNM media supplemented with 4 µg/mL TPCK-trypsin reconstituted in phosphate buffered saline (PBS) to 1 mg/mL for working stock and added 100 µL virus per well.

We incubated the plates at 37°C and monitored them daily for cytopathic effect (CPE) for up to 4 days.

Viral passages

Each passage had 1×10^6 infectious viral particles on 1×10^8 A549 cells to achieve an approximate multiplicity of infection (MOI) of 0.01 $TCID_{50}$/cell. We seeded 8×10^7 A549 cells in a total of 60 mL A549 growth media in three T182 flasks. Seventeen hours later, we suspended 1×10^6 $TCID_{50}$ of virus in 45 mL of A549 viral media with 4 µg/mL freshly added TPCK-trypsin. We aspirated the overnight A549 growth media, rinsed the cells gently with pre-warmed PBS, and added 15 mL of viral dilution to each flask. Three hours after infection, we removed the inoculum, rinsed the cells again with pre-warmed PBS, and replaced the inoculum with 20 mL of fresh A549 viral media per flask with 4 µg/mL TPCK-trypsin. We harvested viral supernatants by centrifugation at 400 × g for 4 minutes, 48 hours after infection, and snap-froze the supernatant in 0.5% glycerol prior to storage at -80°C.

Barcoded subamplicon sequencing

Passaged viruses were concentrated by ultracentrifugation at 27,000 rpm, using Thermo Scientific Sorvall WX Ultra Series Centrifuge with rotor Sorvall AH-629 (DuPont Instruments), for 2 hours at 4°C using Beckman Coulter Centrifuge Tubes (25 × 89 mm, 344058). We then resuspended the viruses in 500 µL of residual media and extracted viral RNA using a QIAamp Viral RNA Mini Kit (QIAGEN, 52906). To accurately measure mutation frequencies, we used a barcoded-subamplicon sequencing strategy described in reference (41) that adds unique sequence barcodes to every DNA molecule in a sample, as follows.

We reverse transcribed the extracted RNA using SuperScript III First-Strand Synthesis System (Invitrogen, 18080-051) and performed PCR to amplify the entire PB1 open reading frame (PCR0). For plasmid samples, we used 2 µL of plasmid DNA at 10 ng/µL as template in PCR0. We cleaned up the PCR0 products using GeneJet PCR clean up kit (GeneJet, K0702) and gel isolated the bands corresponding to full PB1 genome length (~2,341 bp).

Next, we PCR amplified the PB1 gene in eight subamplicons (PCR1). The subamplicons were designed to start and end in full codons, and each subamplicon starts precisely after the previous subamplicon ends. In this way, the nucleotides in one codon in a PB1 DNA molecule will only be calculated once. Forward and reverse primers for PCR1 contained random 8N barcodes at their 5′ termini to uniquely label every cDNA molecule in the template. Theoretically, there would be $4^{16} = 4.29\times10^{9}$ unique barcodes. The template input for PCR1 was limited to ~8×10^{7} molecules such that each was uniquely barcoded. Illumina compatible, sample-specific adapters were added in a subsequent PCR reaction, PCR2. Eight subamplicons for each sample were pooled together, and we used ~1×10^{6} uniquely barcoded molecules from PCR1 as template and unique dual (UD) indexed primers to diminish the issue of index hopping. Finally, we gel-isolated the PCR2 products before sequencing on an Illumina NextSeq 1000, P2 600 cycle (2 × 300 PE), with 20% PhiX. We conducted two sequencing runs with 60 µL of the combined PCR2 products at 5 nM, 30 µL for each run, and merged the reads for analysis. We used KOD Hot Start Master Mix (EMD Millipore, 71842) to perform all PCRs. Primers and cycling programs can be found in Appendix A.

Analysis of deep sequencing data

Sequence files were analyzed using dms_tools2 (42), which groups the paired-end reads with the same PCR1 barcodes. Sequences were discarded if the Q-score of any nucleotide in the barcode was <15. Consensus sequences were generated for barcodes with at least two reads and aligned to the reference genome to record the codon at each site for that molecule. Because mutations are defined at a subamplicon level, it is possible that rare secondary mutations on the same PB1 haplotype could be present on distinct subamplicons.

We calculated the fitness of each mutation based on the enrichment ratio method described in reference (35) with modifications. We calculated the frequency of mutation *i* at site *s* as:

$$frequency_{i,s} = \frac{read\ count_{i,s} + pseudo\ count}{\sum_{k \in s} read\ count_{i,k} + pseudo\ count}$$

where the pseudo count was added to ensure a non-zero denominator and fixed as 1 by default. To offset frequency inflation by the pseudo count, we discarded a mutation if its read count in the variant plasmid library was less than 10. We then discarded the mutations whose frequency in the variant plasmid library was not at least sixfold higher than that in the wild-type plasmid library. With these filters, we calculated the enrichment ratio as:

$$enrichment\ ratio_{i,s} = \frac{frequency_{post-passage_{i,s}}}{frequency_{pre-passage_{i,s}}}$$

and we defined "fitness" as $log_{10}(enrichment\ ratio)$ normalized by the average fitness of silent mutations in the corresponding subamplicon in each individual library:

$$fitness_{i,s} = log_{10}(enrichment\ ratio_{i,s}) - \frac{\sum_{k \in silent,amp} log_{10}(enrichment\ ratio_{k,s})}{N_{silent,amp}}$$

The fitness of a mutation at a certain site used for subsequent analyses was the average fitness of that in all replicates where it was available.

Analysis of naturally occurring influenza sequences

We downloaded the influenza sequences from Global Initiative on Sharing All Influenza Data (GISAID) from 1918 to 2023, with the filtering conditions of "type A," "H1N1," "human host," "required segment PB1," and "complete sequences only." According to CDC's timeline for the 2009 H1N1 pandemic (43), we classified pre-09 strains as all sequences collected before 14 April 2009 and post-09 strains as sequences collected after 12 August 2010. We discarded the sequences collected during the pandemic to avoid the time period when pre- and post-09 strains might be co-circulating. We downloaded the amino acid sequences along with the corresponding metadata and filtered out any sequences that had been passaged in eggs. We aligned the sequences to the wild-type WSN33 amino acid sequence using MAFFT (44). The entropy of a site was measured as the Shannon entropy (45) of all amino acids that appeared at that site:

$$site\ entropy_s = -\sum_{x \in s} p(x) \log p(x)$$

To adjust for uneven sampling over time, we adopted the weighted entropy method described in reference (46). Briefly, we grouped the sequences by collection year, calculated the frequencies of the amino acids in each year, and used the average of amino acid frequencies over all years for the entropy calculation.

Protein structure visualization and analysis

We used UCSF ChimeraX (47) for protein visualizations, including movies. To visualize site entropy on the PB1 protein, we replaced the b-factor column with site entropy data in the PDB files. We identified protein-RNA contacts using LigPlot Plus with default thresholds for the maximum distance between interacting atoms (48). The protein structures used are as follows:

5D9A (apoenzyme), 7NHX (template binding, early), 6T0N (template binding, late), 5M3H (cap-snatching), 6RR7 (pre-initiation), 6QCW (mixed pre-initiation), 6QCV (mixed pre-catalysis), 6QCX (mixed post-incorporation), 6SZV (elongation), and 6SZU (termination).

Measurement of accessible surface area

We measured the Accessible Surface Area (ASA) using PDBePISA ("Protein interfaces, surfaces and assemblies" service PISA at the European Bioinformatics Institute, http://www.ebi.ac.uk/pdbe/prot_int/pistart.html) (49), with the influenza A/Brevig Mission/1/1918(H1N1) polymerase heterotrimer structure (PDB: 7NHX). We chose to perform this and subsequent analyses with 7NHX because this is the only resolved structure for the H1N1 polymerase complex, which may be a closer approximation to WSN33 RdRp. We used a default water probe of 1.4 Å in diameter to roll over the surface of the entire polymerase complex and added up all points in contact with the probe.

Molecular dynamics simulation and measurement of root mean square fluctuation*

We performed a molecular dynamics simulation of A/Brevig Mission/1/1918(H1N1) RdRp (PDB: 7NHX) to measure the relative structural flexibility of the heterotrimer. We removed the RNA molecules in the structure and modeled the missing residues (Chain B, PB1: 187-204 and 645-653) using SWISS-MODEL template-based homology using full sequences from UNIPROT (PA: Q3HM39, PB1: Q3HM40, and PB2: Q3HM41). The global model quality estimate (GMQE) for this homology model is 0.88. Molecular dynamics were simulated using GROMACS on the Princeton University HPC Tiger GPU. The system build parameters used a cubic tip3p water box, charmm27 force field, neutralizing NaCl ions, temperature of 310.15 K

(37°C), and time steps of 0.002 ps. The total system had 2,148 protein residues, 166,370 water residues, and 1,002 ion residues. Energy minimization was performed for a total of 100 ps and converged to a maximum force of less than 1,000kJ/mol in 2,102 steps. Equilibration (both constant number of particles, volume, and temperature (NVT) and constant number of particles, pressure, and temperature (NPT) ensembles) was performed for 200 ps. A 20-ns (10,000,000 steps) production simulation took roughly 19 hours. We analyzed the resulting trajectory for the root mean square fluctuations (RMSF) of the atomic positions at every time point and calculated the RMSF of each residue within the structure from an average of RMSF values of each atom within the residue.

* The molecular dynamics simulation and measurement of root mean square fluctuation was done by Kimberly Sabsay.

2.4 Results

A comprehensive library of single amino acid substitutions in PB1

We used overlap PCR mutagenesis to create a PB1 plasmid library in which every codon in the WSN33 PB1 open reading frame is mutated to code for every other amino acid. We cloned the mutagenized plasmid library three times, independently, to make three replicate plasmid libraries (Figure 2.1). High depth-of-coverage sequencing demonstrated that the three plasmid libraries covered 82%-93% of 24,224 possible codon mutations and 89%-96% of 15,897 possible amino acid substitutions at 757 residues across 758 (stop codon included) sites in PB1 (Table 2.1). After excluding mutations whose frequencies might have been inflated by mutational hotspots during sequencing library preparation, each replicate library covered 64%-70% of all possible amino acid substitutions with 84% of all possible amino acid substitutions present in at least one replicate library.

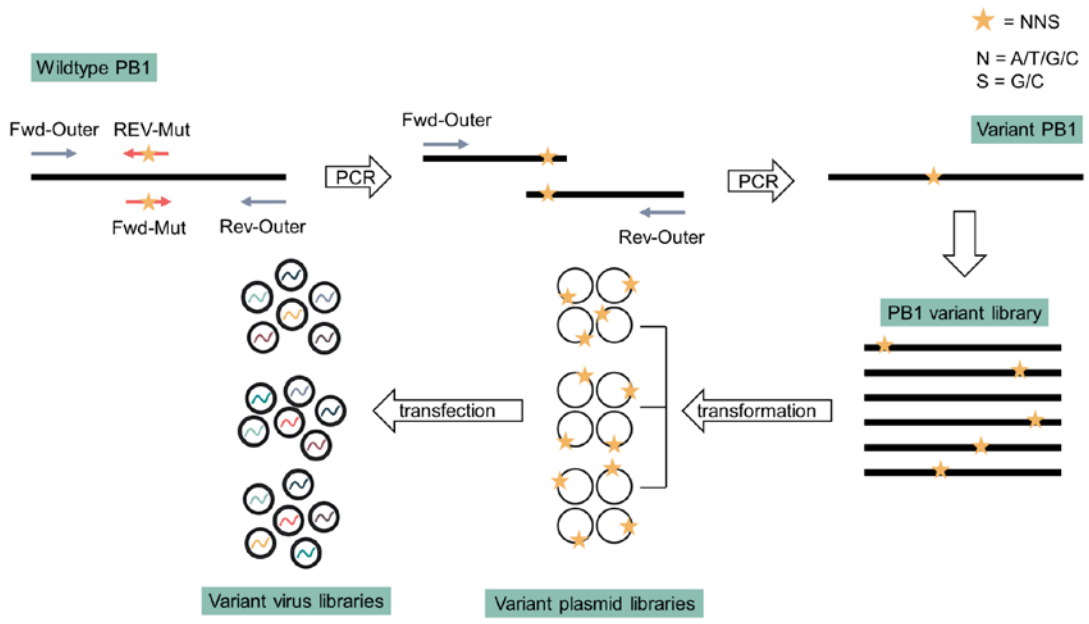


Figure 2.1: Deep mutational scanning of influenza PB1 protein. Scheme of major steps for generating variant virus libraries. We mutagenized wild-type PB1 by overlap PCR, using primers encoding NNS in the codon for the targeted residue. N refers to an equal mixture of A, T, G, and C nucleotides, while S refers to a mixture of only G and C. This coding is able to generate 32 codons, 20 amino acids, and stop codons. The PB1 variant library was ligated and transformed independently three times to make variant plasmid library replicates. Each plasmid library was then transfected independently along with plasmids expressing the other seven influenza segments to make three variant virus library replicates.

Table 2.1 Codon and amino acid variant diversity in plasmid libraries before and after filtering out mutations with low codon counts or under the influence of PCR errors.

	Library	Number of codons	Percentage of codons	Number of amino acids	Percentage of amino acids
Before filtering	Replicate 1	22820	92.6%	15168	95.4%
	Replicate 2	21995	89.2%	14786	93.0%
	Replicate 3	20264	82.2%	14154	89.0%
After filtering	Replicate 1	-[a]	-	11008	69.3%
	Replicate 2	-	-	10565	66.5%
	Replicate 3	-	-	10194	64.1%
	Present in all replicates	-	-	7351	46.2%

	Present in at least one replicate	-	-	13354	84.0%

[a] -, indicates not applicable.

We rescued the corresponding viral variant libraries by transfecting HEK293T cells that stably express PB1 with the plasmid libraries and bidirectional expression plasmids containing the other seven genomic segments from WSN33. The passage 0 (P0) viral stocks exhibited titers of 3.51×10^6 to 5.27×10^7 $TCID_{50}$/mL after 48 hours, slightly lower than those from "wild-type" WSN33 rescues. There were 482 mutations in Replicate 1 (2.1% of total mutations in Replicate 1), 1,371 mutations in Replicate 2 (6.2%), and 175 mutations in Replicate 3 (0.86%) that were present in the plasmid library but not in the P0 viral library, which may indicate lethal mutations. The experimental lethal mutation rate was lower than the expected ~ 25%-30% lethal mutation fraction (50), because the wild-type PB1 protein expressed by the cells partially rescued the variant PB1 proteins with lethal mutations.

We examined the fitness effects of the mutations through serial passage of the variant virus libraries. We passaged the three libraries independently on A549 human lung epithelial carcinoma cells at an MOI of 0.01 for four passages, during which viruses carrying different PB1 substitutions competed against each other. The titers of viruses at each passage decreased slightly to 5×10^6 to 5×10^7 $TCID_{50}$/mL (Figure 2.2A). Forty-three percent of codons and 57% of unique amino acids on average were detected through four passages (Figure 2.2B). We used barcoded-subamplicon sequencing to correct for PCR and sequencing errors and measured the frequencies of individual mutations in each library at passages 1 and 4 (Supplemental Figure 2.1). Throughout passaging, we observed signs of purifying selection, reflected by a relative reduction in the number of non-synonymous mutations (Figure 2.2C) and in codons with two or three nucleotide changes (Figure 2.2D).

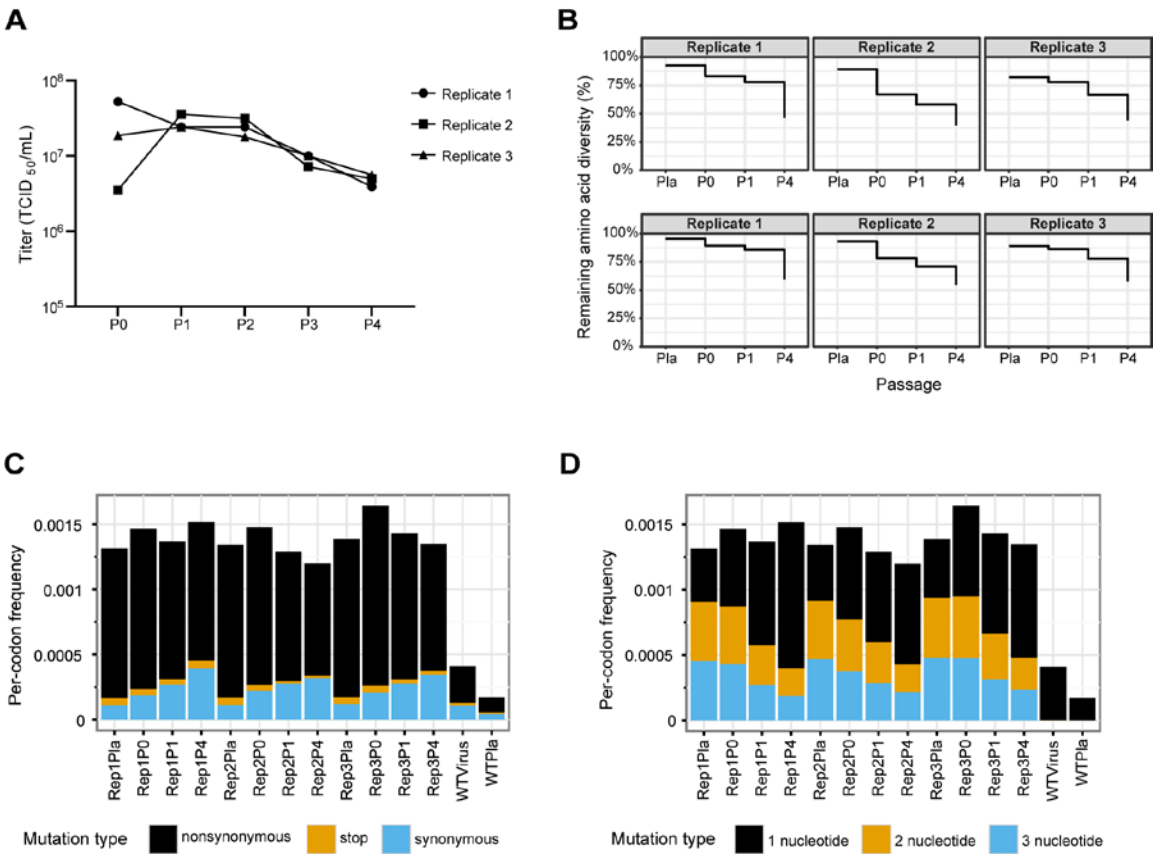
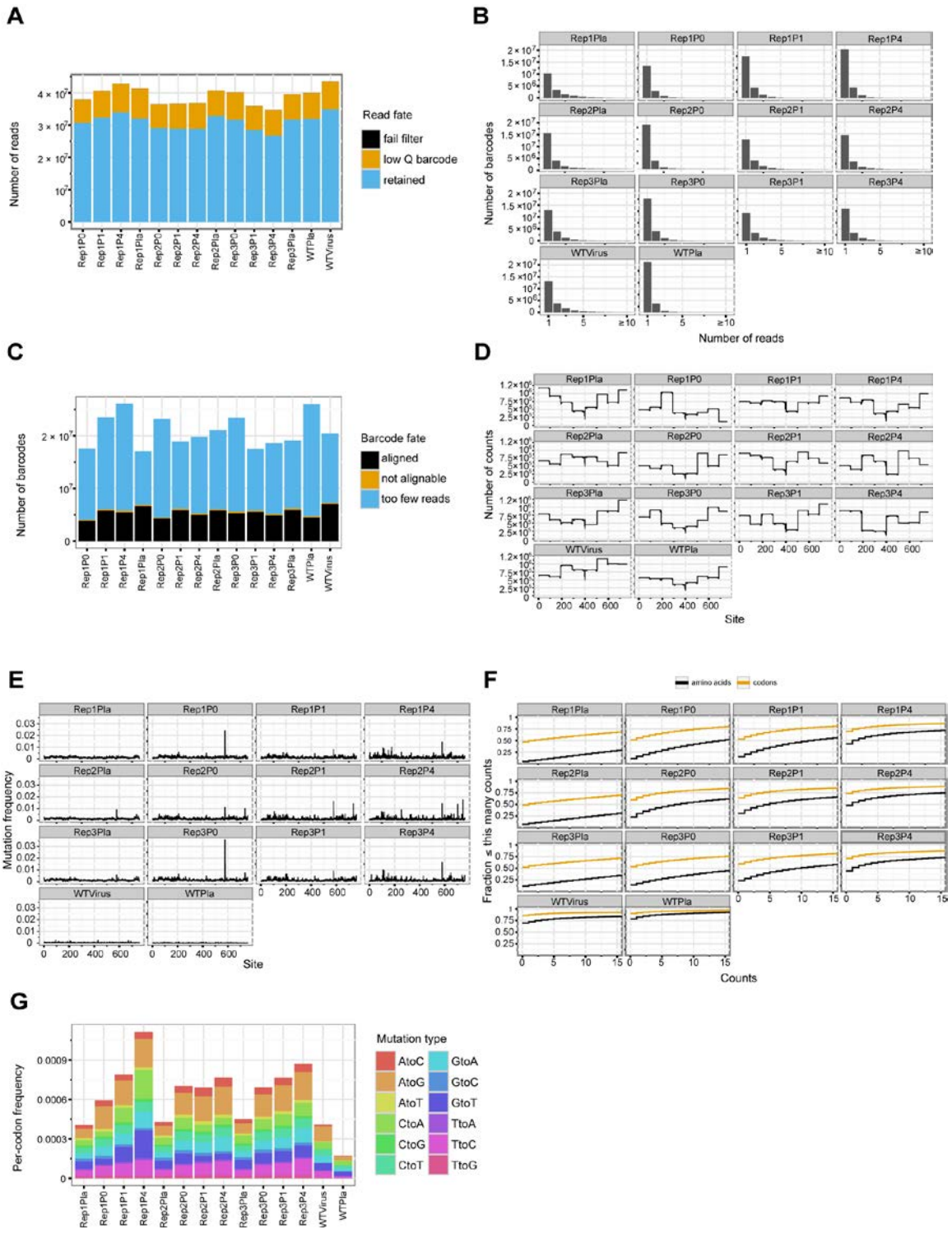


Figure 2.2 Change in codon and amino acid mutations throughout passaging. (**A**) Titers of variant virus libraries before and after each passage. (**B**) Percentage of codon and amino acid variants remaining at each passage. Pla: in plasmid library, before rescue; P0: after rescue, before passaging; P1: after the first passage; P4: after four passages. (**C**) Frequency of synonymous, non-synonymous, and nonsense mutations in replicate (Rep) plasmid libraries, virus libraries after passages, and the wild-type plasmid and virus samples as controls. (**D**) Frequency of codon mutations with 1-, 2-, and 3-nucleotide changes in plasmid libraries, virus libraries after passages, and the wild-type plasmid and virus samples. Frequency in both (**C** and **D**) panels were averaged across the PB1 gene and were prior to filtering and adjustment in fitness calculations, as described in Materials and Methods.



