Globally defining the effects of mutations in a picornavirus capsid

Florian Mattenberger[1], Victor Latorre[1], Omer Tirosh[2], Adi Stern[2], Ron Geller[1]*

[1]Institute for Integrative Systems Biology, I2SysBio (Universitat de València-CSIC), Paterna, Spain; [2]The Shmunis School of Biomedicine and Cancer Research, Tel-Aviv University, Tel-Aviv, Israel

Abstract The capsids of non-enveloped viruses are highly multimeric and multifunctional protein assemblies that play key roles in viral biology and pathogenesis. Despite their importance, a comprehensive understanding of how mutations affect viral fitness across different structural and functional attributes of the capsid is lacking. To address this limitation, we globally define the effects of mutations across the capsid of a human picornavirus. Using this resource, we identify structural and sequence determinants that accurately predict mutational fitness effects, refine evolutionary analyses, and define the sequence specificity of key capsid-encoded motifs. Furthermore, capitalizing on the derived sequence requirements for capsid-encoded protease cleavage sites, we implement a bioinformatic approach for identifying novel host proteins targeted by viral proteases. Our findings represent the most comprehensive investigation of mutational fitness effects in a picornavirus capsid to date and illuminate important aspects of viral biology, evolution, and host interactions.

Introduction

The capsids of non-enveloped viruses are among the most complex of any viral protein. These highly multimeric structures must correctly assemble around the genome from numerous subunits, at times numbering in the hundreds, while avoiding aggregation (*Harrison, 2013*; *Hunter, 2013*; *Perlmutter and Hagan, 2015*). Moreover, the assembled structure must be both sufficiently stable to protect the viral genome during its transition between cells yet readily disassemble upon entry to initiate subsequent infections. For these functions to be achieved, viral capsids must encode the information for interacting with numerous cellular factors that are required to correctly fold and assemble around the genome (*Callaway et al., 2001*; *Fields et al., 2013*; *Geller et al., 2007*; *Jiang et al., 2014*; *Macejak and Sarnow, 1992*). Viral capsids also play key roles in pathogenesis, dictating host and cell tropism by encoding the determinants for binding cellular receptors (*Helenius, 2013*; *Rossmann et al., 2002*) and mediating escape from humoral immune responses (*Cifuente and Moratorio, 2019*; *Heise and Virgin, 2013*). As a result, viral capsids show the highest evolutionary rates among viral proteins.

The picornaviruses constitute a large group of single-stranded, positive-sense RNA viruses and include several pathogens of significant medical and economic impact (*Racaniello, 2013*). Their relative simplicity and ease of culture have made picornaviruses important models for understanding virus biology. Among the many breakthroughs achieved with these viruses was the determination of the first high-resolution structure of the capsid of an animal virus, making the picornavirus capsid the prototypical non-enveloped, icosahedral viral capsid (*Racaniello, 2013*). Picornavirus capsid genesis initiates with the co-translational release of the P1 capsid precursor protein from the viral polyprotein via the proteolytic activity of the viral encoded 2A protease (*Jiang et al., 2014*; *Racaniello, 2013*). Subsequently, the viral encoded 3CD protease (3CD^{pro}) cleaves the P1 capsid precursor to liberate three capsid proteins (VP0, VP3, and VP1), generating the capsid protomer.

eLife digest A virus is made up of genetic material that is encased with a protective protein coat called the capsid. The capsid also helps the virus to infect host cells by binding to the host receptor proteins and releasing its genetic material. Inside the cell, the virus hitchhikes the infected cell's machinery to grow or replicate its own genetic material.

Viral capsids are the main target of the host's defence system, and therefore, continuously change in an attempt to escape the immune system by introducing alterations (known as mutations) into the genes encoding viral capsid proteins. Mutations occur randomly, and so while some changes to the viral capsid might confer an advantage, others may have no effect at all, or even weaken the virus.

To better understand the effect of capsid mutations on the virus' ability to infect host cells, Mattenberger et al. studied the Coxsackievirus B3, which is linked to heart problems and acute heart failure in humans. The researchers analysed around 90% of possible amino acid mutations (over 14,800 mutations) and correlated each mutation to how it influenced the virus' ability to replicate in human cells grown in the laboratory.

Based on these results, Mattenberger et al. developed a computer model to predict how a particular mutation might affect the virus. The analysis also identified specific amino acid sequences of capsid proteins that are essential for certain tasks, such as building the capsid. It also included an analysis of sequences in the capsid that allow it to be recognized by another viral protein, which cuts the capsid proteins into the right size from a larger precursor. By looking for similar sequences in human genes, the researchers identified several ones that the virus may attack and inactivate to support its own replication.

These findings may help identify potential drug targets to develop new antiviral therapies. For example, proteins of the capsid that are less likely to mutate will provide a better target as they lower the possibility of the virus to become resistant to the treatment. They also highlight new proteins in human cells that could potentially block the virus in cells.

Five protomers then assemble to form the pentamer, twelve of which assemble around the viral genome to yield the virion. Finally, in some picornaviruses, VP0 is further cleaved into two subunits, VP4 and VP2, following genomic encapsidation to generate the infectious, 240 subunit particles (*Jiang et al., 2014*; *Racaniello, 2013*). Work over the years has identified numerous host factors that help support capsid formation (*Corbic Ramljak et al., 2018*; *Geller et al., 2007*; *Macejak and Sarnow, 1992*; *Qing et al., 2014*; *Thibaut et al., 2014*), defined antibody neutralization sites (*Cifuente and Moratorio, 2019*), and identified numerous host receptors for many members of this viral family (*Rossmann et al., 2002*).

Despite significant progress in understanding the structure and function of picornavirus capsids, a comprehensive understanding of how mutations affect viral fitness across different structural and functional attributes is lacking. To address this, we perform a comprehensive analysis of mutational fitness effects (MFE) across the complete capsid region of the human picornavirus coxsackievirus B3 (CVB3), analyzing >90% of all possible single amino acid mutations. Furthermore, using these data, we develop models to predict the effect of mutations with high accuracy from available sequence and structural information, improve evolutionary analyses of CVB3, and define the sequence preferences of several viral encoded motifs. Finally, we use the information obtained in our dataset for the sequence requirements of capsid-encoded 3CD protease cleavage sites to identify host targets of this viral protease. Overall, our data comprise the most comprehensive survey of MFE effects in a picornavirus capsid to date and provide important insights into virus biology, evolution, and interaction with the host.

Results

Deep mutational scanning of a CVB3 capsid

To generate CVB3 libraries encoding a large amount of diversity in the capsid region, we used a codon-level PCR mutagenesis method (*Bloom, 2014*). The mutagenesis protocol was performed on

the capsid precursor region P1 in triplicate to generate three independent mutagenized libraries (Mut Library 1–3; *Figure 1A*). From these, three independent viral populations (Mut Virus 1–3) were derived by electroporation of in vitro transcribed viral RNA into HeLa-H1 cells (*Figure 1A*). High-fidelity next-generation sequencing (*Schmitt et al., 2012*) was then used to analyze the mutagenized libraries and resulting viruses, unmutagenized virus populations (WT virus 1–2), as well as controls for errors occurring during PCR (PCR) and reverse transcription (RT-PCR). High coverage was obtained for all samples (>10^6 per codon across all experimental conditions and >6.5×10^5 for the controls; *Supplementary file 2*). Due to the high rate of single mutations within codons observed in the RT-PCR control compared to the mutagenized virus populations (*Supplementary file 2*), all single mutants were omitted from our analysis to increase the signal-to-noise ratio. While this resulted in an inability to analyze 83.4% of synonymous codons in the capsid region (1746/2094), only 2.8% of non-synonymous mutations were lost to analysis (458/16,169). Upon removing single mutations within codons, we obtained a large signal-to-noise ratio in the average mutation rate of 510× (range 449–572) and 245× (range 174–285) for the mutagenized libraries and viruses, respectively, compared to their error controls (*Figure 1B* and *Supplementary file 2*). On average, 0.9 (range 0.8–1.02) codon mutations were observed per genome, which was in agreement with Sanger sequencing of 59 clones (range 18–23 per library; *Figure 1—figure supplement 1* and *Supplementary file 3*). As expected, the rate of stop codons, which should be invariably lethal in the CVB3 capsid, decreased significantly following growth in cells to <0.5% of that observed in the corresponding mutagenized libraries (p<0.005 by paired t-test on log-transformed data; *Supplementary file 2*). No major bias was observed in the position within a codon where mutations were observed (*Figure 1—figure supplement 2*) or in the type of mutation (*Figure 1—figure supplement 2*), except for the WT virus, which had a high rate of A to G transitions in the two independent replicates analyzed. Of all 16,169 possible amino acid mutations in the capsid region (851 AA × 19 AA mutation = 16,169), a total of 14,839 amino acid mutations were commonly observed in all three mutagenized libraries, representing a 91.8% of all possible amino acid mutations in the capsid region, allowing us to globally assess the effects of the vast majority of amino acid mutations on the capsid (*Figure 1C*).

