

Comprehensive fitness landscape of SARS-CoV-2 M^{pro} reveals insights into viral resistance mechanisms

Julia M. Flynn^{1,#}, Neha Samant¹, Gily Schneider-Nachum¹, David T. Barkan², Nese Kurt Yilmaz¹, Celia A. Schiffer¹, Stephanie A. Moquin², Dustin Dovala², and Daniel N.A. Bolon^{1,#}

¹Department of Biochemistry and Molecular Biotechnology
University of Massachusetts Chan Medical School
Worcester, MA 01605, USA

²Novartis Institutes for Biomedical Research
Emeryville, CA 94608, USA

corresponding authors, Julia.Flynn@umassmed.edu, Dan.Bolon@umassmed.edu

2 Abstract

3

4 With the continual evolution of new strains of SARS-CoV-2 that are more virulent, transmissible, and
5 able to evade current vaccines, there is an urgent need for effective anti-viral drugs. SARS-CoV-2 main
6 protease (M^{pro}) is a leading target for drug design due to its conserved and indispensable role in the
7 viral life cycle. Drugs targeting M^{pro} appear promising but will elicit selection pressure for resistance. To
8 understand resistance potential in M^{pro} , we performed a comprehensive mutational scan of the
9 protease that analyzed the function of all possible single amino acid changes. We developed three
10 separate high-throughput assays of M^{pro} function in yeast, based on either the ability of M^{pro} variants
11 to cleave at a defined cut-site or on the toxicity of their expression to yeast. We used deep sequencing
12 to quantify the functional effects of each variant in each screen. The protein fitness landscapes from
13 all three screens were strongly correlated, indicating that they captured the biophysical properties
14 critical to M^{pro} function. The fitness landscapes revealed a non-active site location on the surface that
15 is extremely sensitive to mutation making it a favorable location to target with inhibitors. In addition,
16 we found a network of critical amino acids that physically bridge the two active sites of the M^{pro} dimer.
17 The clinical variants of M^{pro} were predominantly functional in our screens, indicating that M^{pro} is under
18 strong selection pressure in the human population. Our results provide predictions of mutations that
19 will be readily accessible to M^{pro} evolution and that are likely to contribute to drug resistance. This
20 complete mutational guide of M^{pro} can be used in the design of inhibitors with reduced potential of
21 evolving viral resistance.

22

23

24 Introduction

25

26 The COVID-19 pandemic, caused by the Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-
27 2), has had an unprecedented impact on global health, the world economy, and our overall way of life.
28 Despite the rapid deployment of mRNA and traditional vaccines against SARS-CoV-2 which have served
29 to greatly improve patient outcomes and decrease spread of the disease, vaccines remain unavailable
30 in many parts of the world and there is hesitancy to get vaccinated among portions of the population.
31 Additionally, the virus appears to be evolving mutations in the spike protein that reduce immune
32 protection from both vaccines and prior infections. Additional strategies including direct-acting
33 antiviral drugs are needed to combat the SARS-CoV-2 pandemic. The main protease (M^{pro}) of SARS-
34 CoV-2 is a promising target for drug development with many laboratories working collaboratively to
35 develop drugs against this protease, leading to thousands of M^{pro} inhibitors in the pipeline and the first
36 FDA-authorized clinical drug against this target, Paxlovid. The use of drugs that target M^{pro} will apply
37 selection pressure for the evolution of resistance. There is potential to design drugs with reduced
38 likelihood of developing M^{pro} resistance, but these efforts will require an in-depth understanding of the
39 evolutionary potential of the protease.

40

41 SARS-CoV-2 is a highly contagious virus responsible for the ongoing COVID-19 pandemic. SARS-CoV-2
42 belongs to the group of coronaviruses and has a positive-sense single-stranded RNA genome
43 (Macnaughton and Madge 1978). Immediately upon entry into the host cell, the SARS-CoV-2 virus
44 translates its replicase gene (ORF1) into two overlapping large polyproteins produced in tandem by a
45 ribosomal frameshift, pp1a and pp1ab (Herold, Raabe et al. 1993). These polyproteins are cleaved by
46 two cysteine proteases, M^{pro} (also known as the chymotrypsin-like protease, $3CL^{\text{pro}}$, or Nsp5) and the
47 papain-like protease (PL^{pro}) to yield functional replication machinery indispensable to viral replication
48 (Ziebuhr, Herold et al. 1995, Lim, Ng et al. 2000). M^{pro} initiates autoproteolysis from the pp1a and
49 pp1ab polypeptides at its N- and C- terminus, through a poorly understood mechanism (Hsu, Kuo et al.
50 2005). Subsequently, mature M^{pro} cuts at 11 additional cleavage sites in both pp1a and pp1ab (Fan,
51 Wei et al. 2004). The sites cut by M^{pro} all include a conserved Gln at the P1 position, a small amino acid
52 (Ser, Ala or Gly) at the P1' position, and a hydrophobic residue (Leu, Phe, or Val) at the P2 position
53 (Hegyi, Friebe et al. 2002, Thiel, Ivanov et al. 2003). Along with its vital role in the liberation of viral
54 proteins, M^{pro} also cleaves specific host proteins, an activity which has been shown to enhance viral
55 replication (Meyer, Chiaravalli et al. 2021). Through its substrates, M^{pro} function is required for almost
56 every known step in the viral life cycle.

57

58 M^{pro} is a highly attractive target for drug development against SARS-CoV-2 and future coronavirus-
59 mediated pandemics for numerous reasons. M^{pro} plays an essential functional role in the viral life cycle
60 so that blocking its function will impair viral propagation. M^{pro} is highly conserved among all

61 coronaviruses making it likely that inhibitors will have broad efficacy in potential future pandemics.
62 There are no human M^{pro} homologs, and it shares no overlapping substrate specificity with any known
63 human protease, minimizing the possibility of side effects. Additionally, its nucleophilic cysteine active
64 site enables the design of covalent inhibitors that provide advantages such as increased potency,
65 selectivity, and duration of inhibition (Singh, Petter et al. 2011). For these reasons, M^{pro} has become
66 one of the most characterized SARS-CoV-2 drug targets (Jin, Du et al. 2020, Zhang, Lin et al. 2020,
67 Biering, Van Dis et al. 2021, Fischer, Veprek et al. 2021).

68

69 Native M^{pro} is a homodimer, and each monomer is composed of three domains (Jin, Du et al. 2020).
70 Domain I (8-101) and Domain II (102-184) are comprised of antiparallel β -barrel structures. Cys145 and
71 His41 make up M^{pro}'s noncanonical catalytic dyads and are located in a clefts between Domains I and II.
72 Domain III (201-303) is an all α -helical domain that coordinates M^{pro} dimerization, which is essential for
73 M^{pro} function (Tan, Verschueren et al. 2005). Much of the structural and enzymatic knowledge of
74 SARS-CoV-2 M^{pro} has been derived from studies of SARS-CoV-1 that caused the 2003 SARS outbreak
75 (Ksiazek, Erdman et al. 2003), as well as MERS-CoV that caused the 2012 MERS outbreak (Zaki, van
76 Boheemen et al. 2012). M^{pro} from SARS-CoV-1 and SARS-CoV-2 differ in sequence at only 12 residues,
77 however SARS-CoV-2 M^{pro} exhibits increased structural flexibility and plasticity (Bzowka, Mitusinska et
78 al. 2020, Estrada 2020, Kneller, Phillips et al. 2020).

79

80 We performed comprehensive mutational analysis of SARS-CoV-2 M^{pro} to provide functional and
81 structural information to aid in the design of effective inhibitors against the protease. Systematic
82 mutational scanning assesses the consequences of all point mutations in a gene providing a
83 comprehensive picture of the relationship between protein sequence and function (Hietpas, Jensen et
84 al. 2011, Fowler and Fields 2014). Mutational scanning requires a selection step that separates
85 variants based on function. Following selection, the frequency of each variant is assessed by deep
86 sequencing to estimate functional effects. The resulting protein fitness landscape describes how all
87 individual amino acid changes in a protein impact function and provides a detailed guide to the
88 biophysical and biochemical properties that underlie fitness. Protein fitness landscapes identify
89 mutation-tolerant positions that may readily contribute to drug resistance. These studies also elucidate
90 mutation-sensitive residues that are critical to function, making them attractive target sites for
91 inhibitors with reduced likelihood of developing resistance. The work described here focuses on
92 fitness landscapes without drug pressure because these provide critical information regarding M^{pro}
93 mechanism and evolutionary potential that we hope will be useful in the efforts to combat SARS-CoV-
94 2. We are pursuing investigations in the presence of inhibitors, but these experiments will require
95 further optimization steps to make our yeast-based assays compatible with inhibition. Of note,
96 mutational scans of other drug targets including lactamases (Deng, Huang et al. 2012, Firnberg,
97 Labonte et al. 2014) and oncogenes (Choi, Landrette et al. 2014, Ma, Boucher et al. 2017) have
98 demonstrated the potential to accurately identify and predict clinically-relevant resistance evolution.

100 In this study, we used systematic mutational scanning to analyze the functional effects of every
101 individual amino acid change in M^{pro}. We developed three orthogonal screens in yeast to separate
102 M^{pro} variants based on function. The first screen measures M^{pro} activity via loss of Fluorescence
103 Resonance Energy Transfer (FRET) from a genetically-encoded FRET pair linked by the Nsp4/5 cleavage
104 sequence (Figure 1a). The second screen similarly measures cleavage of the Nsp4/5 cut site; however,
105 in this screen M^{pro} cleavage leads to inactivation of a transcription factor driving GFP expression (Figure
106 1b). The final screen leverages the toxicity of wild-type (WT) M^{pro} to yeast that is likely due to cleavage
107 of essential yeast proteins, and leads to depletion of active variants during growth (Figure 1c).
108 Following selection in the three screens, populations were subjected to deep sequencing in order to
109 quantify function based on the enrichment or depletion of each variant.

110

111 We found that the functional scores between screens were correlated, indicating that they all captured
112 key biophysical properties governing function. Our functional scores also correlated well with
113 previously measured catalytic rates of purified individual mutants. Additionally, substitutions in M^{pro}
114 from coronaviruses distantly related to SARS-CoV-2 consistently exhibited high function in our screens
115 indicating that similar biophysical properties underlie the function of genetically-diverse M^{pro}
116 sequences. Our study revealed mutation-sensitive sites distal to the active site and dimerization
117 interface. These sites reveal important communication networks that may be targeted by inhibitors.
118 Our results provide a comprehensive dataset which can be used to design molecules with decreased
119 vulnerability to resistance, by building drug-protein interactions at mutation-sensitive sites while
120 avoiding mutation-tolerant residues.

121

122 Results

123

124 Expression of mature WT M^{pro} in yeast

125 The main protease of SARS-CoV-2 is produced by self-cleavage of polyproteins translated from the viral
126 RNA genome, and its enzymatic activity is inhibited by the presence of additional N- and C-terminal
127 amino acids (Xue, Yang et al. 2007). To express M^{pro} with its authentic N-terminal serine residue, we
128 generated a Ubiquitin-M^{pro} fusion protein. In yeast and other eukaryotes, Ubiquitin (Ub) fusion
129 proteins are cleaved by Ub-specific proteases directly C-terminal to the Ub, revealing the N-terminal
130 residue of the fused protein, regardless of sequence (Bachmair, Finley et al. 1986). Expression of
131 functionally-active M^{pro} is toxic to yeast cells (Alalam, Sigurdardottir et al. 2021). To control the
132 expression level of M^{pro} while limiting its toxic side effects, we placed Ub-M^{pro} under control of the
133 inducible and engineered LexA-ER-AD transcription factor (Ottoz, Rudolf et al. 2014). LexA-ER-AD is a
134 fusion of the bacterial LexA DNA-binding protein, the human estrogen receptor (ER) and the B112

activation domain, and its activity is tightly and precisely regulated by the hormone β -estradiol. We inserted 4 *lexA* boxes recognized by the LexA DNA binding domain upstream of Ub-M^{pro} to control its expression. The Western blot in Figure 1 – figure supplement 1a illustrates both induction of M^{pro} by β -estradiol and successful removal of the Ub moiety, indicating that the protease is being expressed in its mature and functional form. We performed a titration curve with β -estradiol to determine the lowest concentration at which M^{pro} can be expressed without inhibiting yeast cell growth while still enabling measurement of substrate cleavage (Figure 1 – figure supplement 1b).

Engineering of functional screens to monitor intracellular M^{pro} activity

We developed three distinct yeast screens to characterize the effects of M^{pro} variants on function (Figure 1). The first screen utilized a FRET-based reporter of two fluorescent proteins, YPet and CyPet, fused together with the Nsp4/5 M^{pro} cleavage site engineered in the middle (YPet-M^{pro}CS-CyPet) (Figure 1a). The YPet-CyPet pair are derivatives of the YFP-CFP proteins that have been fluorescently optimized by directed evolution for intracellular FRET (Nguyen and Daugherty 2005) and provide a 20-fold signal change upon cleavage. The linker between the two fluorescent proteins contains the M^{pro} cleavage site, TSAVLQ|SGFRK, the cut-site at the N-terminus of the M^{pro} protease. This is the most commonly used cut-site for *in vitro* cleavage assays, which allowed us to directly compare our mutational results to those that were previously published. One advantage of this assay is that the fluorescent readout directly reports on cleavage of a specific cut-site. The plasmid containing Ub-M^{pro} under the control of β -estradiol was transformed into yeast cells expressing a chromosomally integrated copy of YPet-M^{pro}CS-CyPet. Expression of WT M^{pro} led to a β -estradiol-dependent decrease in FRET signal as measured by fluorescence-activated single cell sorting (FACS). Mutation of the essential catalytic cysteine of M^{pro} to alanine (C145A) abolished this change in FRET signal indicating that the change in signal was dependent on the presence of functional M^{pro} (Figure 1 – figure supplement 1c).

The second screen utilized the DNA binding domain and activation domain of the Gal4 transcription factor, separated by the Nsp4/5 cut site (Johnston, Zavortink et al. 1986, Murray, Hung et al. 1993). We used this engineered transcription factor (TF) to drive GFP expression, enabling cells with varying levels of M^{pro} protease activity to be separated by FACS (Figure 1b). One benefit of this system is its signal amplification, as one cut transcription factor can cause a reduction of more than one GFP molecule. However, due to this amplification, the fluorescent signal is indirectly related to cutting efficiency. Expression of Ub-M^{pro} in cells engineered with the split transcriptional factor exhibited a β -estradiol-dependent decrease in GFP reporter activity that required the presence of catalytically-functional M^{pro} protein (Figure 1 – figure supplement 1d). The final screen leverages the toxicity of M^{pro} expression in yeast, which likely results from cleavage of essential yeast proteins by the protease (Alalam, Sigurdardottir et al. 2021) (Figure 1c). Increasing concentrations of β -estradiol correlates with a decrease in yeast growth rate that is dependent on the presence of catalytically-functional M^{pro}

(Figure 1 – figure supplement 1b). At a high expression level induced with 2 μ M of β -estradiol, yeast growth rate becomes tightly coupled to M^{pro} function and can be used as a readout of the function of the expressed M^{pro} variant. While the endogenous yeast substrates are unknown, this assay is likely reporting on M^{pro} cleavage of numerous cellular targets. Sampling of more than one cleavage site may better represent the physiologic role of M^{pro}, which has 11 viral and numerous host cleavage sites.

Comprehensive deep mutational scanning of M^{pro}

We integrated our three screens with a systematic mutational scanning approach to determine the impact of each single amino acid change in M^{pro} on its function (Figure 1d). A comprehensive M^{pro} single site variant library was purchased commercially (Twist Biosciences). Each position of M^{pro} was mutated to all other 19 amino acids plus a stop codon, using the preferred yeast codon for each substitution. We transferred the library to a plasmid under the LexA promoter. To efficiently track each variant of the library using deep sequencing, we employed a barcoding strategy that allowed us to track mutations across the gene using a short sequence readout. We engineered the barcoded library so that each mutant was represented by 20-40 unique barcodes and used PacBio sequencing to associate barcodes with M^{pro} mutations (Figure 1d). 96% of library variants were linked to 10 or greater barcodes (Figure 1 – figure supplement 1e). As a control, the library was doped with a small amount of WT M^{pro} linked to approximately 150 barcodes.

We transformed the plasmid library of M^{pro} mutations into yeast strains harboring the respective reporter for each functional screen. The mutant libraries were amplified in the absence of selection and subsequently β -estradiol was added to induce M^{pro} expression. Variant counts analyzed by sequencing before and after the pre-selection amplification step were correlated, consistent with minimal to no selection prior to induction with β -estradiol (Figure 1 – figure supplement 1f and Figure 1 – figure supplement 1g). For the fluorescent screens, the cells were incubated with β -estradiol at the concentration determined to limit M^{pro} toxicity (125 nM) for the time required for WT M^{pro} to achieve full reporter activity (1.5 hours for the FRET screen and 6 hours for the TF screen). Subsequently cells were separated by FACS into populations with either uncleaved or cleaved reporter proteins (See Figure 1a and Figure 1b). For the growth screen, cells were incubated with a higher concentration of β -estradiol determined to slow yeast growth (2 μ M) (Figure 1 – figure supplement 1b). Populations of cells were collected at the 0- and 16-hour time points. For each cell population in each screen, plasmids encoding the mutated M^{pro} library were recovered, and the barcoded region was sequenced using single end Illumina sequencing. For the TF and FRET screens, the functional score of each mutant was calculated as the fraction of the mutant in the cut population relative to its fraction in both populations. For the growth screen, the functional score was calculated as the fraction of the mutant at the 0-hour time point relative to the fraction in the 0-hour and 16-hour time points. We normalized the functional scores in all three screens to facilitate comparisons, setting the score for the average WT

210 M^{pro} barcode as 1 and the average stop codon as 0 (See Figure 2 – source data 1 for all functional
211 scores).

212

213 To analyze the reproducibility of each screen, we performed biological replicates. For each biological
214 replicate we separately transformed the library into yeast cells, and independently performed
215 competition experiments and sequencing. Functional scores between replicates were strongly
216 correlated ($R^2 > 0.98$ for all three screens, Figure 2a) and we could clearly distinguish between
217 functional scores for WT M^{pro} and those containing stop codons (Figure 2b). There was a narrow
218 distribution of functional scores for stop codons in all three screens across the M^{pro} sequence except at
219 the last seven positions (amino acids 300-306) (Figure 2c), supporting previous experiments showing
220 that these residues are dispensable for M^{pro} activity and the importance of residue Q299 for M^{pro}
221 function (Lin, Chou et al. 2008). We categorized functional scores as WT-like, intermediate, or null-like
222 based on the distribution of WT barcodes and stop codons in each screen (Figure 2d and Figure 2 –
223 figure supplement 1). Heatmap representations of the functional scores determined in replicate 1 of
224 all three screens are shown in Figure 3 (FRET screen), Figure 3 – figure supplement 1 (TF screen), and
225 Figure 3 – figure supplement 2 (growth screen).

226

227 **Comparison between three screens**

228 Comparing the average functional score at each position (a measure of mutational sensitivity) between
229 the three screens shows a strong correlation (Figure 4a-c). The principal differences lie in the
230 sensitivity of the screens to mutation, with the average defective mutation in the growth screen being
231 more exaggerated than that in the fluorescent-based screens (Figure 4c). The scores in the growth
232 screen are likely integrating cutting efficiency over a diverse set of cleavages sites which may
233 contribute to this screen's increased sensitivity to mutation. Despite these differences, there are
234 striking correlations in the mutational patterns of M^{pro} across all three screens as can be visualized in
235 the heatmap of average scores per position and when mapped to M^{pro}'s structure (Figure 4a and b).
236 These similarities indicate that the three screens are reporting the same fundamental biophysical and
237 biochemical constraints of the protein.