Supplemental Figure 2.1 Full description of deep mutational scanning libraries. (**A**) Raw sequencing reads of each sample. The "filter" in "fail filter" refers to general Illumina filters. "Low Q barcode" refers to sequences having any nucleotide with a Q-score below 15 in the 16× N molecule-specific barcodes. Sequences that failed the filter or with low-Q barcodes were discarded in subsequent analyses. (**B**) The number of distinct barcodes observed in each sample. Each barcode needs to be observed at least twice to determine the consensus sequence for that contig. The bar at 1 corresponds either to barcodes that were only observed once or to sequencing errors that gave rise to new barcodes.

(**C**) Barcodes after aligning to wild type WSN33 PB1 sequence. "Too few reads" corresponds to the bar at 1 in panel (**B**). Sequences categorized as "too few reads" were removed from subsequent analyses. (**D**) Sequencing depth at each site in PB1 after removing contigs with too few reads. The number of counts includes the codon counts for both variant and wild type codons. (**E**) The mutational frequency at each site in PB1 after removing contigs with too few reads. The spike at site 577 in library Rep1P0, Rep1P4, Rep3P0, and Rep3P4 is likely an issue with the sequencing library preparation, as other sequencing runs using the same samples did not show such peaks (data not shown). The spikes have little impact on the type of codons present in the libraries. The impact of peaks on fitness measurements is also minimal since we compared the passaged libraries to the plasmid libraries. (**F**) Mutation sampling completeness. The plot shows the fraction of codon and amino acid mutations observed no more than the indicated number of times. This plot describes both variant diversity and sequencing completeness in a library. (**G**) Frequency of different types of nucleotide change. The plot shows nucleotide change among mutations with only one nucleotide change and works as a check for oxidative damage. An excessive number of C to A or G to T mutations suggests potential oxidative damage. The plot shows no over-representation of either mutation in the libraries.

Replicative fitness of amino acid substitutions in PB1

We quantified the fitness of viral mutants at the amino acid level based on an amino acid's frequency before and after passage. All fitness values were measured after passage four unless stated otherwise (Figure 2.3; for full fitness results, see Supplemental data set of published paper). Here, the fitness of an amino acid at a site is the $\log_{10}$ enrichment ratio normalized by the average fitness of silent mutations in the same amplicon (see Materials and methods). Since > 99% of the codons at any given site in the libraries encoded the wild-type amino acid, the change in the frequency of wild-type variants was negligible, and the measured fitness of the wild type ($\log_{10}$ of 1, or 0) was fixed by the experimental design. As expected, the frequency of most mutations decreased after four passages, indicating that most mutations in the influenza virus RdRp are detrimental (fitness < 0, Figure 2.3 and 2.4A). Nonsense mutations never increased in frequency. Fitness measurements were well correlated across biological replicates with Pearson correlation coefficients between 0.788 and 0.864 (Figure 2.4B).

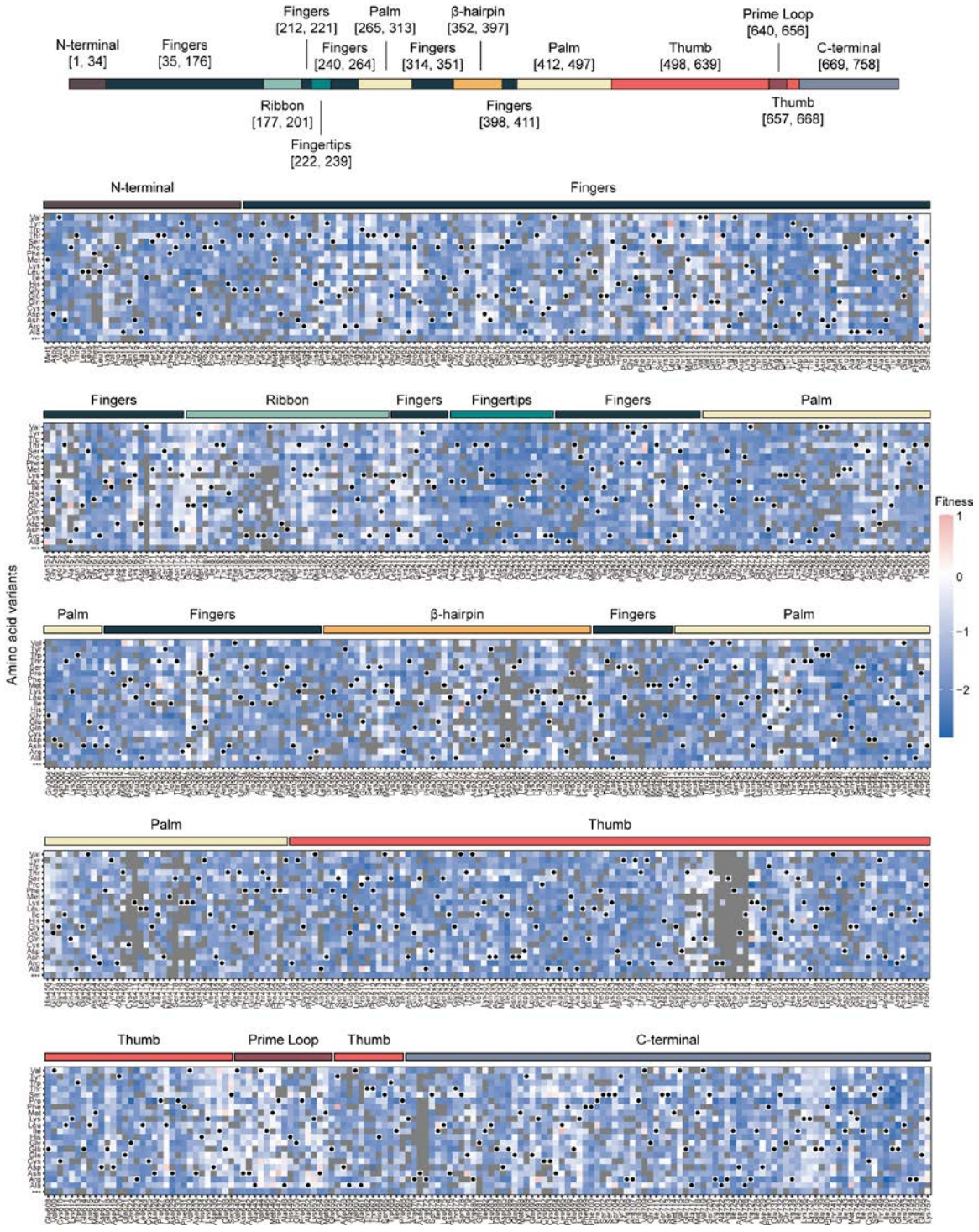


Figure 2.3 Replicative fitness of amino acid substitutions on PB1. The replicative fitness of individual amino acid variants in PB1, with subdomains annotated by the colored bar above the heatmap. Mutations in gray were excluded from the analysis due to low counts in the plasmid library or high occurrence in the wild-type sample, as described in Materials and Methods. Wild-type amino acids are marked by black dots.

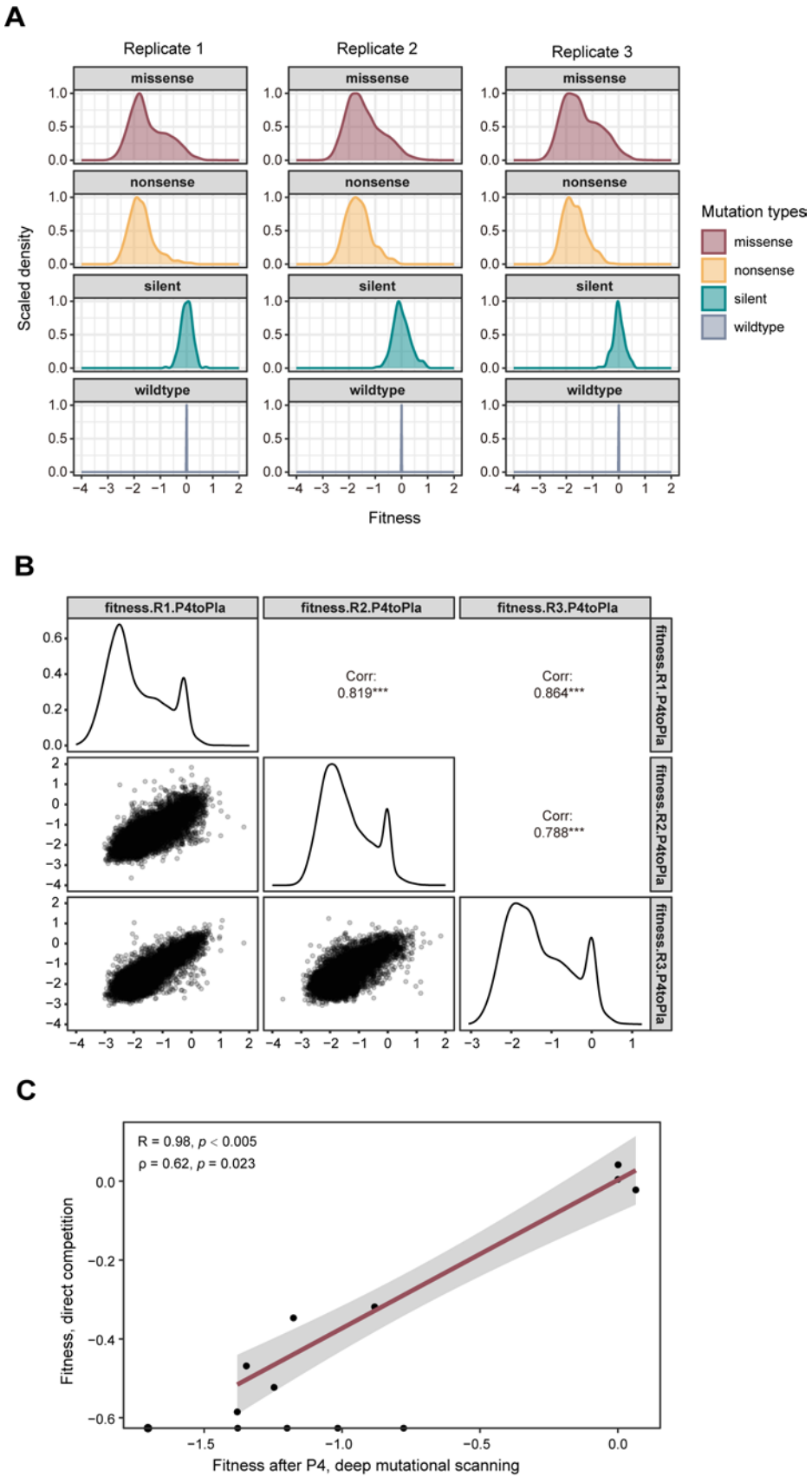
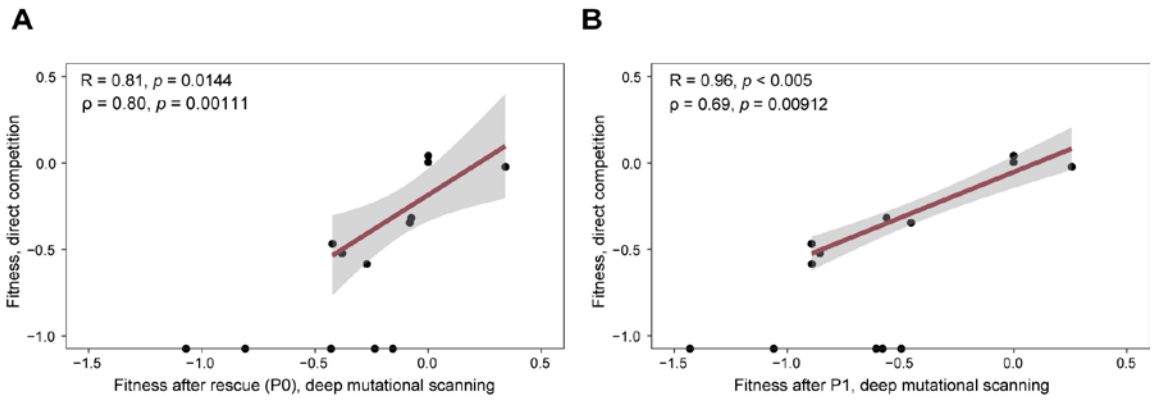

A
Replicate 1
Replicate 2
Replicate 3
missense
nonsense
silent
wildtype
Scaled density
Fitness
Mutation types
missense
nonsense
silent
wildtype
B
fitness.R1.P4toPla
fitness.R2.P4toPla
fitness.R3.P4toPla
Corr:
0.819***
Corr:
0.864***
Corr:
0.788***
C
R = 0.98, p < 0.005
ρ = 0.62, p = 0.023
Fitness, direct competition
Fitness after P4, deep mutational scanning

Figure 2.4 Precision and accuracy of replicative fitness, as measured by deep mutational scanning. (**A**) The fitness distribution of missense, nonsense, and silent mutations, after filtering out mutations caused by potential PCR errors. Code used to make this figure was generated by Sarah Arcos. (**B**) Correlations of variant fitness in three replicates. The upper right panels show the Pearson correlation coefficients of corresponding replicates with the significance level. Diagonal panels show the overall fitness distribution, disregarding the types of mutation. The lower left panels show the fitness values for individual mutants in the indicated replicates. (**C**) The fitness values of 13 selected mutations were measured by deep mutational scanning or pairwise competition with the wild-type virus. Lethal mutations in the competition assay are shown on the x-axis. R indicates the Pearson correlation coefficient among viable variants, while ρ indicates the Spearman correlation coefficient in all variants, including the lethal mutations. The red line shows the trendline using a linear regression model. The gray zone indicates the 95% CI for predictions from the linear model.

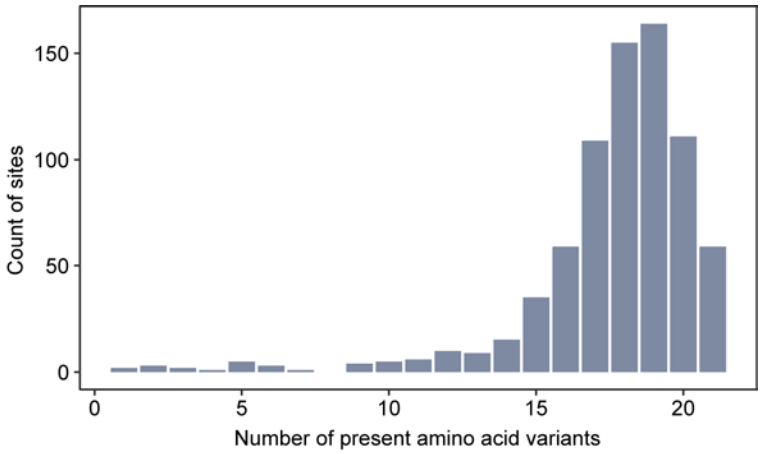
We validated our fitness measurements by comparing the deep mutational scanning fitness of 13 amino acid substitutions to the fitness values we have measured previously by pairwise competition and quantitative RT-PCR (50). These 13 PB1 substitutions were measured in the same genomic background (A/WSN/33 (H1N1)) with pairwise competition assays performed in the same cells (A549), at the same MOI (0.01), and for the same number of passages (four). The fitness values in two experiments were well correlated with a Pearson correlation coefficient of 0.98 ($P < 0.005$) for viable variants and a Spearman correlation coefficient of 0.62 ($P = 0.023$) for all variants including the lethal mutants (Figure 2.4C; Supplemental Figure 2.2A and B). The fitness of two non-lethal (R192K and E751D) and one lethal (E519D) substitutions in the targeted mutagenesis (50) could not be measured in deep mutational scanning after filtering for mutations caused by potential PCR errors. Five other lethal substitutions were identified in passaged DMS libraries, but with very low fitness values.



Supplemental Figure 2.2 Fitness comparison between deep mutational scanning and direct competition in early passages. The comparison between the replicative fitness measured by direct competition with the wild type strain and by deep mutational scanning (**A**) After virus library rescue, before passaging, and (**B**) after one passage on A549 cells. R indicates the Pearson correlation coefficient for viable variants, while ρ indicates the Spearman correlation coefficient for all variants including lethal mutations. The red line shows the trendline using a linear regression model. The gray zone indicates the 95% confidence interval for predictions from the linear model.

Site entropy defines constraints

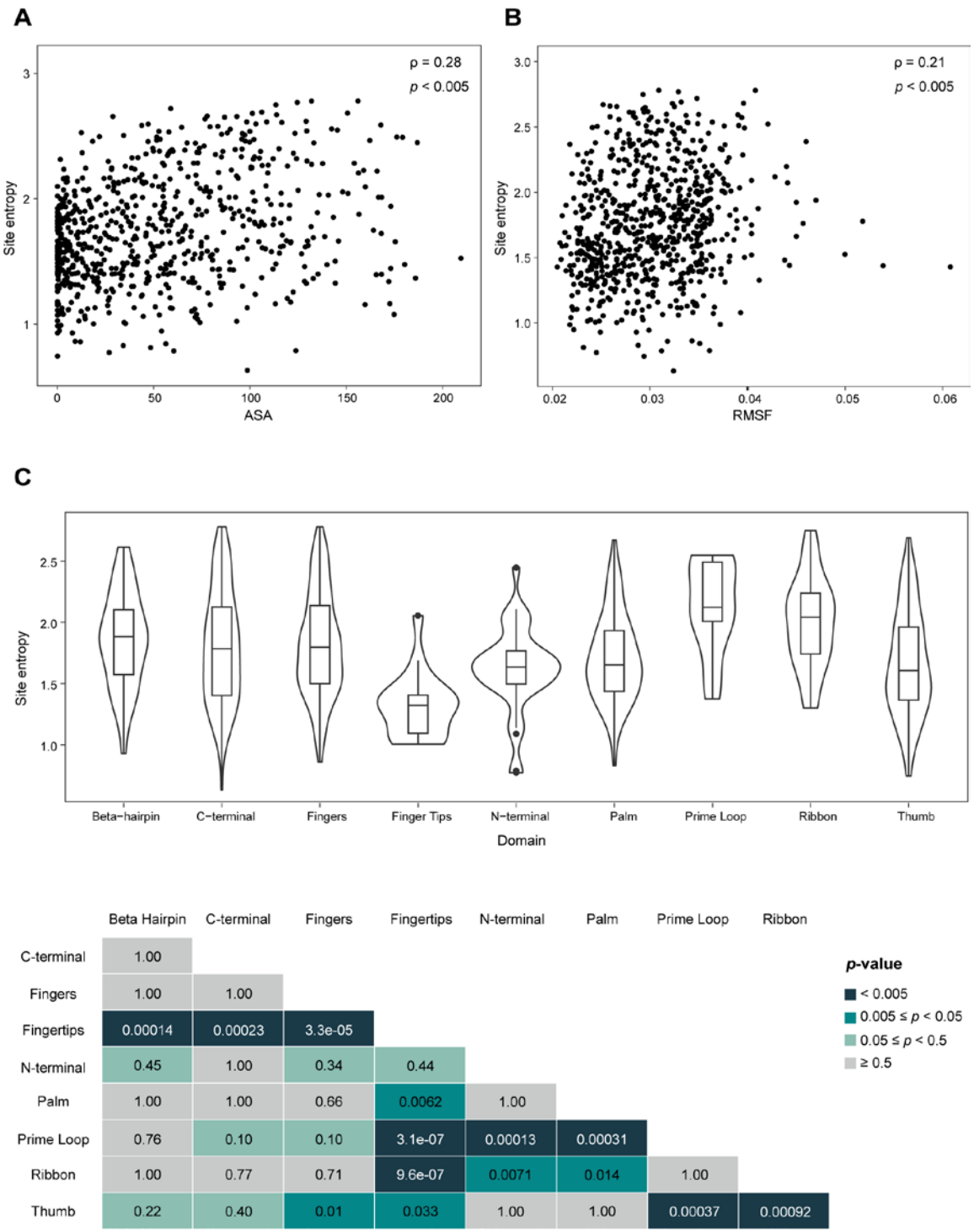
We calculated site entropy, or Shannon entropy at each site, based on the enrichment of all amino acid variants at a site. The enrichment of each amino acid variant in the calculation was determined by its enrichment ratio after four passages and normalized to sum to 1 (see Materials and methods). High site entropy indicates that variation at the amino acid level does not substantially impact viral fitness and/or that several amino acids are equally tolerated at a site. Because the site entropy calculation would be misleading if some amino acids were absent in the initial libraries, we marked and excluded 16 sites with fewer than 40% of amino acid variants (fewer than 9 out of 21 possible variants) generated in the plasmid libraries (Supplemental Figure 2.3).



Supplemental Figure 2.3 Sites with varying mutational representations. Count of sites with different numbers of amino acid variants present at that site in the plasmid libraries, after filtering out the mutations with low sequencing counts or with sequencing library preparation errors. The maximum variation includes twenty amino acid variants plus variants for stop codons at a site.

Site entropy varied across PB1 subdomains. Structural mapping revealed lower site entropy at buried sites and at interfaces between PB1 and RNA and between PB1 and either PA or PB2 (ChimeraX file available at DOI: 10.5061/dryad.p2ngf1vxm). Consistent with this observation, there was a modest, but statistically significant correlation ($\rho = 0.28$, $P < 0.005$) between site entropy and a residue's Accessible Surface Area (Supplemental Figure 2.4A). Residues that are more flexible are often more tolerant to mutation and evolve at a higher rate (23). We performed a molecular dynamics simulation and found the correlation between residue flexibility, captured by the root mean square fluctuation of a 20-ns molecular dynamics simulation of A/Brevig Mission/1/1918(H1N1) RdRp, and site entropy was also weak but significant ($\rho = 0.21$, $P < 0.005$) (Supplemental Figure 2.4B). Using the subdomains defined in (46), we grouped site entropy by subdomain. Residues in the fingertips subdomain exhibited lower entropy ($P < 0.005$ compared with β-hairpin, C-terminal, fingers, prime loop, and ribbon subdomains; $P < 0.05$ compared with palm and thumb subdomains, and $P > 0.05$ compared with N-terminal subdomain), residues in the prime loop and ribbon subdomains exhibited higher entropy (for prime loop subdomain: $P < 0.005$ compared with fingertips, N-terminal, palm, and

thumb subdomains; and $P < 0.005$ compared with fingertips and thumb subdomains), and the distribution of entropy values across other subdomains were largely similar (Supplemental Figure 2.4C).