MFE across the CVB3 capsid

We next derived the MFE of each observed mutation by examining how its frequency changed relative to that of the WT sequence following growth in cells. The preferences for the different amino acids at each position (amino acid preferences [*Bloom, 2015*]) showed a high correlation between biological replicates (Spearman's ρ > 0.83; *Figure 2—figure supplement 1* and *Supplementary file 4* MFE). Overall, most mutations in the capsid were deleterious to growth in cell culture, with only 1.2% of mutations increasing fitness relative to the WT amino acid (*Figure 2A* and *Supplementary file 4*; Interactive heatmap available at https://rgellerlab.github.io/CVB3_capsid_DMS_Interactive_Heatmap/). Hotspots where mutations were tolerated were observed at several regions across the capsid (*Figure 2A*). These hotspots largely overlapped with highly variable regions in natural sequences, as measured by Shannon entropy in the enterovirus B family, indicating that lab measured MFE reflect natural evolutionary processes (*Figure 2A*, top). Indeed, a strong correlation was observed between the average MFE observed at each site and sequence variability for the enterovirus B genus (Spearman's ρ = 0.59, p<10^{-16}; *Figure 2B*). Similarly, antibody neutralization sites overlapped with hotspots for mutations (*Figure 2A*, top), with individual mutations in antibody neutralization sites showing lower MFE (p<10^{-16} by Mann–Whitney test; *Figure 2C*). As expected, mutations were also less deleterious in loops compared to β-strands (p<10^{-16} by Mann–Whitney test; *Figure 2D*), at surface residues compared to core residues (p<10^{-16} by Mann–Whitney test; *Figure 2E*), and for mutations predicted to be destabilizing or aggregation-prone (p<10^{-16} by Mann–Whitney test for both; *Figure 2F*). Importantly, independent validation of the MFE of 10 different mutants using a sensitive qPCR-based competition assay (*Moratorio et al., 2017*) showed a strong correlation with the deep mutational scanning (DMS) results (Spearman's ρ = 0.9, p<0.001; *Figure 2G* and *Supplementary file 5*). It is important to note that laboratory-measured MFE may not always reflect those in nature due to differences in the environments.
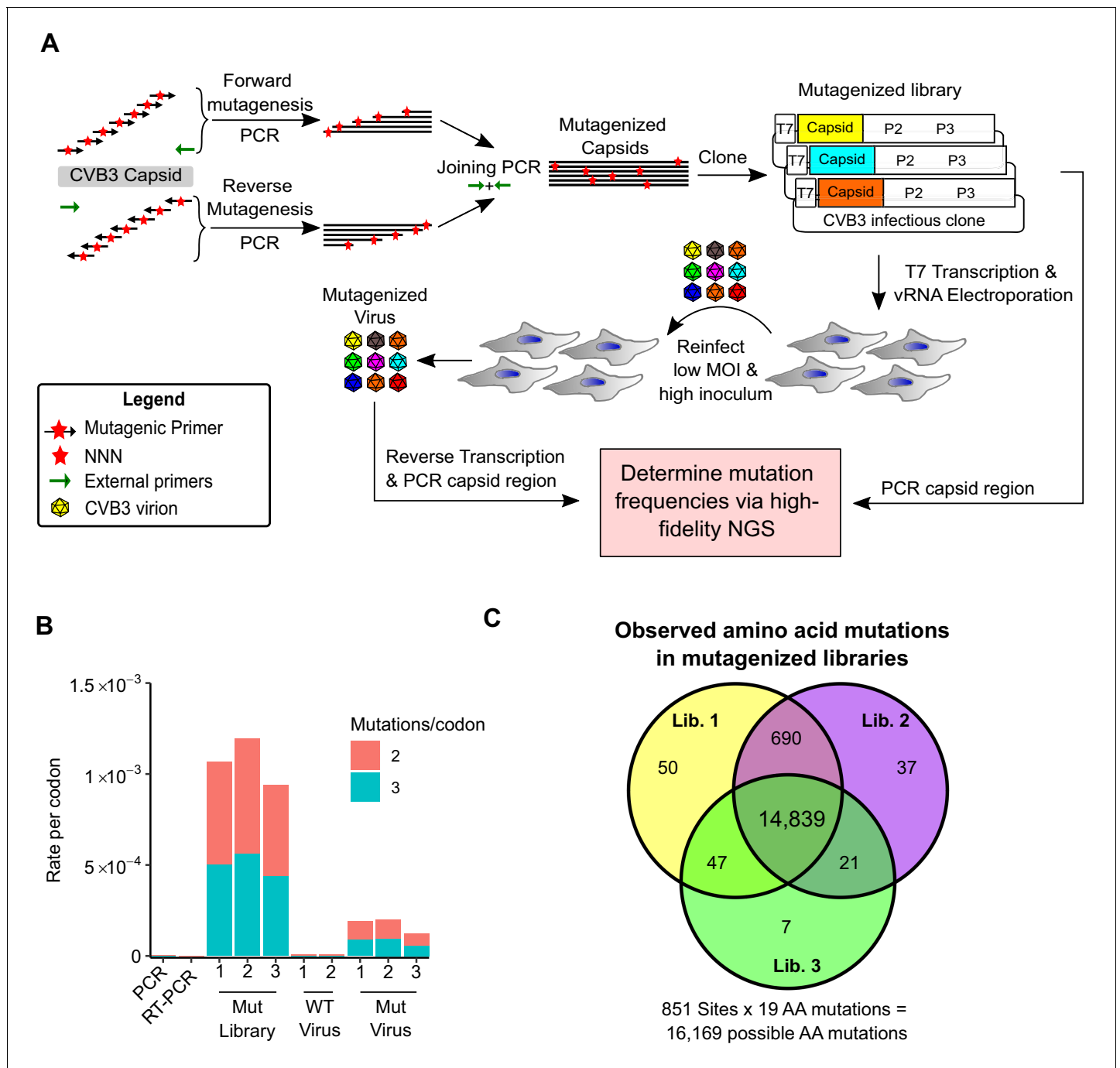
Figure 1. Deep mutational scanning (DMS) of the CVB3 capsid. (A) Overview of the deep mutational scanning experimental approach. A mutagenesis PCR was used to introduce all possible single amino acid mutations across the CVB3 capsid region (Mut Library 1–3). Viral genomic RNA (vRNA) produced from the mutant libraries was then electroporated into cells to generate high diversity CVB3 populations (Mut Virus 1–3). The frequency of each mutation relative to the WT amino acid was then determined in both the mutagenized libraries and the resulting virus populations via high-fidelity duplex sequencing. (B) The average rate of double or triple mutations per codon observed in the mutagenized libraries (Mut Library 1–3), the resulting mutagenized virus (Mut Virus 1–3), as well as controls for the error rate of the amplification and sequencing process (PCR and RT-PCR) or the WT unmutagenized virus (WT Virus 1–2). Single mutations per codon were omitted from the analysis to increase the signal-to-noise ratio. (C) Venn diagram showing the number of amino acid mutations observed in the mutagenized libraries. MOI: multiplicity of infection. NGS: next-generation sequencing. The online version of this article includes the following figure supplement(s) for figure 1:

Figure supplement 1. Sanger analysis of DMS libraries.

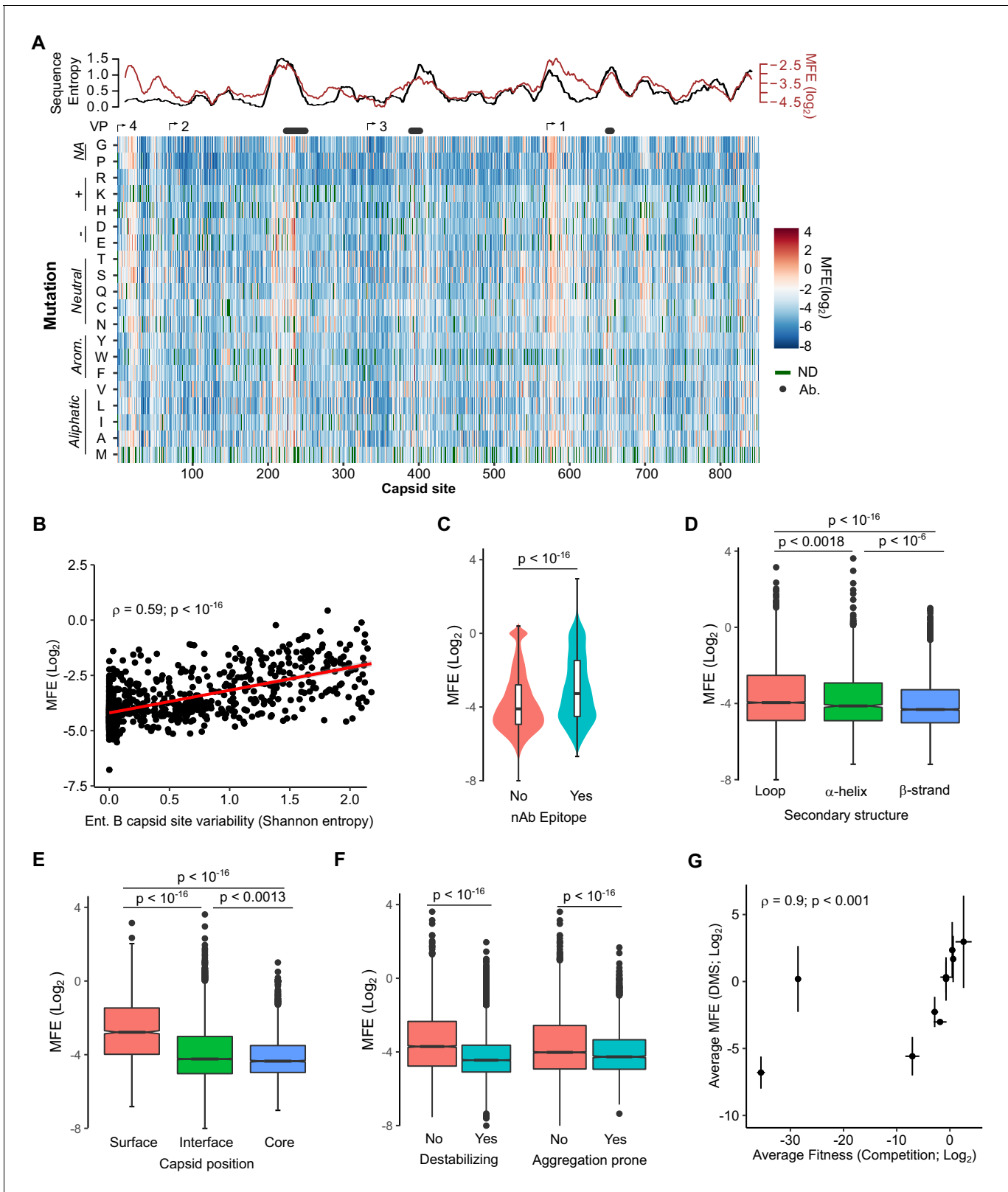**Figure supplement 2.** Results of high-fidelity duplex sequencing.

Figure 2. Mutational fitness effects (MFE) across the CVB3 capsid and their correlation with structural, evolutionary, and immunological attributes. (**A**) Overview of the MFE observed across the CVB3 capsid. Bottom: A heatmap representing the MFE of all mutations observed at each capsid site. Green indicates no data available (ND), and the positions of the mature viral proteins (VP1–4) or antibody neutralization sites (nAb) are indicated above. Top: A 21 amino acid sliding window analysis of the average sequence variation in enterovirus B genomes (Shannon entropy; black line) or a 21 amino acid

Figure 2 continued on next page

Figure 2 continued

sliding window of the average MFE observed at each capsid site (red line). (B) Correlation between the average MFE observed at each capsid site and variation in enterovirus B sequence alignments (Shannon entropy). (C) Violin plot of MFE in antibody neutralization sites versus other capsid sites. (D–F) Boxplots of MFE as a function of secondary structure (D), position in the capsid (E), or the predicted effect of mutations on stability or aggregation propensity (F). (G) Validation of the MFE obtained by DMS using a competition assay. For each mutant, the average and standard deviation of the MFE obtained by DMS (n = 3) is plotted against the average and standard deviation of the fitness derived using the competition assay (n = 4). A two-sided Mann–Whitney test was used for two- category comparisons.