238

239 Several lines of evidence indicate that the functional scores are biochemically and biologically relevant.
240 First, we compared the scores to previously published studies of point mutations (Figure 4d and Figure
241 4 – source data 2). For example, mutating the residues of the catalytic dyad, C145 and H41, inactivates
242 the protease both in our screen and in *in vitro* biochemical assays as expected (Hegyi, Friebe et al.
243 2002). Additionally, *in vitro* assays have shown that residues at the dimer interface including S10, G11
244 and E14 are essential for SARS-CoV-1 M^{pro} dimerization and function (Chen, Zhang et al. 2008).
245 Mutations at these residues are also deleterious to M^{pro} function in our screen. Because of the high
246 sequence and functional similarities between SARS-CoV-1 and CoV-2 M^{pro}, we expect that the majority

247 of the mutational analyses performed previously on SARS-CoV-1 M^{pro} will be valid for SARS-CoV-2 M^{pro}.
 248 We examined how the dynamic range of our screens relate to catalytic measurements. The growth
 249 screen measurements exhibited a linear pattern with relative catalytic rates previously reported for
 250 individual variants (Figure 4d). In contrast, the TF screen results showed a non-linear pattern,
 251 reminiscent of a binding equation. To assess these patterns in a systematic manner, we fit the graphs
 252 to both a linear equation and a non-linear binding equation with initial parameters of 1:1 for the linear
 253 fit, and an inflection point of 0.5 for the non-linear equation. Using this approach, we observed an
 254 apparent non-linear relationship between the functional scores measured in both the FRET and TF
 255 screens and the relative catalytic activity of mutants measured independently for M^{pro} *in vitro* in
 256 various studies ($R^2 = 0.81$ for non-linear fit to TF screen and $R^2 = 0.93$ for non-linear fit to FRET screen)
 257 (Figure 4d). Compared to the fluorescent screens, there is a stronger linear relationship ($R^2 = 0.86$)
 258 between the scores measured in our growth screen and the catalytic efficiencies of the individual
 259 mutants. These analyses indicate that the growth screen more fully captures the dynamic range of
 260 mutations with small functional defects that tend to appear WT-like in the FRET and TF screens. For
 261 the remainder of this paper, we will report the functional scores collected for the FRET and growth
 262 screens in the main figures and the TF screen in the supplementary figures. The advantage of the
 263 functional scores for each mutant from the FRET screen is that they report direct cleavage of a defined
 264 substrate, with the drawback being that they exhibit less sensitivity to mutation. The advantage of the
 265 growth screen is that the functional scores show a more linear relationship with catalytic rate while the
 266 drawback is that the screen reports cleavage of undefined substrates. Because of the correlation
 267 between all three screens, similar overall biophysical conclusions are supported by each screen.

268

269 **Functional characterization of natural M^{pro} variants**

270 To further assess the scores from our screen, we examined the functional scores of the M^{pro} variants
 271 observed in clinical samples. Because M^{pro} is essential for viral replication, deleterious mutations
 272 should be purged from the circulating population. The CoV-Glue-Viz database archives all mutations
 273 observed in the GISAID human SARS-CoV-2 sequences sampled from the ongoing COVID-19 pandemic
 274 (Singer, Gifford et al. 2020). We compared the frequency at which the clinical variants of the M^{pro} gene
 275 (ORF1ab/nsp5A-B) have been observed to their functional scores. The vast majority of the clinical
 276 isolates that have been sequenced to date have either 0 or 1 M^{pro} mutations with fewer than 0.4%
 277 having 2 or greater mutations and thus we did not account for epistasis in our analysis. We found that
 278 the most abundant clinical variants are highly functional in our assays (Figure 5a (FRET and growth
 279 screens) and Figure 5 – figure supplement 1a (TF screen)), however, lower frequency variants in clinical
 280 samples were found to have a wide range of M^{pro} function. Surprisingly, M^{pro} sequences among the
 281 clinical samples include premature stop codons that have been observed up to 100 times (out of over 5
 282 million total isolates to date) (Figure 5a (FRET and growth screens) and Figure 5 – figure supplement 1a
 283 (TF screen)). Because M^{pro} function is required for viral fitness, we assume that the frequency of stop
 284 codons observed in the data is an indication of sequencing error in the clinical samples. Accounting for
 285 this sequencing error, we examined the functional score of the 290 nonsynonymous mutations in the

286 M^{pro} gene that have been observed more than 100 times. The vast majority of these clinical variants
287 exhibit WT-like function with only nine having a score below that of the WT distribution (see Figures
288 5a-c). This observed enrichment for variants with WT-like function in the circulating SARS-CoV-2 virus
289 indicates that M^{pro} is undergoing strong purifying selection in the human population.

290 Additionally, we examined the experimental function of M^{pro} mutations compared with the diversity of
291 M^{pro} in viruses related to SARS-CoV-2. There is a 96% sequence identity between the SARS-CoV-2 and
292 SARS-CoV-1 M^{pro} proteases, with only 12 amino acid differences. In our study, all of the amino acid
293 differences in SARS-CoV-1 M^{pro} are WT-like in SARS-CoV-2, underscoring the credibility of the
294 functional scores (Figure 5b (FRET and growth screens) and Figure 5 – figure supplement 1b (TF
295 screen)). We went on to analyze the diversity in 852 sequences across a set of M^{pro} homologs with an
296 average homology of 47% from genetically diverse coronaviruses. We identified 1207 amino acid
297 changes located at 263 positions of M^{pro} and examined the functional score of these variants in our
298 data. Here again, we saw enrichment towards functional M^{pro} variants with only 6% (77 out of 1207)
299 natural variants having functional scores in the FRET screen below the WT range (Figure 5b and Figure
300 5c (FRET and growth screens) and Figure 5 – figure supplement 1b (TF screen)). Further analysis of
301 these deleterious variants should provide insight into the role epistasis played in the historical
302 evolution of M^{pro}, and these insights may have utility in the generation of future pan-coronavirus
303 inhibitors.

304

305 **Structural distribution of mutationally-sensitive M^{pro} positions**

306 Invariant sites that are essential to M^{pro} function are promising targets for designing inhibitors. 24
307 positions of M^{pro} exhibited low mutation tolerance, defined as 17 or more substitutions with null-like
308 function: P9, S10, G11, E14, R40, H41, T111, S113, R131, C145, G146, S147, G149, F150, H163, G174,
309 G179, G183, D187, D197, N203, D289, E290, and D295 (Figure 6a). Only four of these mutation-
310 sensitive residues contact the substrate: H41 and C145 (the catalytic residues), as well as H163, and
311 D187. H163 interacts with the invariable P1 Gln of the substrate and D187 forms a hydrogen bond
312 with a catalytic water and a salt bridge with R40. A large body of work has previously shown that
313 dimerization is indispensable to M^{pro} function (Chou, Chang et al. 2004, Hsu, Chang et al. 2005, Chen,
314 Zhang et al. 2008, Cheng, Chang et al. 2010). Our study also supports the critical functional role of
315 dimerization as we see prevalent mutation-sensitivity in residues at the dimer interface, including P9,
316 S10, G11, E14, and E290, each of which cannot be altered without complete loss of function.

317

318 Outside of these well-studied critical M^{pro} sites, there are additional clusters of mutation-intolerant
319 residues. R131, D197, N203, D289 and E290 lie at the interface of Domain II and Domain III sandwiched
320 between dimers and make up part of a surface identified by structural modeling as a possible distal
321 drug binding pocket (Bhat, Chitara et al. 2021, Weng, Naik et al. 2021) (Figure 6b). Within this cluster,
322 a dynamic salt bridge is formed between R131 located on the loop of Domain II connecting β 10-11 of
323 the catalytic pocket, and D289 in the α -helical Domain III that has been reported to contribute to the

flexibility and structural plasticity of M^{pro} (Bhat, Chitara et al. 2021). The location of these residues at the interface of the two domains and the dimer interface, combined with the fact that they are critical to M^{pro} function suggests that they are part of a distal regulatory communication network. Our studies clearly indicate the critical function played by this network of residues providing motivation for further examination of their potential as a mutation-resistant target for inhibitor design.

A second cluster of mutation-intolerant residues appear to be part of an allosteric communication network between the active site and the dimerization interface. Prior studies of individual mutations also suggest allosteric connections between the dimerization and active sites. Mutations at both E166 (Cheng, Chang et al. 2010) and S147 (Barrila, Bacha et al. 2006) were found to disrupt dimerization. Both positions E166 and S147 are located distal to the dimerization site, suggesting that the properties of these two sites are interdependent. Our results show that there is a physically-interacting chain of mutation-sensitive residues that bridge from the active site to the dimerization site (Figure 6c). This bridge is composed of H163 that directly contacts the P1 Gln of substrate, S147, L115 and S10 at the dimer interface. Each of these dimer-to-active site bridging residues are critical to M^{pro} function and are strongly conserved among M^{pro} homologs. Based on these observations, we suggest that the physical interactions between H163, S147, L115, and S10 mediate critical communication between the active sites of both subunits in the M^{pro} dimer.

All 24 of the identified mutation-intolerant residues are highly conserved among SARS-CoV-2 M^{pro} homologs (Figure 6d (FRET and growth screens) and Figure 6 – figure supplement 1 (TF screen)). While functional hot spots accurately predict evolutionary conservation, conservation does not accurately predict functional hot spots. There are many residues in M^{pro} that are strongly conserved, but that can be mutated without strong impacts on function. This pattern has been widely observed for other proteins (Hietpas, Jensen et al. 2011, Melamed, Young et al. 2013, Roscoe, Thayer et al. 2013, Starita, Pruneda et al. 2013, Mishra, Flynn et al. 2016). While many features distinguish natural evolution and experimental studies of fitness (Boucher, Bolon et al. 2019) one of the outstanding differences is the strength of selection. While functional hot spots can be defined by strong impacts on function that are experimentally measurable, small fitness changes that may be too small for experimental resolution can drive selection in natural evolution due to large population sizes and timescales (Ohta 1973). Our functional screen captures the mutations that are critical to catalytic function while evolutionary conservation depicts a wide range of mutations including those that make more nuanced contributions to function. When designing drugs to disrupt M^{pro} function, we hypothesize that it will be important to focus on the functionally critical sites which are a subset of the evolutionarily conserved positions.

Functional variability at key substrate and inhibitor-contact positions

360 M^{pro} function is essential for SARS-CoV-2 replication, making it a key drug target. To help further guide
361 inhibitor design, we assessed the mutations that are compatible with function and that should be
362 readily available to the evolution of drug resistance. We focused these analyses on the active site,
363 which is the target binding site for most inhibitors that have been generated against M^{pro} (Cho, Rosa et
364 al. 2021). In Figure 7a and Figure 7 – figure supplement 1a, we highlight all the M^{pro} residues that
365 contact the Nsp4/5 peptide, either through hydrogen bonds or van der Waals interactions (Shaqra,
366 Zvornicanin et al. 2022). In our functional screens, we found dramatic variability in mutational
367 sensitivity at these substrate-contact positions. For example, residues G143, H163, D187 and Q192
368 were extremely sensitive to mutation while residues M49, N142, E166 and Q189 were highly tolerant.
369 Despite the diverse sequence variation amongst M^{pro}'s substrates, they occupy a conserved volume in
370 the active site, known as the substrate envelope, and the interactions between M^{pro}'s residues and all
371 of its substrates are highly conserved (Shaqra, Zvornicanin et al. 2022) indicating that our mutation
372 results from the Nsp4/5 cut-site will likely translate to other cut-sites.

373

374 Even among residues whose side chains make direct hydrogen bonds with substrates are positions that
375 are surprisingly tolerant to mutation, namely N142, E166 and Q189. N142 forms distinct hydrogen
376 bonds with Nsp4/5 and Nsp8/9, which has been proposed as a mechanism of M^{pro} substrate
377 recognition (MacDonald, Frey et al. 2021). Q189 is in a flexible loop that closes over the substrates,
378 allowing accommodation of diverse cut-sites (Shaqra, Zvornicanin et al. 2022). In our screens, we find
379 that these proposed substrate-recognition positions are very tolerant to mutation (Figure 7b (FRET and
380 growth screens) and Figure 7 – figure supplement 1b (TF screen)) and have high potential for
381 developing inhibitor resistance. Our results indicate that mutations at N142, E166 and Q189 are
382 compatible with function and are readily available to the evolution of drug resistance.

383

384 A recent study comprehensively examined 233 X-ray crystal structures of SARS-CoV-2 M^{pro} in complex
385 with a wide range of inhibitors (Cho, Rosa et al. 2021). In 185 of these 233 structures, inhibitors lie in
386 the same binding pocket in the active site, primarily contacting M^{pro} positions T25, H41, M49, N142,
387 S144, C145, H163, H164, E166, P168, H172, Q189 and A191. We therefore went on to determine the
388 mutations at these key inhibitor binding residues that are compatible with M^{pro} function and should
389 likely be available to resistance evolution. Figures 7c and Figure 7 – figure supplement 1c illustrate a
390 representative structure of M^{pro} bound to the N3 inhibitor with the average mutational sensitivity of
391 each position mapped to the structure by color (Jin, Du et al. 2020). In addition, heatmaps are shown
392 detailing the mutations at these positions that are compatible with function (Figure 7 – figure
393 supplement 1d). Of note, residues N142, E166, and Q189 form direct hydrogen bonds with many M^{pro}
394 inhibitors and most mutations at these positions result in a functional protease. Additionally, T25,
395 M49, M164, P168 and A191 form van der Waals interactions with a variety of inhibitors suggesting that
396 mutations at these positions could disrupt inhibitor interactions while maintaining M^{pro} function. In
397 contrast, positions H41, S144, C145, H163 and H172 are highly sensitive in our screen, as well as

strongly conserved in nature, and therefore would be ideal contact positions for inhibitors with reduced likelihood of evolving M^{pro} resistance.

Pfizer has developed the first FDA-authorized M^{pro} inhibitor, PF-07321332 (Owen, Allerton et al. 2021). We examined the structure of M^{pro} bound to PF-07321332 to identify positions with the potential to evolve resistance against this drug (Figure 7d (FRET and growth screens) and Figure 7 – Figure supplement 1e (TF screen)) (Zhao, Fang et al. 2021). Evolutionarily-accessible resistance mutations are single base change mutations that would disrupt inhibitor binding while maintaining WT-like substrate recognition and cleavage. We identified all mutations of M^{pro} that have WT-like function in both the FRET and growth screens, would lead to a predicted decrease in inhibitor binding energy upon mutation of greater than 1 kcal/mol, and are accessible with a single nucleotide base change. These criteria led to the identification of three mutations, Q189E, E166A and E166Q with potential resistance against PF-07321332. These three positions are at sites where the inhibitor protrudes out of the defined substrate envelope, providing further evidence that these residues may evolve inhibitor resistance while maintaining substrate recognition (Shaqra, Zvornicanin et al. 2022). Of note, Q189E is a natural variant in both the avian infectious bronchitis virus (IBV) and the swine coronavirus, HKU15 CoV, widely detected in pigs in Asia and North America and of pandemic concern due to its ability to replicate in human cells (Edwards, Yount et al. 2020). PF-07321332 may have reduced efficacy against these concerning homologs due to its decreased interactions with Q189E M^{pro}.

In addition to the impacts on side-chain properties, mutations in M^{pro} may also impact resistance through changes in main-chain conformation and dynamics, particularly in loops. In-depth structural analyses will be important to extensively assess the potential impacts of mutations on resistance through these mechanisms. Of note, mutations at N142 appears of particular interest for further investigation of conformational changes that may impact resistance evolution. N142 is mutation tolerant and located in a loop over the P1 position of the substrate. The lactam ring on PF-07321332 protrudes outside of the substrate envelope at this location (Shaqra, Zvornicanin et al. 2022). Mutations at position 142 should be readily available to M^{pro} evolution and appear likely to influence loop conformation at a site where PF-07321332 extends beyond the substrate envelope. Together these observations suggest that N142 warrants further attention as a potential contributor to drug resistance.

Discussion

During the SARS-CoV-2 pandemic, intensive efforts have been launched to rapidly develop vaccines and anti-viral drugs to improve human health. In this study, we provide comprehensive functional information on a promising therapeutic target, M^{pro}, with the hopes that these results will be useful in the design of more effective and long-lasting anti-SARS-CoV-2 drugs. We built three yeast screens to

measure the functional effects of all individual amino acid changes in M^{pro}. The resulting fitness landscapes provide information on residues to both target and avoid in the drug design process. In the active site, the primary current target of M^{pro} inhibitors, our results indicate both mutation-sensitive positions that provide ideal anchors for inhibitors, and mutation-tolerant positions to avoid. Among the positions to avoid, Q189 is noteworthy because it forms hydrogen bonds directly with substrates (MacDonald, Frey et al. 2021, Shaqra, Zvornicanin et al. 2022), contacts promising M^{pro} drugs such as PF-07321332 (Cho, Rosa et al. 2021, Owen, Allerton et al. 2021, Zhao, Fang et al. 2021), is a natural variant in coronaviruses of future pandemic concern, and is surprisingly tolerant of mutations in our screen.

We found that the functional scores measured from all three distinct screens were highly correlated, that they identified known critical M^{pro} residues, and that clinical variants were overwhelmingly functional, indicating that the scores successfully capture key biochemical and functional properties of M^{pro}. However, there are a couple of caveats that should be kept in mind when utilizing these data sets. For example, we do not fully understand how M^{pro}'s biochemical function relates to viral fitness. Having some M^{pro} function is essential to the virus, so mutations that destroy M^{pro} function will form non-functional viruses. Function-fitness relationships tend to be non-linear (Heinrich and Rapoport 1974, Kacser and Fell 1995, Jiang, Mishra et al. 2013) and it may be likely that M^{pro} function must be decreased by a large amount in order to cause measurable changes in viral replication efficiency. This relationship between M^{pro} function and SARS-CoV-2 fitness would need to be determined in order to translate our functional scores to fitness scores. Additionally, our TF and FRET screens quantify cleavage at one defined site (Nsp4/5) and it may be important to analyze all sites in order to fully understand the selection pressures acting on M^{pro}. Another important caveat is that our fitness landscape captures single amino acid changes and therefore does not provide information on the potential interdependence or epistasis between double and higher order mutations. Information regarding epistasis will be important for accurately predicting the impacts of multiple mutations on fitness. Despite these caveats, the similarity in fitness landscapes for the TF and FRET screens with the yeast growth screen suggests that all three capture fundamental and general aspects of M^{pro} selection. In addition, the high function of almost all naturally occurring substitutions in the diversity of natural M^{pro} sequences indicates that estimates of fitness effects in different genetic backgrounds can be made based on our results.

We believe that our results will be a useful guide for the continuing intense efforts to develop drugs that target M^{pro} and the interpretation of future M^{pro} evolution in the face of drug pressure. In particular, our results identify amino acid changes that can be functionally tolerated by M^{pro} that are likely to disrupt binding to inhibitors. In a recent study, Shaqra, Schiffer and colleagues mapped the M^{pro} substrate envelope; locations where the inhibitors protrude from this envelope is an indicator of susceptibility to resistance mutations (Shaqra, Zvornicanin et al. 2022). The information in these two studies provides a new view into resistance evolution that can be incorporated into ongoing drug

474 design efforts. Locations in the active site as well as at a likely allosteric site that cannot readily evolve
475 without compromising function are ideal targets for anchoring inhibitors with reduced potential to
476 evolve drug resistance.