Supplemental Figure 2.4 Correlation between site entropy and defined features on RdRp. Correlation between a residue's (**A**) accessible surface area or (**B**) root mean square fluctuation, and its site entropy. The molecular dynamics simulation and measurement of root mean square fluctuation was done by Kimberly Sabsay. Each dot represents a residue on the RdRp. ρ indicates the Spearman correlation coefficient. (**C**) Site entropy distribution in different subdomains of RdRp. The chart below shows the adjusted *p*-values by Bonferroni correction between each pair of subdomain comparisons.

Because ASA, RMSF, and simple subdomain identity may mask important differences by averaging over a number of high and low entropy sites, we focused subsequent analyses on specific sites with defined functions. The PB1 active site consists of the evolutionarily conserved motifs A-G, with the catalytic metal ions being coordinated by motifs A and C at the edge of the central cavity (29, 51). We used logo plots to display the enrichment of each amino acid substitution at residues in motif C (52). The site entropy for the active site was quite low, and there were few alternatives to the wild-type amino acid (Figure 2.5A). Similarly, we evaluated PB1 residues that are bound to RNA by hydrogen bonds or interact with RNA due to proximity at each stage (e.g., apo-enzyme, early template binding, late template binding, cap-snatching, pre-initiation, mixed-initiation, initiation-to-catalysis, catalysis-to-nucleotide-incorporation, elongation, and termination). Here, residues interacting with the template RNA (3′ vRNA) and product (mRNA) had lower site entropy than others. Site entropy at residues that bind to RNA (5′ vRNA) but are not involved in transcription was not significantly different than those of other sites (Figure 2.5B).

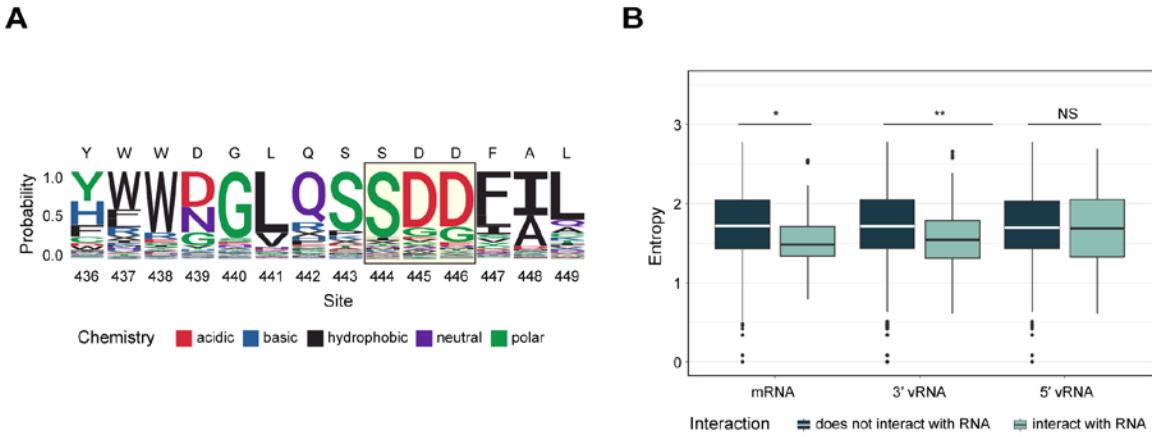
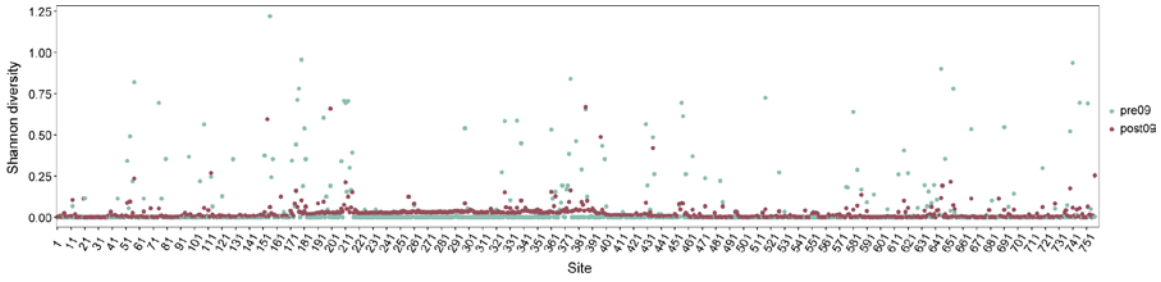


Figure 2.5 Site entropy of key residues. (**A**) Enrichment of amino acid substitutions at each residue in motif C. Residues conserved in all negative sense RNA viruses are marked with the light-yellow box. Amino acids are colored by their biochemical characteristics. A stop codon is represented by "X". (**B**) Site entropy of sites based on their direct interaction with mRNA, 3′ vRNA, and 5′ vRNA, visualized by Tukey boxplot. The line in the boxes represents the median, and the top and bottom of the boxes represent the 25th and 75th percentile. Data points greater than the 75th

percentile + 1.5 × interquartile range (IQR) or less than the 25th percentile – 1.5 × IQR are shown outside the box and the whisker. Wilcoxon test. *P < 0.5 and **P < 0.05; NS: non-significant.

Beneficial residues observed in natural evolution

To gain insights into the relationship between mutational tolerance and the long-term evolution of PB1, we compared our measured site entropy to the "natural" amino acid Shannon diversity of each site. The calculation of Shannon diversity in natural sequences is slightly different from that of site entropy for deep mutational scanning; they use the same equation (see Materials and methods), but the former uses the frequency of each amino acid variant, while the latter uses the enrichment ratio. We divided the records of naturally evolved PB1 sequences from human hosts available on GISAID into pre- and post-2009 subsets, separated by the time period of the 2009 H1N1 pandemic, to minimize the impact of co-circulation of pre- and post-pandemic viruses. After filtering, we evaluated 1,491 PB1 sequences in the pre-2009 data set and 35,501 in the post-2009 data set. Since Shannon diversity is biased for mutations observed in years that have been sampled more densely, we corrected for the uneven sampling of PB1 sequences over time by calculating weighted Shannon diversity as previously described (46, see Materials and methods). In general, there was greater Shannon diversity in the pre-2009 data set (Supplemental Figure 2.5). We found a moderate correlation between DMS site entropy and natural Shannon diversity in the pre- and post-2009 data sets (pre-2009: $\rho = 0.40$, $P < 0.005$; post-2009: $\rho = 0.31$, $P < 0.005$; Figure 2.6A and B).



Supplemental Figure 2.5 Amino acid diversity at sites of naturally occurring influenza H1N1 PB1 sequences. Weighted Shannon diversity for each site in natural PB1 evolution. Diversity in pre- and post-2009 sequences was calculated separately.

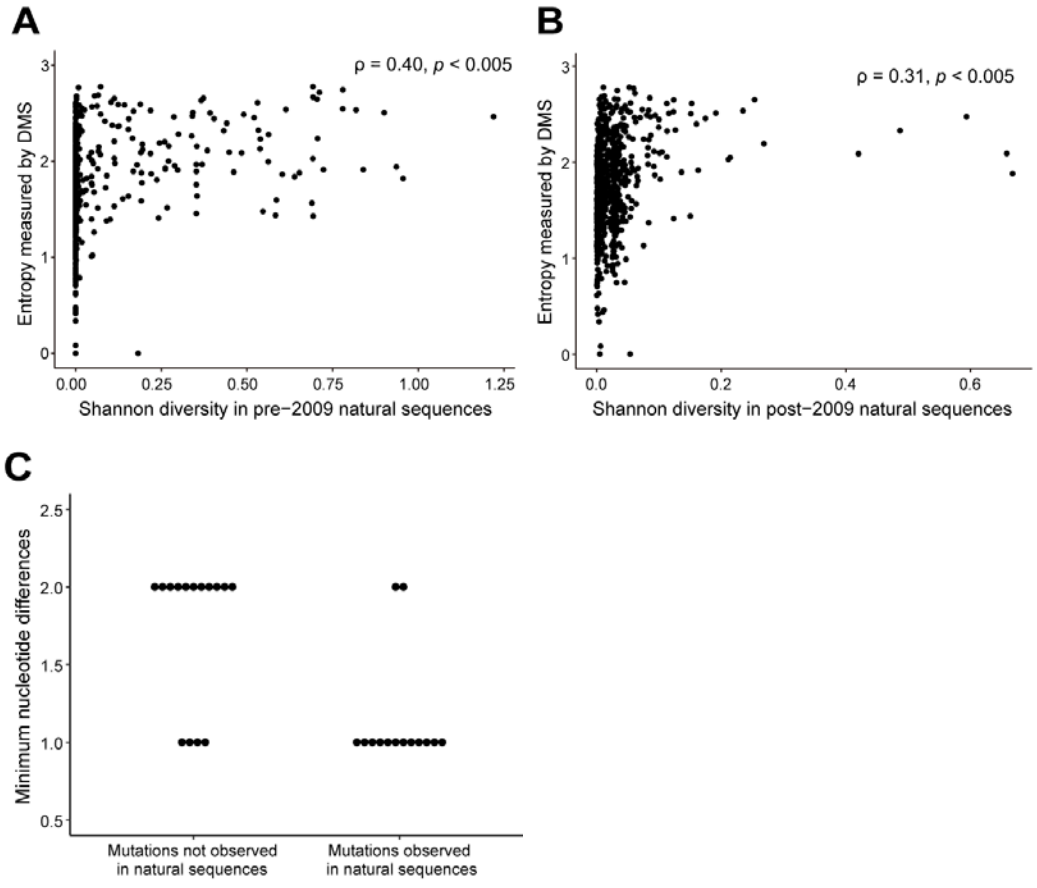


Figure 2.6 Impacts of DMS fitness and mutational tolerance on natural PB1 evolution. Correlation between the Shannon diversity of naturally occurring sequences (**A**) before and (**B**) after 2009 and the site entropy measured by deep mutational scanning. Five hundred five residues in the pre-2009 and one residue in the post-2009 natural sequences are completely conserved. The difference between the number of conserved residues before and after 2009 is potentially due to insufficient sampling prior to 2009. ρ indicates the Spearman correlation coefficient. (**C**) The minimum nucleotide differences between the wild type (or dominant amino acid) in naturally occurring PB1 sequences and beneficial mutations identified by deep mutational scanning. Each dot represents a beneficial mutation.

Similarly, we determined whether mutations identified as beneficial in deep mutational scanning forecast those that appear in the natural evolution of PB1 in human hosts. We defined 29 mutations as beneficial based on a measured fitness greater than two standard deviations (Z-score > 2) above the mean fitness of silent mutations (the neutral, null model) (Table 2.2). All beneficial amino acid mutations had one or two nucleotide changes compared with the corresponding wild-type codon. Fourteen of 29 beneficial mutations have occurred during the evolution of H1N1 PB1, and of these, many have appeared multiple times independently. The other 15 were not observed in the available sequences. The majority of beneficial mutations that did appear in natural evolution are accessible by a single nucleotide substitution from the wild-type codon, while those that did not appear in natural sequences usually required two nucleotide substitutions in the codon (Figure 2.6C). Of the beneficial mutations, M317V, T323M, I637V, K653R, K691R, and M744A have appeared in >0.1% of all sequences collected in at least 1 year when the mutation was present. Notably, although 691K was the wild type in WSN33 at site 691, arginine (R) was the dominant amino acid in pre-2009 strains (Supplemental Figure 2.6), suggesting a true fitness advantage of arginine over lysine at this site. Lysine was again the dominant amino acid in the 2009 pandemic strain, but 691R has been detected every year.

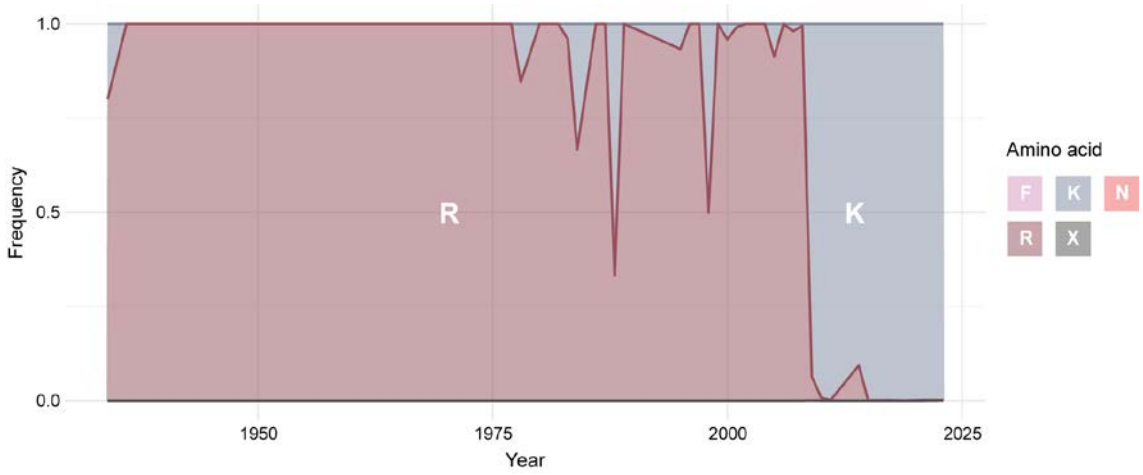
Table 2.2 Natural occurrence of beneficial mutations identified by deep mutational scanning.

Natural frequency	Mutation	DMS fitness Z-score	Nucleotide change from wildtype
Co-existed as dominant amino acids	K691R	3.162489	1
Above 0.1% in natural sequences	T323M	2.386425	1
	M317V	2.209847	1
	K653R	2.047385	1
	I637V	2.025235	1

	M744A	3.443046	2
Appeared but below 0.1%	L108R	2.372813	1
	K577Q	3.535065	1
	V255A	3.021593	1
	Q116V	2.247724	2
	I164L	2.747743	1
	P701L	2.360081	1
	I674L	2.073289	1
	K578T	2.050551	1
Did not appear in natural dataset	L108Y	5.126652	2
	P647N	3.608524	2
	V255S	3.490547	2
	V255T	2.050218	2
	Q116M	3.409729	2
	Q116T	2.281114	2
	Q679N	2.979386	2
	P510A	2.890052	1
	P510G	2.233921	2
	R151L	2.793400	1
	L351R	2.513735	1
	T105R	2.454788	2[a]
	S261F	2.241466	1
	M646A	2.150846	2
	N654D	2.020037	2[b]

[a] Although the wild-type amino acid at site 105 for WSN33 is threonine (T), the dominant amino acid at this site in natural PB1 population is asparagine (N). Therefore, the number of nucleotide change(s) needed for most natural PB1 to have arginine (R) at site 105 should be 2 (from N) rather than 1 (from T).

[b] Wild-type amino acid at site 654 was asparagine (N) before 2009 and became serine (S) after 2009. The minimum nucleotide change needed from N to D is 1, and from S to D is 2.



Supplemental Figure 2.6 Frequency change of amino acid variants at site 691. Frequency of amino acid variants observed at site 691 of naturally occurring PB1 sequences from 1934 to 2023. "X" stands for uncertain/ambiguous amino acid. Dominant amino acid variants were labeled on the plot.

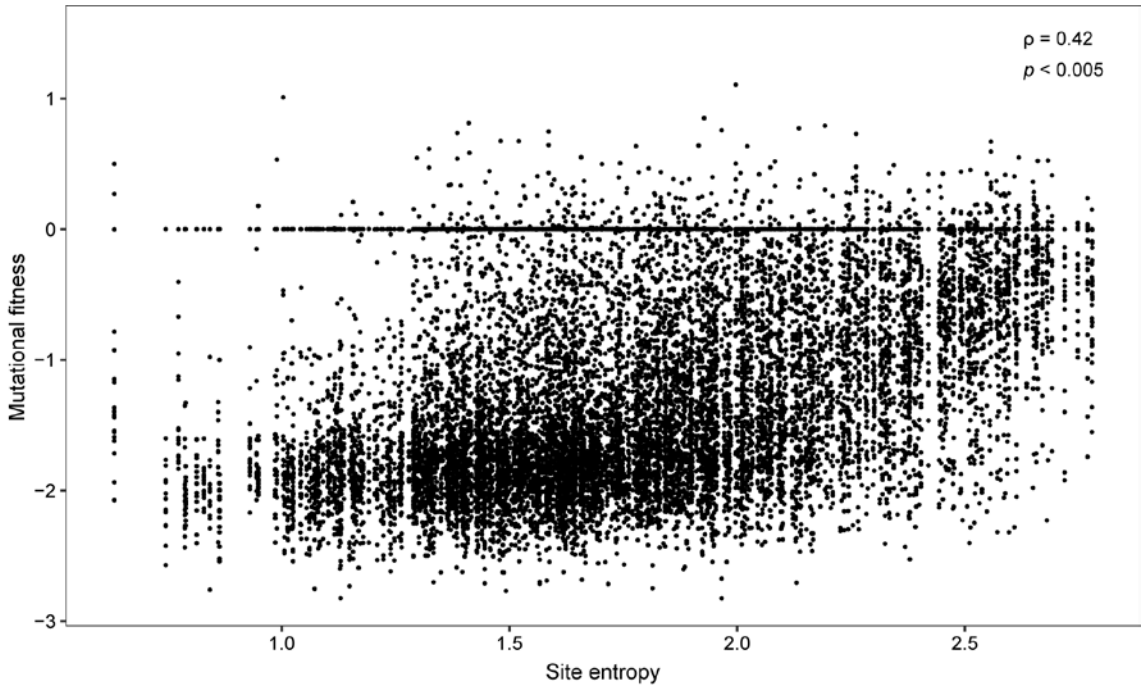
Sites where we identified beneficial mutations have also been found to be relevant to polymerase activity and viral fitness. Mutations at site 317 were identified in the 1997 Hong Kong H5N1 outbreak (53) and were found to be functionally significant for virulence in mammals (54, 55). Site 744 is located in the vRNA-binding region, and M744V was found to be a canine-adaptive mutation of avian H3N2 (56). Site 674 is both part of the contact points between the PB1 C-terminal and PB2 N-terminal subdomains and interacts with the 3′ end of the vRNA promoter; mutations to T, L, and S all increased polymerase activity (57). At the polymerase dimer interface, residue 577 interacts with PA and PB2 (58, 59), and residue 578 orients to a residue in the PB2 unstructured loop; K577E in avian H9N2 increases polymerase activity at a lower replication temperature (60), and serial passage of A/Hong Kong/1/68 (H3N2) in mice also gave rise to K577E/M/Q (61). Lysine 578, the wild type, is a ubiquitination site, and mutations from K578 to both non-charged alanine (A) and positively charged arginine (R) increase polymerase activity but are harmful to viral fitness (62). The neutral side chain of A578

reduced polymerase dimerization, while the positively charged R578 aborted cRNA synthesis and led to the premature assembly of the dimer.

2.5 Discussion

We performed a near complete deep mutational scan of the WSN33 PB1 RdRp subunit, defining the impacts of nearly all amino acid substitutions on replicative fitness in A549 cells. Most substitutions are detrimental, and we identified mutational constraints at sites involved in key polymerase interactions, specifically at sites interacting with the RNA template and product. In contrast, mutations in other regions of the protein are better tolerated. Diversity at these sites was moderately correlated with site diversity as defined in available influenza sequences. A small number of mutations are beneficial, and many of these have been observed in natural evolution. Those that were not observed in natural evolution were generally inaccessible by single nucleotide mutation. Our study was comprehensive, as we interrogated a much larger number of codon and amino acid variants compared with studies that evaluate mutations occurring in natural sequences or generated by error-prone PCR. While prior work on the functional domains and evolutionary constraints on RdRp have largely relied on the analyses of sequence conservation (51, 63, 64), our DMS identified significant, site-specific heterogeneity in the influenza virus polymerase.

Through deep mutational scanning, we find that the fitness of mutations on influenza virus PB1 is moderately and positively correlated with site entropy (Supplemental Figure 2.7). The rise of most beneficial mutations requires some degree of mutational flexibility, and highly detrimental mutations are more commonly seen in sites with low mutational tolerance.



Supplemental Figure 2.7 Correlation between site entropy and mutational fitness. Each dot represents an amino acid substitution at a site. ρ indicates the Spearman correlation coefficient.

Unlike in hemagglutinin (65) and neuraminidase (35), the evolutionary constraints on the influenza virus RdRp are not well defined by protein subdomain. Instead, each subdomain has some sites that are under strict purifying selection and other sites that are more tolerant to mutation. Similar phenomena were observed in naturally occurring genomes, where conservative and variable residues were distributed relatively evenly across major subdomains (66). These findings highlight the importance of local structures and functional interactions in influenza virus replication. As expected, mutations to amino acids with side chains of similar biochemical properties (e.g., charged/uncharged, polar/non-polar) are usually more tolerated. This is consistent with the impact of these biochemical properties on higher-level protein structures: large and non-polar amino acids are more likely to form hydrophobic cores, while polar or charged amino acids are more likely to be surface residues.

We identified beneficial mutations that have been observed in the natural evolution of influenza virus RdRp and found accessibility by single nucleotide substitution to be a key factor

determining whether a beneficial mutation can arise naturally. We also identified several adaptative mutations that arose in nature with more than one nucleotide change, which could imply an indirect evolutionary path involving gain and subsequent loss of intermediate mutations (67). Many of the beneficial mutations identified in our study not only increase polymerase activity but have been shown to be functionally important for host adaptation or by altering post-translation modification. In addition, mutations with higher fitness had a moderate but significant association with sites that have higher mutational tolerance.