The online version of this article includes the following figure supplement(s) for figure 2:

Figure supplement 1. Correlation of amino acid preferences observed in experimental replicates.

Prediction of MFE from available structural and sequence information

As MFE correlated with natural sequence variation and different structural features of the capsid (*Figure 2*), we next investigated if MFE could be predicted from available structural and sequence information. For this, we obtained a dataset of 52 parameters, including structural information derived from the crystal structure of the CVB3 capsid (PDB:4GB3), amino acid properties, and natural variation in available enterovirus sequences (Shannon entropy), and predicted the effects of mutation on stability and aggregation propensity using FoldX (*Schymkowitz et al., 2005*) and TANGO (*Fernandez-Escamilla et al., 2004*), respectively (*Supplementary file 6*). We then employed a random forest algorithm to identify the parameters that can best predict MFE, limiting our analysis to sites that present in the crystal structure and where mutations were observed in at least two replicates to improve accuracy (total of 9685 mutations). Overall, a model trained on 70% of the dataset was able to predict the remaining 30% of the data (2905 mutations) with high accuracy (Spearman's $\rho > 0.75$, Pearson's $r = 0.76$; $p<10^{-16}$; *Figure 3—figure supplement 1*). Surprisingly, a random forest model trained on the top five predictors alone showed similar accuracy (Spearman's $\rho = 0.73$, Pearson's $r = 0.73$; $p<10^{-16}$; *Figure 3B*). Excluding natural sequence variation, amino acid identity, or structural attributes reduced model predictability significantly (>20%; data not shown), suggesting
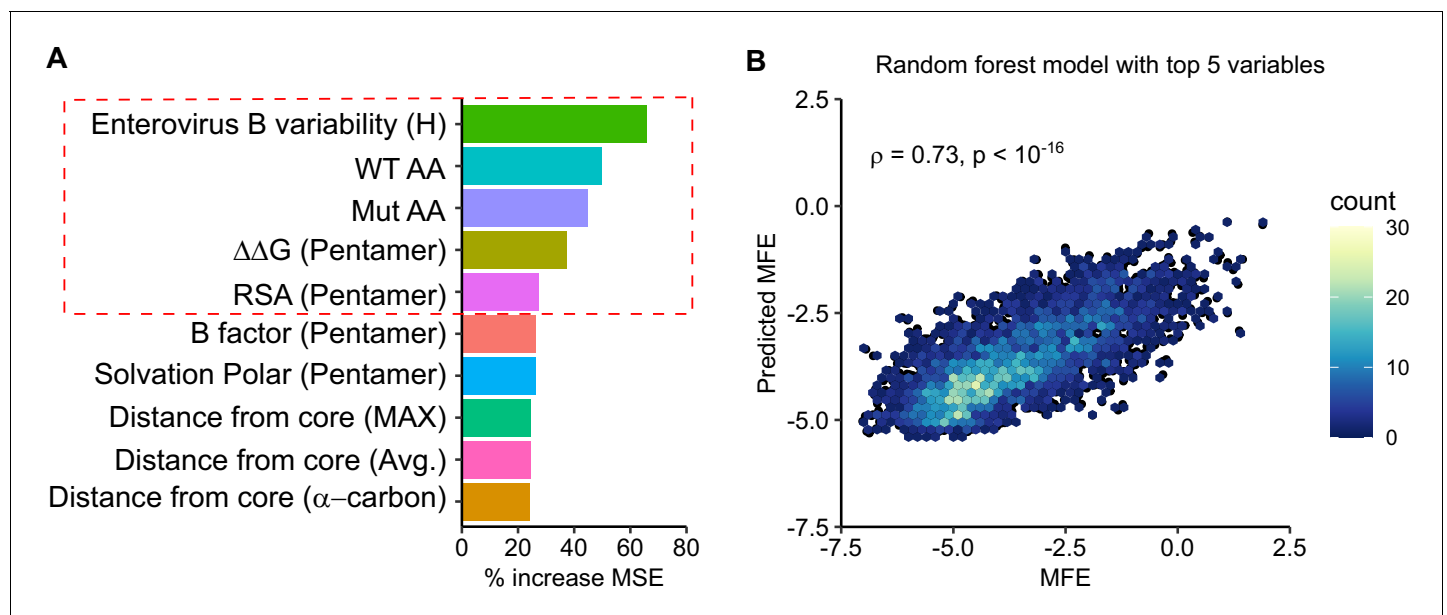
Figure 3. Prediction of MFE based on structural and sequence information. (A) The top 10 predictors identified in a random forest model for explaining MFE in the CVB3 capsid based on the percent of mean squared error (MSE) increase. (B) Hexagonal plot showing the correlation between MFE predicted using a random forest algorithm trained on the top five variables versus observed MFE. The random forest model was trained on 70% of the data and then tested on the remaining 30% (shown). RSA, relative surface area.

The online version of this article includes the following figure supplement(s) for figure 3:

Figure supplement 1. Prediction of mutational fitness effects using random forest or linear models.

a combination of evolutionary, sequence, and structural information best explains MFE. Using an alternative approach, we were able to predict the data with slightly lower accuracy using a linear model with the same five predictors (p<10^{-16}, Spearman's ρ = 0.67, Pearson's r = 0.67; *Figure 3—figure supplement 1*). Together, these results suggest that the prediction of MFE in the CVB3 capsid can be achieved at relatively high accuracy based on available structural and sequence information. Due to the high conservation of capsid structure in picornaviruses, as well as the availability of numerous capsid sequences and structures, these findings are likely generalizable to related picornaviruses.

Experimentally measured MFE inform of natural evolutionary processes

We next examined if our experimentally measured MFE could improve phylogenetic models of CVB3 evolution by incorporating site-specific amino acid preferences using PhyDMS (*Hilton et al., 2017*). Indeed, significant improvement in model fit was observed (*Table 1* PHY; p<10^{-16} using a log-likelihood test compared to non-site-specific codon models), supporting the relevance of our results to understanding evolutionary processes in nature. Nevertheless, selection in nature was significantly more stringent than in the lab (β = 2.18), indicating the presence of additional selection pressures. As laboratory conditions lack selection from antibodies, we used the sum of the absolute differential selection observed at each site (*Bloom, 2017*) to examine whether known antibody neutralization sites show differential selection between the two environments (*Supplementary file 7*). Indeed, antibody neutralization sites showed significantly higher differential selection values compared to other residues (p<10^{-6} by Mann–Whitney test; *Figure 4A*). Moreover, the three sites showing the strongest overall differential selection were found in known antibody neutralization sites: positions 226 and 242 in the EF loop (residues 157 and 173 of VP2) and position 650 in the BC loop (residue 80 of VP1; *Figure 4B–D* and *Supplementary file 7*). In summary, incorporation of experimentally derived amino acid preferences into phylogenetic analyses significantly improved model fit and identified residues in antibody neutralization sites that show differential selection, suggesting these may play important roles in immune evasion in vivo.

Insights into capsid-encoded motifs: myristoylation and protease cleavage

Picornavirus capsids undergo a complex assembly path to generate the infectious particle. These include myristoylation, cleavage by the viral proteases 2A and 3CD^{pro}, as well as interaction with cellular chaperones and glutathione (*Corbic Ramljak et al., 2018*; *Geller et al., 2007*; *Jiang et al., 2014*; *Qing et al., 2014*; *Thibaut et al., 2014*; *Figure 5A*). Having obtained a comprehensive dataset for MFE across the capsid, we next examined the sequence requirements for several of these capsid-encoded motifs. Specifically, myristoylation of the N-terminal glycine is essential for virion assembly (*Corbic Ramljak et al., 2018*). In agreement with this, the N-terminal glycine in the CVB3 capsid showed the strongest average fitness cost upon mutation in the capsid (*Figure 4—figure supplement 1* and *Supplementary file 4*). The remaining sites in the myristoylation motif agreed with the canonical myristoylation motif in cellular proteins (Prosite pattern PDOC00008) (*Bologna et al., 2004*), albeit with increased selectivity at three of the six positions (*Figure 4—figure supplement 1*). On the other hand, a conserved WCPRP motif in the C-terminal region of VP1 that was shown to be important for 3CD^{pro} cleavage of the related foot and mouth disease virus capsid (FDMV; YCPRP motif) (*Kristensen and Belsham, 2019*) was found to be intolerant to mutations

Table 1. Incorporation of DMS results in evolutionary models better describes natural CVB3 evolution compared to standard codon models.