477

478 Our next steps involve developing efficient strategies for assaying M^{pro} fitness landscapes in the
479 presence of potential inhibitors in order to define structure-resistance relationships. This would
480 provide critical guidance for reducing the likelihood of resistance at earlier stages of drug development
481 than is currently possible. For example, it would identify inhibitors with the least likelihood of
482 developing resistance. It would also provide the potential for identifying inhibitors with non-
483 overlapping resistance profiles that if used in combination would not be susceptible to resistance from
484 an individual mutation. There are technical hurdles to overcome in using our yeast-based screens to
485 investigate resistance because many small-molecules are ineffective due to poor permeability and/or
486 export from yeast. We are assessing strategies to both increase the druggability of yeast and porting
487 our assays to mammalian cells (Chinen, Hamada et al. 2017). The results from our current work on M^{pro}
488 in yeast as well as previous studies using fitness landscapes to analyze drug resistance in other proteins
489 (Deng, Huang et al. 2012, Choi, Landrette et al. 2014, Firnberg, Labonte et al. 2014, Ma, Boucher et al.
490 2017) indicates a strong potential of these approaches to improve our understanding and ability to
491 combat resistance evolution.

492

493 **Materials and methods**

494

495

Key Resources Table				
Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Gene (SARS-CoV-2)	ORF1ab/ nsp5A-B	NIH GenBank	NC_045512	M ^{pro}
Strain, Strain background (<i>Saccharomyces cerevisiae</i>)	W303	Saccharomy ces Genome Database	GenBank JRIU000000000	

Antibody	anti-his tag HRP-labelled (Mouse monoclonal)	R&D systems	CAT#: MAB050H	WB (1:4000)
Recombinant DNA reagent	Barcoded UbM ^{pro} plasmid library	This paper	p416LexA- UbM ^{pro} (lib)- N18	See Materials and Methods section “Generating mutant libraries”
Recombinant DNA reagent	Barcoded WT UbM ^{pro} plasmid	This paper	p416LexA- UbM ^{pro} (WT)- N18	See Materials and Methods section “Construction of WT Ub-M ^{pro} vector”
Recombinant DNA reagent	C145A-M ^{pro} - his ₆ plasmid	This paper	p416LexA- UbM ^{pro} (C145A)-his	See Materials and Methods section “Analysis of M ^{pro} expression”
Recombinant DNA reagent	pCyPet-His	Addgene	#14040	
Recombinant DNA reagent	pYPet-His	Addgene	#14031	
Recombinant DNA reagent	CyPet-MproCS- YPet fusion gene	This paper		See Materials and Methods section “Generating FRET strain”
Recombinant DNA reagent	pDK-ATC	PMID 28660202		Integrative bidirectional plasmid with TEF and CUP promoters
Recombinant DNA reagent	pDK-ATG	PMID 28660202		Integrative bidirectional plasmid with TEF and GPD promoters

Recombinant DNA reagent	DBD-M ^{pro} CS-AD fusion gene	This paper		See Materials and Methods section “Generating split TF strain”
Commercial assay or kit	KAPA SYBR FAST qPCR Master Mix	Kapa Biosystems	KK4600	
Commercial assay or kit	BCA protein assay kit	Pierce	CAT# 23225	
Chemical compound, drug	β-Estradiol	Sigma Aldrich	E2768	
Software, algorithm	Scripts to tabulate variant counts	This paper	https://github.com/JuliaFlynn/BolonLab	See Materials and Methods section “Analysis of Illumina sequencing data”
Software, algorithm	Scripts to associate barcodes with variants	This paper	https://github.com/JuliaFlynn/PacBio_barcode_association	See Materials and Methods section “Barcode Association”
Software, algorithm	GraphPad Prism 9	Graphpad.com	RRID: SCR_008520	
Software, algorithm	Flowjo v.10.8.0	BD Biosciences	RRID: SCR_008520	
Software, algorithm	Pymol v. 2.5.2	Schrödinger	RRID: SCR_000305	
Software, algorithm	Matplotlib	http://matplotlib.sourceforge.net	RRID: SCR_008624	

Sequenced-based reagent	Sequencing primers	This paper		See Supplemental file 1
Sequenced-based reagent	Site-directed mutagenesis primers	This paper		See Supplemental file 1

496

497 **Construction of WT Ub-M^{pro} vector (p416LexA_UbM^{pro}(WT)_B112)**

498 The Ubiquitin-M^{pro} gene fusion was constructed using overlapping PCR of the yeast ubiquitin gene and
499 SARS-CoV-2 M^{pro} gene (Jin, Du et al. 2020) and was inserted into the pRS416 vector after digestion with
500 SpeI and BamHI. Four LexA boxes were amplified from the LexAbox4_citrine plasmid (FRP793_insul-
501 (lexA-box)4-PminCYC1-Citrine-TCYC1 was a gift from Joerg Stelling; Addgene plasmid # 58434;
502 <http://n2t.net/addgene:58434>)(Ottoz, Rudolf et al. 2014) and inserted between the SacI and SpeI sites
503 upstream of the ubiquitin-M^{pro} gene. The LexA_ER_B112 transcription factor was amplified from
504 Addgene_58437 (FRP880_PACT1(-1-520)-LexA-ER-haB112-TCYC1 was a gift from Joerg Stelling;
505 Addgene plasmid # 58437; <http://n2t.net/addgene:58437>)(Ottoz, Rudolf et al. 2014) and inserted into
506 the KpnI site. The resulting vector is named (p416LexA-UbM^{pro}(WT)-B112). A destination vector was
507 generated by removing the M^{pro} sequence and replacing it with a restriction site for SphI.

508

509 **Generating mutant libraries**

510 The SARS-CoV-2 M^{pro} (ORF1ab polyprotein residues 3264-3569, GenBank code: MN908947.3) single
511 site variant library was synthesized by Twist Biosciences (twistbioscience.com) by massively parallel
512 oligonucleotide synthesis. In the library, each amino acid position was modified to all 19 amino acid
513 variants plus a premature termination encoded by a stop codon, using the preferred yeast codon for
514 each substitution. All 306 amino acids of M^{pro} were modified yielding 6120 total variants. Due to
515 challenges in construction, positions 27 and 28 were missing from the library. 35 bp of sequence
516 homologous to the destination vector was added to both termini of the library during synthesis to
517 enable efficient cloning. The library was combined via Gibson assembly (NEB) with the destination
518 vector. To avoid bottlenecking the library, sufficient transformations were performed to recover more
519 than 50 independent transformants for each designed M^{pro} variant in the library. To improve efficiency
520 and accuracy of deep sequencing steps during bulk competition, each variant of the library was tagged
521 with a unique barcode. A pool of DNA constructs containing a randomized 18 bp barcode sequence
522 (N18) was cloned into the NotI and Ascl sites upstream of the LexA promoter sequence via restriction
523 digestion, ligation and transformation into chemically competent *E. coli*. These experiments were

524 performed at a scale designed to have each M^{pro} variant represented by 10-20 unique barcodes. The
525 resulting library is named p416LexA-UbM^{pro}(lib)-B112.

526

527 **Barcode association**

528 To associate barcodes with M^{pro} variants, we digested the p416-UbM^{pro}(lib)-B112 plasmid upstream of
529 the N18 sequence and downstream of the M^{pro} sequence with NotI and SalI enzymes (NEB). The
530 resulting 1800 bp fragment containing the barcoded library was isolated by Blue Pippen selecting for a
531 1 to 4 kB range. Of note, we determined it was important to avoid PCR to prepare the DNA for PacBio
532 sequencing, as PCR led to up to 25% of DNA strands recombining, leading to widespread mismatch
533 between the barcode and M^{pro} variant. DNA was prepared for sequencing with the Sequel II Binding
534 Kit v2.1 and the libraries were sequenced on a Pacific Biosciences Sequel II Instrument using a 15-hour
535 data collection time, with a 0.4-hour pre-extension time (PacBio Core Enterprise, UMass Chan Medical
536 School, Worcester, MA). PacBio circular consensus sequences (CCS) were generated from the raw
537 reads using SMRTLink v.10.1 and standard Read-Of-Insert (ROI) analysis parameters. After filtering low-
538 quality reads (Phred scores < 10), the data was organized by barcode sequence using custom analysis
539 scripts that have been deposited on GitHub (<https://github.com>, see Key Resource Table). For each
540 barcode that was read more than three times, we generated a consensus of the M^{pro} sequence that we
541 compared to WT to call mutations.

542

543 As a control for library experiments, the WT Ub-M^{pro} gene was also barcoded with approximately 150
544 unique barcode sequences. The randomized 18 bp barcode sequence (N18) was cloned between the
545 NotI and Ascl sites upstream of the LexA promoter sequence in the p416LexA-Ub-M^{pro}(WT)-B112
546 vector with the goal of the WT sequence being represented by approximately 100 barcodes. The
547 barcoded region of the plasmid was amplified by PCR using the primers listed in Supplementary file 1
548 (for the WT barcoding it was not necessary to avoid strand recombination) and sequenced by EZ
549 Amplicon deep sequencing (www.genewiz.com).

550

551 **Generating split transcription factor strain**

552 The GFP reporter strain was generated by integration of GFP driven by a Gal1 promoter together with a
553 HIS3 marker into the HO genomic locus. The Gal4, Gal80 and Pdr5 genes were disrupted to create the
554 following strain: W303 *HO::Gal1-GFP-v5-His3; gal4::trp1; gal80::leu2 pdr5::natMX*.

555 The Gal4 DNA binding domain-M^{pro}CS-activation domain fusion gene (DBD-M^{pro}CS-AD) was generated
556 by overlapping PCR. The Gal4 DNA binding domain (DBD) was amplified by PCR with a forward primer
557 containing the EcoRI site and a reverse primer containing the extending M^{pro}CS overhang sequence.
558 The Gal4 activation domain (AD) was amplified by PCR with a forward primer containing the M^{pro}CS
559 overhang sequence and a reverse primer containing the SacI site (SacI_R). The DBD-M^{pro}CS-AD fusion

gene was generated using the overlapping DBD-M^{pro}CS and M^{pro}CS-AD products from above as templates and the EcoRI_F and SacI_R primers. The resulting DBD-M^{pro}CS-AD fusion gene was inserted between the EcoRI and SacI sites downstream of the CUP promoter in the integrative bidirectional pDK-ATC plasmid (kindly provided by D. Kaganovich) (Amen and Kaganovich 2017). The mCherry gene was subsequently cloned into the XhoI/BamHI sites downstream of the TEF promoter in the opposite orientation to create the plasmid pDK-CUP-DBD-M^{pro}CS-AD-TEF-mCherry. The fragment for genomic integration was generated by PCR with the primers listed in Supplementary file 1, was transformed into the reporter strain using LiAc/PEG transformation (Gietz, Schiestl et al. 1995), and successful integration of the module into the adenine biosynthesis gene was verified by PCR.

569

570 Bulk Split transcription factor (TF) competition experiment

Barcoded WT UbM^{pro} (p416LexA-UbM^{pro}(WT)-N18) plasmid was mixed with the barcoded UbM^{pro} library (p416LexA-UbM^{pro}(lib)-N18) at a ratio of 20-fold WT to the average library variant. The blended plasmid library was transformed using the lithium acetate procedure into the reporter strain (*W303 ade::CUP-DBD-M^{pro}CS-AD-TEF-mCherry; ho::gal1-gfp-v5-his3; gal4::trp1; gal80::leu2; pdr5::natMX*). Sufficient transformation reactions were performed to attain about 5 million independent yeast transformants representing a 50-fold sampling of the average barcode. Each biological replicate represents a separate transformation of the library. Following 12 hours of recovery in synthetic dextrose lacking adenine (SD-A), transformed cells were washed three times in SD-A-U media (SD lacking adenine and uracil to select for the presence of the M^{pro} variant plasmid) to remove extracellular DNA and grown in 500 mL SD-A-U media at 30°C for 48 hours with repeated dilution to maintain the cells in log phase of growth and to expand the library. At least 10⁷ cells were passed for each dilution to avoid population bottlenecks. Subsequently, the library was diluted to early log phase in 100 mL of SD-A-U, grown for two hours, the culture was split in half, and 125 nM β-estradiol (from a 10 mM stock in 95% ethanol, Sigma-Aldrich) was added to one of the cultures to induce Ub-M^{pro} expression. Cultures with and without β-estradiol were grown with shaking at 180 rpm for 6 hours at which point samples of ~10⁷ cells were collected for FACS analysis.

587

588 FACS sorting of TF screen yeast cells

A sample of 10⁷ cells were washed three times with 500 μL of Tris-Buffered Saline containing 0.1% Tween and 0.1% bovine serum albumin (TBST-BSA). Cells were diluted to 10⁶/mL and transferred to polystyrene FACS tubes. Samples were sorted for GFP and mCherry expression on a FACS Aria II cell sorter with all cells expressing cut TF (low GFP expression) in one population and uncut TF (high GFP expression) in a second population. To ensure adequate library coverage, we sorted at least 1.5 million cells of each population and collected them in SD-A-U media. For the first replicate, sorted yeast cells were amplified in 20 mL SD-U-A media for 10 hours at 30°C. These yeast samples were collected by centrifugation and cell pellets were stored at -80°C. It was observed that different populations of cells recovered at different rates during this amplification period, so in the second replicate cells were

immediately spun down and stored at -80°C. Functional scores between the two replicates correlated well indicating that the amplification step was dispensable.

600

601

602 **Generating FRET strain**

603 The YPet-CyPet FRET pair is a YFP-CFP fluorescent protein pair that has been fluorescently optimized by
604 directed evolution for intracellular FRET (Nguyen and Daugherty 2005). The YPet- M^{pro}CS-CyPet fusion
605 gene was generated by overlapping PCR as follows. The CyPet gene was amplified by PCR from the
606 pCyPet-His vector (pCyPet-His was a gift from Patrick Daugherty; Addgene plasmid # 14030 ;
607 <http://n2t.net/addgene:14030>) with a forward primer containing the BamHI site (BamHI_F) and a
608 reverse primer containing the extending M^{pro}CS overhang sequence. The YPet gene was amplified by
609 PCR from the pYPet-His vector (pYPet-His was a gift from Patrick Daugherty; Addgene plasmid # 14031
610 ; <http://n2t.net/addgene:14031>) with a forward primer containing the extending M^{pro}CS overhang
611 sequence and a reverse primer containing the XhoI site (XhoI_R). The CyPet-M^{pro}CS-YPet fusion gene
612 was generated using the overlapping CyPet-M^{pro}CS and M^{pro}CS-YPet products from above as templates
613 and BamHI_F and XhoI_R primers. The resulting CyPet- M^{pro}CS-YPet gene was inserted between the
614 BamHI and XhoI sites downstream of the TEF promoter in the integrative bidirectional pDK-ATG
615 plasmid (kindly provided by D. Kaganovich) (Amen and Kaganovich 2017). The fragment for genomic
616 integration was generated by PCR with the primers listed in Supplementary file 1, was transformed into
617 W303 (*leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15*) using LiAc/PEG transformation (Gietz,
618 Schiestl et al. 1995), and successful integration of the module into the adenine biosynthesis gene was
619 verified by PCR.

620

621 **Bulk FRET competition experiment**

622 The plasmid library including the barcoded WT plasmid was transformed as above using the lithium
623 acetate procedure into W303 Ade::TEF-CyPet-M^{pro}CS-YPet cells. Sufficient transformation reactions
624 were performed to attain about 5 million independent yeast transformants representing a 50-fold
625 sampling of the average barcode. Cultures were grown and induced with β-estradiol as above for the
626 transcription factor screen with the exception that cells were induced for 1.5 hours. Samples of 10⁷
627 cells were collected for FACS analysis.

628

629 **FACS sorting of FRET screen yeast cells**

630 A sample of 10⁷ cells were washed three times with 500 μL of TBST-BSA. Cells were diluted to 10⁶/mL
631 and transferred to polystyrene FACS tubes. Samples were sorted for YFP and CFP expression on a FACS
632 Aria II cell sorter with all cells expressing cut FRET pair (low FRET) in one population and uncut FRET
633 pair (high FRET) in a second population. To ensure adequate library coverage, we sorted at least 3

million cells of each population and collected them in SD-A-U media. Yeast samples were collected by centrifugation and cell pellets were stored at -80°C.

Growth strain

The plasmid library including the barcoded WT plasmid was transformed as above using the lithium acetate procedure into W303 cells. Sufficient transformation reactions were performed to attain about 5 million independent yeast transformants representing a 50-fold sampling of the average barcode. Each biological replicate represents a separate transformation of the library. Following 12 hours of recovery in synthetic dextrose media (SD), transformed cells were washed three times in SD-U media (SD lacking uracil to select for the presence of the M^{pro} variant plasmid) to remove extracellular DNA and grown in 500 mL SD-U media at 30°C for 48 hours with repeated dilution to maintain the cells in log phase of growth (OD₆₀₀ = 0.05-1) and to expand the library. At least 10⁷ cells were passed for each dilution to avoid population bottlenecks. Subsequently, the library was diluted to early log phase (OD₆₀₀=0.05) in 100 mL of SD-U, grown for two hours, the culture was split in half, and 2 μM β-estradiol (from a 10 mM stock in 95% ethanol) was added to one of the cultures to induce Ub-M^{pro} expression. Cultures with and without β-estradiol were grown with shaking at 180 rpm for 16 hours with dilution after 8 hours to maintain growth in exponential phase. Samples of ~10⁸ cells were collected by centrifugation and cell pellets were stored at -80°C.

DNA preparation and sequencing

We isolated plasmid DNA from each FACS cell population and the time points from the growth experiment as described (Jiang, Mishra et al. 2013). Additionally, we sequenced the original barcoded plasmid library to evaluate the collateral effects on variants during the pre-selection library expansion stages. Purified plasmid DNA was linearized with *Ascl*. Barcodes were amplified with 22 cycles of PCR using Phusion polymerase (NEB) and primers that add Illumina adapter sequences and a 6 bp identifier sequence used to distinguish cell populations. PCR products were purified two times over silica columns (Zymo Research) and quantified using the KAPA SYBR FAST qPCR Master Mix (Kapa Biosystems) on a Bio-Rad CFX machine. Samples were pooled and sequenced on an Illumina NextSeq instrument in single-end 75 bp mode.

Analysis of Illumina sequencing data

We analyzed the Illumina barcode reads using custom scripts that have been deposited on GitHub (<https://github.com>, see Key Resource Table). Illumina sequence reads were filtered for Phred scores > 10 and strict matching of the sequence to the expected template and identifier sequence. Reads that passed these filters were parsed based on the identifier sequence. For each screen/cell population, each unique N18 read was counted. The unique N18 count file was then used to identify

the frequency of each mutant using the variant-barcode association table. To generate a cumulative count for each codon and amino acid variant in the library, the counts of each associated barcode were summed.

Determination of functional scores

To determine the functional score for each variant in the two FACS-based screens, the fraction of each variant in the cut and uncut windows was first calculated by dividing the sequencing counts of each variant in a window by the total counts in that window. The functional score was then calculated as the fraction of the variant in the cut window divided by the sum of the fraction of the variant in the cut and uncut windows. The functional score for the growth screen was calculated by the fraction of the variant at the 0 hour time point divided by the sum of the fraction of the variant in the 0 and 16 hour time points. Functional scores were not calculated for variants with less than 100 total reads. The functional scores were normalized setting the score for the average WT M^{pro} barcode as 1 and the average stop codon as 0. Both the unnormalized and normalized scores are reported in Figure 2 – source data 1. For comparison, the counts for the growth-based screen were fit to selection coefficients (slope of $\log_2(\text{variant}/\text{WT counts})$). We chose to report the functional scores as opposed to the selection coefficients in this paper so they would be directly comparable to the TF and FRET functional scores.