Our work is subject to several limitations. First, while our deep mutational scan provides comprehensive fitness measurements in the WSN33 genetic background, the measured mutational effects may not be recapitulated in the genetic background of other H1N1 strains. Second, we performed our DMS on A549 cells, which allow for high-volume infections and the robust viral replication necessary for a comprehensive screen with a large library. It is possible that fitness values may differ in a more physiologically relevant replication system, such as primary airway epithelial cells. Third, we focused on the mutational effects of single amino acid substitutions and did not account for epistatic interactions within PB1 and between PB1 and other viral proteins. In natural evolution, interacting sites often co-evolve (46, 68), and an adaptive mutation towards a stimulus is commonly accompanied by compensatory mutations that maintain effective replication (5). Finally, we only examined fitness in terms of replication, but various treatments can be applied to the variant virus library and future research can examine mutational fitness under specific conditions such as with drug selection or altered baseline mutational rates.

Overall, we have developed a comprehensive map of the local fitness landscape for the influenza A virus PB1 protein. In doing so, we identified how specific amino acid substitutions

affect the replicative fitness of the virus and the degree of evolutionary constraint at each site. Our work provides a foundation for subsequent studies of influenza virus replication and host adaptation and may prove to be a valuable addition to genomic surveillance efforts.

2.6 Acknowledgements

We thank Jesse Bloom and Shirleen Soh for making their analysis code available and for the helpful suggestions, and Aaron King, Gideon Bradburd, and Kayla Peck for the helpful discussion. We further acknowledge the contributions of all submitters to GISAID. We performed molecular graphics with UCSF ChimeraX, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from the National Institutes of Health (NIH) R01-GM129325 and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases. The MD simulations were performed on computational resources managed and supported by Princeton Research Computing, a consortium of groups including the Princeton Institute for Computational Science and Engineering (PICSciE) and the Office of Information Technology's High Performance Computing Center and Visualization Laboratory at Princeton University.

This work was supported by NIH R01 AI170520 and a Burroughs Wellcome Fund Investigator in the Pathogenesis of Infectious Diseases Award, both to A.S.L., and NIH DP2 AI175474 to A.T.V.

2.7 Data Availability

Raw sequence reads are available in the NCBI Sequence Read Archive under Bioproject #PRJNA1009589.

2.8 References

1. Pappas C, Aguilar PV, Basler CF, Solorzano A, Zeng H, Perrone LA, Palese P, Garcia-Sastre A, Katz JM, Tumpey TM. 2008. Single gene reassortants identify a critical role for PB1, HA, and NA in the high virulence of the 1918 pandemic influenza virus. Proceedings of the National Academy of Sciences 105:3064–3069.

2. Zhang X, Li Y, Jin S, Zhang Y, Sun L, Hu X, Zhao M, Li F, Wang T, Sun W, Feng N, Wang H, He H, Zhao Y, Yang S, Xia X, Gao Y. 2021. PB1 S524G mutation of wild bird-origin H3N8 influenza A virus enhances virulence and fitness for transmission in mammals. Emerging Microbes & Infections 10:1038–1051.

3. Feng X, Wang Z, Shi J, Deng G, Kong H, Tao S, Li C, Liu L, Guan Y, Chen H. 2016. Glycine at Position 622 in PB1 Contributes to the Virulence of H5N1 Avian Influenza Virus in Mice. Journal of Virology 90:1872–1879.

4. Xu C, Hu W-B, Xu K, He Y-X, Wang T-Y, Chen Z, Li T-X, Liu J-H, Buchy P, Sun B. 2012. Amino acids 473V and 598P of PB1 from an avian-origin influenza A virus contribute to polymerase activity, especially in mammalian cells. Journal of General Virology 93:531–540.

5. Goldhill DH, te Velthuis AJW, Fletcher RA, Langat P, Zambon M, Lackenby A, Barclay WS. 2018. The mechanism of resistance to favipiravir in influenza. Proceedings of the National Academy of Sciences 115:11613–11618.

6. Pauly MD, Lyons DM, Fitzsimmons WJ, Lauring AS. 2017. Epistatic Interactions within the Influenza A Virus Polymerase Complex Mediate Mutagen Resistance and Replication Fidelity. mSphere 2.

7. Naito T, Shirai K, Mori K, Hidetaka Muratsu, Hiroshi Ushirogawa, Ohniwa RL, Hanada K, Saito M. 2019. Tyr82 Amino Acid Mutation in PB1 Polymerase Induces an Influenza Virus Mutator Phenotype. Journal of Virology 93.

8. Cheung PPH, Watson SJ, Choy K-T, Fun Sia S, Wong DDY, Poon LLM, Kellam P, Guan Y, Malik Peiris JS, Yen H-L. 2014. Generation and characterization of influenza A viruses with altered polymerase fidelity. Nature Communications 5.

9. Li J, Liang L, Jiang L, Wang Q, Wen X, Zhao Y, Cui P, Zhang Y, Wang G, Li Q, Deng G, Shi J, Tian G, Zeng X, Jiang Y, Liu L, Chen H, Li C. 2021. Viral RNA-binding ability conferred by SUMOylation at PB1 K612 of influenza A virus is essential for viral pathogenesis and transmission. PLOS Pathogens 17:e1009336.

10. Varga ZT, Ramos I, Hai R, Schmolke M, García-Sastre A, Fernandez-Sesma A, Palese P. 2011. The Influenza Virus Protein PB1-F2 Inhibits the Induction of Type I Interferon at the Level of the MAVS Adaptor Protein. PLoS Pathogens 7:e1002067.

11. Hai R, Schmolke M, Varga ZT, Manicassamy B, Wang TT, Belser JA, Pearce MB, García-SastreA, Tumpey TM, Palese P. 2010. PB1-F2 Expression by the 2009 Pandemic H1N1 Influenza Virus Has Minimal Impact on Virulence in Animal Models. Journal of Virology 84:4442–4450.

12. Laporte M, Stevaert A, Raeymaekers V, Boogaerts T, Nehlmeier I, Chiu W, Benkheil M, Vanaudenaerde B, Pöhlmann S, Naesens L. 2019. Hemagglutinin Cleavability, Acid Stability, and Temperature Dependence Optimize Influenza B Virus for Replication in Human Airways. Journal of Virology 94.

13. Goñi N, Iriarte A, Comas V, Martín Soñora, Moreno P, Moratorio G, Musto H, Cristina J. 2012. Pandemic influenza A virus codon usage revisited: biases, adaptation and implications for vaccine strain development. Virology Journal 9.

14. Fan RZ, Eric, Chloe, Olive, Nicholls JM, Rabadan R, Peiris M, Leo L.M. Poon. 2015. Generation of Live Attenuated Influenza Virus by Using Codon Usage Bias. Journal of Virology 89:10762–10773.

15. Kumar N, Bera BC, Greenbaum BD, Bhatia S, Sood R, Selvaraj P, Anand T, Tripathi BN, Virmani N. 2016. Revelation of Influencing Factors in Overall Codon Usage Bias of Equine Influenza Viruses. PLOS ONE 11:e0154376.

16. Mintseris J, Weng Z. 2005. Structure, function, and evolution of transient and obligate protein-protein interactions. Proceedings of the National Academy of Sciences 102:10930–10935.

17. Aharoni A, Gaidukov L, Khersonsky O, Gould SM, Roodveldt C, Tawfik DS. 2005. The “evolvability” of promiscuous protein functions. Nature Genetics 37:73–76.

18. Andreeva A, Murzin AG. 2006. Evolution of protein fold in the presence of functional constraints. Current Opinion in Structural Biology 16:399–408.

19. Hom N, Gentles L, Bloom JD, Lee KK. 2019. Deep Mutational Scan of the Highly Conserved Influenza A Virus M1 Matrix Protein Reveals Substantial Intrinsic Mutational Tolerance. Journal of Virology 93:e00161-19.

20. Goldman N, Thorne JL, Jones DT. 1998. Assessing the Impact of Secondary Structure and Solvent Accessibility on Protein Evolution. Genetics 149:445–458.

21. Velázquez-Muriel J, Rueda M, Cuesta I, Pascual-Montano A, Orozco M, José María Carazo. 2009. Comparison of molecular dynamics and superfamily spaces of protein domain deformation. BMC Structural Biology 9.

22. Friedland GD, Nils-Alexander Lakomek, Griesinger C, Meiler J, Kortemme T. 2009. A Correspondence Between Solution-State Dynamics of an Individual Protein and the Sequence and Conformational Diversity of its Family. PLOS Computational Biology 5:e1000393–e1000393.

23. Marsh JA, Teichmann SA. 2013. Parallel dynamics and evolution: Protein conformational fluctuations and assembly reflect evolutionary changes in sequence and structure. BioEssays 36:209–218.

24. Mintseris J, Weng Z. 2005. Structure, function, and evolution of transient and obligate protein-protein interactions. Proceedings of the National Academy of Sciences 102:10930–10935.

25. Eames M, Kortemme T. 2007. Structural Mapping of Protein Interactions Reveals Differences in Evolutionary Pressures Correlated to mRNA Level and Protein Abundance. Structure 15:1442–1451.

26. Worth CL, Gong S, Blundell TL. 2009. Structural and functional constraints in the evolution of protein families. Nature Reviews Molecular Cell Biology 10:709–720.

27. Franzosa EA, Xia Y. 2009. Structural Determinants of Protein Evolution Are Context-Sensitive at the Residue Level. Molecular Biology and Evolution 26:2387–2395.

28. Ramsey DC, Scherrer MP, Zhou T, Wilke CO. 2011. The Relationship Between Relative Solvent Accessibility and Evolutionary Rate in Protein Evolution. Genetics 188:479–488.

29. te Velthuis AJW, Fodor E. 2016. Influenza virus RNA polymerase: insights into the mechanisms of viral RNA synthesis. Nature Reviews Microbiology 14:479–493.

30. Kouba T, Drncová P, Cusack S. 2019. Structural snapshots of actively transcribing influenza polymerase. Nature Structural & Molecular Biology 26:460–470.

31. York A, Hengrung N, Vreede FT, Huiskonen JT, Fodor E. 2013. Isolation and characterization of the positive-sense replicative intermediate of a negative-strand RNA virus. Proceedings of the National Academy of Sciences of the United States of America 110:E4238-4245.

32. Soh YS, Moncla LH, Eguia R, Bedford T, Bloom JD. 2019. Comprehensive mapping of adaptation of the avian influenza polymerase protein PB2 to humans. eLife 8:e45079.

33. Sourisseau M, Lawrence DA, Schwarz MC, Storrs C, Veit EC, Bloom JD, Evans M. 2019. Deep Mutational Scanning Comprehensively Maps How Zika Envelope Protein Mutations Affect Viral Growth and Antibody Escape. Journal of Virology 93.

34. Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KHD, Dingens AS, Navarro MJ, Bowen JE, Tortorici MA, Walls AC, King NP, Veesler D, Bloom JD. 2020. Deep Mutational Scanning

of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. Cell 182:1295-1310.e20.

35. Lei R, Milena A, Tan TC, Qi Wen Teo, Wang Y-Q, Zhang X, Luo S, Nair SK, Peng J, Wu NC. 2023. Mutational fitness landscape of human influenza H3N2 neuraminidase. Cell Reports 42:111951–111951.

36. Doud MB, Hensley SE, Bloom JD. 2017. Complete mapping of viral escape from neutralizing antibodies. PLOS Pathogens 13:e1006271.

37. Bloom JD, Dingens A. 2019. Tiling primers for codon mutagenesis. Github. https://github.com/jbloomlab/CodonTilingPrimers

38. Bloom JD. 2014. An Experimentally Determined Evolutionary Model Dramatically Improves Phylogenetic Fit. Molecular Biology and Evolution 31:1956–1978.

39. Dingens AS, Haddox HK, Overbaugh J, Bloom JD. 2017. Comprehensive Mapping of HIV-1 Escape from a Broadly Neutralizing Antibody. Cell Host & Microbe 21:777-787.e4.

40. Peck K. 2017. RE_check. Github. https://github.com/kmpeck/RE_check

41. Doud M, Bloom J. 2016. Accurate Measurement of the Effects of All Amino-Acid Mutations on Influenza Hemagglutinin. Viruses 8:155.

42. Bloom JD. 2015. Software for the analysis and visualization of deep mutational scanning data. BMC Bioinformatics 16.

43. Centers for Disease Control and Prevention. 2019. 2009 H1N1 Pandemic Timeline . Centers for Disease Control and Prevention. https://www.cdc.gov/flu/pandemic-resources/2009-pandemic-timeline.html.

44. Katoh K. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research 30:3059–3066.

45. Shannon CE. 1948. A Mathematical Theory of Communication. Bell System Technical Journal 27:379–423.

46. Arcos S, Han AX, te W, Russell CA, Lauring AS. 2023. Mutual information networks reveal evolutionary relationships within the influenza A virus polymerase. Virus Evolution 9.

47. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH, Ferrin TE. 2020. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Science 30:70–82.

48. Laskowski RA, Swindells MB. 2011. LigPlot+: Multiple Ligand–Protein Interaction Diagrams for Drug Discovery. Journal of Chemical Information and Modeling 51:2778–2786.

49. Krissinel E, Henrick K. 2007. Inference of Macromolecular Assemblies from Crystalline State. Journal of Molecular Biology 372:774–797.

50. Visher E, Whitefield SE, McCrone JT, Fitzsimmons W, Lauring AS. 2016. The Mutational Robustness of Influenza A Virus. PLoS Pathogens 12.

51. Chu C, Fan S, Li C, Macken C, Kim JH, Hatta M, Neumann G, Kawaoka Y. 2012. Functional Analysis of Conserved Motifs in Influenza Virus PB1 Protein. PLoS ONE 7:e36113.

52. Wagih O. 2017. ggseqlogo: a versatile R package for drawing sequence logos. Bioinformatics 33:3645–3647.

53. Katz JM, Lu X, Tumpey TM, Smith CB, Shaw MW, Subbarao K. 2000. Molecular Correlates of Influenza A H5N1 Virus Pathogenesis in Mice. Journal of Virology 74:10807–10810.

54. Lycett SJ, Ward MJ, Lewis FI, Poon AFY, Kosakovsky Pond SL, Brown AJL. 2009. Detection of Mammalian Virulence Determinants in Highly Pathogenic Avian Influenza H5N1 Viruses: Multivariate Analysis of Published Data. Journal of Virology 83:9901–9910.

55. Nao N, Kajihara M, Manzoor R, Maruyama J, Yoshida R, Muramatsu M, Miyamoto H, Igarashi M, Eguchi N, Sato M, Kondoh T, Okamatsu M, Sakoda Y, Kida H, Takada A. 2015. A Single Amino Acid in the M1 Protein Responsible for the Different Pathogenic Potentials of H5N1 Highly Pathogenic Avian Influenza Virus Strains. PLOS ONE 10:e0137989.

56. Li X, Liu J, Qiu Z, Liao Q, Peng Y, Chen Y, Shu Y. 2021. Host-Adaptive Signatures of H3N2 Influenza Virus in Canine. Frontiers in Veterinary Science 8:740472.

57. Welkers MRA, Pawestri HA, Fonville JM, Sampurno OD, Pater M, Holwerda M, Han AX, Russell CA, Jeeninga RE, Setiawaty V, de Jong MD, Eggink D. 2019. Genetic diversity and host adaptation of avian H5N1 influenza viruses during human infection. Emerging Microbes & Infections 8:262–271.

58. Fan H, Walker AP, Carrique L, Keown JR, Serna Martin I, Karia D, Sharps J, Hengrung N, Pardon E, Steyaert J, Grimes JM, Fodor E. 2019. Structures of influenza A virus RNA polymerase offer insight into viral genome replication. Nature 573:287–290.

59. Kuang Yu Chen, Dos E, Enouf V, Isel C, Naffakh N. 2019. Influenza virus polymerase subunits co-evolve to ensure proper levels of dimerization of the heterotrimer. Plos Pathogens 15:e1008034–e1008034.

60. Kamiki H, Matsugo H, Kobayashi T, Ishida H, Takenaka-Uema A, Murakami S, Horimoto T. 2018. A PB1-K577E Mutation in H9N2 Influenza Virus Increases Polymerase Activity and Pathogenicity in Mice. Viruses 10:653.

61. Ping J, Keleta L, Forbes NE, Dankar SK, Stecho W, Tyler S, Zhou Y, Babiuk LA, Weingartl HM, Halpin RA, Boyne A, Bera J, Hostetler J, Fedorova N, Proudfoot K, Katzel DA, Stockwell T, Elodie Ghedin, Spiro DM, Brown EG. 2011. Genomic and Protein Structural Maps of Adaptive Evolution of Human Influenza A Virus to Increased Virulence in the Mouse. PLOS ONE 6:e21740–e21740.

62. Günl F, Krischuns T, Schreiber JA, Henschel L, Wahrenburg M, Drexler HCA, Leidel SA, Cojocaru V, Seebohm G, Mellmann A, Schwemmle M, Ludwig S, Brunotte L. 2023. The ubiquitination landscape of the influenza A virus polymerase. Nature Communications 14:787.

63. Wu NC, Olson CA, Du Y, Le S, Tran K, Remenyi R, Gong D, Al-Mawsawi LQ, Qi H, Wu T-T, Sun R. 2015. Functional Constraint Profiling of a Viral Protein Reveals Discordance of Evolutionary Conservation and Functionality. PLOS Genetics 11:e1005310.

64. Chan JJ-S, Tang Y-S, Lo C-Y, Shaw P-C. 2023. Functional Importance of the Hydrophobic Residue 362 in Influenza A PB1 Subunit. Viruses 15:396.

65. Lee JM, Huddleston J, Doud MB, Hooper KA, Wu NC, Bedford T, Bloom JD. 2018. Deep mutational scanning of hemagglutinin helps predict evolutionary fates of human H3N2 influenza variants. Proceedings of the National Academy of Sciences 115.

66. Figueiredo-Nunes I, Trigueiro-Louro J, Rebelo-de-Andrade H. 2023. Exploring new antiviral targets for influenza and COVID-19: Mapping promising hot spots in viral RNA polymerases. Virology 578:45–60.

67. Wu NC, Dai L, Olson CA, Lloyd-Smith JO, Sun R. 2016. Adaptation in protein fitness landscapes is facilitated by indirect paths. eLife 5:e16965.

68. Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW. 2002. Evolutionary Rate in the Protein Interaction Network. Science 296:750–752.

Chapter 3 Anti-Influenza Virus Effects of Mutagenic Drugs

3.1 Introduction

Replicative fidelity is central to influenza virus evolution. The population of variants created through influenza virus's error-prone replication defines the spectrum of virus adaptability, and fidelity itself is associated with other viral characteristics such as virulence and transmissibility.

Single amino acid substitutions within viral RdRp can alter a virus' replicative fidelity. A PB1 mutation that contributed to the host switch of pdmH1N1, S216G, also leads to higher mutational frequency in H1N1 viruses (1). Mutation T123A on PB1 increases the rate of C-to-U and G-to-A mutations (2). Tyrosine at residue 82 of PB1 is crucial for controlling replicative fidelity; PB1 Y82C is less capable of distinguishing correct and incorrect nucleotides during replication, and therefore confers a mutator phenotype. Mutation at residue 82 of PB1 to many other amino acids (C, S, T, N, A, F, G, V, L, I, K, R, H) also increases the mutation frequency to varying degrees (3). The influence of mutations on replicative fidelity depends on specific genetic backgrounds. The mutation PB1 V43I was reported to reduce the mutational frequency in A/Wuhan/359/95 (H3N2) and A/Vietnam/1203/04 (H5N1) (4) but did not alter the baseline mutation rate in the A/Puerto Rico/8/1934 (H1N1) genetic background (2). While mutation rates evolve over time, RNA viruses replicate within a limited range of fidelity to maintain fitness (5). Influenza and other RNA viruses with higher or lower replicative fidelity both exhibited attenuated phenotypes (6-9), suggesting that the current fidelity is a result of multiple finely tuned evolutionary forces.

Mutagenic drugs are an important tool in studies of replicative fidelity. Since most spontaneous mutations are detrimental (10), there is a theory that the high mutation rates of RNA

viruses have approached an upper limit that allows the viruses to maintain fitness with replication. A higher-than-normal mutation rate would lead to the accumulation of detrimental mutations in the virus genome and make the virus less fit or even inviable. Guided by this theory, mutagenic drugs have been used in lethal mutagenesis - a broad-spectrum antiviral strategy that drives viral populations to extinction by elevating their mutation rates over their viability threshold (11). The variants that originally have a lower mutation rate / high fidelity phenotype are likely to be resistant to mutagenic drugs and start to exhibit a fitness edge.

With the single amino acid substitution variant library generated by deep mutational scanning, we want to elucidate the mechanisms of replicative fidelity by probing the variant library with mutagenic drugs. When passaging with mutagens of proper concentrations, variants with low replicative fidelity are expected to be less fit than those with high fidelity and decrease in frequency, while high-fidelity variants are expected to increase in frequency.

Although the literature often refers to a single mutation for a given virus, each class of nucleotide substitution (A-to-T, -C, -G; C-to-A, -T, -G; T-to-A, -C, -G; G-to-A, -C, -T) has a distinct rate. For influenza A viruses, A-to-G and U-to-C transition mutations have the highest rates, while the rates of the other two transitions (C-to-U and G-to-A) and the 8 transversions are six-fold lower (12). Most mutagenic drugs are nucleoside analogs and work by incorporation into replicating genomes and inducing mismatch in the next round of replication (13). Therefore, different mutagenic drugs increase the frequency of a different subset of mutations, and viruses have distinct sensitivity against every drug.

Ribavirin (1-β-D-ribofuranosyl-1-H-1,2,4-triazole-3-carboxamide) is a purine analog. It was approved by the U.S. Food and Drug Administration (FDA) in 1986 for serious infections of respiratory syncytial virus (RSV) in infants (14) and later used to treat hepatitis C virus as well

as influenza A and B viruses in immunodeficient patients (15). Ribavirin has a structure similar to guanosine. In cells, ribavirin is converted to ribavirin monophosphate (RMP) by adenosine kinase and further phosphorylated to ribavirin diphosphate (RDP) and ribavirin triphosphate (RTP). The derivatives of ribavirin can be incorporated into RNA and form ambiguous base pairs with both cytidine and uridine (Figure 3.1), therefore increasing C-U and G-A transitions (16). Since the catalytic efficiency of ribavirin-cytidine/uridine pairing is exceedingly low (200-300 times slower than that of natural nucleotide pairing) (17), RTP also works as a competitive inhibitor of both ATP and GTP and inhibit RNA synthesis (18). The antiviral activity of ribavirin is achieved through both direct and indirect mechanisms. Direct mechanisms include inhibiting RNA capping activity, direct inhibition of viral polymerases, and lethal mutagenesis. Indirect mechanisms include nucleotide depletion (reducing the guanosine triphosphate (GTP) pools) and immunomodulatory effects.