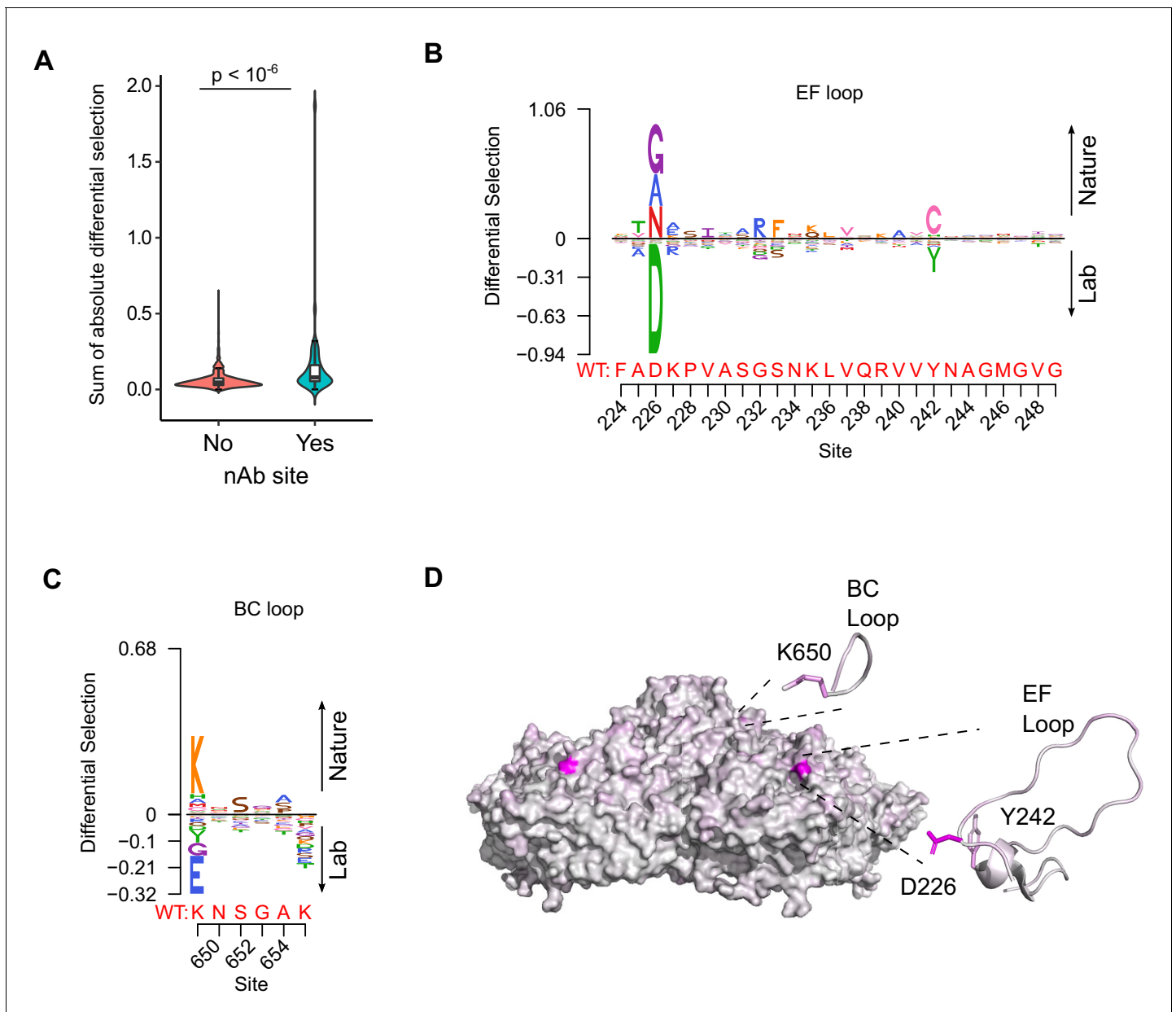| Model	ΔAIC	Log-likelihood	Parameters	Parameter values
ExpCM	0.00	−14,580.51	6	Beta = 2.18, kappa = 7.47, omega = 0.16
Goldman-Yang M5	4187.56	−16,668.29	12	Alpha_omega = 0.30, beta_omega = 10.00, kappa = 7.15
Averaged ExpCM	4303.74	−16,732.38	6	Beta = 0.61, kappa = 7.55, omega = 0.02
Goldman-Yang M0	4371.26	−16,761.14	11	Kappa = 7.14, omega = 0.02

Figure 4. Antibody neutralization sites show differential selection between laboratory conditions and nature. (**A**) Violin plot showing the sum of absolute differential selection observed at capsid sites comprising antibody neutralization epitopes (nAb) versus all other capsid sites. (**B–C**) Logoplots showing the observed differential selection of sites in the EF loop or BC loop. The WT sequence is indicated in red. (**D**) The CVB3 capsid pentamer (PDB:4GB3), colored according to the amount of differential selection. The BC and EF loops are shown next to the structure together with the side chains for sites showing the highest differential selection.

The online version of this article includes the following figure supplement(s) for figure 4:

Figure supplement 1. Sequence preferences of capsid-encoded motifs.

compared to other capsid residues (p<0.05 versus all other positions by Mann–Whitney test; sites 815–819 in CVB3). Moreover, within this motif, the sites showing the highest average fitness cost in our DMS dataset were identical to analogous positions in FMDV that resulted in a loss of viability upon mutation to alanine (*Figure 4—figure supplement 1*; *Kristensen and Belsham, 2019*), highlighting the conservation of this motif across different picornaviruses.

The viral 3C protease (3C^{pro}) cleaves the picornavirus capsid at two conserved glutamine–glycine (QG) pairs to liberate the viral capsid proteins VP0, VP3, and VP1 (*Figure 5A*). Previous work has
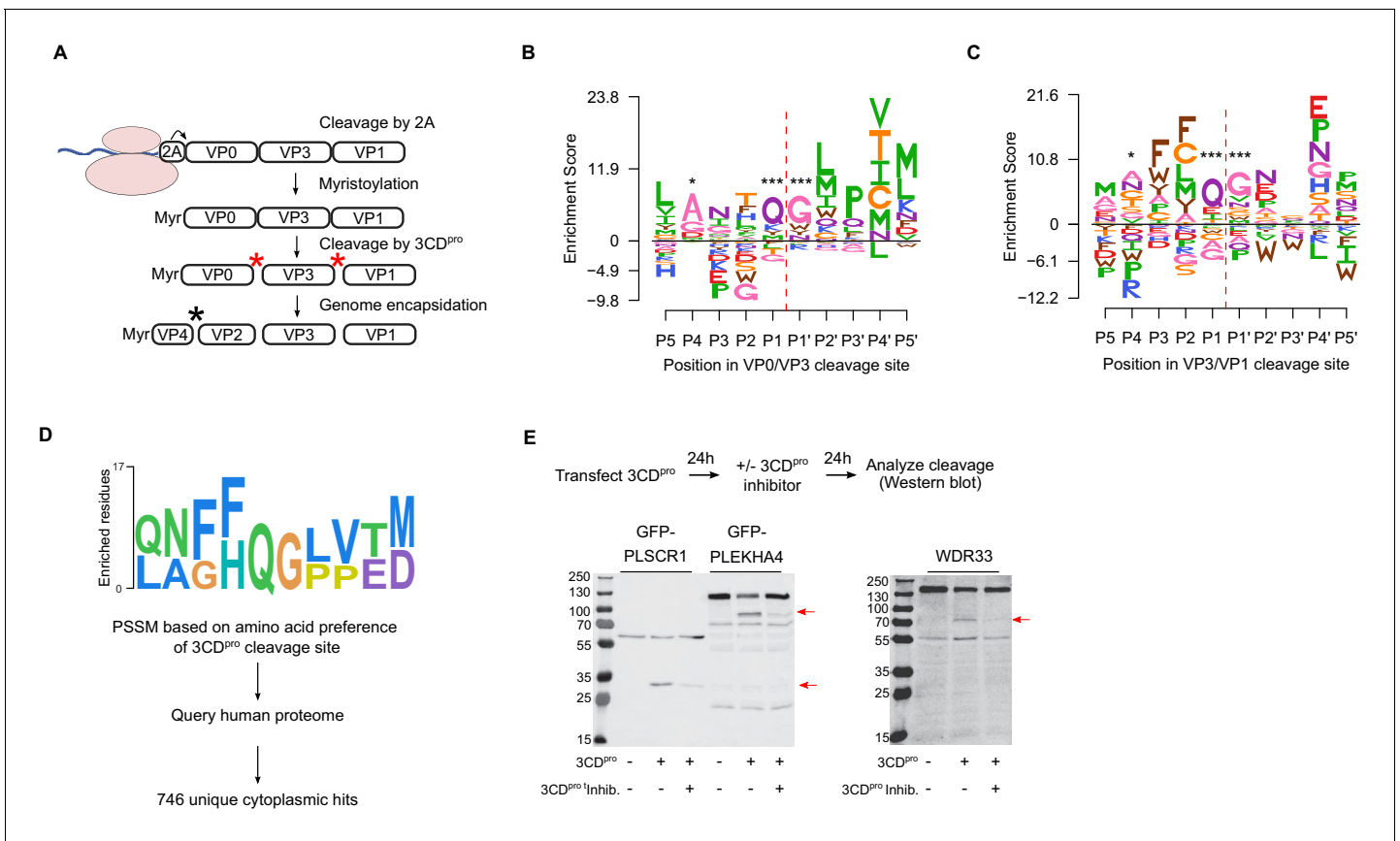
Figure 5. Sequence preference of capsid 3CD^{pro} cleavage sites and their use for the identification of novel cellular targets of the viral protease. (**A**) Overview of the CVB3 capsid maturation pathway. The CVB3 capsid precursor P1 is co-translationally cleaved by the viral 2A protease. P1 is then myristoylated and cleaved by the viral 3CD^{pro} to generate the capsid proteins VP0, VP3, and VP1. Finally, upon assembly and genome encapsidation, VP0 is further cleaved into VP4 and VP2 in a protease-independent manner to generate the mature capsid. Red and black asterisks indicated 3CD^{pro} or protease-independent cleavage events, respectively. (**B,C**) Logoplots showing amino acid preferences for the 10 amino acid regions spanning the 3CD^{pro} cleavage sites (P1–P'1) of both VP0/VP3 and VP3/VP1 in the DMS dataset. (**D**) Overview of the bioinformatic pipeline for identification of novel 3CD^{pro} cellular targets using the amino acid preferences for the capsid cleavage sites from our DMS study. A position-specific scoring matrix (PSSM) was generated based on the amino acid preferences for the 10 amino acid regions spanning the two 3CD^{pro} cleavages sites. This PSSM was then used to query the human genome for potential cellular targets, and non-cytoplasmic proteins were filtered out, yielding 746 proteins. (**E**) The cellular proteins PLSCR1, PLEKHA4, and WDR33 are cleaved by 3CD^{pro}. Western blot analysis of cells cotransfected with 3CD^{pro} and GFP-PLSCR1 or GFP-PLEKHA4 and probed with a GFP antibody or transfected with 3CD^{pro} and probed using a WDR33 antibody. When indicated, the 3CD^{pro} inhibitor rupintrivir was included to ensure cleavage was mediated by the viral protease. Red arrows indicate cleavage products of the expected size (GFP-PLSCR1 full length = 64 kDa, cleaved N-terminus = 36 kDa; GFP-PLEKHA4 full length = 118 kDa, cleaved N-terminus = 72 kDa; WDR33 full length = 146 kDa, cleaved N-terminus = 72 kDa). *p<0.05, ***p<0.001.