Analysis of M^{pro} expression and Ubiquitin removal by Western Blot

To facilitate analysis of expression levels of M^{pro} and examine effective removal of Ubiquitin, a his tag was fused to the C-terminus of M^{pro} to create the plasmid p416LexA-UbM^{pro}-his₆-B112. In addition, the C145A mutation was created by site-directed mutagenesis to ensure cleavage by Ub specific proteases and to reduce the toxicity caused by WT M^{pro} expression. W303 cells were transformed with the p416LexA-UbM^{pro}(C145A)-his₆ construct and the resulting yeast cells were grown to exponential phase in SD-U media at 30°C. 2 μM β-estradiol was added when indicated and cells were grown for an additional eight hours. 10⁸ yeast cells were collected by centrifugation and frozen as pellets at –80°C. Cells were lysed by vortexing the thawed pellets with glass beads in lysis buffer (50 mM Tris-HCl pH 7.5, 5 mM EDTA and 10 mM PMSF), followed by addition of 2% Sodium dodecyl sulfate (SDS). Lysed cells were centrifuged at 18,000 g for 1 min to remove debris, and the protein concentration of the supernatants was determined using a BCA protein assay kit (Pierce) compared to a Bovine Serum Albumin (BSA) protein standard. 15 μg of total cellular protein was resolved by SDS-PAGE, transferred to a PVDF membrane, and probed using an anti-his HRP-conjugated antibody (R&D systems). Purified M^{pro}-his₆ protein was a gift from the Schiffer laboratory. There is a slight size difference on the Western blot between the purified M^{pro}-his₆ protein and the C145A M^{pro}-his₆ in the yeast lysate. We do not completely understand the origin of this mobility shift, but possible causes are an abnormal gel shift due to the C145A mutation, a mobility difference due to buffer, nucleic acids or additional proteins in the lysate, or an unknown modification of M^{pro} in bacteria compared to yeast.

708

709 **Sequence and structure analysis**

710 Evolutionary conservation was calculated with an alignment of homologs from diverse species using
711 the ConSurf server (Ashkenazy H, Abadi, S.). The effects of single mutations on protein-ligand
712 interactions were predicted by calculating the binding affinity changes using PremPLI
713 (<https://lilab.iysw.suda.edu.cn/research/PremPLI/>) (Sun, T., Chen Y et al). The figures were generated
714 using Matplotlib (Hunter 2007), PyMOL and GraphPad Prism version 9.3.1.

715

716 **Identifying mutations in circulating SARS-COV-2 sequences**

717 The complete set of SARS-COV-2 isolate genome sequences was downloaded from the GISAID
718 database. The SARS-COV-2 M^{pro} reference sequence (NCBI accession NC_045512.2) was used as a
719 query in a tBLASTn search against the translated nucleotide sequences of these isolates to identify the
720 M^{pro} region and its protein sequence for each isolate, if present. M^{pro} sequences were discarded if they
721 contained 10 or more ambiguous “X” amino acids or had amino acid length less than 290. A multiple
722 sequence alignment was performed and for each of the twenty standard amino acids, the number of
723 times it was observed at each position in the M^{pro} sequence was calculated.

724

725 **Data availability**

726 Next generation sequencing data has been deposited to the NCBI short read archive (PRJNA842255).
727 Tabulated raw counts of all variants in all replicates are included in Figure 2 – source data 1. Source
728 data files have been provided in Figure 2, 3, 4 and 5.

729

730 **Main Figure Legends**

731

732 **Figure 1. Experimental strategy to measure the function of all individual mutations of M^{pro}.** A. FRET-
733 based reporter screen. M^{pro} variants were sorted based on their ability to cleave at the M^{pro} cut-site,
734 separating the YFP-CFP FRET pair. Cells were separated by FACS into cleaved (low FRET) and uncleaved
735 (high FRET) populations. B. Split transcription factor screen. M^{pro} variants were sorted based on their
736 ability to cleave at the M^{pro} cut-site, separating the DNA binding domain (DBD) and activation domain
737 (AD) of the Gal4 transcription factor. The transcription factor drives GFP expression from a galactose
738 promoter. Cells were separated by FACS into cleaved (low GFP expression) and uncleaved (high GFP
739 expression) populations. C. Growth screen. Yeast cells expressing functional M^{pro} variants that cleave
740 essential yeast proteins grow slowly and are depleted in bulk culture, while yeast cells expressing non-

functional M^{pro} variants are enriched. D. Barcoding strategy to measure frequency of all individual mutations of M^{pro} in a single experiment.

743

Figure 2. M^{pro} functional scores are reproducible, and variants can be clearly distinguished based on function. A. Correlation between biological replicates of functional scores of all M^{pro} variants for each screen. Red line indicates best fit. B. Distribution of functional scores for all variants (grey), stop codons (red) and WT barcodes (blue) in each screen. C. The functional scores for all variants (grey) and stop codons (red) at each position of M^{pro} in the FRET screen. D. Distribution of all functional scores (grey) in each screen. Functional scores are categorized as WT-like, intermediate, or null based on the distribution of WT barcodes (blue) and stop codons (red) in each screen. See Figure 2 – source data 1

Figure 3. Heatmap representation of the M^{pro} functional scores measured in the FRET screen (replicate 1). Arrows represent positions that form β -sheets, coils represent α -helices, and red triangles indicate the catalytic dyad residues H41 and C145

Figure 4. Functional scores reflect fundamental biophysical constraints of M^{pro}. A. Heatmap representation of the average functional score at each position (excluding stops) in replicate 1 of each screen (see Figure 4 – source data 1). B. The average functional score at each position mapped to M^{pro} structure for each screen. The Nsp4/5 substrate peptide is shown in green (PDB 7T70). C. The average functional score at each position compared between the three screens. The diagonal is indicated with a blue dashed line. D. Comparison between relative catalytic rates measured independently in various studies and functional scores measured in each screen (see Figure 4 – source data 2). Each graph is fit with a non-linear and linear regression with the best of the two fits represented with a black solid line and the worst fit represented with a red dashed line. The non-linear regression is fit to the equation $Y = Y_m - (Y_0 - Y_m)e^{-kx}$.

764

Figure 5. Functional scores indicate that natural amino acid variants of M^{pro} are generally fit. A. Comparison of functional scores in the FRET screen (left panel) and growth screen (right panel) to the number of observations among clinical samples. All missense mutations excluding stops are indicated with black circles and stop codons are indicated with red x's. (See Figure 5 – source data 1) B. The distribution of functional scores of all variants in the FRET and growth screens compared to the observed clinically-relevant M^{pro} variants (human SARS-CoV-2 variants, blue), 12 amino acid differences between SARS-CoV-2 and SARS-CoV-1 (green), and the different amino acids in a broad sample of M^{pro} SARS-CoV-2 homologs (natural variants, pink). Distributions are significantly different as measured by a two-sample Kolmogorov-Smirnov (KS) (All FRET vs. human SARS-CoV-2 variants: N = 6044, 289, p<0.0001, D = 0.3258; All FRET vs. SARS-CoV-1 variants: N=6044, 12, p=0.0398, D=0.4223; All FRET vs. natural variants: N = 6044, 1205, p<0.0001, D = 0.2984; All Growth vs. human SARS-CoV-2 variants: N = 6044, 289, p<0.0001, D = 0.3938; All growth vs. SARS-CoV-1 variants: N=6044, 12, p=0.0024, D=0.5533; All growth vs. natural variants: N=6044,1205, p<0.0001, D = 0.3462) C. Histogram of functional scores of all variants (grey) compared to that of human SARS-CoV-2 variants (blue), SARS-CoV-1 variants (green), and natural variants (pink).

Figure 6. Structural distribution of M^{pro} positions that are intolerant to mutation. A. M^{pro} positions that are intolerant of mutations with 17 or more substitutions having null-like function are represented by red spheres on chain A (shown in grey) and pink spheres on chain B (shown in white). The Nsp4/5 substrate peptide is shown in green (PDB 7T70). B. Representation of a cluster of the mutation-intolerant positions (red spheres) at a site distal to the active site. C. A cluster of mutation-intolerant residues (red spheres) appear to be part of a distal communication network between the active site and the dimerization interface. D. Comparison of the average functional score of each position to conservation observed in a broad sample of SARS-CoV-2 M^{pro} homologs. The 24 mutation-intolerant positions shown as red spheres in part A are highlighted in red. Positions exhibiting the strongest evolutionary conservation exhibit a broad range of experimental sensitivity to mutation while the most evolutionary variable positions are experimentally tolerant to mutations.

Figure 7. Substrate and inhibitor binding sites are variably sensitive to mutation. A. All M^{pro} positions that contact the Nsp4/5 substrate peptide are represented in spheres and colored by their average FRET functional score (left panel) and growth functional score (right panel) (PDB 7T70). The Nsp4/5 peptide is shown in green. B. M^{pro} positions that form hydrogen bonds with the Nsp4/5 substrate are shown in sticks and colored by their average FRET functional score (left panel) and growth functional score (right panel) (PDB 7T70). Oxygens are shown in red and nitrogens in cyan. Water molecules are represented as red spheres and hydrogen bonds as yellow dashed lines. C. M^{pro} positions shown to contact over 185 inhibitors in crystal structures (Cho, Rosa et al. 2021) are shown in sticks and are colored by their average FRET functional score (left panel) and average growth functional score (right panel). Shown is a representative structure of M^{pro} bound to the N3 inhibitor (PDB 6LU7) (Jin, Du et al. 2020). The N3 inhibitor is shown in green, oxygens in red, and nitrogens in cyan. D. M^{pro} positions that form hydrogen bonds with the Pfizer inhibitor, PF-07321332, are represented by sticks and colored by their average FRET functional score (left panel) or growth functional score (right panel) (PDB 7VH8) (Owen, Allerton et al. 2021, Zhao, Fang et al. 2021). PF-07321332 is shown in green, oxygens in red, nitrogens in cyan, fluorines in pink. Hydrogen bonds less than 4 Å are represented with thick yellow dashed lines and greater than 4 Å with a thin yellow dashed line. The table below lists the mutations with highest potential for being resistant against PF-07321332.

Supplemental Figure Legends

Figure 1 – figure supplement 1. M^{pro} expression in cells harboring the LexA-UbM^{pro} plasmid construct. A. Yeast cells transformed with a plasmid expressing C145A Ub-M^{pro}-his₆ under the LexA promoter were grown to exponential phase followed by the addition of 2 μM β-estradiol to induce expression for 8 hours. M^{pro} levels were monitored by Western blot with an anti-his₆ antibody and the correct size was measured against purified M^{pro}-his₆ protein (control). B. The plasmid expressing WT Ub-M^{pro} under control of the LexA promoter was transformed into cells expressing the split transcription factor. Cells were grown to exponential phase followed by addition of the indicated concentration of β-estradiol. Cell density was monitored based on absorbance at 600 nm at the times indicated (left panel). At the same time points, cells were washed, diluted to equal cell number, and GFP fluorescence was monitored at 525 nm (right panel). C. FACS analysis of cells expressing the CFP-

819 M^{pro}CS-YFP FRET pair and either WT Ub-M^{pro} (left) or C145A Ub-M^{pro} (right). Cell samples were
 820 collected before and after induction of M^{pro} expression with 125 nM β -estradiol for 1.5 hours. D. FACS
 821 analysis of cells expressing the split transcription factor separated by the M^{pro} cut-site and either WT
 822 Ub-M^{pro} (left) or C145A Ub-M^{pro} (right). Cell samples were collected before and after induction of M^{pro}
 823 expression with 125 nM β -estradiol for 6 hours. E. Distribution of number of barcodes associated with
 824 all M^{pro} variants. F. Correlation between total counts of each variant in the M^{pro} plasmid library
 825 (plasmid count) and the total counts of that variant before M^{pro} induction (pre-induction count). G.
 826 M^{pro} variants present at low frequency in the library showed a wider variance between plasmid library
 827 counts and counts in the pre-induction sample, consistent with lower sampling.

828 **Figure 2 – figure supplement 1. Cumulative frequency distributions for all variants (grey), stops (red)**
 829 **and WT barcodes (blue) for all three screens.**

830 **Figure 3 – figure supplement 1. Heatmap representation of scores from the TF screen (replicate 1).**
 831 Arrows represent positions that form beta sheets, coils represent α -helices, and red triangles indicate
 832 the catalytic dyad residues H41 and C145.

833 **Figure 3 – figure supplement 2. Heatmap representation of scores from the growth screen (replicate**
 834 **1).** Arrows represent positions that form beta sheets, coils represent α -helices, and red triangles
 835 indicate the catalytic dyad residues H41 and C145.

836 **Figure 5 – figure supplement 1. Functional scores indicate that natural amino acid variants of M^{pro}**
 837 **are generally fit.** A. Comparison of functional scores in the TF screen to the number of observations
 838 among clinical samples. All missense mutations excluding stops are indicated with black circles and
 839 stop codons are indicated with red x's. (See Figure 5 – source data 1) B. The distribution of functional
 840 scores of all variants in the TF screen compared to the observed clinically-relevant M^{pro} variants
 841 (human SARS-CoV-2 variants, blue), 12 amino acid differences between SARS-CoV-2 and SARS-CoV-1
 842 (green), and the different amino acids in a broad sample of M^{pro} SARS-CoV-2 homologs (natural
 843 variants, pink). Distributions are significantly different as measured by a two-sample Kolmogorov-
 844 Smirnov (KS) (All TF vs. human SARS-CoV-2 variants: N = 6038, 289, p<0.0001, D = 0.2845; All TF vs.
 845 SARS-CoV-1 variants: N=6038, 12, p=0.0196, D=0.4589; All TF vs. natural variants: N = 6038, 1205,
 846 p<0.0001, D = 0.2608).

847 **Figure 6 – figure supplement 1. Comparison of the average TF functional score of each position to**
 848 **conservation observed in a broad sample of SARS-CoV-2 M^{pro} homologs.**

849 **Figure 7- figure supplement 1.** A. All M^{pro} positions that contact the Nsp4/5 substrate peptide are
 850 represented in spheres and colored by their average TF functional score (PDB 7T70). The Nsp4/5
 851 peptide is shown in green. B. M^{pro} positions that form hydrogen bonds with the Nsp4/5 substrate are
 852 shown in sticks and colored by their average TF functional score (PDB 7T70). Oxygens are shown in red
 853 and nitrogens in cyan. Water molecules are represented as red spheres and hydrogen bonds as yellow
 854 dashed lines. C. M^{pro} positions shown to contact over 185 inhibitors in crystal structures (Cho, Rosa et
 855 al. 2021) are shown in sticks and are colored by their average TF functional score. Shown is a
 856 representative structure of M^{pro} bound to the N3 inhibitor (PDB 6LU7) (Jin, Du et al. 2020). The N3

inhibitor is shown in green, oxygens in red, and nitrogens in cyan. D. Heatmap representation of functional scores for the FRET screen (left panel), TF screen (middle panel) and the growth screen (right panel) at key inhibitor-contact positions as illustrated in Figure 7c. E. M^{pro} positions that form hydrogen bonds with the Pfizer inhibitor, PF-07321332, are represented by sticks and colored by their average TF functional score (PDB 7VH8) (Owen, Allerton et al. 2021, Zhao, Fang et al. 2021). PF-07321332 is shown in green, oxygens in red, nitrogens in cyan. Hydrogen bonds less than 4 Å are represented with thick yellow dashed lines and greater than 4 Å with a thin yellow dashed line.

Figure 2 – source data 1. Sequencing counts and functional scores for each amino acid of M^{pro} in both replicates of all three screens. For each data set, the sequencing counts, unnormalized functional scores, and normalized functional scores (normalized to average stop = 0, average wild-type barcode = 1) are reported. For the growth screens, the selection coefficients are also reported. All figures in this paper use the data from replicate 1 of each screen.

Figure 4 – source data 1. Average functional score (excluding stops) at each position of M^{pro} in replicate 1 of each screen.

Figure 4- source data 2. Comparison of previously measured relative catalytic rates of individual mutations to functional scores.

Figure 5 – source data 1. Frequency at which the clinical variants of the M^{pro} gene have been observed.

Supplementary file 1. List of oligomers used in this study.

Acknowledgements

This work was sponsored by Novartis Institutes for BioMedical Research. We would like to thank the UMass Chan Medical School Pacific Biosciences Core Enterprise for providing the PacBio NGS services and the UMass Chan Medical School Flow Cytometry Core Facility for providing the FACS services. We would also like to thank Ala Shaqra, Sarah Zvornicanin, and Qiu Yu Huang for conceptual discussions of the manuscript.

890

891 **Competing interests**

892 DTB, SAM, and DD are employees of Novartis Institutes for Biomedical Research

893

894

895 **References**

896

- 897 Alalam, H., S. Sigurdardottir, C. Bourgard, I. Tiukova, R. D. King, M. Grotli and P. Sunnerhagen (2021). "A Genetic
898 Trap in Yeast for Inhibitors of SARS-CoV-2 Main Protease." mSystems: e0108721.
- 899 Amen, T. and D. Kaganovich (2017). "Integrative modules for efficient genome engineering in yeast." Microb Cell
900 **4**(6): 182-190.
- 901 Bachmair, A., D. Finley and A. Varshavsky (1986). "In vivo half-life of a protein is a function of its amino-terminal
902 residue." Science **234**(4773): 179-186.
- 903 Barrila, J., U. Bacha and E. Freire (2006). "Long-range cooperative interactions modulate dimerization in SARS
904 3CLpro." Biochemistry **45**(50): 14908-14916.
- 905 Bhat, Z. A., D. Chitara, J. Iqbal, B. S. Sanjeev and A. Madhumalar (2021). "Targeting allosteric pockets of SARS-
906 CoV-2 main protease M(pro)." J Biomol Struct Dyn: 1-16.
- 907 Biering, S. B., E. Van Dis, E. Wehri, L. H. Yamashiro, X. Nguyenla, C. Dugast-Darzacq, T. G. W. Graham, J. R.
908 Stroumza, G. R. Golovkine, A. W. Roberts, D. M. Fines, J. N. Spradlin, C. C. Ward, T. Bajaj, D. Dovala, U. Schulze-
909 Gamen, R. Bajaj, D. M. Fox, M. Ott, N. Murthy, D. K. Nomura, J. Schaletzky and S. A. Stanley (2021). "Screening a
910 Library of FDA-Approved and Bioactive Compounds for Antiviral Activity against SARS-CoV-2." ACS Infect Dis
911 **7**(8): 2337-2351.
- 912 Boucher, J. I., D. N. A. Bolon and D. S. Tawfik (2019). "Quantifying and understanding the fitness effects of
913 protein mutations: Laboratory versus nature." Protein Sci **28**(3): 671.
- 914 Bzowka, M., K. Mitusinska, A. Raczynska, A. Samol, J. A. Tuszyński and A. Gora (2020). "Structural and
915 Evolutionary Analysis Indicate That the SARS-CoV-2 Mpro Is a Challenging Target for Small-Molecule Inhibitor
916 Design." Int J Mol Sci **21**(9).
- 917 Chen, S., J. Zhang, T. Hu, K. Chen, H. Jiang and X. Shen (2008). "Residues on the dimer interface of SARS
918 coronavirus 3C-like protease: dimer stability characterization and enzyme catalytic activity analysis." J Biochem
919 **143**(4): 525-536.
- 920 Cheng, S. C., G. G. Chang and C. Y. Chou (2010). "Mutation of Glu-166 blocks the substrate-induced dimerization
921 of SARS coronavirus main protease." Biophys J **98**(7): 1327-1336.
- 922 Chinen, T., K. Hamada, A. Taguchi, Y. Asami, K. Shiomi, Y. Hayashi and T. Usui (2017). Multidrug Sensitive Yeast
923 Strains, Useful Tools for Chemical Genetics. The Yeast Role in Medical Applications. W. M. H. Abdulkhair.
- 924 Cho, E., M. Rosa, R. Anjum, S. Mehmood, M. Soban, M. Mujtaba, K. Bux, S. T. Moin, M. Tanweer, S. Dantu, A.
925 Pandini, J. Yin, H. Ma, A. Ramanathan, B. Islam, A. Mey, D. Bhowmik and S. Haider (2021). "Dynamic Profiling of
926 beta-Coronavirus 3CL M(pro) Protease Ligand-Binding Sites." J Chem Inf Model **61**(6): 3058-3073.
- 927 Choi, J., S. F. Landrette, T. Wang, P. Evans, A. Bacchiocchi, R. Bjornson, E. Cheng, A. L. Stiegler, S. Gathiaka, O.
928 Acevedo, T. J. Boggon, M. Krauthammer, R. Halaban and T. Xu (2014). "Identification of PLX4032-resistance
929 mechanisms and implications for novel RAF inhibitors." Pigment Cell Melanoma Res **27**(2): 253-262.
- 930 Chou, C. Y., H. C. Chang, W. C. Hsu, T. Z. Lin, C. H. Lin and G. G. Chang (2004). "Quaternary structure of the
931 severe acute respiratory syndrome (SARS) coronavirus main protease." Biochemistry **43**(47): 14958-14970.
- 932 Deng, Z., W. Huang, E. Bakkalbasi, N. G. Brown, C. J. Adamski, K. Rice, D. Muzny, R. A. Gibbs and T. Palzkill (2012).
933 "Deep sequencing of systematic combinatorial libraries reveals beta-lactamase sequence constraints at high
934 resolution." J Mol Biol **424**(3-4): 150-167.
- 935 Edwards, C. E., B. L. Yount, R. L. Graham, S. R. Leist, Y. J. Hou, K. H. Dinnon, 3rd, A. C. Sims, J. Swanstrom, K.
936 Gully, T. D. Scobey, M. R. Cooley, C. G. Currie, S. H. Randell and R. S. Baric (2020). "Swine acute diarrhea
937 syndrome coronavirus replication in primary human cells reveals potential susceptibility to infection." Proc Natl
938 Acad Sci U S A **117**(43): 26915-26925.
- 939 Estrada, E. (2020). "Topological analysis of SARS CoV-2 main protease." Chaos **30**(6): 061102.