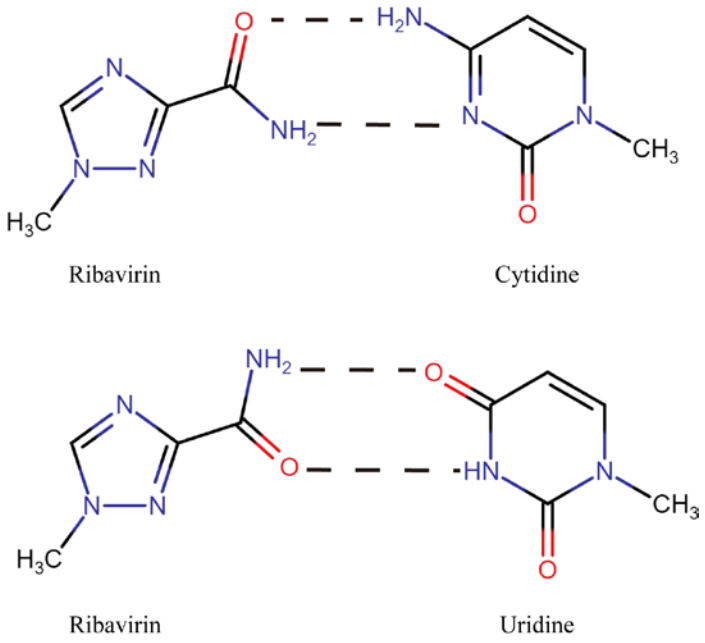


Figure 3.1 Ribavirin can pair with both cytidine and uridine. Phosphate side chains not shown. The misincorporation of ribavirin derivatives has two effects: increasing transition mutations and slow down RNA synthesis.

Favipiravir (T-705; 6-fluoro-3-hydroxy-2-pirazinecarboxamide) is another purine analog. It shows broad-spectrum antiviral activity towards RNA viruses, including influenza virus, rhinovirus, RSV, and Ebola virus, but is not effective against DNA viruses. At 10 μg/mL, favipiravir inhibited plaque formation of influenza virus A/Puerto Rico/8/1934(H1N1) for over 72 hours, compared to a 48-hour inhibition when treating with oseltamivir at the same concentration. The viral yields were lower with 1 μg/mL favipiravir treatment compared to the same concentration of oseltamivir at an MOI from 0.0001 to 1. The survival rates of mice challenged with over 1000-fold of the LD_{50} of the influenza virus were also significantly higher under favipiravir treatment compared to the same dosage of oseltamivir, at both 200 and 400 mg/kg/day (19). Similar to ribavirin, favipiravir can be converted to T-705-4-ribofuranosyl-5-triphosphate (T-705RTP) and incorporated into RNA in place of guanosine or adenosine but primarily inhibits viral RNA synthesis as a chain terminator. T-705RTP has a higher affinity for the influenza virus RdRp than GTP. When T-705RTP is incorporated into the nascent RNA strand, it inhibits the subsequent incorporation of nucleotides and inhibits RNA strand elongation (20). The chain termination function of favipiravir can partially explain its lower inhibitory concentration for influenza virus RdRp activity compared to ribavirin (21). Favipiravir primarily acts as a guanine analog and secondarily as an adenine analog, causing the accumulation of transitions (22). Mutational bias appears in favor of G-to-U, C-to-U, A-to-G, U-to-C, and G-to-A both in cell cultures and *in vivo* when applying favipiravir to dengue virus, foot-and-mouth disease virus, hepatitis C virus, influenza virus, etc. (23-27).

5-Fluorouracil (2,4-dihydroxy-5-fluoropyrimidine) is an analog of uracil. It can be converted to various active metabolites and incorporated into both RNA and DNA. Therefore, 5-fluorouracil is not only used as an antiviral drug but also to treat cancer. Due to structural

similarity, 5-fluorouracil can enter some metabolic pathways in replacement of uracil. However, the fluor (F) atom of 5-fluorouracil would decrease the length of the nearby C=O bond and the negative charge on the oxygen atom, altering the capacity of the oxygen atom to form hydrogen bonds with other molecules (28). A 5-fluorouracil derivative, 5-fluorodeoxyuridine monophosphate (FdUMP), inhibits an essential enzyme for DNA synthesis, thymidylate synthase, through such a mechanism. The incorporation of 5-fluorouridine triphosphate (FUTP) into cellular RNA also disrupts processes such as pre-rRNA maturation, post-transcriptional modification of tRNA, and the splicing of pre-mRNA (29). As a mutagen, fluorouridine monophosphate (FUMP) can pair with both A and G (30). If the original nucleotide is A, FUMP can induce an A-to-G transition after one round of replication; if the original nucleotide is U, then FUMP can induce a U-to-C transition after two rounds of replication. The antiviral effect of 5-fluorouracil results from the combination of inhibitory and mutagenic activities.

5-Azacytidine (4-amino-1-(beta-D-ribofuranosyl)-1,3,5-triazine-2(1H)-one) is a ribonucleoside analog that has been shown to inhibit human immunodeficiency virus type 1 (HIV-1) since 1990. 5-Azacytidine was later shown to be active against human immunodeficiency virus type 2 (HIV-2) (31), foot-and-mouth disease virus (32), and hepatitis B virus (33). Lethal mutagenesis is its major antiviral mechanism. The ribonucleotide form of 5-azacytidine, 5-aza-CTP, can be incorporated into viral RNA, and its 2′-deoxy form, 5-aza-2′-deoxycytidine triphosphate (5-aza-dCTP), can be incorporated into viral DNA through reverse transcription. The latter form has been shown to be the major contributor to mutagenesis (34, 35). Treatment of 5-azacytidine to HIV-1 increases the average number of mutations per nucleotide by approximately 2.3-fold and dramatically enriches the G-C transversion mutation (36).

β-D-N^4-hydroxycytidine (NHC, or EIDD-1931) is a ribonucleoside analog and the active metabolite form of the prodrug, molnupiravir (EIDD-2801). EIDD-1931 was initially developed as a potential treatment for influenza viruses and encephalitic alphaviruses (37) but ultimate was authorized as an orally available antiviral drug against SARS-CoV-2 infection (38). Unlike other RNA viruses, the coronavirus RdRp has an exonucleolytic proofreading activity, which can repair the aborted elongation induced by many chain-terminators. Therefore, lethal mutagenesis became an important strategy to combat SARS-CoV-2. Molnupiravir does not terminate RNA synthesis. The active form of molnupiravir is NHC triphosphate (MTP). MTP competes most effectively with cytidine triphosphate (CTP) for incorporation into RNA. It can also be incorporated instead of uridine triphosphate (UTP), but much less efficiently (39). When MTP is incorporated into RNA, it forms stable base pairs with G or A at the RdRp active site, which leads to G-to-A and C-to-U transition mutations (40). A-to-G and U-to-C mutations are also observed but in much lower frequencies (41, 42).

Many studies have examined the antiviral efficacy of mutagenic drugs. The efficacy of mutagenic drugs is often measured as the concentration that inhibits viral growth. The commonly used measurements are half-maximal effective concentration (EC_{50}), or, the half-maximal inhibition concentration (IC_{50}), and the concentrations for 1-, 2-, or 3-$\log_{10}$ reduction in viral growth. I summarize the findings regarding the concentrations of drugs required to inhibit viral growth from various literature in Appendix B for better comparison. However, since the measurements were performed on various cells, obtained through distinct assays, and used different readouts, it is difficult to compare the activity of mutagenic drugs across studies.
To identify important PB1 residues and amino acids to replicative fidelity while studying the replicative fidelity of each class of mutation separately, we plan to probe the variant library with

mutagenic drugs that target different classes of mutations. In this chapter, I determined the inhibitory concentrations of five mutagenic drugs – ribavirin, favipiravir, 5-fluorouracil, 5-azacytidine, and molnupiravir. These experiments serve as a pilot experiment for the deep mutational scanning of influenza replicative fitness, aiming to find an appropriate concentration for each drug as an intermediate selective pressure for variants with normal-to-lower replicative fidelity.

3.2 Materials and Methods

Cells, virus, media, and drugs

MDCK-SIAT1-TMPRSS2 and A549 cells were used in mutagen sensitivity assays and maintained in D10 and A549 growth media, respectively. As described in previous sections, IGM+ media was used for viral infection in MDCK-SIAT1-TMPRSS2 cells and A549 viral media for infection in A549 cells. WNM media was used for $TCID_{50}$ assays. Influenza A virus H1N1 (A/WSN/1933) was used for all infections. All incubations were done at 37°C and 5% CO_2 in a humidified incubator.

The inhibition kinetics of five mutagenic drugs were tested in the study. To prepare the working solutions, ribavirin (1-β-D-Ribofuranosyl-1,2,4-triazole-3-carboxamide) (Sigma-Aldrich, R9644) was dissolved in PBS at 100 mM. 5-fluorouracil (2,4-Dihydroxy-5-fluoropyrimidine) (Sigma-Aldrich, F6627) and 5-azacytidine (4-Amino-1-(β-D-ribofuranosyl)-1,3,5-triazine-2(1H)-one) (Sigma-Aldrich, A2385) were dissolved in dimethyl sulfoxide (DMSO) at 100 mM. Favipiravir (T-705, 6-Fluoro-3-hydroxy-2-pyrazinecarboxamide, 6-fluoro-3,4-dihydro-3-oxo-2-pyrazinecarboxamide) (Selleckchem, S7975) came from the manufacturer at 10 mM in DMSO. Molnupiravir (EIDD-2801, (2R,3S,4R,5R)-3,4-Dihydroxy-5-(4-(hydroxyamino)-2-oxopyrimidin-1(2H)-yl)tetrahydrofuran-2-yl)methyl isobutyrate) (Sigma

Aldrich, SML2873) was dissolved in DMSO at 10 mM. All working solutions were aliquoted and stored at -20℃.

Mutagenic drug concentrations

The ability of five mutagenic drugs to inhibit viral growth was evaluated. For each drug, six concentrations were picked based on existing research regarding their effective concentrations or inhibition concentrations on different cells. The examined concentrations for MDCK-SIAT1-TMPRSS2 and A549 cells are listed in the following tables. All units are in μM. The assay was performed in A549 cells twice, where the concentrations examined in the second experiment were slightly different based on the results of the first experiment.

- **A549**

Ribavirin	5-Azacytidine	5-Fluorouracil	Favipiravir	Molnupiravir
40	30	80	80	50
30	20	60	40	20
20	10	40	20	10
10	5	20	10	5
5	3	10	5	3
0	0	0	0	0

- **MDCK**

Ribavirin	5-Azacytidine	5-Fluorouracil	Favipiravir	Molnupiravir
40	30	100	80	50
30	25	80	40	20

20	20	60	20	10
10	10	40	10	5
5	5	20	3	2
0	0	0	0	0

Mutagen sensitivity assay

On MDCK-SIAT1-TMPRSS2 cells

Twelve-well plates were seeded with 1×10^5 MDCK-SIAT1-TMPRSS2 cells per well in 1 mL of D10 media. Twenty-four hours later, I aspirated the old D10 media and washed the cell monolayer with pre-warmed PBS. One mL of IGM+ media containing mutagenic drugs of different concentrations was added to each well, with the same concentration within each row. The cells were incubated with mutagens for 3 hours before being infected with 2.5×10^3 PFU viral particles at an MOI of 0.01 in 500 μL of IGM+ containing drug of the same concentration. After one hour of incubation, the media containing the virus was aspirated, and the cells were washed again with pre-warmed PBS. One mL of IGM+ containing drugs of corresponding concentrations was added back to each well, and the plates were incubated for 24 hours.

On A549 cells

Twelve-well plates were seeded with 2×10^5 A549 cells per well in 1 mL of A549 growth media. Twenty-four hours later, I gently took out the overnight A549 growth media with a P1000 pipette and washed the cell monolayer with pre-warmed PBS. Then I added back 1 mL of A549 viral media containing mutagenic drugs of different concentrations but without TPCK-trypsin to each well. Four replicate wells in the same row contain drugs of the same concentration. After 3-

hour incubation, I removed the A549 viral media without TPCK-trypsin and infected each well with 2.5×10^3 PFU viral particles at an MOI of 0.01 in 500 μL of A549 viral media with 4 μg/mL TPCK-trypsin containing drugs. After one hour of incubation, I removed the media containing the virus with a P1000 pipette, and slowly washed the cells with pre-warmed PBS. One mL of A549 viral media without TPCK-trypsin, containing drugs of corresponding concentrations, was added back to each well, and the plates were incubated for 24 hours.

Twenty-four hours after infection, I collected the cell-free supernatant by centrifuging for 5 min at 3,000 ×g and added 0.5% glycerol followed by snap-freezing the samples with dry ice and ethanol. All samples were stored at -80°C for long-term storage. When measuring the viral titers, 100 μL of the supernatant sample was taken from each of the four replicate wells of the same drug concentration and combined as the representation sample. The viral titers were then measured by $TCID_{50}$ assay as described in the previous sections, with wells scored for CPE on day four.

3.3 Results and Discussion

To examine mutagenic drugs' antiviral activity against the influenza virus, I performed mutagen sensitivity assays with ribavirin, 5-azacytidine, 5-fluorouracil, favipiravir, and EIDD-1931 (Figure 3.2). Without drug treatment, influenza virus A/WSN/1933 (H1N1) on MDCK-SIAT1-TMPRSS2 cells can reach a titer of over 10^8 $TCID_{50}$/mL. 5-Azacytidine showed the strongest inhibitory activity, with a 2-$\log_{10}$ reduction concentration at less than 5 μM. The concentration inducing 2-$\log_{10}$ growth reduction for ribavirin was between 5 to 10 μM. Favipiravir, EIDD, and 5-fluorouracil are less active, with 2-$\log_{10}$ reduction concentrations slightly above 20 μM, between 20 to 30 μM, and slightly lower than 40 μM, respectively.

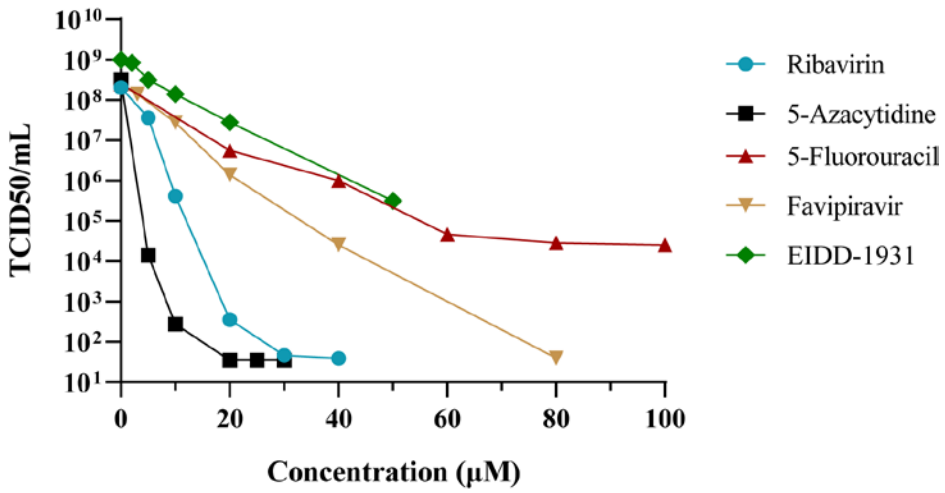


Figure 3.2 Antiviral efficacy of mutagenic drugs on WSN33 in MDCK infection. Each datapoint represents a $TCID_{50}$ measurement of four replicate infections combined.

I repeated the sensitivity assays twice for each mutagenic drug on A549 cells. Since A549 cells are less adherent than MDCK cells, it was difficult to maintain an intact monolayer of cells during the infection. When the intact monolayer was well preserved, the virus titer reached 10^7 $TCID_{50}$/mL without drug treatment (Figure 3.3A); when the monolayer was disrupted, the viral titer was 2- to 4×10^5 $TCID_{50}$/mL (Figure 3.3B). Despite the different starting titers, inhibition curves exhibited similar trends for infections in intact or disrupted cell monolayers (Figure 3.4).

Comparing infections in MDCK or A549 cells, 5-azacytidine and EIDD-1931 showed similar inhibition curves, while ribavirin, favipiravir, and 5-fluorouracil had more dramatic differences. WSN33 replicates more slowly in A549 cells. Although I tested 5-azacytidine at lower concentrations, it induced a 2-log_{10} reduction in viral growth at 2 μM on A549 cells. The concentration of EIDD for 2-log_{10} reduction on A549 cells was also slightly above 20 μM. On the other hand, the reduction in WSN33 growth was much less impacted by ribavirin or favipiravir during A549 infections. Neither drug was able to induce a 2-log_{10} reduction in viral growth at their highest tested concentration, 40 and 80 μM, respectively. 5-Fluorouracil got close

to reaching the 2-$\log_{10}$ reduction in viral growth on A549 at its highest tested concentration, 100 μM. However, this concentration is greatly higher than that of MDCK-SIAT1-TMPRSS2 infection.

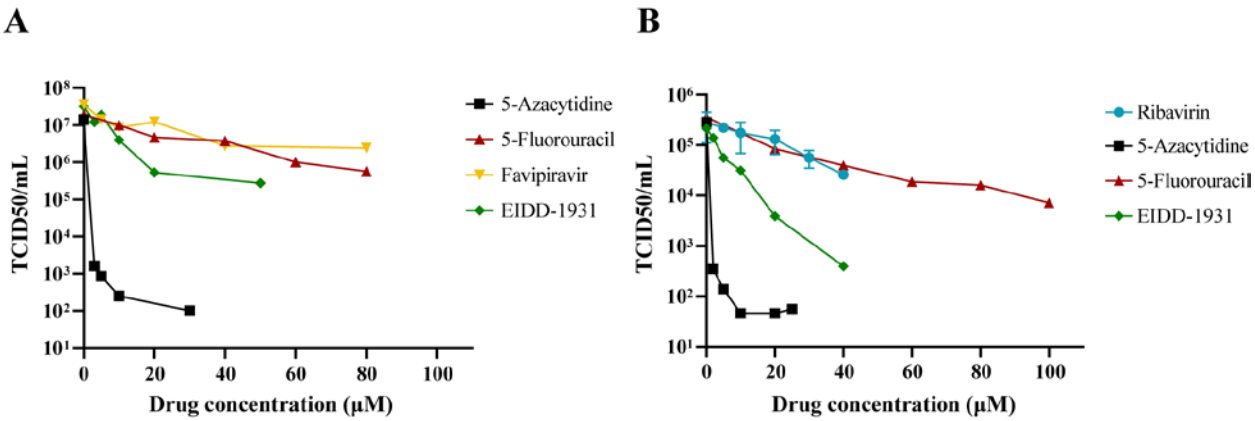


Figure 3.3 Antiviral efficacy of mutagenic drugs in A549 infections. Each datapoint represents a $TCID_{50}$ measurement of four replicate infections combined. (**A**) shows the viral yield when the cell monolayer was well preserved, while (**B**) shows the viral yield when the cell monolayer was disrupted.

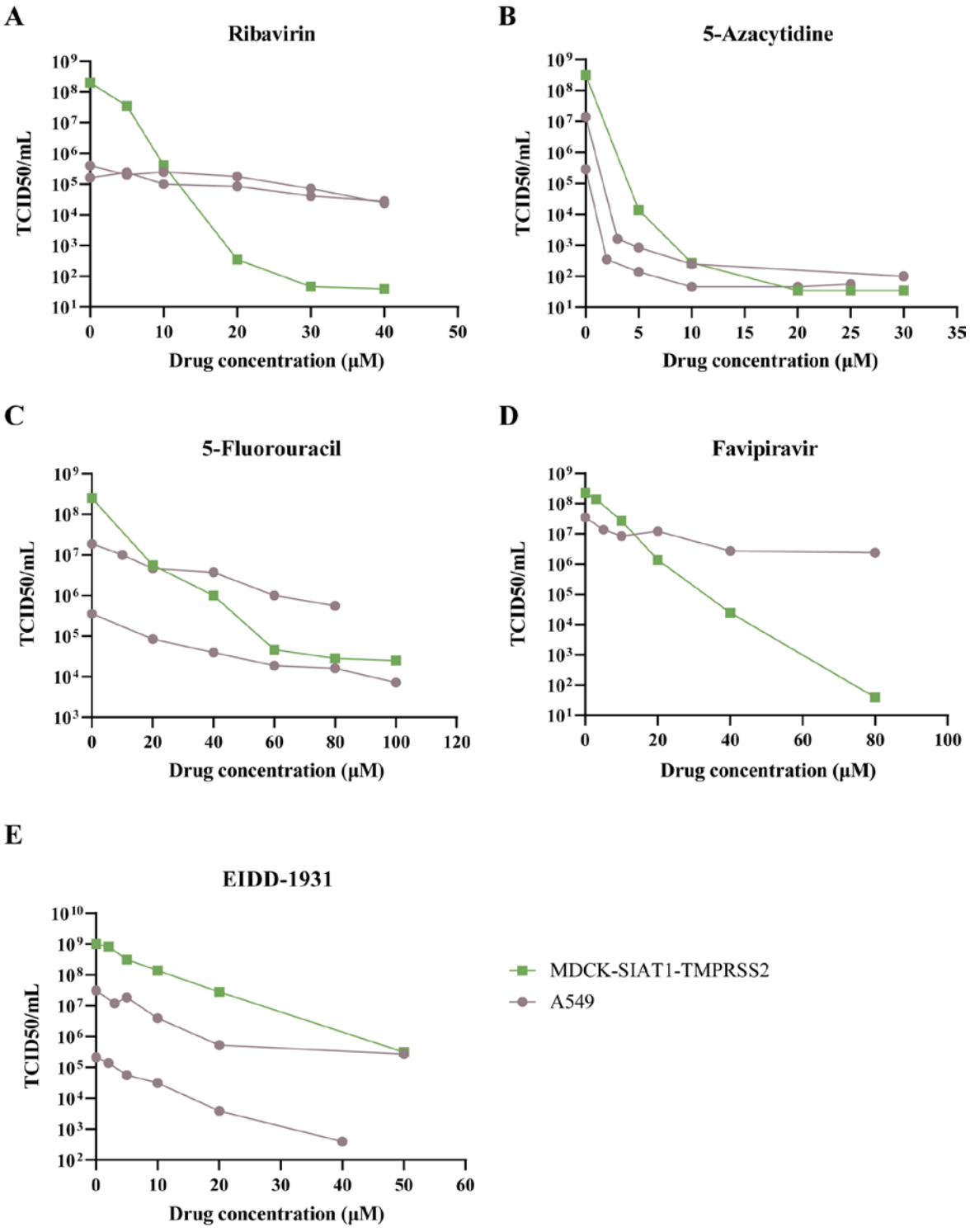


Figure 3.4 Comparison of antiviral efficacy of mutagenic drugs in MDCK and A549 infections. Each datapoint represents a $TCID_{50}$ measurement of four replicate infections in the same experiment, combined. Different lines represent different experiments.