The online version of this article includes the following figure supplement(s) for figure 5:

Figure supplement 1. Evaluation of select hits identified as potential 3CD^{pro} target proteins.

defined the sequence specificity of several picornavirus 3C^{pro} enzymes by examining both natural sequence variation and in vitro cleavage assays using synthetic peptides (*Laitinen et al., 2016*). However, unlike other 3C^{pro}-mediated cleavage events in the viral polyprotein, the capsid is only efficiently cleaved by the precursor protein 3CD^{pro} (*Ypma-Wong et al., 1988*). To gain insights into the sequence specificity of 3CD^{pro}, we examined the amino acid preferences for a 10 amino acid region surrounding the protease cleavage site (P5–P5'). As expected based on the known specificity of the 3C protease (*Laitinen et al., 2016*), a strong preference for the presence of QG was observed at both 3CD^{pro} cleavage sites in our dataset (positions P1 and P1' in the cleavage site; *Figure 5B,C*). Interestingly, significant correlation in amino acid preferences between the two cleavage sites was observed only at P1–P1' (Pearson's ρ > 0.99, p<10^{-16}) and P4 (Pearson's ρ > 0.49, p<0.05), as was the case in the enterovirus B alignments (Pearson's ρ > 0.84 and p<10^{-6} for positions P4, P1, and

P'1; data not shown). Hence, the low agreement in amino acid preferences observed for most positions across the two 3CD^pro cleavage sites suggests cleavage is strongly dictated by positions P4, P1, and P1'.

Identification of 3CD^pro cellular targets based on the sequence preferences of capsid-encoded protease cleavage sites

In addition to cleaving the viral polyprotein, the picornavirus proteases cleave cellular factors to facilitate viral replication, including both antiviral factors and cellular factors that favor viral IRES-driven translation mechanism over cellular cap-dependent translation (e.g. DDX58, eIF4G, and PABP) (*Laitinen et al., 2016*; *Sun et al., 2016*). As the canonical 3C/3CD^pro QG cleavage site occurs on average 1.6 times per protein in the human proteome (~33,000,000 times), we sought to examine whether the rich dataset we obtained for the amino acid preferences of the capsid 3CD^pro cleavage sites can be used to identify novel cellular factors that are targeted by the viral protease. Specifically, a position-specific score matrix (PSSM) was generated for the 10 amino acid regions spanning the two protease cleavage sites in the CVB3 capsid (P5–P5') based on the amino acid preferences identified in our study (*Figure 5D*). This PSSM was then used to query the human proteome for potential cleavage sites, yielding a total of 746 cytoplasmic proteins (*Figure 5D*; *Supplementary file 8*). Eleven cellular factors that are known to be cleaved during enterovirus infection were identified using this approach, including the viral sensor Probable ATP-dependent RNA helicase DDX58 (RIG1), the immune transcription factors p65 (RELA) and interferon regulatory factor 7 (IRF7), and polyadenylate-binding protein 1 (PABPC1), an important factor in translation initiation and mRNA stability (*Supplementary file 8*; *Jagdeo et al., 2018*; *Laitinen et al., 2016*).

To evaluate whether our approach can identify novel cellular targets for the viral protease, we examined the ability of 3CD^pro to cleave eight different proteins found in the data set, focusing on those with cellular functions of potential relevance to CVB3 biology and which could be readily detected in our cell culture assay (e.g. availability of antibodies or tagged-variants, cleavage fragments of observable size, and high expression level). These included four interferon-inducible proteins (Pleckstrin homology domain containing A4, PLEKHA4; phospholipid scramblase 1, PLSCR1; NOD-like receptor family CARD domain containing 5, NLRC5; zinc finger CCCH-type containing, antiviral 1, ZC3HAV1) and four proteins involved in various cellular functions, namely apoptosis (MAGE family member D1, MAGED1), RNA processing (WD repeat domain 33, WDR33), and vesicle transport (cyclin G-associated kinase, GAK; tumor susceptibility 101, TSG101). Of these, three proteins were cleaved upon expression of the viral protease to generate fragments of the expected size (PLSCR1, PLEKHA4, and WDR33; *Figure 5E* and *Supplementary file 8*). Of note, while WDR33 was predicted to harbor two potential cleavage sites, only a single cleavage event was observed. Treatment with a specific 3CD^pro inhibitor, rupintrivir (*Dragovich et al., 1999*), blocked the cleavage of these proteins, indicating the effect was due to the viral protease (*Figure 5D*). In contrast, five of the proteins were found to not be cleaved upon 3CD^pro expression, suggesting additional determinants are involved in the cleavage of host factors (*Figure 5—figure supplement 1*). Hence, our approach correctly identified 30% of the predicted cleavage sites (three of the nine different cleavage sites), indicating a strong enrichment of cellular targets of the 3CD^pro in the dataset.

Discussion

The picornavirus capsid is a highly complex structure that plays key roles in viral biology and pathogenesis. In the current study, we employ a comprehensive approach to define the effects of single amino acid mutations in the CVB3 capsid, measuring the effects of >90% of all possible mutations. We find that most mutations in the capsid are deleterious to growth in cell culture, with very few mutations showing higher fitness than the WT sequence (1.2% of all mutations). Similar results have been reported in other non-enveloped capsid proteins (*Acevedo et al., 2014*; *Hartman et al., 2018*; *Ogden et al., 2019*) as well as non-capsid viral proteins (*Ashenberg et al., 2017*; *Bloom, 2014*; *Doud and Bloom, 2016*; *Du et al., 2016*; *Haddox et al., 2016*; *Hom et al., 2019*; *Thyagarajan and Bloom, 2014*; *Wu et al., 2015*). In light of these results, it is likely that the large population sizes of RNA viruses help maintain viral fitness in the face of high mutation rates and strong mutational fitness costs. It is important to note that the effect of a particular mutation on

fitness observed under laboratory conditions may not always reflect its effect in nature due to inherent differences between these two environments.

Investigation of the factors that influence MFE in the capsid revealed a strong correlation with various structural and functional attributes. These included computationally predicted effects on stability and aggregation propensity, secondary structure, and surface exposure (*Figure 2*). Surprisingly, we find that MFE can be predicted with relatively high accuracy using only five parameters: natural sequence variation, the identity of the original and mutant amino acids, the predicted effect on protein stability, and relative solvent accessibility (*Figure 3*). A recent study examined the ability of 46 different variant effect prediction tools to predict MFE from 31 different DMS datasets of both viral and non-viral proteins (*Livesey and Marsh, 2020*). Overall, viral proteins showed the lowest predictability (Spearman's correlation of <0.5). In contrast, we were able to predict MFE using a random forest model using these above-mentioned five parameters with an accuracy similar to the best prediction obtained in this analysis for any viral or non-viral protein (Pearson's r = 0.73; Spearman's ρ = 0.73; *Figure 3B*). Interestingly, SNAP2 (*Hecht et al., 2015*), a neural network-based classifier of mutational effects that was shown to correlate well with MFE in other studies (*Gray et al., 2018*; *Livesey and Marsh, 2020*; *Reeb et al., 2020*), correlated poorly with our data ($R^2 = -0.26$). Overall, considering the relative conservation of capsid structure in picornaviruses as well as the availability of both capsid sequences and high-resolution structures for numerous members of this family, it is likely that these findings can be extrapolated to additional picornaviruses.

Incorporating site-specific amino acid preferences obtained from our DMS results into phylogenetic models was found to significantly improve model accuracy. This has been observed in DMS studies with other RNA viruses (*Bloom, 2017*; *Doud and Bloom, 2016*; *Haddox et al., 2018*) and indicated that our laboratory-measured MFE capture additional information that cannot be obtained from sequence analysis alone. In addition, this approach allowed us to assess which sites show differential selection patterns as a result of the distinct environments encountered in nature and the laboratory. As expected, pressure from the adaptive immune system was found to be the major difference between these environments, with residues in antibody neutralization sites showing higher differential selection compared to other sites in the capsid (*Figure 4A*). Moreover, the sites showing the highest degree of differential selection were found in known antibody neutralization sites (*Figure 4B–D*). However, why these particular residues within antibody neutralization sites show differential selection, while others do not, remain to be elucidated. It has been shown that one, or a few, sites within antibody binding regions can have strong effects on escape from antibody neutralization (*Lee et al., 2019*), potentially explaining these findings. Interestingly, while the top three sites showing differential selection were in antibody neutralization sites, the mutation showing the fourth-highest differential selection was found in the HI loop of VP1. While not classically considered an antibody epitope, this loop has been shown to interact with an antibody fragment in the picornavirus coxsackievirus A6 (*Xu et al., 2017*), is known to mediate receptor binding in different picornaviruses (*Belnap et al., 2000*; *Xing et al., 2000*), and to interact with host cyclophilin A to facilitate uncoating (*Qing et al., 2014*). Whether these factors or others are responsible for the observed differential selection remains to be elucidated.

The CVB3 capsid encodes the information for directing myristoylation, protease cleavage, and interaction with host factors. We took advantage of our data to examine the sequence specificity and mutational tolerance of several known capsid-encoded motifs. First, we examined the amino acid preferences of the CVB3 capsid myristoylation motif. We observe a strong correlation with the canonical myristoylation pattern (Prosite pattern PDOC00008), although with greater intolerance to mutations in three of the six residues in the capsid (*Figure 4—figure supplement 1*). This is likely to stem from additional constraints imposed by capsid structure. On the other hand, we examined the amino acid preference of a conserved motif in VP1 that is required for 3CD^pro-mediated cleavage of picornavirus capsids (*Kristensen and Belsham, 2019*). Our data showed a higher cost to mutation in this motif relative to other capsid positions (*Figure 4—figure supplement 1*), highlighting its importance for capsid function. Finally, we examined the sequence preferences surrounding the two 3CD^pro cleavage sites. We find a strong dependence on the cleavage site residues (positions P1 and P1'; *Figure 5*) and to a lesser degree position P4, with large variation in the sequence preferences across the remaining positions between the two cleavage sites. Overall, our experimentally measured MFE are congruent with existing information regarding the sequence preferences of the

examined capsid motifs, yet provide in-depth insights into sequence specificity that cannot be obtained from examining natural sequence variation.