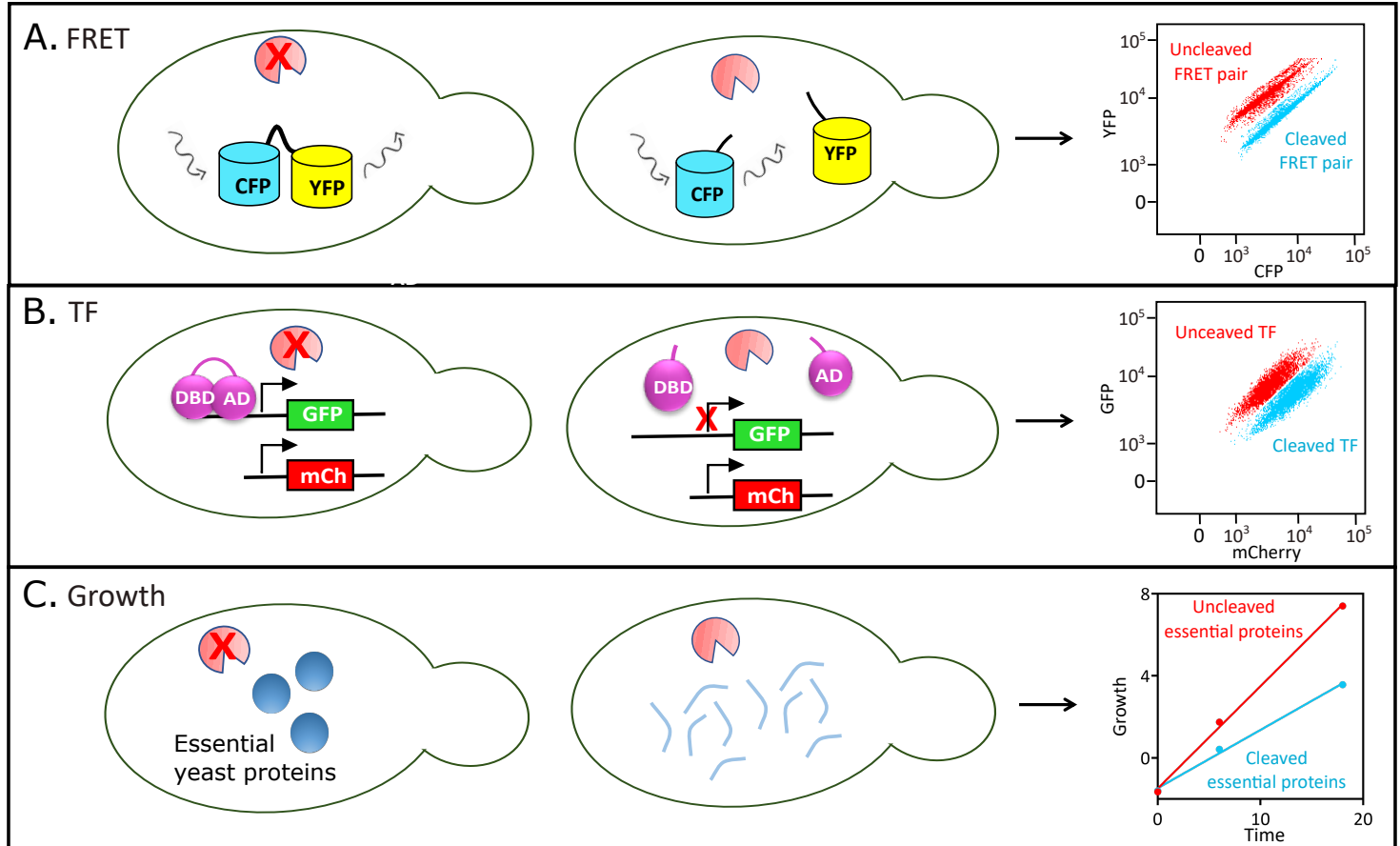
940 Fan, K., P. Wei, Q. Feng, S. Chen, C. Huang, L. Ma, B. Lai, J. Pei, Y. Liu, J. Chen and L. Lai (2004). "Biosynthesis,
 941 purification, and substrate specificity of severe acute respiratory syndrome coronavirus 3C-like proteinase." J
 942 Biol Chem **279**(3): 1637-1642.
 943 Firnberg, E., J. W. Labonte, J. J. Gray and M. Ostermeier (2014). "A comprehensive, high-resolution map of a
 944 gene's fitness landscape." Mol Biol Evol **31**(6): 1581-1592.
 945 Fischer, C., N. A. Veprek, Z. Peitsinis, K. Ruhmann, C. Yang, J. N. Spradlin, D. Dovala, D. K. Nomura, Y. Zyang and F.
 946 Trauner (2021). "De novo Design of SARS-CoV-2 Main Protease Inhibitors." Synlett: [https://doi.org/10.1055/a-](https://doi.org/10.1055/a-1582-0243)
 947 1582-0243.
 948 Fowler, D. M. and S. Fields (2014). "Deep mutational scanning: a new style of protein science." Nat Methods
 949 **11**(8): 801-807.
 950 Gietz, R. D., R. H. Schiestl, A. R. Willems and R. A. Woods (1995). "Studies on the transformation of intact yeast
 951 cells by the LiAc/SS-DNA/PEG procedure." Yeast **11**(4): 355-360.
 952 Hegyi, A., A. Friebe, A. E. Gorbalenya and J. Ziebuhr (2002). "Mutational analysis of the active centre of
 953 coronavirus 3C-like proteases." J Gen Virol **83**(Pt 3): 581-593.
 954 Heinrich, R. and T. A. Rapoport (1974). "A linear steady-state treatment of enzymatic chains. General properties,
 955 control and effector strength." Eur J Biochem **42**(1): 89-95.
 956 Herold, J., T. Raabe, B. Schelle-Prinz and S. G. Siddell (1993). "Nucleotide sequence of the human coronavirus
 957 229E RNA polymerase locus." Virology **195**(2): 680-691.
 958 Hietpas, R. T., J. D. Jensen and D. N. Bolon (2011). "Experimental illumination of a fitness landscape." Proc Natl
 959 Acad Sci U S A **108**(19): 7896-7901.
 960 Hsu, M. F., C. J. Kuo, K. T. Chang, H. C. Chang, C. C. Chou, T. P. Ko, H. L. Shr, G. G. Chang, A. H. Wang and P. H.
 961 Liang (2005). "Mechanism of the maturation process of SARS-CoV 3CL protease." J Biol Chem **280**(35): 31257-
 962 31266.
 963 Hsu, W. C., H. C. Chang, C. Y. Chou, P. J. Tsai, P. I. Lin and G. G. Chang (2005). "Critical assessment of important
 964 regions in the subunit association and catalytic action of the severe acute respiratory syndrome coronavirus
 965 main protease." J Biol Chem **280**(24): 22741-22748.
 966 Hunter, J. D. (2007). "Matplotlib: A 2D Graphics Environment". Computing in Science and Engineering **9**(3): 90-
 967 95.
 968 Jiang, L., P. Mishra, R. T. Hietpas, K. B. Zeldovich and D. N. Bolon (2013). "Latent effects of Hsp90 mutants
 969 revealed at reduced expression levels." PLoS Genet **9**(6): e1003600.
 970 Jin, Z., X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Y. Duan, J. Yu, L. Wang, K. Yang, F.
 971 Liu, R. Jiang, X. Yang, T. You, X. Liu, X. Yang, F. Bai, H. Liu, X. Liu, L. W. Guddat, W. Xu, G. Xiao, C. Qin, Z. Shi, H.
 972 Jiang, Z. Rao and H. Yang (2020). "Structure of M(pro) from SARS-CoV-2 and discovery of its inhibitors." Nature
 973 **582**(7811): 289-293.
 974 Johnston, S. A., M. J. Zavortink, C. Debouck and J. E. Hopper (1986). "Functional domains of the yeast regulatory
 975 protein GAL4." Proc Natl Acad Sci U S A **83**(17): 6553-6557.
 976 Kacser, H. and D. A. Fell (1995). "The control of flux: 21 years on." Biochemical Society Transactions **23**(2): 341-
 977 366.
 978 Kneller, D. W., G. Phillips, H. M. O'Neill, R. Jedrzejczak, L. Stols, P. Langan, A. Joachimiak, L. Coates and A.
 979 Kovalevsky (2020). "Structural plasticity of SARS-CoV-2 3CL M(pro) active site cavity revealed by room
 980 temperature X-ray crystallography." Nat Commun **11**(1): 3202.
 981 Ksiazek, T. G., D. Erdman, C. S. Goldsmith, S. R. Zaki, T. Peret, S. Emery, S. Tong, C. Urbani, J. A. Comer, W. Lim, P.
 982 E. Rollin, S. F. Dowell, A. E. Ling, C. D. Humphrey, W. J. Shieh, J. Guarner, C. D. Paddock, P. Rota, B. Fields, J.
 983 DeRisi, J. Y. Yang, N. Cox, J. M. Hughes, J. W. LeDuc, W. J. Bellini, L. J. Anderson and S. W. Group (2003). "A novel
 984 coronavirus associated with severe acute respiratory syndrome." N Engl J Med **348**(20): 1953-1966.
 985 Lim, K. P., L. F. Ng and D. X. Liu (2000). "Identification of a novel cleavage activity of the first papain-like
 986 proteinase domain encoded by open reading frame 1a of the coronavirus Avian infectious bronchitis virus and
 987 characterization of the cleavage products." J Virol **74**(4): 1674-1685.

988 Lin, P. Y., C. Y. Chou, H. C. Chang, W. C. Hsu and G. G. Chang (2008). "Correlation between dissociation and
 989 catalysis of SARS-CoV main protease." *Arch Biochem Biophys* **472**(1): 34-42.
 990 Ma, L., J. I. Boucher, J. Paulsen, S. Matuszewski, C. A. Eide, J. Ou, G. Eickelberg, R. D. Press, L. J. Zhu, B. J. Druker,
 991 S. Branford, S. A. Wolfe, J. D. Jensen, C. A. Schiffer, M. R. Green and D. N. Bolon (2017). "CRISPR-Cas9-mediated
 992 saturated mutagenesis screen predicts clinical drug resistance with improved accuracy." *Proc Natl Acad Sci U S A*
 993 **114**(44): 11751-11756.
 994 MacDonald, E. A., G. Frey, M. N. Namchuk, S. C. Harrison, S. M. Hinshaw and I. W. Windsor (2021). "Recognition
 995 of Divergent Viral Substrates by the SARS-CoV-2 Main Protease." *ACS Infect Dis* **7**(9): 2591-2595.
 996 Macnaughton, M. R. and M. H. Madge (1978). "The genome of human coronavirus strain 229E." *J Gen Virol*
 997 **39**(3): 497-504.
 998 Melamed, D., D. L. Young, C. E. Gamble, C. R. Miller and S. Fields (2013). "Deep mutational scanning of an RRM
 999 domain of the *Saccharomyces cerevisiae* poly(A)-binding protein." *RNA* **19**(11): 1537-1551.
 1000 Meyer, B., J. Chiaravalli, S. Gellenoncourt, P. Brownridge, D. P. Bryne, L. A. Daly, A. Grauslys, M. Walter, F. Agou,
 1001 L. A. Chakrabarti, C. S. Craik, C. E. Eyers, P. A. Eyers, Y. Gambin, A. R. Jones, E. Sierecki, E. Verdin, M. Vignuzzi and
 1002 E. Emmott (2021). "Characterising proteolysis during SARS-CoV-2 infection identifies viral cleavage sites and
 1003 cellular targets with therapeutic potential." *Nat Commun* **12**(1): 5553.
 1004 Mishra, P., J. M. Flynn, T. N. Starr and D. N. A. Bolon (2016). "Systematic Mutant Analyses Elucidate General and
 1005 Client-Specific Aspects of Hsp90 Function." *Cell Rep* **15**(3): 588-598.
 1006 Murray, M. G., W. Hung, I. Sadowski and B. Das Mahapatra (1993). "Inactivation of a yeast transactivator by the
 1007 fused HIV-1 proteinase: a simple assay for inhibitors of the viral enzyme activity." *Gene* **134**(1): 123-128.
 1008 Nguyen, A. W. and P. S. Daugherty (2005). "Evolutionary optimization of fluorescent proteins for intracellular
 1009 FRET." *Nat Biotechnol* **23**(3): 355-360.
 1010 Ohta, T. (1973). "Slightly deleterious mutant substitutions in evolution." *Nature* **246**(5428): 96-98.
 1011 Ottoz, D. S., F. Rudolf and J. Stelling (2014). "Inducible, tightly regulated and growth condition-independent
 1012 transcription factor in *Saccharomyces cerevisiae*." *Nucleic Acids Res* **42**(17): e130.
 1013 Owen, D. R., C. M. N. Allerton, A. S. Anderson, L. Aschenbrenner, M. Avery, S. Berritt, B. Boras, R. D. Cardin, A.
 1014 Carlo, K. J. Coffman, A. Dantonio, L. Di, H. Eng, R. Ferre, K. S. Gajiwala, S. A. Gibson, S. E. Greasley, B. L. Hurst, E.
 1015 P. Kadar, A. S. Kalgutkar, J. C. Lee, J. Lee, W. Liu, S. W. Mason, S. Noell, J. J. Novak, R. S. Obach, K. Ogilvie, N. C.
 1016 Patel, M. Pettersson, D. K. Rai, M. R. Reese, M. F. Sammons, J. G. Sathish, R. S. P. Singh, C. M. Steppan, A. E.
 1017 Stewart, J. B. Tuttle, L. Updyke, P. R. Verhoest, L. Wei, Q. Yang and Y. Zhu (2021). "An oral SARS-CoV-2 M(pro)
 1018 inhibitor clinical candidate for the treatment of COVID-19." *Science*: eabl4784.
 1019 Roscoe, B. P., K. M. Thayer, K. B. Zeldovich, D. Fushman and D. N. Bolon (2013). "Analyses of the effects of all
 1020 ubiquitin point mutants on yeast growth rate." *J Mol Biol* **425**(8): 1363-1377.
 1021 Shaqra, A. M., S. Zvornicanin, Q. Y. Huang, G. J. Lockbaum, M. Knapp, L. Tandeske, D. T. Barkan, J. M. Flynn, D. N.
 1022 A. Bolon, S. Moquin, D. Dovala, N. K. Yilmaz and C. A. Schiffer (2022). "Defining the Substrate Envelope of SARS-
 1023 CoV-2 Main Protease to Predict and Avoid Drug Resistance." *bioRxiv*.
 1024 Singer, J., R. Gifford, M. Cotten and D. Robertson (2020). "CoV-GLUE: A Web Application for Tracking SARS-CoV-2
 1025 Genomic Variation." *Preprints*
 1026 Singh, J., R. C. Petter, T. A. Baillie and A. Whitty (2011). "The resurgence of covalent drugs." *Nat Rev Drug Discov*
 1027 **10**(4): 307-317.
 1028 Starita, L. M., J. N. Pruneda, R. S. Lo, D. M. Fowler, H. J. Kim, J. B. Hiatt, J. Shendure, P. S. Brzovic, S. Fields and R.
 1029 E. Klevit (2013). "Activity-enhancing mutations in an E3 ubiquitin ligase identified by high-throughput
 1030 mutagenesis." *Proc Natl Acad Sci U S A* **110**(14): E1263-1272.
 1031 Tan, J., K. H. Verschuere, K. Anand, J. Shen, M. Yang, Y. Xu, Z. Rao, J. Bigalke, B. Heisen, J. R. Mesters, K. Chen, X.
 1032 Shen, H. Jiang and R. Hilgenfeld (2005). "pH-dependent conformational flexibility of the SARS-CoV main
 1033 proteinase (M(pro)) dimer: molecular dynamics simulations and multiple X-ray structure analyses." *J Mol Biol*
 1034 **354**(1): 25-40.