Two factors might contribute to the variation in inhibition by ribavirin, favipiravir, and 5-fluorouracil between infections on MDCK-SIAT1-TMPRSS2 cells and A549 cells: influenza virus's replication kinetics in the two cell lines and toxicity of the drugs to the cells. In MDCK cells, influenza viruses quickly enter the exponential growth phase and keep growing until reaching a plateau of 6-$\log_{10}$ expansion in RNA copy numbers at 48~72 hours post-infection. In A549 cells, influenza viruses, especially influenza A viruses, only have a 2- to 3-$\log_{10}$ expansion in RNA copies during a 96-hour infection, and there is no clear distinction between the growth phase and plateau phase (43). Some drugs, such as ribavirin and favipiravir, inhibit viral growth through multiple mechanisms and target different steps of genome replication. Besides lethal mutagenesis, ribavirin reduces the cellular synthesis of guanine nucleotides by inhibiting host inosine monophosphate dehydrogenase, inhibits the 5′ end capping of viral genomic RNAs, and was found to block the elongation of RNA synthesis mediated by Hepatitis C virus RdRp (44, 45, 46); the active form of favipiravir, favipiravir ribofuranosyl-triphosphate (FTP), can form non-productive binding with RdRp, terminating or slowing down the RNA replication depending on virus types (11, 47, 48). Therefore, these drugs might be more impacted by different replication kinetics than others.

Additionally, MDCK and A549 cells may respond differently to a drug, indirectly affecting viral yields. Josset *et al.* also reported the significance of the cellular effect on the drug's antiviral functions (49). I did not have a chance to examine the cytotoxicity of each drug but summarized data from previous studies in Appendix C. Based on existing data, 5-azacytidine exhibited a low 50% cytotoxicity concentration (CC_{50}) for both A549 and MDCK cells, and favipiravir seemed to be well tolerated by both cell types. All drugs except for favipiravir showed cytotoxicity at a lower concentration for A549 than for MDCK cells. These results

highlighted the complexity of virus and cell responses to mutagens as well as the necessity of additional pilot experiments to establish a more consistent and reproducible virus inhibition model before using lethal mutagenesis as a selective pressure for deep mutational scanning. They also raised the possibility that the variant viruses increased in frequency during the passaging under mutagenic drugs may not only be high-fidelity strains. Indeed, although lowering the baseline mutation rates is a common mechanism to combat lethal mutagenesis (4, 50), there are other strategies to establish resistance, including building up overall genetic robustness (51), controlling the affinity and misincorporation to specific nucleosides (52), and even increasing the rate of back, or reverse, mutations (2). As an example, influenza virus with PB1 D27N mutation is resistant to ribavirin because of its better recognition of nucleotide in low concentrations and higher polymerase activity that counteracts ribavirin's inhibition of *de novo* purine synthesis (53).

Overall, this study evaluated the efficacy of five mutagenic drugs regarding inhibiting the growth of influenza A virus on the cells that are commonly used in labs (MDCK) and the cells that are more physiologically relevant (A549). This experiment provides a more controlled and straightforward comparison of mutagenic drug efficacy and highlights the importance of cellular factors on the antiviral effects of different drugs. This work sets a foundation for future research regarding influenza viruses' replicative fidelity, which may be helpful for further explaining the evolutionary trajectory of influenza viruses and predicting how resistance may arise against anti-RdRp drugs.

3.4 References

1. Lin R-W, Chen G-W, Sung H-H, Lin R-J, Yen L-C, Tseng Y-L, Chang Y-K, Lien S-P, Shih S-R, Liao C-L. 2019. Naturally occurring mutations in PB1 affect influenza A virus replication fidelity, virulence, and adaptability. Journal of Biomedical Science 26:55.

2. Pauly MD, Lyons DM, Fitzsimmons WJ, Lauring AS. 2017. Epistatic Interactions within the Influenza A Virus Polymerase Complex Mediate Mutagen Resistance and Replication Fidelity. mSphere 2.

3. Naito T, Shirai K, Mori K, Muratsu H, Ushirogawa H, Ohniwa RL, Hanada K, Saito M. 2019. Tyr82 Amino Acid Mutation in PB1 Polymerase Induces an Influenza Virus Mutator Phenotype. Journal of Virology 93.

4. Cheung PPH, Watson SJ, Choy K-T, Fun Sia S, Wong DDY, Poon LLM, Kellam P, Guan Y, Malik Peiris JS, Yen H-L. 2014. Generation and characterization of influenza A viruses with altered polymerase fidelity. Nature Communications 5.

5. Smith EC, Sexton NR, Denison MR. 2014. Thinking Outside the Triangle: Replication Fidelity of the Largest RNA Viruses. Annual Review of Virology 1:111–132.

6. Li C, Wang H, Yuan T, Woodman A, Yang D, Zhou G, Cameron CE, Yu L. 2018. Foot-and-mouth disease virus type O specific mutations determine RNA-dependent RNA polymerase fidelity and virus attenuation. Virology 518:87–94.

7. Graham RL, Becker MM, Eckerle LD, Bolles M, Denison MR, Baric RS. 2012. A live, impaired-fidelity coronavirus vaccine protects in an aged, immunocompromised mouse model of lethal disease. Nature Medicine 18:1820–1826.

8. Pfeiffer JK, Kirkegaard K. 2005. Increased Fidelity Reduces Poliovirus Fitness and Virulence under Selective Pressure in Mice. PLoS Pathogens 1:e11.

9. Naito T, Mori K, Ushirogawa H, Takizawa N, Nobusawa E, Odagiri T, Tashiro M, Ohniwa RL, Nagata K, Saito M. 2017. Generation of a Genetically Stable High-Fidelity Influenza Vaccine Strain. Journal of Virology 91.

10. Sanjuán R, Moya A, Elena SF. 2004. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. Proceedings of the National Academy of Sciences 101:8396–8401.

11. Shiraki K, Daikoku T. 2020. Favipiravir, an anti-influenza drug against life-threatening RNA virus infections. Pharmacology & Therapeutics 209:107512.

12. Pauly MD, Procario MC, Lauring AS. 2017. A novel twelve class fluctuation test reveals higher than expected mutation rates for influenza A viruses. eLife 6.

13. Ewald B, Sampath D, Plunkett W. 2008. Nucleoside analogs: molecular mechanisms signaling cell death. Oncogene 27:6522–6537.

14. Patterson JL, Fernandez-Larsson R. 1990. Molecular Mechanisms of Action of Ribavirin. Clinical Infectious Diseases 12:1139–1146.

15. Fernandez H, Banks GB, Smith E. 1986. Ribavirin: A clinical overview. European Journal of Epidemiology 2:1–14.

16. Moreno H, Gallego I, Sevilla N, Torre JC de la, Domingo E, Martín V. 2011. Ribavirin Can Be Mutagenic for Arenaviruses. Journal of Virology 85:7246–7255.

17. Te HS, Randall G, Jensen DM. 2007. Mechanism of action of ribavirin in the treatment of chronic hepatitis C. Gastroenterology & hepatology 3:218–25.

18. Rankin JT, Eppes SB, Antczak JB, Joklik WK. 1989. Studies on the mechanism of the antiviral activity of ribavirin against reovirus. Virology 168:147–158.

19. Takahashi K, Furuta Y, Fukuda Y, Kuno M, Kamiyama T, Kozaki K, Nomura N, Egawa H, Minami S, Shiraki K. 2003. In Vitroandin VivoActivities of T-705 and Oseltamivir against Influenza Virus. Antiviral Chemistry and Chemotherapy 14:235–241.

20. Sangawa H, Komeno T, Nishikawa H, Yoshida A, Takahashi K, Nomura N, Furuta Y. 2013. Mechanism of Action of T-705 Ribosyl Triphosphate against Influenza Virus RNA Polymerase. Antimicrobial Agents and Chemotherapy 57:5202–5208.

21. Furuta Y, Takahashi K, Kuno-Maekawa M, Sangawa H, Uehara S, Kozaki K, Nomura N, Egawa H, Shiraki K. 2005. Mechanism of Action of T-705 against Influenza Virus. Antimicrobial Agents and Chemotherapy 49:981–986.

22. Goldhill DH, Pinky Langat, Xie H, Galiano M, Miah S, Kellam P, Zambon M, Lackenby A, Barclay W. 2019. Determining the Mutation Bias of Favipiravir in Influenza Virus Using Next-Generation Sequencing. Journal of Virology 93.

23. Qiu L, Patterson SE, Laurent Bonnac, Geraghty R. 2018. Nucleobases and corresponding nucleosides display potent antiviral activities against dengue virus possibly through viral lethal mutagenesis. PLOS Neglected Tropical Diseases 12:e0006421–e0006421.

24. Isabel A, Moreno E, Perales C, Domingo E. 2017. Favipiravir can evoke lethal mutagenesis and extinction of foot-and-mouth disease virus. Virus Research 233:105–112.

25. de Ávila AI, Gallego I, Soria ME, Gregori J, Quer J, Esteban JI, Rice CM, Domingo E, Perales C. 2016. Lethal Mutagenesis of Hepatitis C Virus Induced by Favipiravir. PLoS ONE 11.

26. Baranovich T, Wong S-S ., Armstrong J, Marjuki H, Webby RJ, Webster RG, Govorkova EA. 2013. T-705 (Favipiravir) Induces Lethal Mutagenesis in Influenza A H1N1 Viruses In Vitro. Journal of Virology 87:3741–3751.

27. Guedj J, Piorkowski G, Jacquot F, Madelain V, Nguyen THT, Rodallec A, Gunther S, Carbonnelle C, Mentré F, Raoul H, de Lamballerie X. 2018. Antiviral efficacy of favipiravir against Ebola virus: A translational study in cynomolgus macaques. PLOS Medicine 15:e1002535.

28. Agudo R, Arias A, Domingo E. 2009. 5-fluorouracil in lethal mutagenesis of foot-and-mouth disease virus. Future Medicinal Chemistry 1.

29. Longley DB, Harkin DP, Johnston PG. 2003. 5-fluorouracil: mechanisms of action and clinical strategies. Nature reviews Cancer 3:330–8.

30. Ruiz-Jarabo CM, Ly C, Domingo E, Torre JC de la. 2003. Lethal mutagenesis of the prototypic arenavirus lymphocytic choriomeningitis virus (LCMV). Virology 308:37–47.

31. Beach LB, Rawson JM, Kim B, Patterson SE, Mansky LM. 2014. Novel inhibitors of human immunodeficiency virus type 2 infectivity. Journal of General Virology 95:2778–2783.

32. Sierra S, DávilaM, Lowenstein PR, Domingo E. 2000. Response of Foot-and-Mouth Disease Virus to Increased Mutagenesis: Influence of Viral Load and Fitness in Loss of Infectivity. Journal of Virology 74:8316–8323.

33. Wang Y, Hao J, Liu X, Wang H, Zeng X, Yang J, Li L, Kuang X, Zhang T. 2016. The mechanism of apoliprotein A1 down-regulated by Hepatitis B virus. Lipids in Health and Disease 15.

34. Rawson JM, Daly MB, Xie J, Clouser CL, Landman SR, Reilly C, Laurent Bonnac, Kim B, Patterson S, Mansky LM. 2016. 5-Azacytidine Enhances the Mutagenesis of HIV-1 by Reduction to 5-Aza-2′-Deoxycytidine. Antimicrobial Agents and Chemotherapy 60:2318–2325.

35. Roth M, McDaniel YZ, Daly MB, Talledge N, Greggs WM, Patterson S, Kim B, Mansky LM. 2021. Distinct Antiretroviral Mechanisms Elicited by a Viral Mutagen. Journal of Molecular Biology 433:167111–167111.

36. Dapp MJ, Clouser CL, Patterson S, Mansky LM. 2009. 5-Azacytidine Can Induce Lethal Mutagenesis in Human Immunodeficiency Virus Type 1. Journal of Virology 83:11950–11958.

37. Agostini ML, Pruijssers AJ, Chappell JD, Gribble J, Lu X, Andres EL, Bluemling GR, Lockwood MA, Sheahan TP, Sims AC, Natchus MG, Saindane M, Kolykhalov AA, Painter GR, Baric RS, Denison MR. 2019. Small-Molecule Antiviral β-D-N4-Hydroxycytidine Inhibits a Proofreading-Intact Coronavirus with a High Genetic Barrier to Resistance. Journal of Virology 93.

38. Sheahan TP, Sims AC, Zhou S, Graham RL, Pruijssers AJ, Agostini ML, Leist SR, Schäfer A, Dinnon KH, Stevens LJ, Chappell JD, Lu X, Hughes TM, George AS, Hill CS, Montgomery

SA, Brown AJ, Bluemling GR, Natchus MG, Saindane M. 2020. An orally bioavailable broad-spectrum antiviral inhibits SARS-CoV-2 in human airway epithelial cell cultures and multiple coronaviruses in mice. Science Translational Medicine 12:eabb5883.

39. Kabinger F, Stiller C, Schmitzová J, Dienemann C, Kokic G, Hillen HS, Höbartner C, Cramer P. 2021. Mechanism of molnupiravir-induced SARS-CoV-2 mutagenesis. Nature Structural & Molecular Biology 28:740–746.

40. Gordon CJ, Tchesnokov EP, Schinazi RF, Götte M. 2021. Molnupiravir promotes SARS-CoV-2 mutagenesis via the RNA template. Journal of Biological Chemistry 297:100770.

41. Malone B, Campbell EA. 2021. Molnupiravir: coding for catastrophe. Nature Structural & Molecular Biology 28:706–708.

42. Sanderson T, Hisner R, Donovan-Banfield I, Hartman H, Løchen A, Peacock TP, Ruis C. 2023. A molnupiravir-associated mutational signature in global SARS-CoV-2 genomes. medRxiv. https://doi.org/10.1101/2023.01.26.23284998. Retrieved 30 October 2023.

43. Hussain AI, Cordeiro M, Sevilla E, Liu J. 2010. Comparison of egg and high yielding MDCK cell-derived live attenuated influenza virus for commercial production of trivalent influenza vaccine: In vitro cell susceptibility and influenza virus replication kinetics in permissive and semi-permissive cells. Vaccine 28:3848–3855.

44. Graci JD, Cameron CE. 2005. Mechanisms of action of ribavirin against distinct viruses. Reviews in Medical Virology 16:37–48.

45. Maag D, Castro C, Hong Z, Cameron CE. 2001. Hepatitis C Virus RNA-dependent RNA Polymerase (NS5B) as a Mediator of the Antiviral Activity of Ribavirin. Journal of Biological Chemistry 276:46094–46098.

46. Vo NV, Young K-C, Lai MMC. 2003. Mutagenic and inhibitory effects of ribavirin on hepatitis C virus RNA polymerase. Biochemistry 42:10462–10471.

47. Rocha-Pereira J, Jochmans D, Dallmeier K, Leyssen P, Nascimento MSJ, Neyts J. 2012. Favipiravir (T-705) inhibits in vitro norovirus replication. Biochemical and Biophysical Research Communications 424:777–780.

48. Naydenova K, Muir KW, Wu L-F, Zhang Z, Coscia F, Peet MJ, Castro-Hartmann P, Qian P, Sader K, Dent K, Kimanius D, Sutherland JD, Löwe J, Barford D, Russo CJ. 2021. Structure of the SARS-CoV-2 RNA-dependent RNA polymerase in the presence of favipiravir-RTP. Proceedings of the National Academy of Sciences 118:e2021946118.

49. Josset L, Textoris J, Loriod B, Ferraris O, Moules V, Lina B, Nguyen C, Diaz J-J, Rosa-Calatrava M. 2010. Gene Expression Signature-Based Screening Identifies New Broadly Effective Influenza A Antivirals. PLOS ONE 5:e13169–e13169.

50. Pfeiffer JK, Kirkegaard K. 2003. A single mutation in poliovirus RNA-dependent RNA polymerase confers resistance to mutagenic nucleotide analogs via increased fidelity. Proceedings of the National Academy of Sciences 100:7289–7294.

51. Pereira-Gomez M, Sanjuan R. 2014. Delayed Lysis Confers Resistance to the Nucleoside Analogue 5-Fluorouracil and Alleviates Mutation Accumulation in the Single-Stranded DNA Bacteriophage X174. Journal of Virology 88:5042–5049.

52. Agudo R, Ferrer-Orta C, Arias A, de la Higuera I, Perales C, Perez-Luque R, Verdaguer N, Domingo E. 2010. A Multi-Step Process of Viral Adaptation to a Mutagenic Nucleoside Analogue by Modulation of Transition Types Leads to Extinction-Escape. PLOS Pathogens 6:e1001072–e1001072.

53. Nguyen TB, Wakai C, Kawaguchi A, Nagata K. 2014. Involvement of the N-terminal portion of influenza virus RNA polymerase subunit PB1 in nucleotide recognition. Biochemical and Biophysical Research Communications 443:975–979.

Chapter 4 Discussion

Understanding the evolution of viruses is a fundamental topic of virology. It provides us with the implications on the prophylaxis and treatment of viruses as pathogens. With advanced technologies to manipulate viral genomes and high-throughput sequencing, we are no longer satisfied with studying the mutations that appeared naturally: the evolutionary trajectory of a virus or a strain reveals an individual path chosen by nature, yet we want to explore the massive alternative possibilities, which will become our knowledge basis to interpret the significance of future mutations and to rationally design countermeasures. In this work, we want to examine the two most important phenotypes of influenza polymerase – replicative fitness and fidelity – aiming to uncover the evolutionary constraints that prevent the influenza virus from taking those alternative paths. We performed deep mutational scanning on the RdRp subunit of influenza virus polymerase and created a library of single amino acid variant viruses covering 95.4% of all possible amino acid substitutions on that protein. We were able to calculate the fitness of 84% of all possible amino acid substitutions and validate them by comparing the measurements from pairwise competition assays with the wildtype strain. We found that many beneficial mutations revealed by deep mutational scanning had also appeared in the natural evolution of influenza virus PB1, and most of those that had not appeared in nature were not accessible by single nucleotide change. In parallel, we characterized the mutational tolerance of each site on influenza virus PB1. We found that the evolutionary constraints on PB1 could be narrowed down to specific sites, and maintaining interactions with other molecules might be one of the most

important constraints. This work contributes to the overall understanding of the RNA virus polymerase by providing a comprehensive map to link genotype and phenotype. In this chapter, I will discuss some lingering questions and possible future directions that can maximize the findings of this work.

4.1 Future Directions of the Project

We calculated the fitness of variant viruses as the enrichment ratio before and after passages. This is a straightforward method and not computationally demanding. However, the accuracy of this method is obscured by the mutational hotspots during the preparation for sequencing, especially when the sequencing coverage of a codon is low. To ensure the rigor of the fitness calculation, we filtered out 11% of the amino acid substitutions that might be under the impact of artifact mutations. A more finely tuned method to calculate fitness is to use Bayesian inference (1). While the enrichment ratio method draws an empirical, hard line to define that a certain codon count is too heavily impacted by error, the Bayesian inference method sets up variables for the rates of error in pre- and post-selection samples. The error rates can be estimated with the codon counts in the wild type controls. The priors necessary for the Bayesian inference can come from a null assumption, such as all possible amino acids having equal preferences. The preferences of each amino acid at every site can then be inferred by the Markov chain Monte Carlo (MCMC) method. It is worth noticing that since we used only 32 codons (rather than all 64 possible codons) to generate all amino acid substitutions during library creation, either the priors need to be modified accordingly, or the preferences need to be inferred at the codon level and rescaled before converting into amino acid preferences. The Bayesian inference method is most beneficial when the sequencing depth is not very high. By estimating to

what extent a variant codon might have come from an error, the Bayesian inference method can preserve more information while calculating fitness accurately.

We found that one of the most important evolutionary constraints on influenza virus RdRp is the requirement to maintain proper interactions with other molecules. In this dissertation, I explored the constraints brought by the interactions between PB1 and different types of RNAs. Another interesting topic would be to see how the interactions between PB1, PB2, and PA shape the evolution of influenza virus polymerase. PB1 interacts with PB2 and PA extensively when performing polymerase functions. During the transcription process, 29% of PB1 residues interact with PA and 24% interact with PB2 (calculated by PDBePISA, data not shown). The polymerase proteins also show strong signs of co-evolution (2, 3). Mismatch of the polymerase subunits from different strains often causes non-functional RNP complexes (4, 5). Variant viruses with a mutation on one polymerase subunit often have compensatory mutations in other subunits to maintain fitness (6, 7). Consolidating the deep mutational scanning data on PB1 (this project), PB2 (8), and PA (9) can reveal interactions that are essential for polymerase functions and viral fitness (Figure 4.1). Hannon and Bloom (10) have developed an interactive program visualizing the DMS results on a protein's 3D structure, which would be helpful for cross-comparing the mutation-based data from different research. Additionally, we can examine the fitness of deep mutational scanning variants of one polymerase protein when having a mutation of interest in another subunit.