Finally, we used the amino acid preferences observed in 3CD[pro] cleavage sites within the capsid to query the human genome for potential cellular targets of this protease (*Figure 5D*). Using this approach, we identify 746 cytoplasmic proteins that harbor a potential 3CD[pro] target sequence, including 11 proteins previously shown to be cleaved by different picornavirus 3C proteases. We then validated our approach using eight proteins, comprising nine predicted cleavage sites. Six of the predicted cleavage sites were not affected by 3CD[pro] expression (*Figure 5—figure supplement 1*). On the other hand, three proteins were observed to be specifically cleaved by the viral protease (*Figure 5E*): WD repeat domain 33 (WDR33), an important factor for polyadenylation of cellular pre-mRNAs (*Chan et al., 2014*) that has been shown to act as a restriction factor during influenza infection (*Brass et al., 2009*); the interferon-induced protein phospholipid scramblase 1 (PLSCR1), which is involved in the replication of numerous viruses, likely due to its ability to enhance the expression of certain interferon-stimulated genes (*Kodigepalli et al., 2015*); and the interferon-induced Pleckstrin homology domain containing A4 (PLEKHA4), a plasma membrane-localized signaling modulator (*Shami Shah et al., 2019*) that is currently not known to play a role in viral infection. Overall, our approach correctly predicts 30% of the identified cleavage sites. It is likely that incorporating additional selection criteria, such as accessibility of the cleavage peptide in the folded structure, can be used to further reduce false positives. Nevertheless, extrapolating our validation results to the larger dataset suggests >200 new host targets of the protease are identified, some of which could play key roles in viral biology and pathogenesis.

Materials and methods

Key resources table

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Strain, strain background (coxsackievirus B3)	pCVB3-XhoI-P1-Kpn21	10.1016/j.celrep.2019.09.014		Infectious CVB3 clone based on the Nancy strain (Taxon identifier 103903)
Strain, strain background (coxsackievirus B3)	pCVB3-XhoI-ΔP1-Kpn21	This paper		Infectious CVB3 clone without P1 region
Strain, strain background (coxsackievirus B3)	Marked reference CVB3 virus	10.1038/nmicrobiol.2017.88		Infectious CVB3 clone with silent mutations in the polymerase region used as a reference for fitness assays
Strain, strain background (*Escherichia coli*)	NZY5α	NZY Tech	MB004	Competent cells, standard cloning
Strain, strain background (*Escherichia coli*)	MegaX DH10B T1R Electrocomp cells	ThermoFisher	C6400-03	Electrocompetent cells, library cloning
Cell line (*Homo sapiens*)	HeLa-H1	ATCC	CRL-1958; RRID:CVCL_3334	Cell line for CVB3 infection and DMS library production
Cell line (*Homo sapiens*)	HEK293	ATCC	CRL-1573; RRID:CVCL_0045	Cell line used for production of CVB3 mutants and for protease cleavage
Antibody	Anti-GFP (Mouse monoclonal)	SantaCruz	Sc-9996	Western blot (1:2000)

Continued on next page

Continued

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Antibody	Anti-FLAG (Mouse monoclonal)	SantaCruz	Sc-166335	Western blot (1:2000)
Antibody	Anti-HA (Mouse monoclonal)	SantaCruz	Sc-7392	Western blot (1:2000)
Antibody	Anti-WDR33 (Mouse monoclonal)	SantaCruz	Sc-374466	Western blot (1:1000)
Antibody	Anti-TSG101 (Mouse monoclonal)	SantaCruz	Sc-136111	Western blot (1:1000)
Antibody	Anti-GAK (Mouse monoclonal)	SantaCruz	Sc-137053	Western blot (1:1000)
Antibody	Anti-MAGED1 (Mouse monoclonal)	SantaCruz	Sc-393291	Western blot (1:1000)
Recombinant DNA reagent	DMS libraries (1–3)	This paper		CVB3 infectious clone libraries with mutagenized capsid region
Recombinant DNA reagent	pUC19-HiFi-P1 (plasmid)	This paper		CVB3 capsid region used as template for DMS cloned into SalI digested pUC19 vector. Used for site-directed mutagenesis
Recombinant DNA reagent	T7 encoding plasmid (plasmid)	10.1128/jvi.02583–14	RRID:Addgene_65974	Plasmid encoding T7 polymerase for transfection
Recombinant DNA reagent	pIRES-3CDpro (plasmid)	This paper		CVB3 3 CD protease region cloned into XhoI and NotI pIRES plasmid (Clonetech)
Recombinant DNA reagent	peGFP_PLEKHA4	10.1016/j.celrep.2019.04.060		Kind gift from Dr. Jeremy Baskin GFP-PLEKHA4 expression plasmid
Recombinant DNA reagent	peGFP_PLSCR1	10.1371/journal.pone.0005006		Kind gift from Dr. Serfe Benichou GFP-PLSCR1 expression plasmid
Recombinant DNA reagent	pAcGFP-C1 WDR33	https://doi.org/10.1016/j.molcel.2018.11.036		Kind gift from Dr. Matthias Altmeyer pAcGFP-C1 WDR33 expression plasmid
Recombinant DNA reagent	FLAG-NLCR5	Addgene	RRID:Addgene_37521	NLCR5 expression plasmid
Recombinant DNA reagent	HA-ZC3HAV1	Addgene	RRID:Addgene_45907	HA-ZC3HAV1 expression plasmid
Recombinant DNA reagent	Fluc-eGFP	Addgene	RRID:Addgene_90170	Fluc-eGFP expression plasmid

Continued on next page

Continued

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Sequence-based reagent	HiFi_F	IDT	PCR primer	For generating PCR to clone libraries and sequencing: CTTTGTTGGGTTT ATACCACTTAGC TCGAGAGAGG
Sequence-based reagent	HiFi_R	IDT	PCR primer	For generating PCR to clone libraries and sequencing: CCTGTAGTTCCCCA CATACACTGCTCCG
Sequence-based reagent	DMS primers	IDT	PCR primer	Primers spanning the full coding region of the CVB3 capsid to perform codon mutagenesis. Listed in *Supplementary file 1*.
Sequence-based reagent	2045_F	IDT	PCR primer	Primer used for Sanger sequencing. TCGAGTGTTTTTA GTCGGACG
Sequence-based reagent	2143_R	IDT	PCR primer	Primer used for Sanger sequencing. TCGAGTGTTTTT TAGTCGGACG
Sequence-based reagent	3450_RT	IDT	PCR primer	Primer used for Sanger sequencing and RT-PCR. TCGAGTGTTTTT AGTCGGACG
Sequence-based reagent	qPCR_F	10.1038/nmicrobiol. 2017.88	PCR primer	qPCR primer for competition assays. GATCGCATATG GTGATGATGTGA
Sequence-based reagent	qPCR_R	10.1038/nmicrobiol. 2017.88	PCR primer	qPCR primer for competition assays. AGCTTCAGCGAGT AAAGATGCA
Sequence-based reagent	MGB_CVB3_wt	10.1038/nmicrobiol. 2017.88	TaqManProbe	qPCR probe for competition assays. 6FAM-CGCATCGTA CCCATGG-TAMRA
Sequence-based reagent	MGB_CVB3_Ref	10.1038/nmicrobiol. 2017.88	TaqManProbe	qPCR probe for competition assays. HEX-CGCTAGCTA CCCATGG-TAMRA
Sequence-based reagent	Q8D_F	IDT	PCR primer	Primer for site-directed mutagenesis: gtatcaacgGAT aagactggg
Sequence-based reagent	Q8D_R	IDT	PCR primer	Primer for site-directed mutagenesis: ttgag ctcccattttgctgt

Continued on next page

Continued

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Sequence-based reagent	K829L_F	IDT	PCR primer	Primer for site-directed mutagenesis: gagaaggcaCTA aacgtgaac
Sequence-based reagent	K829L_R	IDT	PCR primer	Primer for site-directed mutagenesis: gtattg gcagagtctaggtgg
Sequence-based reagent	K235D_F	IDT	PCR primer	Primer for site-directed mutagenesis: gggtcc aacGATttggtacag
Sequence-based reagent	K235D_R	IDT	PCR primer	Primer for site-directed mutagenesis: gga tgcgaccggtttgtccgc
Sequence-based reagent	R16G_F	IDT	PCR primer	Primer for site-directed mutagenesis: catga gaccGGActgaatgct
Sequence-based reagent	R16G_R	IDT	PCR primer	Primer for site-directed mutagenesis: tgccc cagtcttttgcgttg
Sequence-based reagent	K827G_F	IDT	PCR primer	Primer for site-directed mutagenesis: caatacgagGG Ggcaaagaac
Sequence-based reagent	K827G_R	IDT	PCR primer	Primer for site-directed mutagenesis: gcaga gtctaggtggtctagg
Sequence-based reagent	Q566M_F	IDT	PCR primer	Primer for site-directed mutagenesis: atttcgcagATGaactttttc
Sequence-based reagent	Q566M_R	IDT	PCR primer	Primer for site-directed mutagenesis: gaaaggagtgt ccttcaatag
Sequence-based reagent	T315P_F	IDT	PCR primer	Primer for site-directed mutagenesis: attacgg tcCCCatagcccca
Sequence-based reagent	T315P_R	IDT	PCR primer	Primer for site-directed mutagenesis: tgggacgtacgtggtgga
Sequence-based reagent	N395H_F	IDT	PCR primer	Primer for site-directed mutagenesis: gagaaggtcCAT tctatggaa

Continued on next page

Continued

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Sequence-based reagent	N395H_R	IDT	PCR primer	Primer for site-directed mutagenesis: tccaacatttt ggactgggac
Sequence-based reagent	T849A_F	IDT	PCR primer	Primer for site-directed mutagenesis: actacaatgGTC aatacgggc
Sequence-based reagent	T849A_R	IDT	PCR primer	Primer for site-directed mutagenesis: gatgctttgcct agtagtgg
Sequence-based reagent	K235D_F	IDT	PCR primer	Primer for site-directed mutagenesis: gggtccaacGAT ttggtacag
Sequence-based reagent	K235D_R	IDT	PCR primer	Primer for site-directed mutagenesis: ggatgcgacc ggtttgtccgc
Sequence-based reagent	3C_For	IDT	PCR primer	Primer for cloning CVB3 3 CD into pIRES: TATTCTCGAGACC ATGGGCCCTGC CTTTGAGTTCG
Sequence-based reagent	3D_Rev	IDT	PCR primer	Primer for cloning CVB3 3 CD into pIRES: TATTGCGGCCGCC TAGAAGGAGTCC AACCATTTCCT
Commercial assay or kit	NEBuilder HiFi DNA Assembly kit	NEB	E2621X	Seamless cloning
Commercial assay or kit	TranscriptAid T7 High Yield Transcription Kit	ThermoFisher Scientific	K0441	T7 in vitro transcription kit
Commercial assay or kit	*Quick*-RNA Viral kit	Zymo Research	R1035	RNA purification
Commercial assay or kit	*DNA Clean andConcentrator-5*	Zymo Research	D4013	DNA purification, gel purification
Commercial assay or kit	Luna Universal Probe One-Step RT-qPCR kit	NEB	E3006X	One-step qPCR master mix
Chemical compound, drug	Rupintivir	Tocris Biosciences	Cat. #: 6414	CVB3 3C protease inhibitor