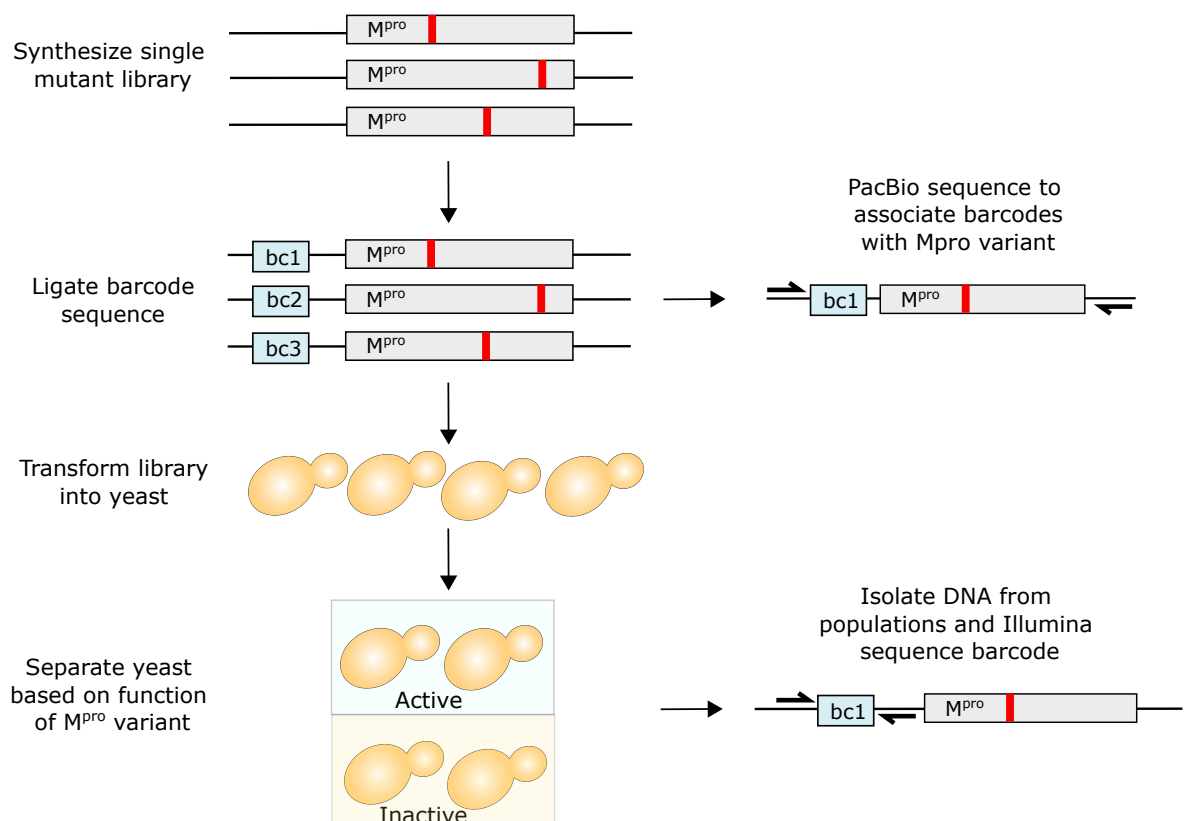
1035 Thiel, V., K. A. Ivanov, A. Putics, T. Hertzog, B. Schelle, S. Bayer, B. Weissbrich, E. J. Snijder, H. Rabenau, H. W.
 1036 Doerr, A. E. Gorbalenya and J. Ziebuhr (2003). "Mechanisms and enzymes involved in SARS coronavirus genome
 1037 expression." J Gen Virol **84**(Pt 9): 2305-2315.
 1038 Weng, Y. L., S. R. Naik, N. Dingelstad, M. R. Lugo, S. Kalyanamoorthy and A. Ganesan (2021). "Molecular
 1039 dynamics and in silico mutagenesis on the reversible inhibitor-bound SARS-CoV-2 main protease complexes
 1040 reveal the role of lateral pocket in enhancing the ligand affinity." Sci Rep **11**(1): 7429.
 1041 Xue, X., H. Yang, W. Shen, Q. Zhao, J. Li, K. Yang, C. Chen, Y. Jin, M. Bartlam and Z. Rao (2007). "Production of
 1042 authentic SARS-CoV M(pro) with enhanced activity: application as a novel tag-cleavage endopeptidase for
 1043 protein overproduction." J Mol Biol **366**(3): 965-975.
 1044 Zaki, A. M., S. van Boheemen, T. M. Bestebroer, A. D. Osterhaus and R. A. Fouchier (2012). "Isolation of a novel
 1045 coronavirus from a man with pneumonia in Saudi Arabia." N Engl J Med **367**(19): 1814-1820.
 1046 Zhang, L., D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox and R. Hilgenfeld (2020). "Crystal
 1047 structure of SARS-CoV-2 main protease provides a basis for design of improved alpha-ketoamide inhibitors."
 1048 Science **368**(6489): 409-412.
 1049 Zhao, Y., C. Fang, Q. Zhang, R. Zhang, X. Zhao, Y. Duan, H. Wang, Y. Zhu, L. Feng, J. Zhao, M. Shao, X. Yang, L.
 1050 Zhang, C. Peng, K. Yang, D. Ma, Z. Rao and H. Yang (2021). "Crystal structure of SARS-CoV-2 main protease in
 1051 complex with protease inhibitor PF-07321332." Protein Cell.
 1052 Ziebuhr, J., J. Herold and S. G. Siddell (1995). "Characterization of a human coronavirus (strain 229E) 3C-like
 1053 proteinase activity." J Virol **69**(7): 4331-4338.
 1054
 1055

Expressing non-functional
 M^{pro} variants

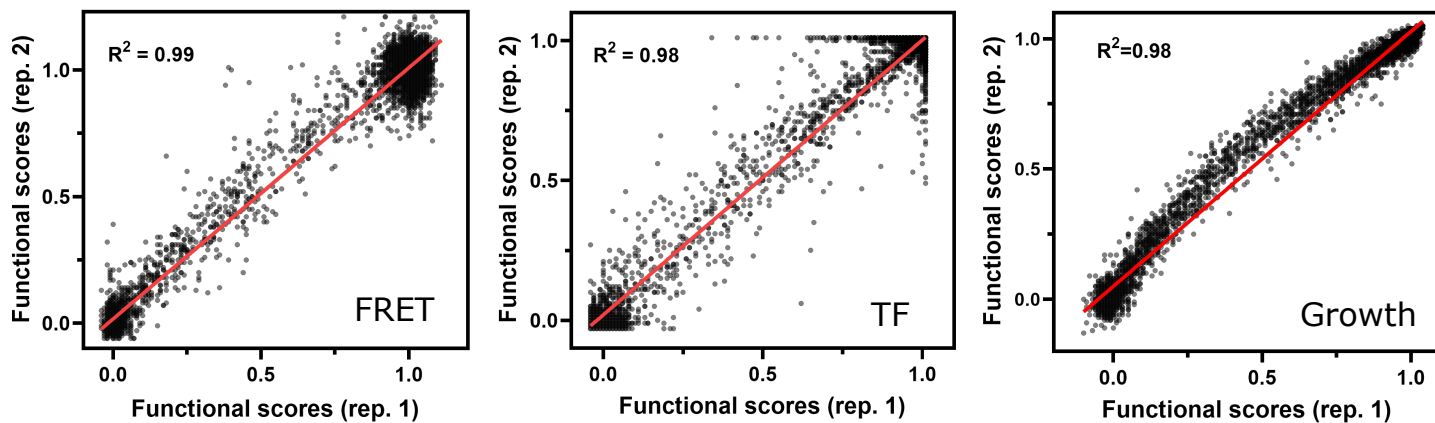
Expressing functional
 M^{pro} variants



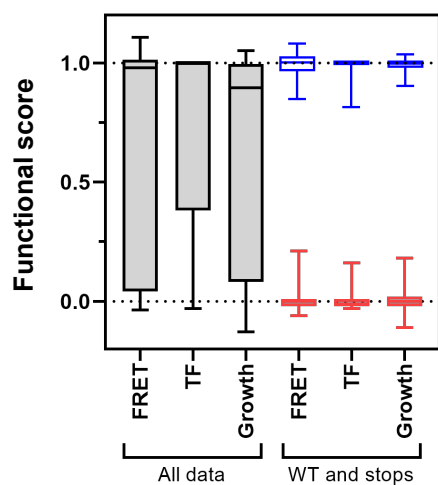
D.



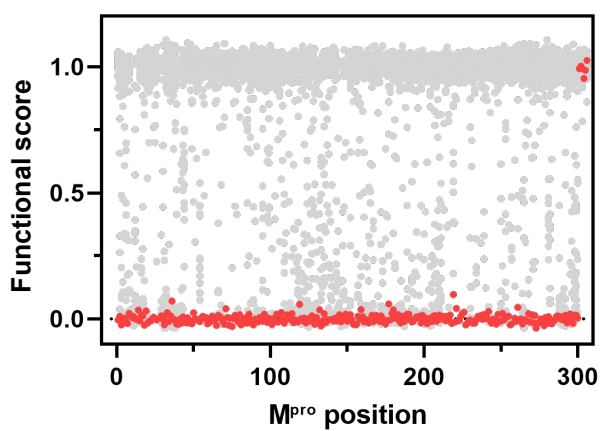
A.



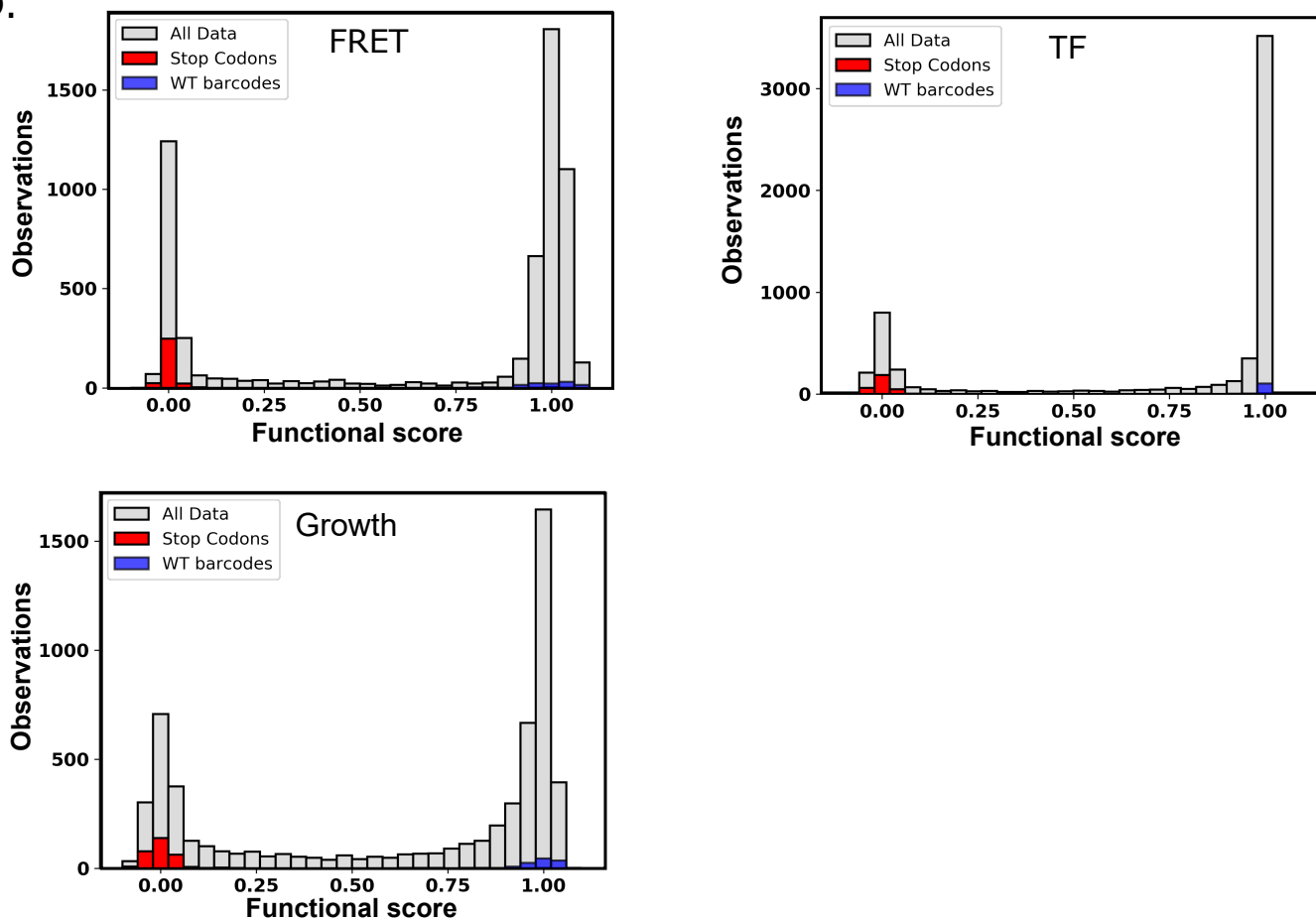
B.



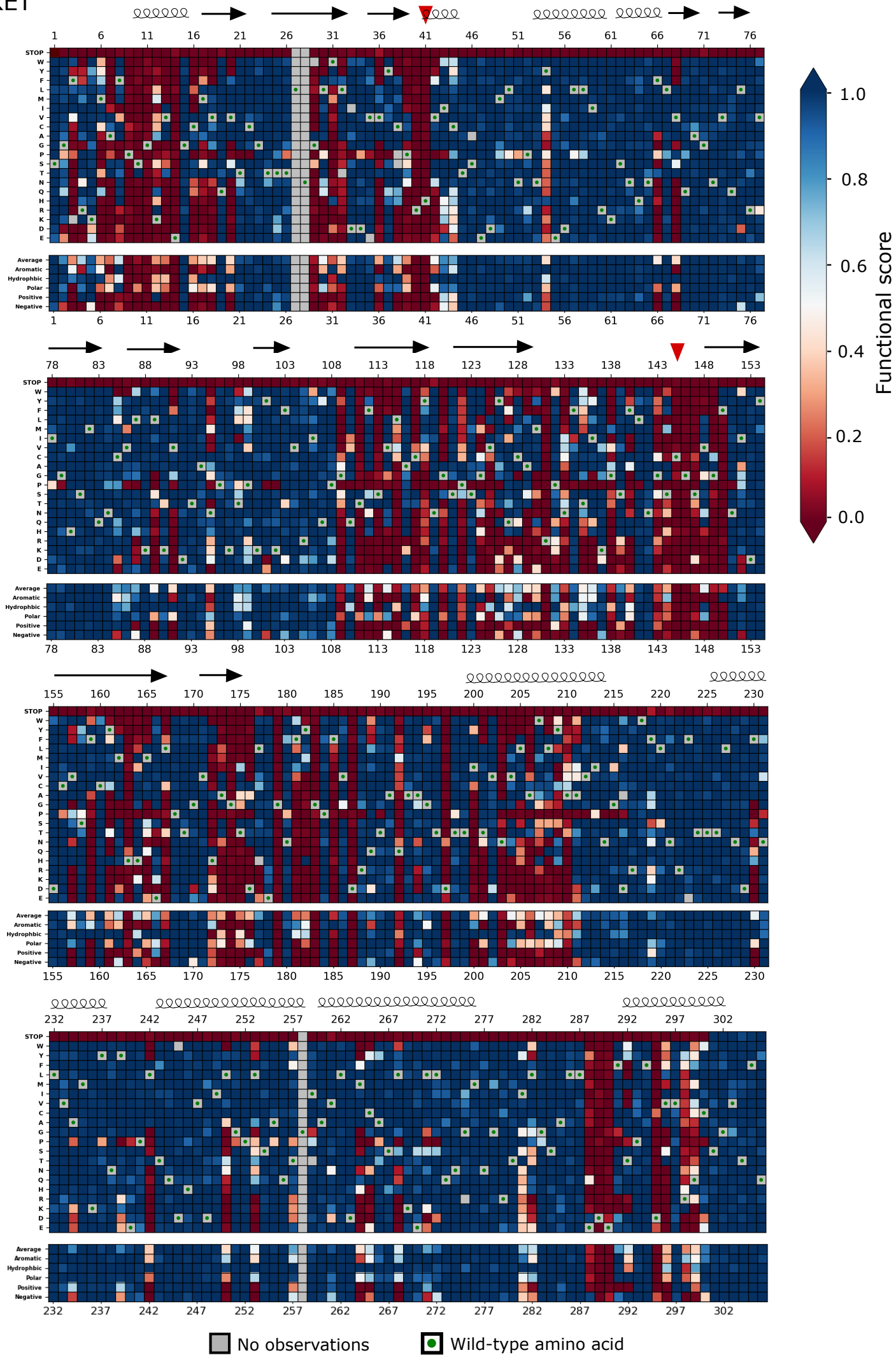
C.



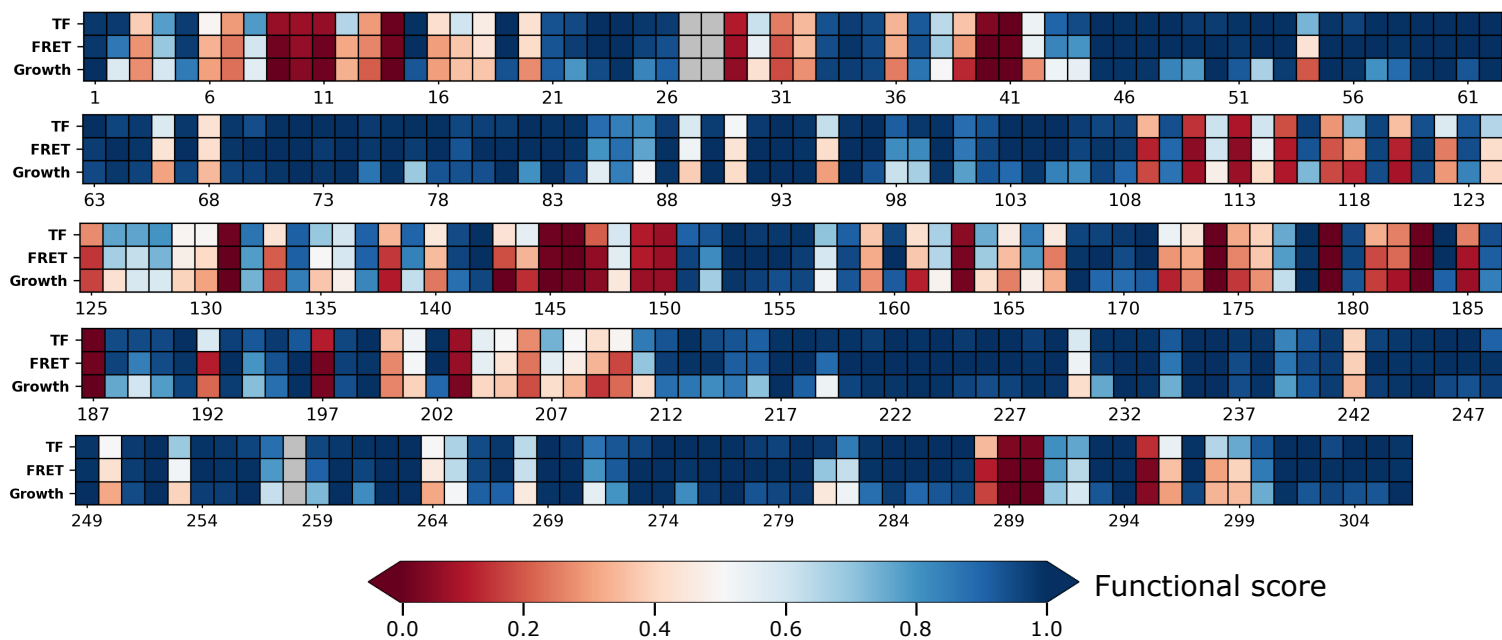
D.