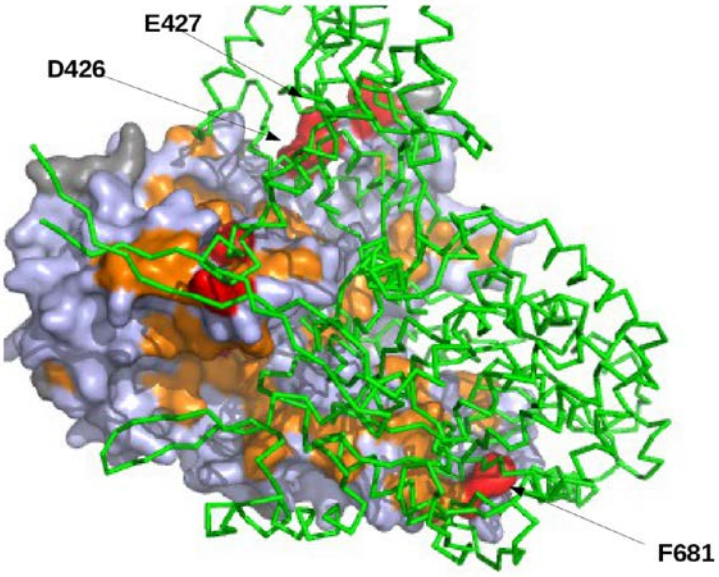


Figure 4.1 Functional residues on PA identified by large-scale mutagenesis interact with PB1. Adapted from Wu et. al., 2015 (9). PA is shown as surface. PB1 is shown in green stick form. Mutations that were individually analyzed are labelled. Predicted functional residues are colored in red; residues that carry significantly deleterious mutations are colored in orange; residues that are not covered in the profiling data are colored in gray.

In this dissertation, we studied the mutational effects and tolerance with the genetic background of A/WSN/1933 but also observed that some beneficial mutations identified by deep mutational scanning exhibited fitness advantages in other strains. We are interested in knowing to what degrees the mutational effects and tolerance are background-dependent and which part of the results apply to other strains of influenza A virus or even other types of influenza viruses. We are currently performing deep mutational scanning of PB1 from an H3N2 strain, A/Aichi/2/1968, which will be the first step for elucidating the background specificity of mutational effects.

To probe the structural basis for nucleotide misincorporation and influenza virus replication fidelity, we plan to perform selections on DMS-generated variant libraries with a panel of drugs that tend to induce different classes of mutations. Selections using different mutagens may identify residues and amino acid substitutions changing the mutation rates of different classes. It is worth noticing that the amino acid preferences in the presence of drug will reflect fitness advantages of three types: general replicative fitness, sensitivity to mutagenesis,

and resistance to other inhibitory mechanisms of the drug (Figure 4.2). High-fidelity strains are included but do not equal to variants that are less sensitive to mutagenesis, since some variants can counteract mutagenesis by increasing the mutation rate of reverse mutations (11). Therefore, the replicative fidelity of residues and amino acids screened out by mutagenic drugs needs to be further validated. We can compare the fitness of a variant with and without drug treatment. Variants showing high fitness with drugs but low fitness without drugs are likely to have a mutation related to replicative fidelity, while variants showing high fitness in both experiments may have a different or additional strategy/strategies to obtain evolutionary advantage. The accurate mutation rate of a variant can be validated by a fluctuation test based on green fluorescent proteins (GFP) (12). This assay inserts a GFP protein into the variant virus genome partly replacing the coding region for hemagglutinin (HA). The inserted GFP has a single nucleotide mutation so that it will only be fluorescent when a reversion mutation is made during viral replication, and the mutation rate of a specific class of nucleotide substitution can be calculated using the null-class model modified based on the classic Luria-Delbrück fluctuation test.

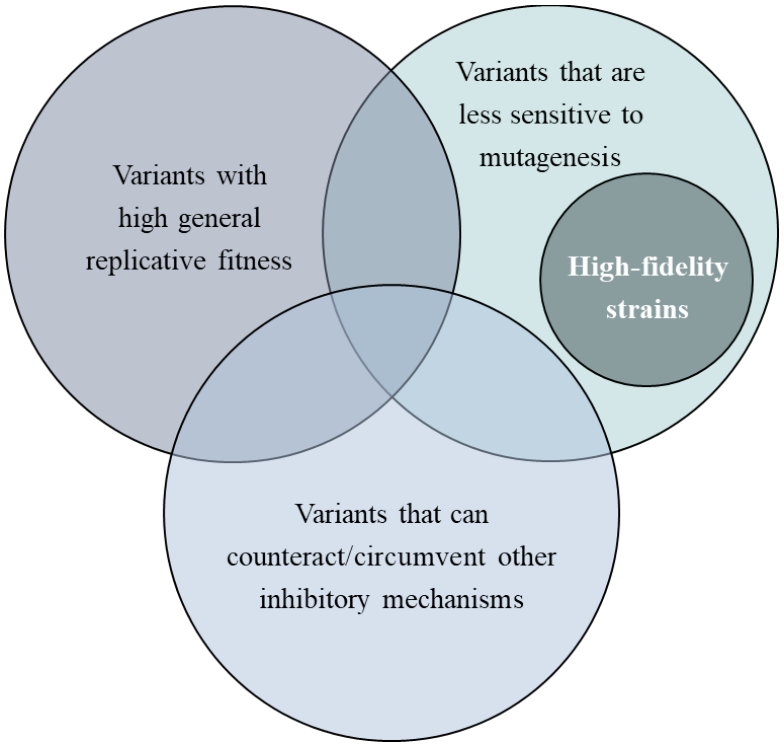


Figure 4.2 Different types of variant viruses exhibiting high fitness under mutagenic drugs. Scheme of possible reasons for a variant to increase in frequency during passaging with mutagens. Only a subset of high-fitness variants would have altered replicative fidelity.

In conclusion, this dissertation performed deep mutation scanning of influenza virus RdRp, identified impactful amino acid substitutions on viral replicative fitness, and characterized the evolutionary constraints and potential of influenza virus polymerase. It provided a comprehensive map of mutational effects and constraints on a RdRp, which will be a valuable resource for future studies of influenza virus evolution and adaptation as well as our genomic surveillance efforts to prevent potential pandemics. This dissertation also examined influenza virus replication under the treatment of various mutagenic drugs, laying a foundation for future research on the mechanisms of nucleotide incorporation and replication fidelity of influenza viruses. The structure of RdRp is not only conserved in influenza viruses. Instead, most RNA viruses possess a polymerase that resembles a cupped right hand and consists of fingers, palm, and thumb subdomains. Unraveling the evolutionary constraints and potential of the influenza

virus RdRp in high resolution will also add to our understanding of the shared features of RNA virus pathogenicity and evolution.

4.2 Broader Applications of Deep Mutational Scanning

This project validated the potential of deep mutational scanning in informing viral surveillance and the rational design of vaccines and anti-viral drugs. Viral surface proteins, such as glycoproteins and envelope proteins, often have critical roles for host cell attachment, membrane fusion, or receptor-mediated entry to the host cells. They are also major targets of the humoral immune response. Deep mutational scanning on the surface proteins can reveal viral evolution on host tropism (13), selections by receptor binding (14), antibody escape (15, 16), and drug resistance (17), all of which are important for viral surveillance. Additionally, deep mutational scanning can proactively assess the likelihood of escape mutations selected by commercially developed antibodies (18, 19, 20), which adds another dimension to evaluating a drug during and after the development pipeline. In vaccine design, deep mutational scanning would be helpful in screening for a vaccine with optimal thermal stability, antigenicity, and immunogenicity.

Viral antigens are recognized by the lymphocytes, activating adaptive immune response; and the epitope is the smallest structural and functional unit for the activation (21). Rationalized selection of candidate antigens is therefore crucial for the development of subunit vaccines with high effectiveness (22). Combined with peptide display assays, deep mutational scanning can help identify viral epitopes at a large scale (23). The control of seasonal influenza viruses has been hampered by the frequent antigenic changes that permit the virus to evade adaptive immune response. The protection against pandemic strains is challenging due to the absence of pre-

existing antibodies against the emerging strains. Therefore, cellular immunity induced by conserved influenza virus antigens has been studied as an alternative vaccination strategy. Compared to traditional vaccines that focus on eliciting neutralizing antibodies, this strategy tries to activate cytotoxic T-lymphocytes that are broadly reactive to most strains in a subtype or even across subtypes (24). The epitopes recognized by the cross-protective cytotoxic T-cells are often located in inner proteins such as NP, M1, PB1, and PB2 (25, 26, 27). While the deep mutational scanning of PB1 may not offer direct help to vaccine design targeting neutralizing antibodies, it may be helpful in the identification of T-cell epitopes that are conserved across strains.

4.3 Technical Difficulties of Deep Mutational Scanning

To realize the potential of deep mutational scanning in viral surveillance and vaccine design, there are several technical difficulties to be resolved.

The major bottleneck for deep mutational scanning to create a variant virus library with high diversity lies in the step of creating viruses from mutagenized nucleic acid materials. For creating influenza viruses, the DNA transfection system using eight plasmids (28) is a reliable and efficient method to generate infectious viral particles in high yield. However, the eight plasmids encoding the eight gene segments are not guaranteed to enter the same cell. The PB1 variant plasmids that did not enter a host cell during the infection period, or entered a cell that did not receive all other seven plasmids would not produce corresponding virus variants. We attempted to saturate the host cells with the seven non-PB1 influenza virus genome segments by infecting the cells with high concentrations of “helper virus” which does not encode functional PB1. However, it was difficult to purge the helper virus after transfection. The helper virus in the variant virus libraries interfered with the replication of polymerase segments for all viruses and disrupted the production of infectious particles during the passage. We partially solved this

problem by performing 36 independent transfection reactions for creating each virus variant library and created three replicate libraries to minimize the randomness of which mutations yielded viruses. In addition, PB1 variants with different degrees of replication defect would have difficulties in generating viral particles even in the library before passage, making the baseline of competition unbalanced. We performed the transfection in 293T cells that had been modified to express the wild type PB1 protein so that the job of producing variant PB1 vRNA could be carried out by a fully functional polymerase during transfection. The methods for creating variant virus libraries vary depending on the type of viral genome. Poliovirus, a positive-strand RNA virus, can be created by *in vitro* transcription followed by mRNA transfection. Our lab was able to create poliovirus libraries with single amino acid mutations on RdRp ($3D^{pol}$) but encountered obstacles in achieving high mutational diversity toward the end of the $3D^{pol}$ gene.

Other genomic features that can affect the generation of variant viruses and the interpretation of the deep mutational scanning results include RNA structures and overlapping open reading frames. Nucleotide substitutions on the low-nucleoprotein binding regions of influenza virus PB1, PB2, and NS can alter their RNA secondary structures and result in reduced viral growth, probably due to compromised packaging (29). Overlapping coding sequences exist in all known virus groups and are particularly frequent in DNA viruses (30). In RNA viruses, overlapping genes can help maintain mutational robustness under high mutation rates (31). Mutations in the overlapping regions may change amino acids in multiple proteins or alter the non-coding region of the second gene, therefore requiring additional design during mutagenesis and mutational frequency calculation.

4.4 References

1. Bloom JD. 2015. Software for the analysis and visualization of deep mutational scanning data. BMC Bioinformatics 16.

2. Chen W, Xu Q, Zhong Y, Yu H, Shu J, Ma T, Li Z. 2017. Genetic variation and co-evolutionary relationship of RNA polymerase complex segments in influenza A viruses. Virology 511:193–206.

3. Arcos S, Han AX, te Velthuis AJW, Russell CA, Lauring AS. 2023. Mutual information networks reveal evolutionary relationships within the influenza A virus polymerase. Virus Evolution 9.

4. Li C, Hatta M, Watanabe S, Neumann G, Kawaoka Y. 2008. Compatibility among Polymerase Subunit Proteins Is a Restricting Factor in Reassortment between Equine H7N7 and Human H3N2 Influenza Viruses. Journal of Virology 82:11880–11888.

5. Iwatsuki-Horimoto K, Hatta Y, Hatta M, Muramoto Y, Chen H, Kawaoka Y, Horimoto T. 2008. Limited compatibility between the RNA polymerase components of influenza virus type A and B. Virus Research 135:161–165.

6. Goldhill DH, te Velthuis AJW, Fletcher RA, Langat P, Zambon M, Lackenby A, Barclay WS. 2018. The mechanism of resistance to favipiravir in influenza. Proceedings of the National Academy of Sciences 115:11613–11618.

7. Octaviani CP, Goto H, Kawaoka Y. 2011. Reassortment between Seasonal H1N1 and Pandemic (H1N1) 2009 Influenza Viruses Is Restricted by Limited Compatibility among Polymerase Subunits. Journal of Virology 85:8449–8452.

8. Soh YS, Moncla LH, Eguia R, Bedford T, Bloom JD. 2019. Comprehensive mapping of adaptation of the avian influenza polymerase protein PB2 to humans. eLife 8:e45079.

9. Wu NC, Olson CA, Du Y, Le S, Tran K, Remenyi R, Gong D, Al-Mawsawi LQ, Qi H, Wu T-T, Sun R. 2015. Functional Constraint Profiling of a Viral Protein Reveals Discordance of Evolutionary Conservation and Functionality. PLOS Genetics 11:e1005310.

10. Hannon WW, Bloom JD. 2023. dms-viz: Structure-informed visualizations for deep mutational scanning and other mutation-based datasets. bioRxiv (Cold Spring Harbor Laboratory) https://doi.org/10.1101/2023.10.29.564578.

11. Pauly MD, Lyons DM, Fitzsimmons WJ, Lauring AS. 2017. Epistatic Interactions within the Influenza A Virus Polymerase Complex Mediate Mutagen Resistance and Replication Fidelity. mSphere 2.

12. Pauly MD, Procario MC, Lauring AS. 2017. A novel twelve class fluctuation test reveals higher than expected mutation rates for influenza A viruses. eLife 6.

13. Setoh YX, Amarilla AA, Peng G, Griffiths RE, Carrera J, Freney ME, Nakayama E, Ogawa S, Watterson D, Modhiran N, Nanyonga FE, Torres FJ, Slonchak A, Periasamy P, Prow NA, Tang B, Harrison JJ, Hobson-Peters J, Cuddihy T, Cooper-White JJ. 2019. Determinants of Zika virus host tropism uncovered by deep mutational scanning. Nature Microbiology 4:876–887.

14. Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KHD, Dingens AS, Navarro MJ, Bowen JE, Tortorici MA, Walls AC, King NP, Veesler D, Bloom JD. 2020. Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. Cell 182:1295-1310.e20.

15. Sourisseau M, Lawrence DA, Schwarz MC, Storrs C, Veit EC, Bloom JD, Evans M. 2019. Deep Mutational Scanning Comprehensively Maps How Zika Envelope Protein Mutations Affect Viral Growth and Antibody Escape. Journal of Virology 93.

16. Wang S, Zhang TH, Hu M, Tang K, Sheng L, Hong M, Chen D, Chen L, Shi Y, Feng J, Qian J, Sun L, Ding K, Sun R, Du Y. 2023. Deep mutational scanning of influenza A virus neuraminidase facilitates the identification of drug resistance mutations in vivo. MSystems https://doi.org/10.1128/msystems.00670-23.

17. Alcantara MC, Higuchi Y, Kirita Y, Matoba S, Hoshino A. 2023. Deep Mutational Scanning to Predict Escape from Bebtelovimab in SARS-CoV-2 Omicron Subvariants. Vaccines 11:711.

18. Tsai K-C, Lee Y-C, Tseng T-S. 2021. Comprehensive Deep Mutational Scanning Reveals the Immune-Escaping Hotspots of SARS-CoV-2 Receptor-Binding Domain Targeting Neutralizing Antibodies. Frontiers in Microbiology 12.

19. Starr TN, Czudnochowski N, Liu Z, Zatta F, Park Y-J, Addetia A, Pinto D, Beltramello M, Hernandez P, Greaney AJ, Marzi R, Glass WG, Zhang I, Dingens AS, Bowen JE, Tortorici MA, Walls AC, Wojcechowskyj JA, De Marco A, Rosen LE. 2021. SARS-CoV-2 RBD antibodies that maximize breadth and resistance to escape. Nature 597:97–102.

20. Radford CE, Schommers P, Gieselmann L, Crawford KHD, Dadonaite B, Yu TC, Dingens AS, Overbaugh J, Klein F, Bloom JD. 2023. Mapping the neutralizing specificity of human anti-HIV serum by deep mutational scanning. Cell Host & Microbe 31:1200-1215.e9.

21. Hu D, Irving AT. 2023. Massively-multiplexed epitope mapping techniques for viral antigen discovery. Frontiers in Immunology 14.

22. Parvizpour S, Pourseif MM, Razmara J, Rafi MA, Omidi Y. 2020. Epitope-based vaccine design: a comprehensive overview of bioinformatics approaches. Drug Discovery Today 25.

23. Davidson E, Doranz BJ. 2014. A high-throughput shotgun mutagenesis approach to mapping B-cell antibody epitopes. Immunology 143:13–20.

24. Effros RB, Doherty PC, Gerhard W, Bennink JR. 1977. Generation of both cross-reactive and virus-specific T-cell populations after immunization with serologically distinct influenza A viruses. Journal of Experimental Medicine 145:557–568.

25. Staneková Z, Varečková E. 2010. Conserved epitopes of influenza A virus inducing protective immunity and their prospects for universal vaccine development. Virology Journal 7.

26. Terajima M, Babon JAB, Co MDT, Ennis FA. 2013. Cross-reactive human B cell and T cell epitopes between influenza A and B viruses. Virology Journal 10.

27. Belz GT, Xie W, Doherty PC. 2001. Diversity of Epitope and Cytokine Profiles for Primary and Secondary Influenza A Virus-Specific CD8+T Cell Responses. The Journal of Immunology 166:4627–4633.

28. Hoffmann E, Neumann G, Kawaoka Y, Hobom G, Webster RG. 2000. A DNA transfection system for generation of influenza A virus from eight plasmids. Proceedings of the National Academy of Sciences of the United States of America 97:6108–6113.

29. Williams GD, Townsend D, Wylie KM, Kim PJ, Amarasinghe GK, Kutluay SB, Boon ACM. 2018. Nucleotide resolution mapping of influenza A virus nucleoprotein-RNA interactions reveals RNA features required for replication. Nature Communications 9.

30. Wright BW, Molloy MP, Jaschke PR. 2022. Overlapping genes in natural and engineered genomes. Nature Reviews Genetics 23:154–168.

31. Fernandes JD, Faust TB, Strauli N, Smith CL, Crosby DA, Nakamura RM, Hernandez RD, Frankel AD. 2016. Functional Segregation of Overlapping Genes in HIV. Cell 167:1762-1773.e12.

Appendices

Appendix A: Primers and Cycling Programs for Barcoded Subamplicon Sequencing Preparation

Primers for PCR0

Rnd0_Fwd	AGCGAAAGCAGGCAAACCATTTGA
Rnd0_Rev	GGCATTTTTTCATGAAGGACAAGCTAAATTCA

Primers for PCR1

Rnd1_Fwd1	CTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNAGCGAAAGCAGGCAAACCATTTGA
Rnd1_Rev1	GGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNACCAGGATGGGATTCCTCAAGGAA
Rnd1_Fwd2	CTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNGATTGTGTATTGGAAGCAATGGCC
Rnd1_Rev2	GGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNCTTAGTCATATTGTCTCTCACTCG
Rnd1_Fwd3	CTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNGAGATCACAACTCATTTTCAGAGAAAGAGA
Rnd1_Rev3	GGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNGAAATTTCAGTGTCCTGAGAATTGGT
Rnd1_Fwd4	CTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNGGCAAATGTTGTAAGGAAGATGATG
Rnd1_Rev4	GGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNCATTCCAGGGCTCAATGATGC
Rnd1_Fwd5	CTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNCGGCCGCTCTTAATAGATGGGACT
Rnd1_Rev5	GGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNGCTGGGAAGCTCCATGCTGAAATT
Rnd1_Fwd6	CTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNTTCTATCGTTATGGGTTTGTTGCC
Rnd1_Rev6	GGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNGTATAAATTTGGGCCTCCGTC
Rnd1_Fwd7	CTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNAAAGCTGGACTGCTGGTCTCC

Rnd1_Rev7	GGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNGTTGCAGCACT TTTGGTACATTTG
Rnd1_Fwd8	CTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNGCCAAAGAGGA ATACTTGAAGATGAA
Rnd1_Rev8	GGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNGGCATTTTTTC ATGAAGGACAAGCTAAATTCA

Primers for PCR2

Rnd2_Fwd1	AATGATACGGCGACCACCGAGATCTACACTCGTGGAGCGACACTCT TTCCCTACACGACGCTCTTCCGATCT
Rnd2_Rev1	CAAGCAGAAGACGGCATACGAGATCGCTCAGTTCGTGACTGGAGTT CAGACGTGTGCTCTTCCGATCT
Rnd2_Fwd2	AATGATACGGCGACCACCGAGATCTACACCTACAAGATAACACTCT TTCCCTACACGACGCTCTTCCGATCT
Rnd2_Rev2	CAAGCAGAAGACGGCATACGAGATTATCTGACCTGTGACTGGAGTT CAGACGTGTGCTCTTCCGATCT
Rnd2_Fwd3	AATGATACGGCGACCACCGAGATCTACACTATAGTAGCTACACTCTT TCCCTACACGACGCTCTTCCGATCT
Rnd2_Rev3	CAAGCAGAAGACGGCATACGAGATATATGAGACGGTGACTGGAGTT CAGACGTGTGCTCTTCCGATCT
Rnd2_Fwd4	AATGATACGGCGACCACCGAGATCTACACACCAGCGACAACACTCT TTCCCTACACGACGCTCTTCCGATCT
Rnd2_Rev4	CAAGCAGAAGACGGCATACGAGATTCGTCTGACTGTGACTGGAGTT CAGACGTGTGCTCTTCCGATCT
Rnd2_Fwd5	AATGATACGGCGACCACCGAGATCTACACCATACACTGTACACTCTT TCCCTACACGACGCTCTTCCGATCT
Rnd2_Rev5	CAAGCAGAAGACGGCATACGAGATGAACATACGGGTGACTGGAGTT CAGACGTGTGCTCTTCCGATCT
Rnd2_Fwd6	AATGATACGGCGACCACCGAGATCTACACTCGGCAGCAAACACTCT TTCCCTACACGACGCTCTTCCGATCT
Rnd2_Rev6	CAAGCAGAAGACGGCATACGAGATAACCATTCTCGTGACTGGAGTT CAGACGTGTGCTCTTCCGATCT
Rnd2_Fwd7	AATGATACGGCGACCACCGAGATCTACACCTAATGATGGACACTCT TTCCCTACACGACGCTCTTCCGATCT
Rnd2_Rev7	CAAGCAGAAGACGGCATACGAGATGGTTGCCTCTGTGACTGGAGTT CAGACGTGTGCTCTTCCGATCT
Rnd2_Fwd8	AATGATACGGCGACCACCGAGATCTACACGGTTGCCTCTACACTCTT TCCCTACACGACGCTCTTCCGATCT
Rnd2_Rev8	CAAGCAGAAGACGGCATACGAGATCTAATGATGGGTGACTGGAGTT CAGACGTGTGCTCTTCCGATCT
Rnd2_Fwd9	AATGATACGGCGACCACCGAGATCTACACCGCACATGGCACACTCT TTCCCTACACGACGCTCTTCCGATCT
Rnd2_Rev9	CAAGCAGAAGACGGCATACGAGATTCGGCCTATCGTGACTGGAGTT CAGACGTGTGCTCTTCCGATCT