Continued on next page

Continued

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Software, algorithm	Codon TilingPrimers	https://doi.org/10.1016/j.chom.2017.05.003		Software to design primers for mutagenesis (https://github.com/jbloomlab/CodonTilingPrimers)
Software, algorithm	Sanger Mutant Library Analysis	Dr. Jesse Bloom		Software to assess library mutagenesis by Sanger sequencing (https://github.com/jbloomlab/SangerMutantLibraryAnalysis)
Software, algorithm	Samtools	http://www.htslib.org/	version 1.5	Suite of programs for interacting with high-throughput sequencing data
Software, algorithm	Fastp	10.1093/bioinformatics/bty560		Software for NGS read trimming and QC
Software, algorithm	PicardTools, FastqToSam	https://broadinstitute.github.io/picard/	Version 2.2.4	Used to generate Bam files from Fastq files
Software, algorithm	Duplex pipeline	https://github.com/KennedyLabUW/Duplex-Sequencing; *Kennedy et al., 2014*	Version 3.0	Analysis pipeline for duplex sequencing (UnifiedConsensusMaker.py)
Software, algorithm	VariantBam	10.1093/bioinformatics/btw111		Software to filter Bam files
Software, algorithm	BWA	https://sourceforge.net/projects/bio-bwa/files/	Version 0.7.16	Software to align NGS reads
Software, algorithm	Fgbio	http://fulcrumgenomics.github.io/fgbio/	version 1.1.0	Software used to hard-clip NGS reads
Software, algorithm	VirVarSeq	10.1093/bioinformatics/btu587	version 1.1.0	Software used to identify codons in each NGS read
Software, algorithm	Custom R scripts	This paper		Custom R scripts to process output of VirVarSeq script. Available at https://github.com/RGellerLab/CVB3_Capsid_DMS
Software, algorithm	DMS_tools2	10.1186/s12859-015-0590-4		Software to determine amino acid preferences and mutational fitness effects
Software, algorithm	TANGO	10.1038/nbt1012		Software to determine the effect of mutations on aggregation
Software, algorithm	FoldX	10.1093/nar/gki387		Software to determine the effect of mutations on stability

Continued on next page

Continued

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Software, algorithm	DSSP	http://swift.cmbi.ru.nl/gv/dssp/		Software used to obtain secondary structure and RSA within DMS_tools2
Software, algorithm	ViprDB	http://viperdb.scripps.edu/ *Carrillo-Tripp et al., 2009*		Software used to obtain structural information on capsid sites
Software, algorithm	DECIPHER Package	10.32614/RJ-2016–025		R package for performing codon alignments
Software, algorithm	PhyDMS	doi: 10.7717/peerj.3657		For phylogenetic and differential selection analyses. https://jbloomlab.github.io/phydms/index.html
Software, algorithm	Custom R scripts	This paper		Custom R script to generate in silico peptides spanning 10AA 3 CD protease cleavage site. Available at https://github.com/RGellerLab/CVB3_Capsid_DMS
Software, algorithm	PSSMSearch	10.1093/nar/gky426		Used to generate position-specific scoring matrix and search human proteome for hits. http://slim.icr.ac.uk/pssmsearch/
Software, algorithm	Peptides R package	ISSN 2073–4859	Version 2.4.2	R package to predict molecular weight of proteins
Software, algorithm	RandomForest R package	10.1023/A:1010933404324	Version 4.6–16	R package for random forest prediction
Software, algorithm	Logolas	10.1186/s12859-018-2489-3		Package to generate logo plots in R

Viruses, cells, and plaque assays

HeLa-H1 (CRL-1958; RRID:CVCL_3334) and HEK293 (CRL-1573; RRID:CVCL_0045) cells were obtained from ATCC and were periodically validated to be free of mycoplasma. All work with CVB3 was based on the Nancy infectious clone (kind gift of Dr. Marco Vignuzzi, Institute Pasteur). Cells were cultured in culture media (Dulbecco's modified Eagle's medium [DMEM] with 10% heat-inactivated fetal bovine serum (FBS), Pen-Strep, and L-glutamine) with FBS concentrations of 2% during infection. For plaque assays, serial dilutions of the virus were used to infect confluent HeLa-H1 cells in six-well plates for 45 min, followed by overlaying the cells with a 1:1 mixture of 56℃ 1.6% agar (Arcos Organics 443570010) and 37℃ 2× DMEM with 4% FBS. Two days later, plates were fixed with formaldehyde (2% final concentration) after which the agar was removed and the cells stained with crystal violet to visualize plaques.

Deep mutational scanning

The infectious clone was modified by site-directed mutagenesis to remove an XhoI site present in the capsid region (P1) and introduce an XhoI site at position 692 as well as a Kpn2I site at position 3314, generating a pCVB3-XhoI-P1-Kpn2I clone (*Bou et al., 2019*). In addition, a pCVB3-XhoI-ΔP1-Kpn2I plasmid was generated by replacing the region between the XhoI and Kpn2I sites in pCVB3-XhoI-P1-Kpn2I with a short linker. To generate the template for DMS, the capsid region was amplified by PCR from pCVB3-XhoI-P1-Kpn2I with Phusion polymerase (Thermo Scientific) and primers HiFi-F (CTTTGTTGGGTTTATACCACTTAGCTCGAGAGAGG) and HiFi-R (CCTGTAGTTCCCCACA TACACTGCTCCG) and gel purified (Zymoclean Gel DNA Recovery Kit). Primers spanning the full coding region of the capsid region were designed using the CodonTilingPrimers software from the Bloom lab (https://github.com/jbloomlab/CodonTilingPrimers; *Dingens et al., 2017*) with the default parameters and synthesized by IDT (*Supplementary file 1*). These primers were used to perform the mutagenesis PCR on the capsid template together with the HiFi-F or HiFi-R primers in triplicate following published protocols (*Dingens et al., 2017*) with the exception that 10 rounds of mutagenesis were performed for libraries 1 and 2, while a second round of seven mutagenesis cycles was performed for library three to increase the number of mutation per clone. The products were gel purified and ligated to an XhoI and Kpn2I digested and gel purified pCVB3-XhoI-ΔP1-Kpn2I using NEBuilder HiFi DNA Assembly reaction (NEB) for 25 min. Mutagenesis efficiency was evaluated by the transformation of the assembled plasmids into NZY5α competent cells (NZY Tech), Sanger sequencing of 18–23 clones per library, and mutation analysis using the Sanger Mutant Library Analysis script (https://github.com/jbloomlab/SangerMutantLibraryAnalysis; *Bloom, 2014*). Subsequently, the assembled plasmid reactions were purified using a Zymo DNA Clean and Concentrator-5 kit (Zymo Research) and used to electroporate MegaX DH10B T1R Electrocomp cells (Thermo-Fisher) using a Gene Pulser XCell electroporator (Bio-Rad) according to the manufacturer's protocol. Cells were then grown overnight in a 50 mL liquid culture at 33°C and DNA purified using the Pure-Link HiPure plasmid midiprep kit (Invitrogen). Transformation efficiency was estimated by plating serial dilutions of the transformation on agar plates. In total, 4.44×10^5, 1.46×10^5, and 2.19×10^5 transformants were obtained for lines 1, 2, and 3, respectively. Viral genomic RNA was then transcribed from SalI linearized, gel-purified full-length plasmids using the TranscriptAid T7 kit (Thermo Scientific), and four electroporations were performed using 4×10^6 HeLa-H1 cells in a 4 mm cuvette in 400 µL of calcium- and magnesium-free phosphate-buffered saline (PBS) using with 8 µg of RNA in a Gene Pulser XCell (Bio-Rad) set to 240 V and 950 µF. Electroporated cells were then pooled, and one-fourth was cultured for 9 hr to produce the passage 0 virus (P0). Following three freeze–thaw cycles, 2×10^6 plaque-forming units (PFU) were used to infect a 90% confluent 15 cm plate in 2.5 mL of infection media for 1 hr. Cells were then washed with PBS and incubated in 12 mL of infection media for 9 hr. Finally, cells were subjected to three freeze–thaw cycles, debris removed by centrifugation at $500 \times g$, and the supernatants collected to generate P1 virus stocks. All infections produced $>2.38 \times 10^6$ PFU in P0 and $>1.2 \times 10^7$ PFU in P1 as judged by plaque assay.

Next-generation sequencing analysis

Libraries were prepared following published protocols (*Kennedy et al., 2014*), and each library was run on a Novaseq6000 2×150 at a maximum of 30G per lane to reduce potential index hopping. Reads trimming was performed using fastp (*Chen et al., 2018*) (command: –max_len1 150 –max_-len2 150 –length_required 150 -x -Q -A), unsorted bam files were generated from fastq files using Picard tools FastqToSam (version 2.2.4) and merged into a single bam using the cat command of Samtools (version 1.5). The duplex pipeline was then implemented (https://github.com/KennedyLa-bUW/Duplex-Sequencing/UnifiedConsensusMaker.py; *Kennedy et al., 2014*). using the UnifiedConsensusMaker.py script and a minimum family size of 3, a cutoff of 0.9 for consensus calling, and an N cutoff of 0.3. The single-stranded consensus files (SSCS) were then aligned using BWA mem (version 0.7.16), sorted using Samtools, size selected to be 133 bp long using VariantBam (*Wala et al., 2016*), unaligned reads were discarded (Samtools view command with -F 4), and the resulting bam file indexed with Samtools. Subsequently, fgbio (http://fulcrumgenomics.github.io/fgbio/; version 1.1.0) was used to hard-clip 10 bp from each end and upgrade all clipping to hard-clip (-c Hard –upgrade-clipping true –read-one-five-prime 10 –read-one-three-prime 10 –read-two-five-prime 10 –read-two-three-prime 10). Variant bam was then used to keep all reads

that were between 50 and 150 bp, well-mapped, and had either no indels and less than five mutations (command –r {'':{'rules':[{'ins':[0,0],'del':[0,0],'nm':[0,4], 'mate_mapped':true,'fr':true,'length': [50,150]}]}}"). Finally, the codons in each read were identified using the VirVarSeq (*Verbist et al., 2015*) Codon_table.pl script using a minimum read quality of 20. A custom R script was then used to generate a codon counts table for each codon position by eliminating all codons containing ambiguous nucleotides and codons with a strong strand bias (StrandOddsRatio > 4), as well as all codons that are reached via a single mutation (available at https://github.com/RGellerLab/CVB3_Capsid_ DMS); *Mattenberger, 2021*; copy archived at swh:1:rev:29dd205182f0886dc5bad3e6b4dd-d6e786c58a75). Amino acid preferences and MFE were determined using DMStools2 (*Bloom, 2015*), with the Bayesian option and the default settings.