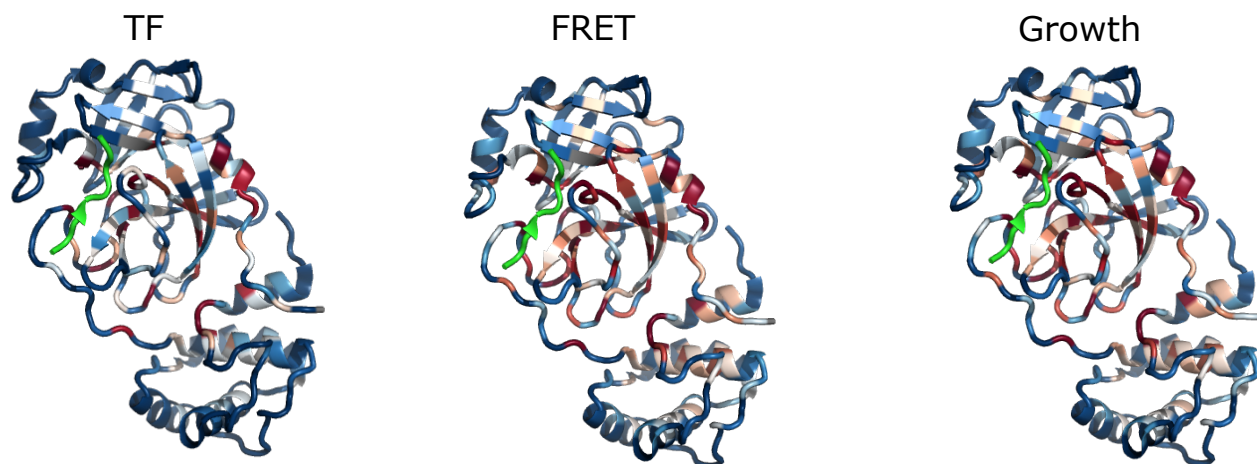
FRET



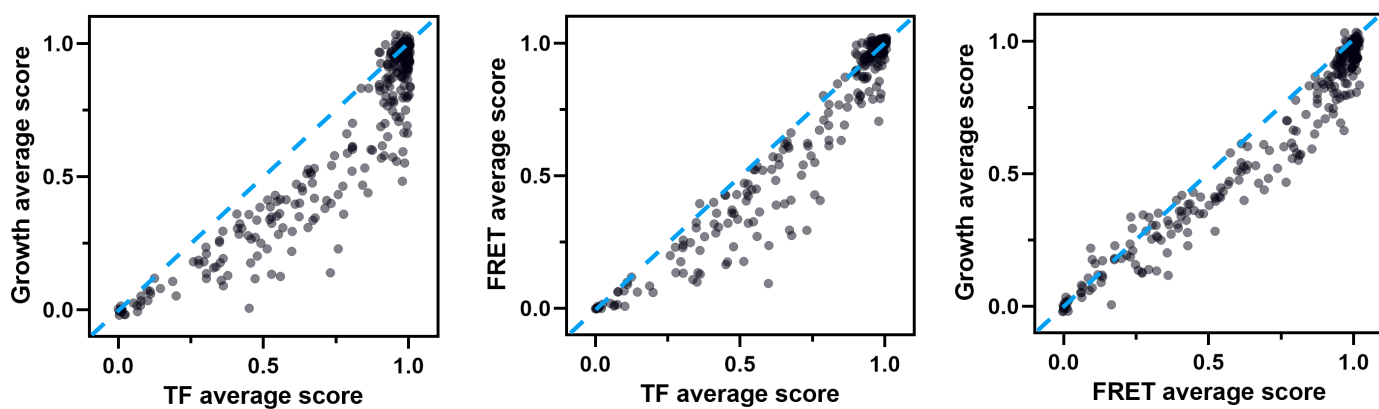
A.



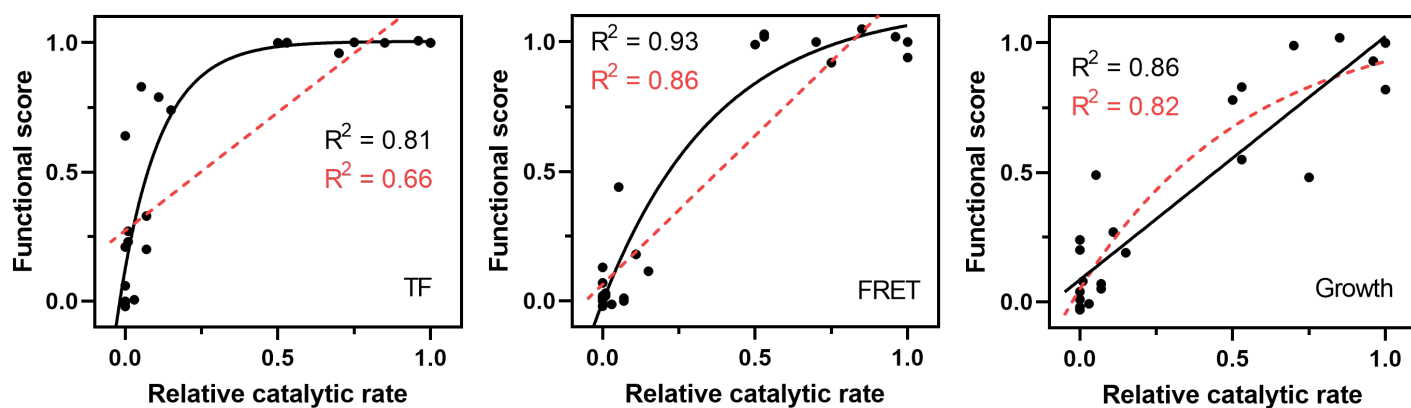
B.

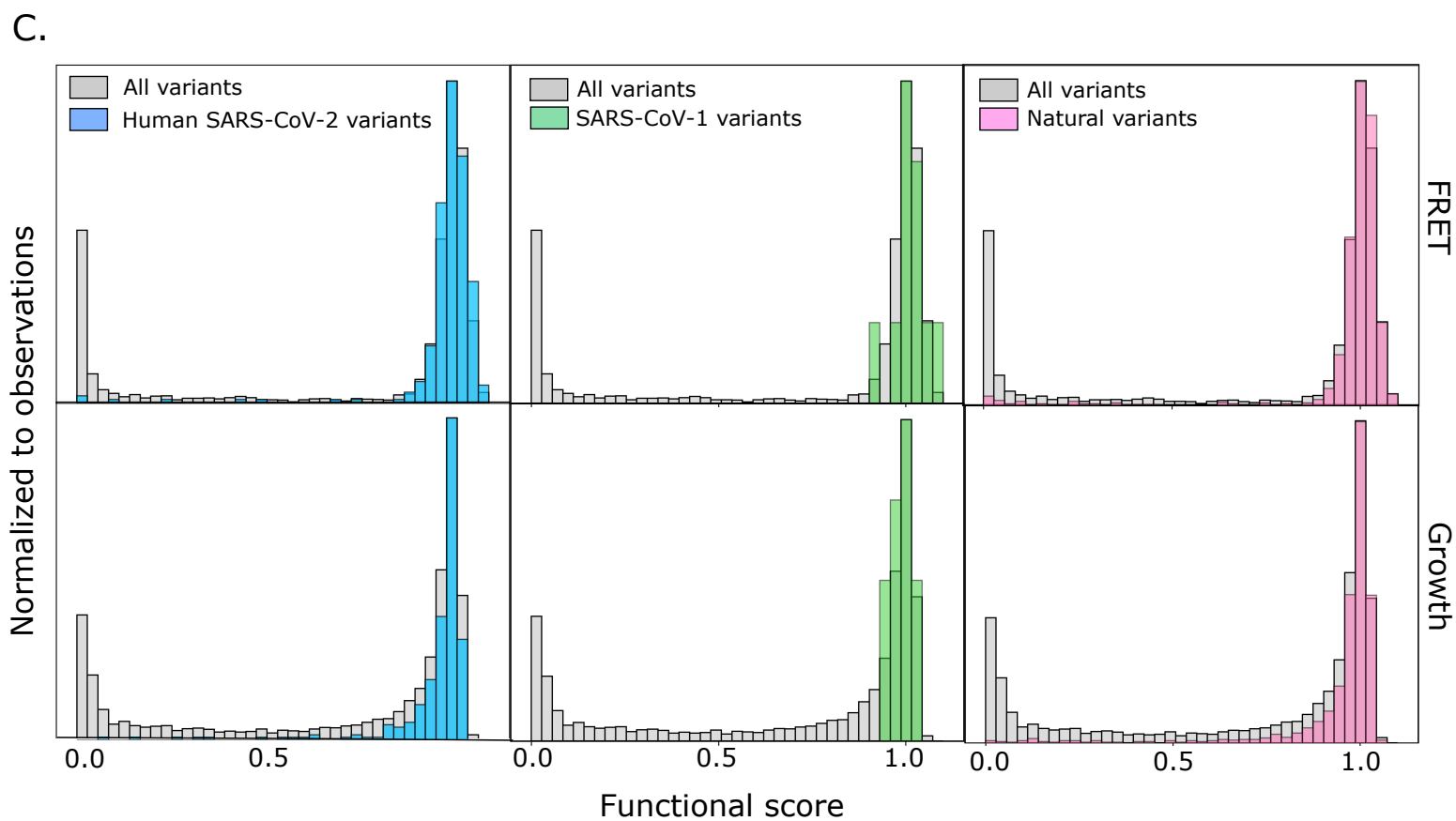
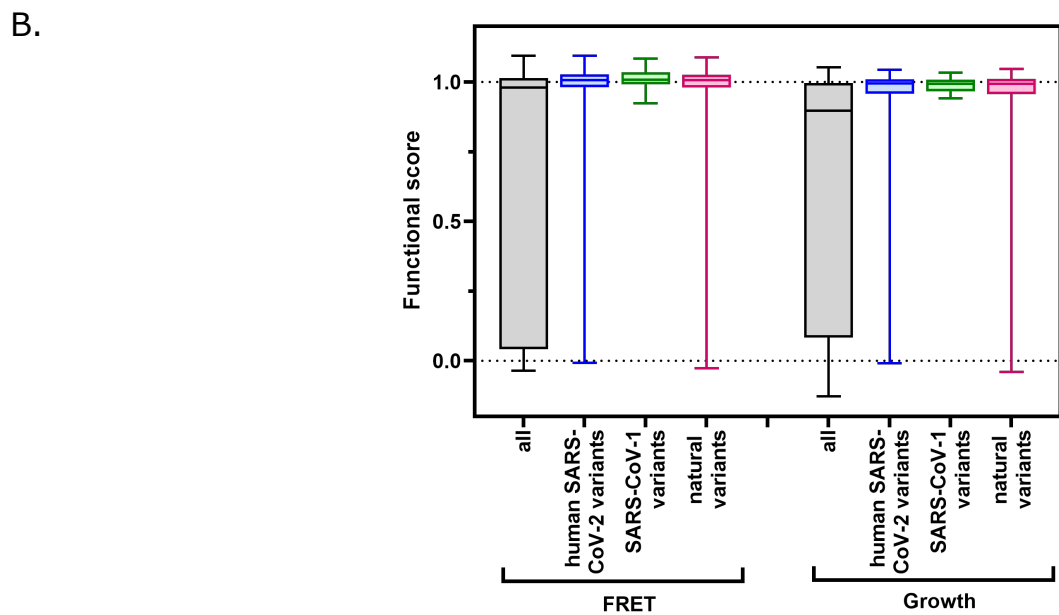
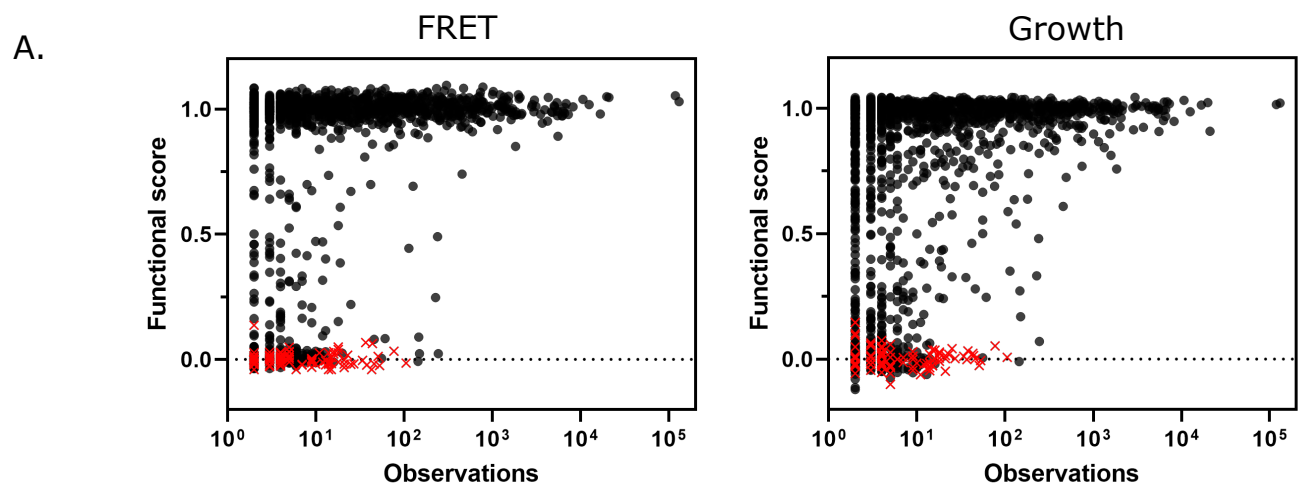


C.

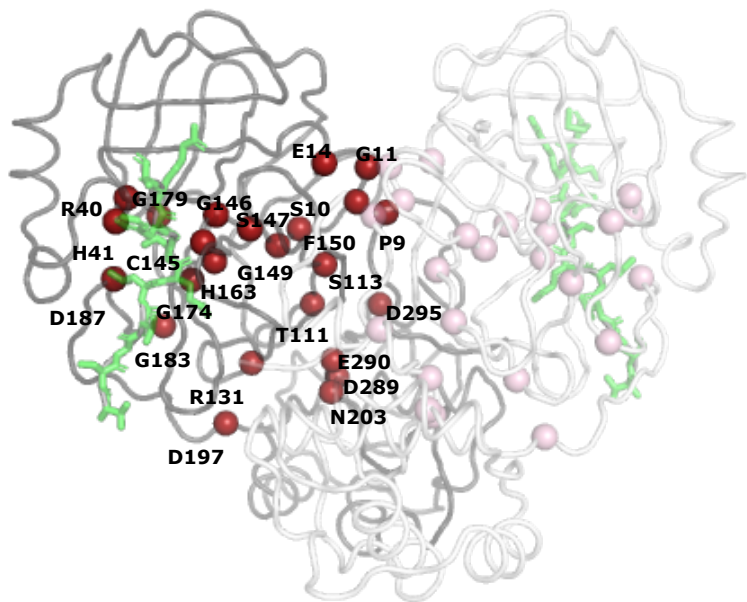


D.

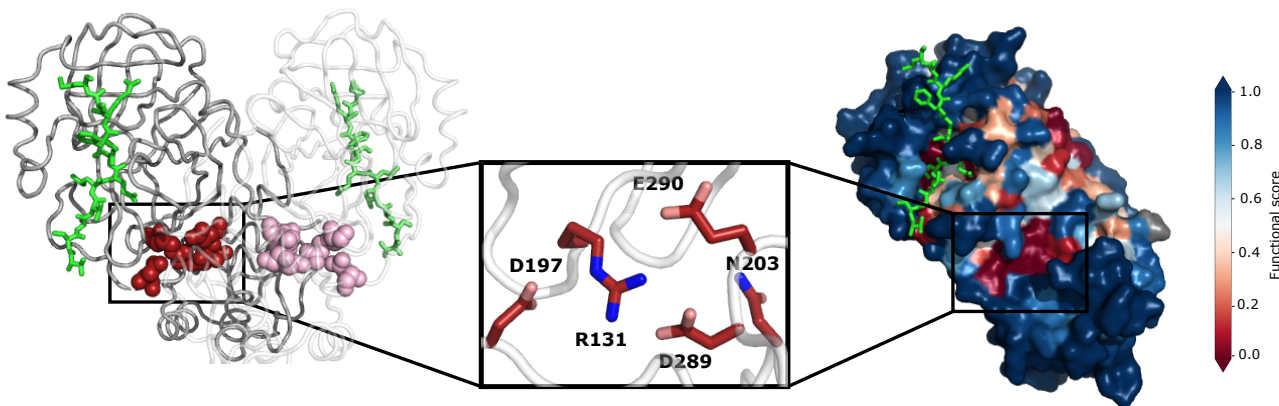




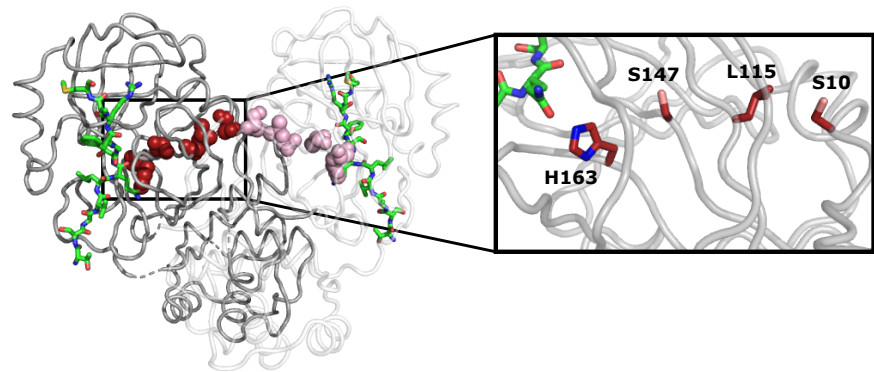
A.



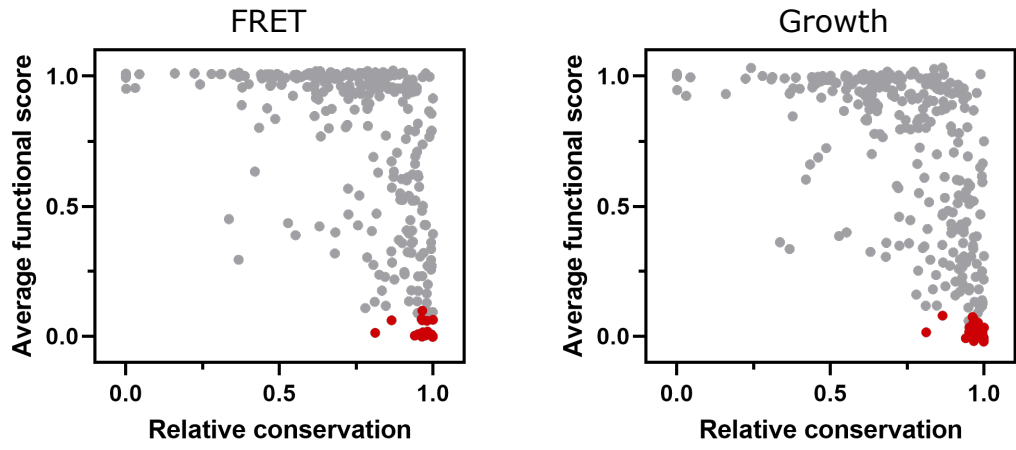
B.



C.

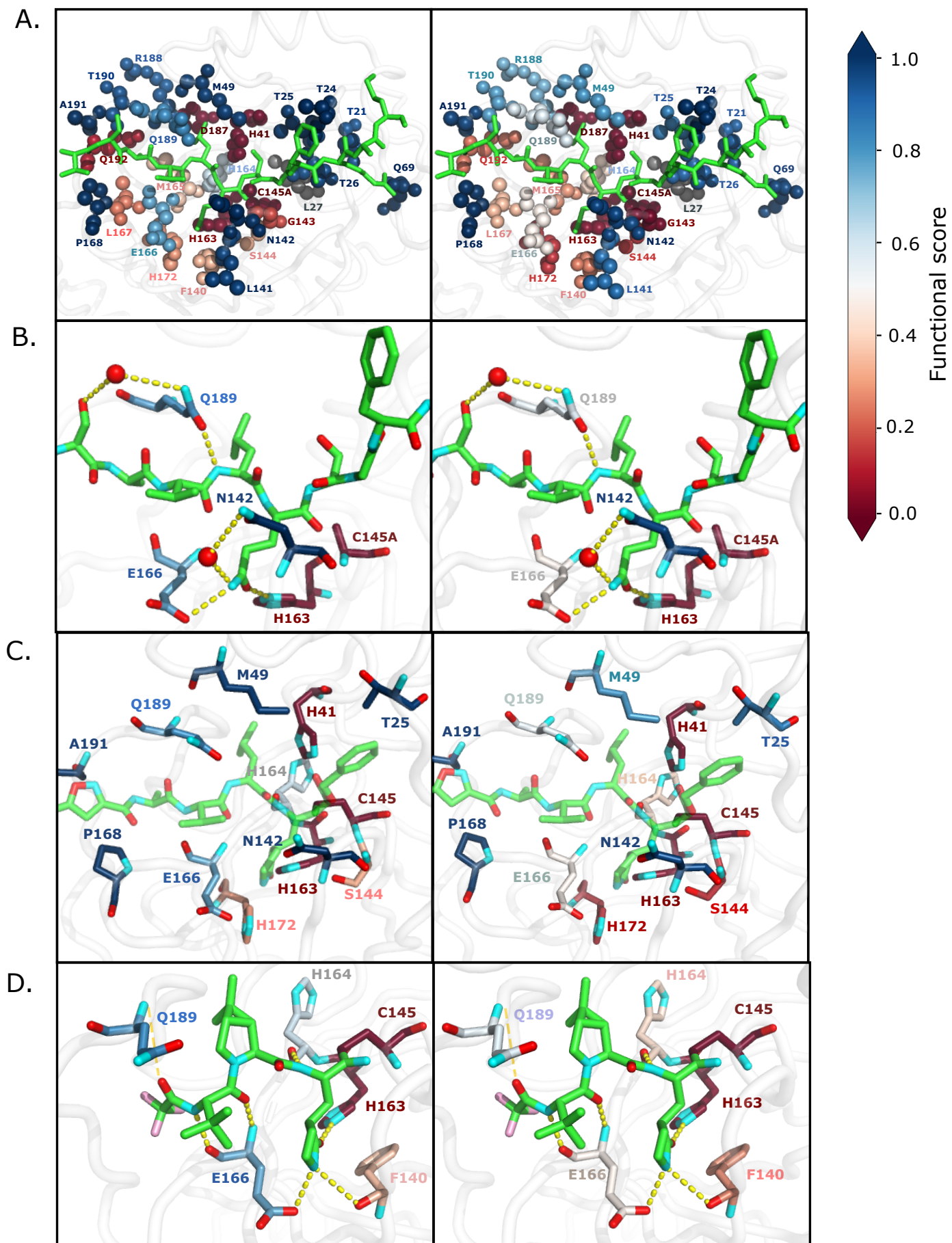


D.