Rnd2_Fwd10	AATGATACGGCGACCACCGAGATCTACACGGCGAGATGGACACTCTTTCCCTACACGACGCTCTTCCGATCT
Rnd2_Rev10	CAAGCAGAAGACGGCATACGAGATTTCTATGGTTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Rnd2_Fwd11	AATGATACGGCGACCACCGAGATCTACACAATAGAGCAAACACTCTTTCCCTACACGACGCTCTTCCGATCT
Rnd2_Rev11	CAAGCAGAAGACGGCATACGAGATCCTCGCAACCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Rnd2_Fwd12	AATGATACGGCGACCACCGAGATCTACACTCGTATGCGGACACTCTTTCCCTACACGACGCTCTTCCGATCT
Rnd2_Rev12	CAAGCAGAAGACGGCATACGAGATATGTCGTGGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Rnd2_Fwd13	AATGATACGGCGACCACCGAGATCTACACGTCGATTACAACACTCTTTCCCTACACGACGCTCTTCCGATCT
Rnd2_Rev13	CAAGCAGAAGACGGCATACGAGATCGTATAATCAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Rnd2_Fwd14	AATGATACGGCGACCACCGAGATCTACACAGTGGTCAGGACACTCTTTCCCTACACGACGCTCTTCCGATCT
Rnd2_Rev14	CAAGCAGAAGACGGCATACGAGATGCCATTAGACGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

Cycling program for PCR0

Step	**Temperature (°C)**	**Time (sec)**
1. polymerase activation	95	120
2. denaturing	95	20
	70	1
3. annealing	52	30
4. extension	70	50
Repeat step 2-4	For 22 cycles in total	
5. hold	4	

Cycling program for PCR1

Step	**Temperature (°C)**	**Time(sec)**
1. polymerase activation	95	120
2. denaturing	95	20
	70	1

3. annealing	54	20
4. extension	70	20
Repeat step 2-4	For 9 cycles in total	
5. termination	95	60
6. hold	4	

Cycling program for PCR2

Step	**Temperature (°C)**	**Time(sec)**
1. polymerase activation	95	120
2. denaturing	95	20
	70	1
3. annealing	55	20
4. extension	70	20
Repeat step 2-4	For 24 cycles in total	
5. hold	4	

Appendix B: Inhibitory Concentrations of Mutagenic Drugs

Appendix Table B.1 Inhibition concentrations for influenza viruses (*in vitro* experiments)

Drug	Concentration (µg/mL)	Concentration (µM)	Measurement	Assay	Virus strain	Cell line	Citation
ribavirin		8.1 ± 1.3	EC50	TCID50	HA, NA from A/Aichi/2/68 (H3N2); other segments from A/Puerto Rico/8/1934 (H1N1)	MDCK	1
ribavirin		9.1 ± 1.9	1-log10 reduction	RT-qPCR	HA, NA from A/Aichi/2/68 (H3N2); other segments from A/Puerto Rico/8/1934 (H1N2)	MDCK	1
ribavirin		14 ± 2	2-log10 reduction	RT-qPCR	HA, NA from A/Aichi/2/68 (H3N2); other segments from A/Puerto Rico/8/1934 (H1N3)	MDCK	1
favipiravir		11 ± 0	EC50	TCID50	HA, NA from A/Aichi/2/68 (H3N2); other segments from A/Puerto Rico/8/1934 (H1N4)	MDCK	1
favipiravir		9.2 ± 1.1	1-log10 reduction	RT-qPCR	HA, NA from A/Aichi/2/68 (H3N2); other segments from A/Puerto Rico/8/1934 (H1N5)	MDCK	1
favipiravir		13 ± 1	2-log10 reduction	RT-qPCR	HA, NA from A/Aichi/2/68 (H3N2); other segments from A/Puerto Rico/8/1934 (H1N6)	MDCK	1
ribavirin		40	2-log10 reduction	TCID50	A/Wuhan/359/95 (H3N2)	MDCK	2
ribavirin	2.5-5.5		EC50	neutral red CPE	A/H1N1	MDCK	3
ribavirin	2.7-5.5		EC50	neutral red CPE	A/H3N2	MDCK	3
ribavirin	2.3-4.3		EC50	neutral red CPE	A/H5N1	MDCK	3
ribavirin	1.5-4.2		EC50	neutral red CPE	IBV	MDCK	3
favipiravir	0.029-0.20		IC50	plaque assay	A/H1N1	MDCK	4
favipiravir	0.013-0.30		IC50	plaque assay	A/H2N2	MDCK	4
favipiravir	0.078-0.48		IC50	plaque assay	A/H3N2	MDCK	4
favipiravir	0.039-0.089		IC50	plaque assay	IBV	MDCK	4
favipiravir	0.030-0.095		IC50	plaque assay	ICV	MDCK	4

favipiravir		6.61 ± 0.02	IC50	polymerase inhibition assay	A/Brisbane/59/2007	A549	5
favipiravir		5.28 ±0.14	IC50	polymerase inhibition assay	A/New Jersey/15/2007	A549	5
favipiravir		3.01 ±0.03	IC50	polymerase inhibition assay	A/Denmark/524/2009	A549	5
favipiravir		17.05 ± 0.71	EC50	plaque assay	A/Brisbane/59/2007	MDCK	5
favipiravir		15.07 ± 0.14	EC50	plaque assay	A/New Jersey/15/2007	MDCK	5
favipiravir		15.54 ± 0.56	EC50	plaque assay	A/Denmark/524/2009	MDCK	5
EIDD-1931		3.1 (2.25–3.82)	EC50	plaque assay	A/WSN/33 (H1N1)	MDCK	6
EIDD-1931		1.1 (0.86–1.22)	EC50	TCID50-HA assay	A/WSN/33 (H1N1)	MDCK	6
EIDD-1931		3.1 (1.49–6.23)	EC50	TCID50-HA assay	A/California/7/2009 (H1N1) pdm09	MDCK	6
EIDD-1931		3.4 (2.92–3.9)	EC50	TCID50-HA assay	A/Georgia/M5081/2012 (H1N1)	MDCK	6
EIDD-1931		1.8 (1.17–2.55)	EC50	TCID50-HA assay	A/Netherlands/602/2009 (H1N1) pdm09	MDCK	6
EIDD-1931		0.14	EC50	TCID50	A/Vietnam/1203/2004 (H5N1)	MDCK	6
EIDD-1931		0.13	EC50	TCID50	A/Anhui/1/2013 (H7N9)	MDCK	6
EIDD-1931		3.2 (2.68–3.88)	EC50	TCID50-HA assay	A/Aichi/2/68 (H3N2)	MDCK	6
EIDD-1931		1.7 (1.27–2.33)	EC50	TCID50-HA assay	A/Wisconsin/67/2005 (H3N2)	MDCK	6
EIDD-1931		1.2 (0.05–2.0)	EC50	TCID50-HA assay	A/Panama/2007/99 (H3N2)	MDCK	6
EIDD-1931		3.2 (2.52–4.05)	EC50	TCID50-HA assay	A/swine/Ohio/sw10-132/2010 (H3N2)	MDCK	6
EIDD-1931		0.015 (0.011–0.019)	EC50	TCID50	B/Yamagata/16/88	MDCK	6
EIDD-1931		0.006 (0.003–0.008)	EC50	TCID50	B/Brisbane/60/08	MDCK	6
5-azacytidine		around 7.5	2-log10 reduction	TCID50	A/WSN/33 (H1N1), A/Puerto Rico/8/1934 (H1N1), A/Panama/2007/1999 (H3N2), A/Wyoming/03/2003 (H3N2)	MDCK	7
5-fluorouracil		around 60	2-log10 reduction	TCID50	A/WSN/33 (H1N1), A/Puerto Rico/8/1934 (H1N1), A/Wyoming/03/2003 (H3N2)	MDCK	7
5-fluorouracil		over 80	2-log10 reduction	TCID50	A/Panama/2007/1999 (H3N2)	MDCK	7

ribavirin		around 7.5	2-log10 reduction	TCID50	A/WSN/33 (H1N1)	MDCK	7
ribavirin		around 7.5	2-log10 reduction	TCID50	A/Puerto Rico/8/1934 (H1N1)	MDCK	7
ribavirin		over 20	2-log10 reduction	TCID50	A/Panama/2007/1999 (H3N2)	MDCK	7
ribavirin		around 10	2-log10 reduction	TCID50	A/Wyoming/03/2003 (H3N2)	MDCK	7

Appendix Table B.2 Inhibition concentrations for influenza viruses (*in vivo* experiments)

Drug	Concentration (mg/kg)	Virus inoculation titer (PFU)	Measurement	Assay	Virus strain	Animal	Citation
ribavirin	37.5	90 µl virus at LD100 concentration	ability to prevent death	survival rate	A/WSN/33 (H1N1)	mice	3
ribavirin	75	90 µl virus at LD100 concentration	ability to prevent death	survival rate	A/Victoria/3/75 (H3N2)	mice	3
favipiravir	200	300	ability to prevent death	survival rate	A/Puerto Rico/8/1934 (H1N1)	mice	4
EIDD-1931	100	300	1-log10 lung viral load reduction	TCID50	A/Puerto Rico/8/1934 (H1N1)	mice	6
EIDD-1931	400	300	2-log10 lung viral load reduction	TCID50	A/Puerto Rico/8/1934 (H1N1)	mice	6

Appendix Table B.3 Inhibition concentrations for other viruses (*in vitro* experiments)

Drug	Concentration (µg/mL)	Concentration (µM)	Measurement	Assay	Virus	Cell line	Citation
ribavirin	2.5		reduce replication by 64% ± 6.1%	qRT-PCR	human norovirus	HG23	8
ribavirin		135 ± 74	EC50	neutral red CPE	enterovirus D68/US/KY/14-18593	RD human rhabdomyosarcoma	9
ribavirin		147 ± 74	EC50	neutral red CPE	enterovirus D68/Fermon	RD human rhabdomyosarcoma	9
ribavirin		106 ± 26	EC50	neutral red CPE	rhinovirus 87	RD human rhabdomyosarcoma	9
ribavirin		93 ± 23	1-log10 reduction	virus yield reduction	enterovirus D68/US/KY/14-18593	RD human rhabdomyosarcoma	9
5-azacytidine		20.2 (18.7–21.8)	EC50	flow cytomery	human immunodeficiency virus type 1	Magi-U373-CXCR4 CEM	10
5-azacytidine		14.5 (10.3–18.8)	EC50	flow cytomery	human immunodeficiency virus type 2	Magi-U373-CXCR4 CEM	10

ribavirin		50	2-log10 reduction	TCID50	bursal disease virus (FVSKG2)	chicken embryo fibroblasts (CEFs)	11
5-azacytidine		200	2-log10 reduction	TCID50	bursal disease virus (FVSKG2)	chicken embryo fibroblasts (CEFs)	11
5-fluorouracil		500	2-log10 reduction*	TCID50	bursal disease virus (FVSKG2)	chicken embryo fibroblasts (CEFs)	11
5-azacytidine		180	IC50	inhibition of relaxed circular DNA synthesis	hepatitis B virus	HepAD38	12
5-azacytidine		50	63% reduction in infectivity	nano luciferase activity	hepatitis B virus	HepG2-NTCP	12
ribavirin		around 200	2-log10 reduction	microtitration infectivity assay (based on CPE)	porcine reproductive and respiratory syndrome virus	MARC-145	13
5-azacytidine		around 100	2-log10 reduction	microtitration infectivity assay (based on CPE)	porcine reproductive and respiratory syndrome virus	MARC-145	13
5-fluorouracil		around 200	2-log10 reduction	microtitration infectivity assay (based on CPE)	porcine reproductive and respiratory syndrome virus	MARC-145	13
EIDD-1931	0.217 (0.142–0.323)		EC50	plaque assay	SARS-CoV-2, beta variant	ACE2-A549	14
EIDD-1931	0.506 (0.483–0.531)		EC50	plaque assay	SARS-CoV-2, delta variant	ACE2-A549	14
EIDD-1931	0.080 (0.290–0.383)		EC50	plaque assay	SARS-CoV-2, omicron BA.2 variant	ACE2-A549	14

* Only at 96 hours post infection

Citation

1. Vanderlinden E, Vrancken B, Van Houdt J, Rajwanshi VK, Gillemot S, Andrei G, Lemey P, Naesens L. 2016. Distinct Effects of T-705 (Favipiravir) and Ribavirin on Influenza Virus Replication and Viral RNA Synthesis. Antimicrobial Agents and Chemotherapy 60:6679–6691.

2. Cheung PPH, Watson SJ, Choy K-T, Fun Sia S, Wong DDY, Poon LLM, Kellam P, Guan Y, Malik Peiris JS, Yen H-L. 2014. Generation and characterization of influenza A viruses with altered polymerase fidelity. Nature Communications 5.

3. Sidwell RW, Bailey KW, Wong M-H., Barnard DL, Smee DF. 2005. In vitro and in vivo influenza virus-inhibitory effects of viramidine. Antiviral Research 68:10–17.

4. Furuta Y, Takahashi K, Fukuda Y, Kuno M, Kamiyama T, Kozaki K, Nomura N, Egawa H, Minami S, Watanabe Y, Narita H, Shiraki K. 2002. In Vitro and In Vivo Activities of Anti-Influenza Virus Compound T-705. Antimicrobial Agents and Chemotherapy 46:977–981.

5. Baranovich T, Wong S-S., Armstrong J, Marjuki H, Webby RJ, Webster RG, Govorkova EA. 2013. T-705 (Favipiravir) Induces Lethal Mutagenesis in Influenza A H1N1 Viruses In Vitro. Journal of Virology 87:3741–3751.

6. Yoon J-J, Toots M, Lee S, Lee M-E, Ludeke B, Luczo JM, Ganti K, Cox RM, Sticher ZM, Edpuganti V, Mitchell DG, Lockwood MA, Kolykhalov AA, Greninger AL, Moore ML, Painter GR, Lowen AC, Tompkins SM, Fearns R, Natchus MG. 2018. Orally Efficacious Broad-Spectrum Ribonucleoside Analog Inhibitor of Influenza and Respiratory Syncytial Viruses. Antimicrobial Agents and Chemotherapy 62.

7. Pauly MD, Lauring AS. 2015. Effective Lethal Mutagenesis of Influenza Virus by Three Nucleoside Analogs. Journal of Virology 89:3584–3597.

8. Wen D, Yin Y, Wang Y, Wang W, Su J, Sprengers D, van der Laan LJW, Felczak K, Pankiewicz KW, Chang K, Koopmans M, Metselaar HJ, Peppelenbosch MP. 2017. Inhibition of Calcineurin or IMP Dehydrogenase Exerts Moderate to Potent Antiviral Activity against Norovirus Replication. Antimicrobial Agents and Chemotherapy 61.

9. Smee DF, Evans WJ, Nicolaou KC, Tarbet EB, Day CW. 2016. Susceptibilities of enterovirus D68, enterovirus 71, and rhinovirus 87 strains to various antiviral compounds. Antiviral Research 131:61–65.

10. Beach LB, Rawson JM, Kim B, Patterson SE, Mansky LM. 2014. Novel inhibitors of human immunodeficiency virus type 2 infectivity. Journal of General Virology 95:2778–2783.

11. Akram T, Gul I, Parveez Zia M, Hassan A, Khatun A, Shah RA, Ahmad SM, Ganai NA, Chikan NA, Kim W-I, Shabir N. 2023. Ribavirin inhibits the replication of infectious bursal disease virus predominantly through depletion of cellular guanosine pool. Frontiers in Veterinary Science 10:1192583.

12. McDaniel YZ, Patterson SE, Mansky LM. 2019. Distinct dual antiviral mechanism that enhances hepatitis B virus mutagenesis and reduces viral DNA synthesis. Antiviral Research 170:104540.

13. Khatun A, Shabir N, Yoon K-J, Kim W-I. 2015. Effects of ribavirin on the replication and genetic stability of porcine reproductive and respiratory syndrome virus. BMC veterinary research 11:21.

14. Franco EJ, Drusano GL, Hanrahan KC, Warfield KL, Brown AN. 2023. Combination Therapy with UV-4B and Molnupiravir Enhances SARS-CoV-2 Suppression. Viruses 15:1175.

Appendix C: Mutagenic Drugs Cell Toxicity

Appendix Table C.1 Cytotoxicity concentrations of mutagenic drugs

Drug	Cell	Minimum cytotoxic concentration (μM)	50% Cytotoxicity concentration (CC_{50}) (μM)	Citation
favipiravir	A549		>1000	1
favipiravir	MDCK		>1000	1
favipiravir	MDCK	>600		2
ribavirin	MDCK	100 ± 0		2
ribavirin	MDCK		3074.8*	3
favipiravir	A549		>6369.4*	4
favipiravir	MDCK		>6369.4*	4
ribavirin	A549		307.4*	4
5-azacytidine	MDCK		20-25	5
5-fluorouracil	MDCK	modest decrease at 480		5
ribavirin	MDCK	modest decrease at 120		5
EIDD-1931	ACE2-A549		>1	6
5-azacytidine	A549		6.3 ± 1.1	7
5-azacytidine	MDCK		54 ± 5	8
5-fluorouracil	A549	210 (24 hours *p.i.*)		9
5-fluorouracil	A549	38.4 (48 hours *p.i.*)		9
EIDD-1931	ACE2-A549	>13.5*		10
EIDD-1931	vero		>100	11
favipiravir	vero		2837 ± 86	11
ribavirin	vero		250 ± 81	11
EIDD-2749	MDCK	200-500		12
ribavirin	A549	24% inhibition at 1158.3*		13

* Concentration was converted by reviewer, with formula:

$$[\mu M] = [\mu g/mL]/Molecular\ Weight\ in\ [Kilo\ Dalton]$$

Citation

1. Baranovich T, Wong S-S., Armstrong J, Marjuki H, Webby RJ, Webster RG, Govorkova EA. 2013. T-705 (Favipiravir) Induces Lethal Mutagenesis in Influenza A H1N1 Viruses In Vitro. Journal of Virology 87:3741–3751.

2. Vanderlinden E, Vrancken B, Van Houdt J, Rajwanshi VK, Gillemot S, Andrei G, Lemey P, Naesens L. 2016. Distinct Effects of T-705 (Favipiravir) and Ribavirin on Influenza Virus Replication and Viral RNA Synthesis. Antimicrobial Agents and Chemotherapy 60:6679–6691.

3. Sidwell RW, Bailey KW, Wong M-H., Barnard DL, Smee DF. 2005. In vitro and in vivo influenza virus-inhibitory effects of viramidine. Antiviral Research 68:10–17.

4. Furuta Y, Takahashi K, Fukuda Y, Kuno M, Kamiyama T, Kozaki K, Nomura N, Egawa H, Minami S, Watanabe Y, Narita H, Shiraki K. 2002. In Vitro and In Vivo Activities of Anti-Influenza Virus Compound T-705. Antimicrobial Agents and Chemotherapy 46:977–981.

5. Pauly MD, Lauring AS. 2015. Effective Lethal Mutagenesis of Influenza Virus by Three Nucleoside Analogs. Journal of Virology 89:3584–3597.

6. Franco EJ, Drusano GL, Hanrahan KC, Warfield KL, Brown AN. 2023. Combination Therapy with UV-4B and Molnupiravir Enhances SARS-CoV-2 Suppression. Viruses 15:1175.

7. Nguyen AN, Hollenbach PW, Richard N, Luna-Moran A, Brady H, Heise C, MacBeth KJ. 2010. Azacitidine and decitabine have different mechanisms of action in non-small cell lung cancer cell lines. Lung Cancer: Targets and Therapy 1:119.

8. Rius M, Keller D, Brom M, Hummel-Eisenbeiss J, Lyko F, Keppler D. 2010. Vectorial Transport of Nucleoside Analogs from the Apical to the Basolateral Membrane in Double-Transfected Cells Expressing the Human Concentrative Nucleoside Transporter hCNT3 and the Export Pump ABCC4. Drug Metabolism and Disposition 38:1054–1063.

9. Jastrzebska E, Flis S, Rakowska A, Chudy M, Jastrzebski Z, Dybko A, Brzozka Z. 2013. A microfluidic system to study the cytotoxic effect of drugs: the combined effect of celecoxib and 5-fluorouracil on normal and cancer cells. Microchimica Acta 180:895–901.

10. Brown AN, Lang Y, Zhou J, Franco EJ, Hanrahan KC, Bulitta JB, Drusano GL. 2022. Why Molnupiravir Fails in Hospitalized Patients. MBio 13.

11. Langendries L, Abdelnabi R, Neyts J, Delang L. 2021. Repurposing Drugs for Mayaro Virus: Identification of EIDD-1931, Favipiravir and Suramin as Mayaro Virus Inhibitors. Microorganisms 9:734.

12. Sourimant J, Lieber CM, Aggarwal M, Cox RM, Wolf JD, Yoon J-J, Toots M, Ye C, Sticher Z, Kolykhalov AA, Martinez-Sobrido L, Bluemling GR, Natchus MG, Painter GR, Plemper RK. 2022. 4′-Fluorouridine is an oral antiviral that blocks respiratory syncytial virus and SARS-CoV-2 replication. Science 375:161–167.

13. Tang W-D, Tang H, Peng H, Ren R, Zhao P, Zhao L. 2023. Inhibition of tick-borne encephalitis virus in cell cultures by ribavirin. Frontiers in Microbiology 14.