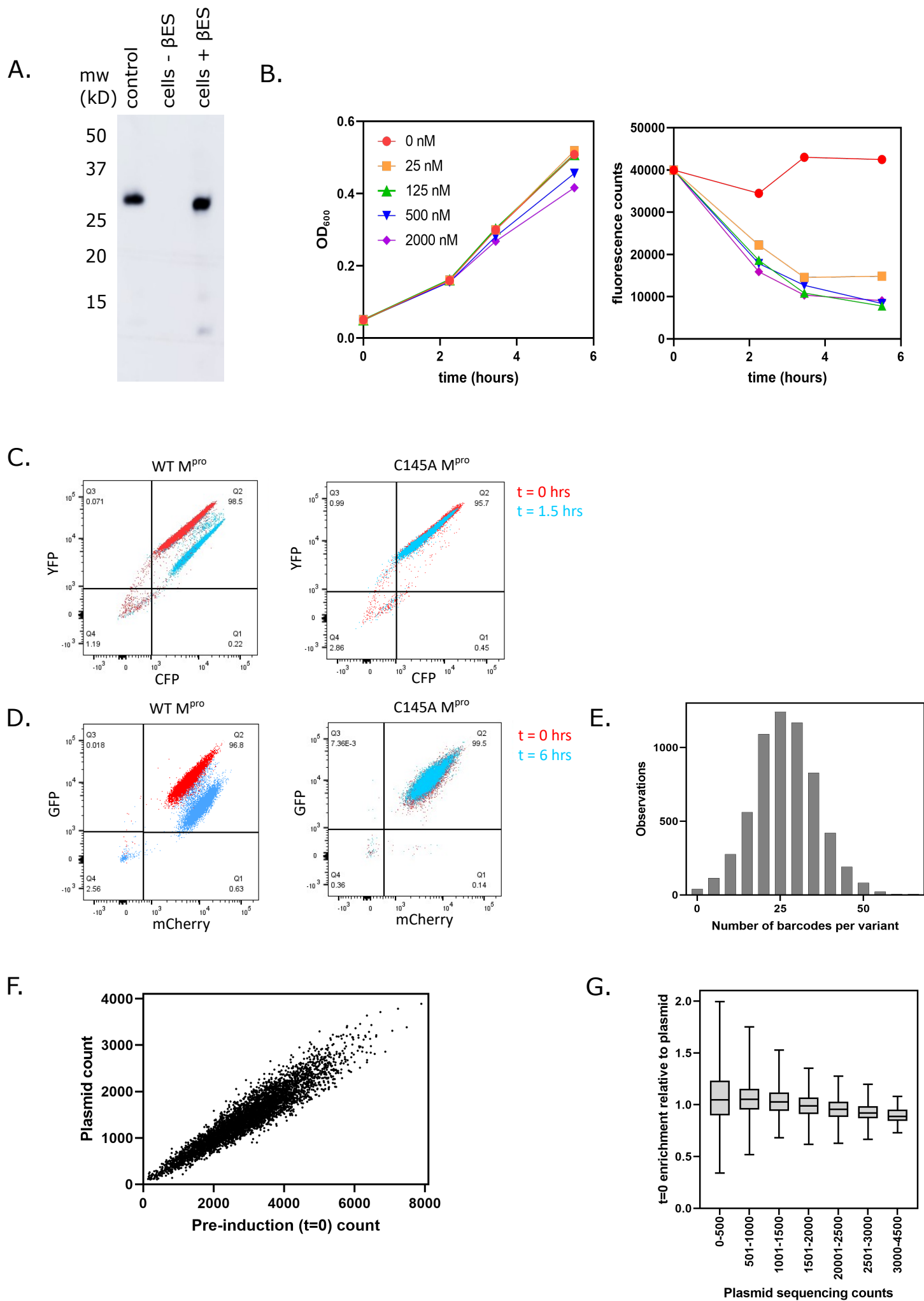


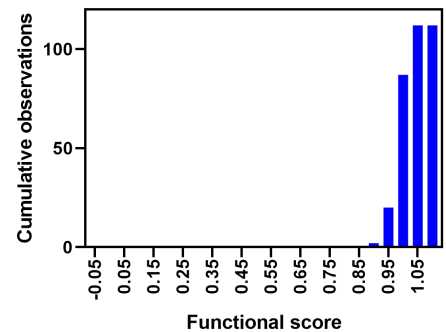
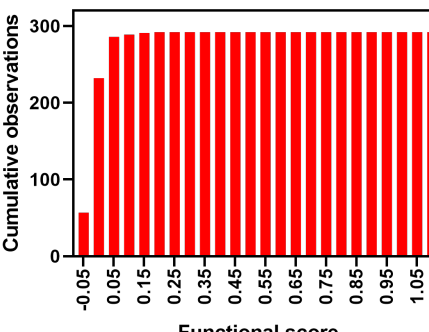
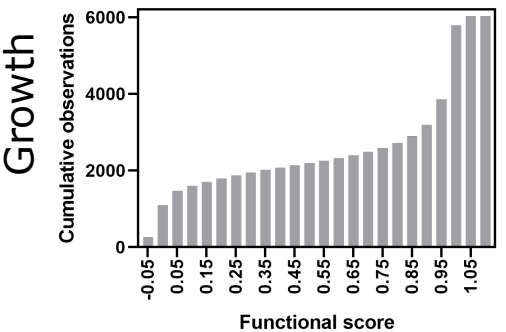
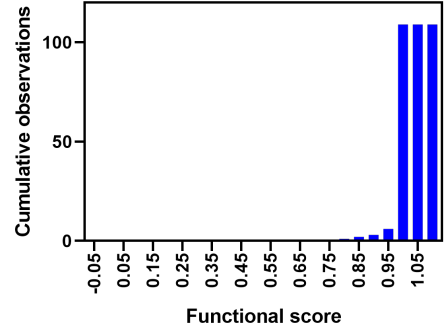
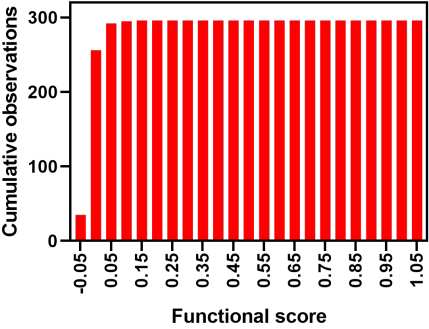
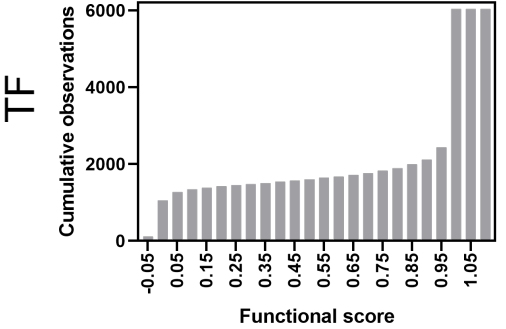
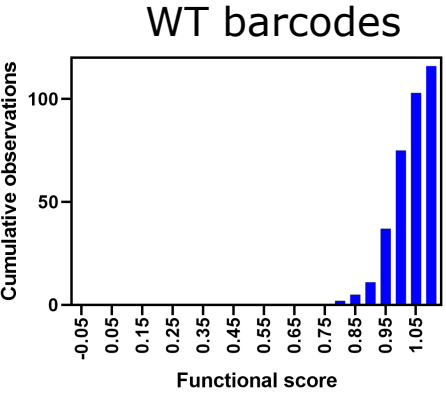
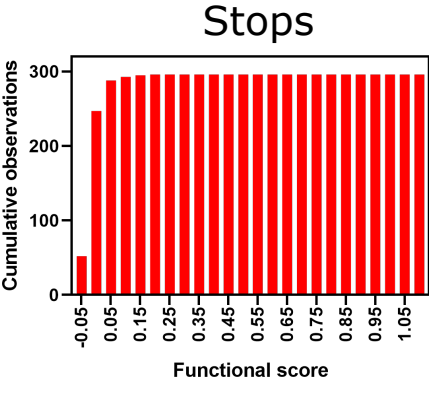
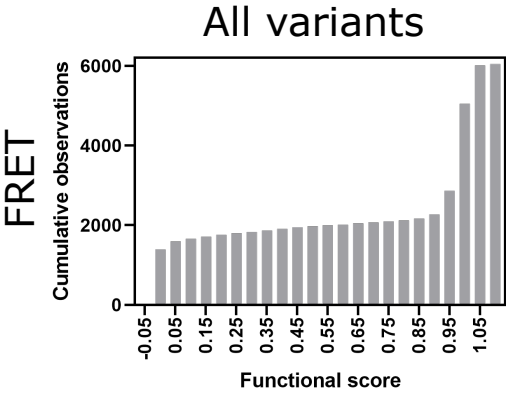
FRET

Growth

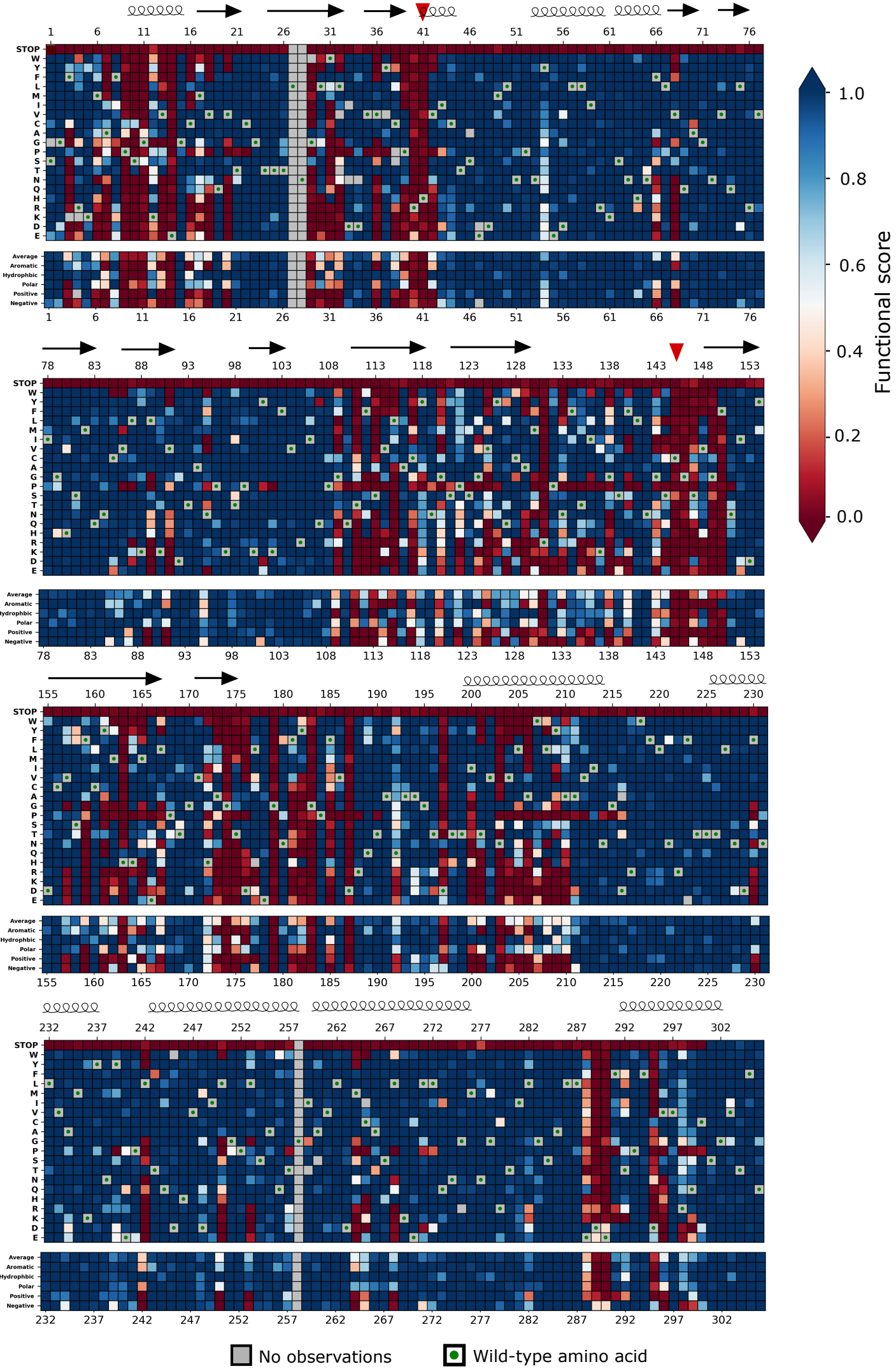


Mutation	$\Delta\Delta G$	Codon change	FRET score	Growth score
Q189E	1.11	CAA -> GAA	1.04	1.0
E166A	1.0	GAA -> GCA	1.0	0.88
E166Q	1.0	GAA -> CAA	1.0	0.96

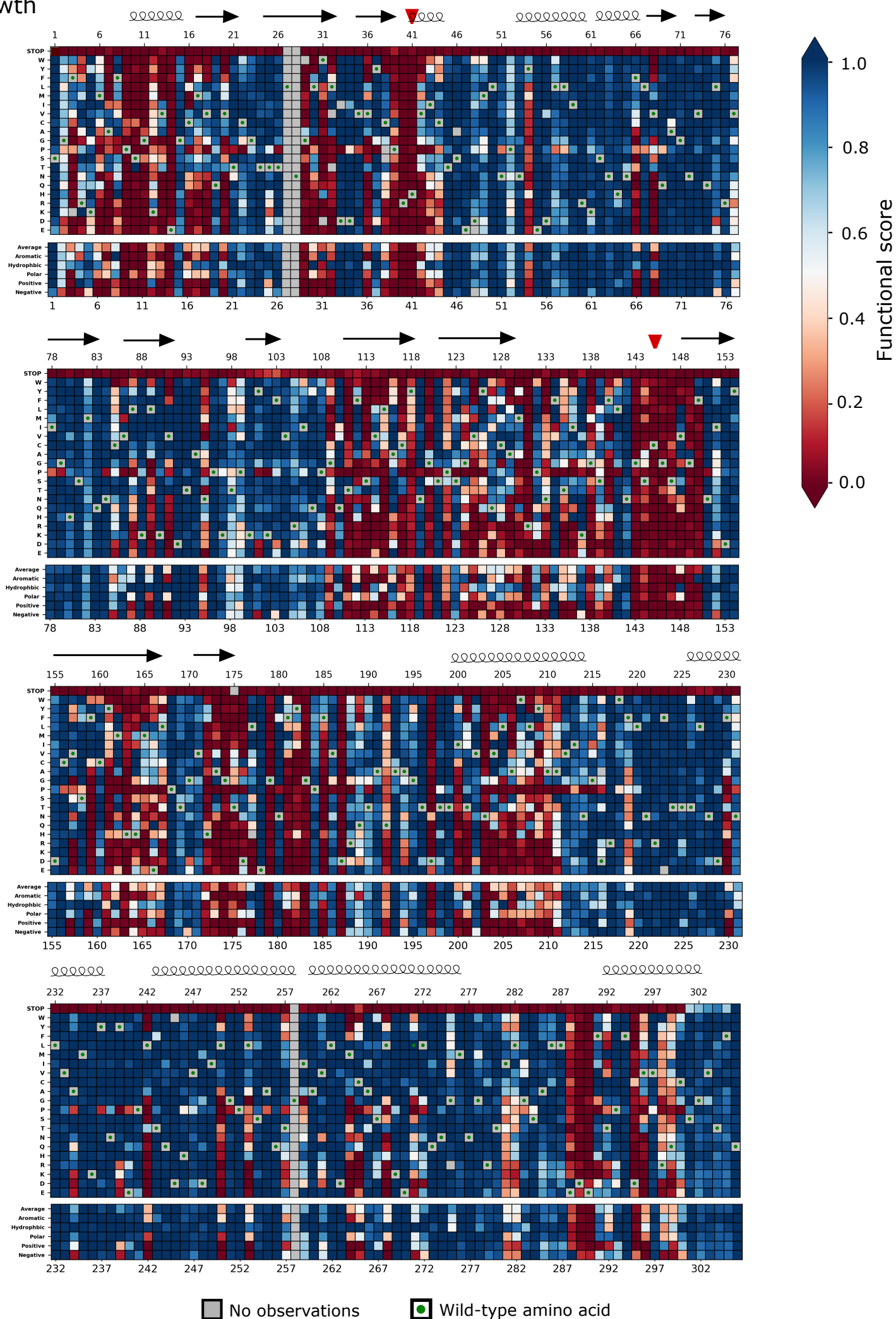




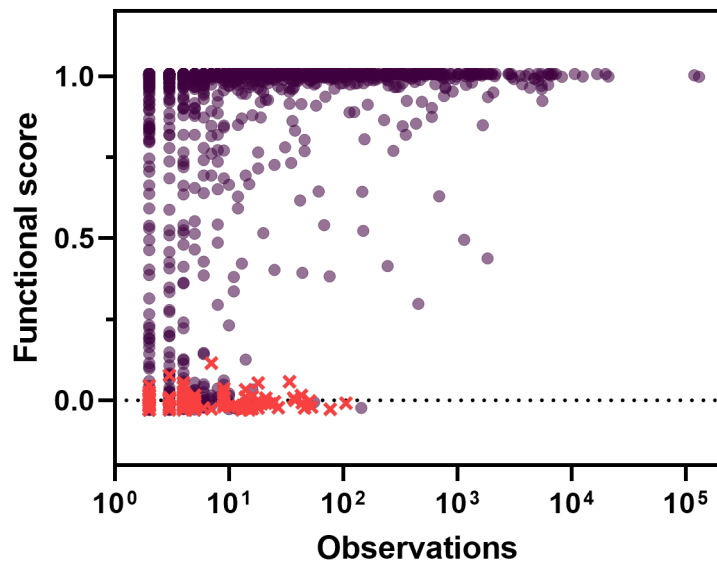
TF



Growth



A.



B.