Structural analyses

The crystal structure PDB:4GB3 (*Yoder et al., 2012*) was used for all structural analyses. The effects of mutations on aggregation were determined using TANGO version 2.3.1 (*Fernandez-Escamilla et al., 2004*) using the default settings, and the effect on stability on the monomer and pentamer was determined using FoldX 4 (*Schymkowitz et al., 2005*) using the default settings. For the latter, the pentamer subunits were renamed to unique letters, all mutations between the reference sequence and the structure sequence were introduced using the BuildModel command, the structure was optimized using the RepairPDB command 5 or 10 times for the pentamer or monomer, respectively, and then the effects of the mutations were predicted using the BuildModel command (modified PDB files can be found at https://github.com/RGellerLab/CVB3_Capsid_DMS). Secondary structure and RSA were obtained from DSSP (http://swift.cmbi.ru.nl/gv/dssp/) using the dms_tools2. dssp function of dms_tools2, while interface, surface, and core residues as well as residue contact number, and presence in the twofold, threefold, and fivefold axes were obtained from ViprDB (http://viperdb.scripps.edu/) (*Carrillo-Tripp et al., 2009*). Distance from the center was calculated with Pymol using the Distancetoatom.py script on the monomer or pentamer. Finally, the location of antibody neutralization sites in CVB3 was obtained from an analysis of the CVB3 capsid structure in a previous publication (*Muckelbauer et al., 1995*).

Generation and evaluation of CVB3 capsid mutants

With the exception of mutant N395H (kind gift of Rafael Sanjuan) (*Bou et al., 2019*), all other mutants were generated by site-directed mutagenesis. For this, the PCR of the capsid region used as a template for DMS was phosphorylated and cloned into a SmaI digested pUC19 vector for use in the mutagenesis reactions (pUC19-HiFi-P1). For each mutant, non-overlapping primers containing the mutation in the middle of the forward primer were used to introduce the mutation with Phusion polymerase, followed by DpnI (Thermo Scientific) treatment, phosphorylation, ligation, and transformation of chemically competent bacteria. Successful mutagenesis was verified by Sanger sequencing. Subsequently, the capsid region was subcloned into pCVB3-XhoI-ΔP1-Kpn2I using XhoI and Kpn2I sites. Plasmids were then linearized with MluI, and 2 µg of plasmid was transfected into 5×10^5 HEK293 cells, together with a plasmid encoding the T7 polymerase (*Yun et al., 2015*) (Addgene 65974) using calcium phosphate. Briefly, an equal volume of 2× HBS (274 mM NaCl, 10 mM KCl, 1.4 mM Na_2HPO_4) was added dropwise to DNA containing 0.25M $CaCl_2$ while mixing, incubated 15 min at RT, and then added dropwise to cells. Following 48 hr, passage 0 (P0) virus was collected and titered by plaque assay. From this, 10^5 PFU were used to infect 90% confluent six-well HeLa-H1 cells (multiplicity of infection (MOI) 0.1) for 1 hr at 37°C, after which the cells were washed twice with PBS and 2 mL of infection media added. Cells were then incubated until cytopathic effect (CPE) was observed. Emerging viral populations were titered by plaque assay and the capsid region sequenced to ensure no compensatory mutations or reversions arose during replication. The fitness of these mutants was then tested by direct competition with a marked reference virus using a Taqman RT-PCR method (*Moratorio et al., 2017*). Briefly, using four biological replicates, confluent HeLa-H1 cells in a 24-well plate were infected with 200 µL of a 1:1 mixture of 4×10^3 PFU (MOI 0.01) of the test and marked reference viruses for 45 min. Subsequently, the inoculum was removed, the cells were washed twice with PBS, 200 µL of infection media was added, and the cells were incubated for 24 hr at 37°C. Finally, cells were subjected to three freeze–thaw cycles, debris removed by centrifugation at $500 \times$ g, the supernatants collected and treated with 2 µL of RNase-Free DNaseI

(ThermoFisher) for 15 min at 37°C, and viral RNA extracted using the *Quick*-RNA Viral Kit (Zymo Research), eluting in 20 µL. Quantification of the replication of each mutant versus the reference was performed using Luna Universal Probe One-Step RT-qPCR kit (New England BioLabs) containing 3 µL of total RNA, 0.4 µM of each qPCR primers, and 0.2 µM of each probe. The standard curve was performed using 10-fold dilutions of RNA extracted from 10^7 PFU of wild-type and reference viruses. All samples were performed with three technical replicates. The relative fitness (W) of each mutant versus the common marked reference virus was calculated using the following formula: $W = [R(t)/R(0)]^{1/t}$, where R(0) and R(t) represent the ratio of the mutant to the reference virus genomes in the initial mixture used for the infection and after 1 day (t = 1), respectively (*Carrasco et al., 2007*; *Moratorio et al., 2017*).

Sequence variability and phylogenetic analyses

Amino acid variability was assessed using Shannon entropy. Briefly, all available, non-identical, full-genome CVB3, CVB, or enterovirus B sequences were downloaded from Virus Pathogen Resource (*Pickett et al., 2012*) (http://www.viprbrc.org) and codon-aligned using the DECIPHER package in R (available at https://github.com/RGellerLab/CVB3_Capsid_DMS). All alignment positions not present in our reference strain were removed, and a custom R script was used to calculate Shannon entropy. For phylogenetic and differential selection analyses, PhyDMS was run using the default settings on an alignment of CVB3 genomes that was processed with the phydms_prealignment module and using the average preferences from the three DMS replicates.

Identification of 3CD^{pro} cleavage sites in the human proteome

The amino acid preferences (the relative enrichment of each amino acid at each position standardized to 1) was used to generate in silico 1000 peptides spanning the 10 amino acid regions surrounding each cleavage site using a custom R script (available at https://github.com/RGellerLab/CVB3_Capsid_DMS). Specifically, for each peptide position, 100 peptides were generated that encoded each amino acid at a frequency corresponding to its preference observed in the DMS results, with the remaining positions unchanged. The resulting 1000 peptides from each cleavage site were uploaded to PSSMSearch (*Krystkowiak et al., 2018*) (http://slim.icr.ac.uk/pssmsearch/) using the default setting (psi_blast IC). Results were filtered to remove proteins indicated to be secreted, lumenal, or extracellular in the Warnings column. To test whether proteins were cleaved by the viral 3 CD protease, the corresponding region was PCR amplified from the Nancy infectious clone (primers 3C-For: TATTCTCGAGACCATGGGCCCTGCCTTTGAGTTCG and 3D-Rev: TATTGCGGCCGCC TAGAAGGAGTCCAACCATTTCCT) and cloned into the pIRES plasmid (Clonetech) using the restriction sites XhoI and NotI (pIRES-3CD^{pro}). For analysis of fusion proteins, HEK293 cells were transfected with GFP-PLEKHA4 (kind gift of Dr. Jeremy Baskin, Cornell University), GFP-PLSCR1 (kind gift of Dr. Serge Benichou, Institut Cochin), pAcGFP-WDR33 (Kind gift of Dr. Matthias Altmeyer, University of Zurich), FLAG-NLCR5 (Addgene #37521), HA-ZC3HAV1 (Addgene #45907), or the control plasmid FLuc-eGFP (Addgene #90170), together with the pIRES-3CD^{pro} plasmid using Lipofectamine 2000. Following 24 hr, proteins were collected by lysing in lysis buffer (50 mM Tris–HCl, 150 mM NaCl, 1% NP40, and protease inhibitor cocktail [Complete Mini EDTA-free, Roche]) and subjected to western blotting with the corresponding antibody (anti-GFP, Santa Cruz sc-9996; anti–FLAG, Santa Cruz sc-166335; anti-HA, Santa Cruz, sc-7392). For analysis of endogenous proteins, 3CD^{pro} was expressed for 48 hr before cell lysis, and western blotting using antibodies against WDR33 (Santa Cruz sc-374466), TSG101 (Santa Cruz sc-136111), GAK (Santa Cruz sc-137053), and MAGED1 (Santa Cruz sc-393291). When indicated, the 3C^{pro} inhibitor rupintrivir (Tocris Biosciences) was added at a concentration of 2 µM for the last 24 hr before collection. The predicted molecular weight of cleaved fragments was calculated using the mw function of the Peptides R package (version 2.4.2).

Statistical analyses

All experiments were performed with at least three biological replicates with the exception of the analysis of protein cleavage by western blotting, which was performed in duplicate. All statistical analyses were performed in R and were two tailed. For random forest prediction, the R RandomForest package (version 4.6–14) was employed using the default setting with an mtry of 10, and for the linear model, the formula lm(MFE ~ enterovirus B entropy + WT amino acid * mutant amino acid +

predicted effect of mutations on stability in the pentamer + relative surface exposure) was used (available at https://github.com/RGellerLab/CVB3_Capsid_DMS). Sequence logoplots were producing using Logolas (*Dey et al., 2018*).

Data availability

Unaligned bam files have been uploaded to SRA (Bioproject PRJNA643896, SRA SRP269871, Accession SRX8663374-SRX8663384). The scripts and data required to obtain the codon count tables for all samples, to perform the random forest and linear model predictions, to generate the peptides for use with PSSMsearch, as well as the sequence alignments and modified structure files for FoldX analysis, can be found on GitHub (https://github.com/RGellerLab/CVB3_Capsid_DMS). Finally, the interactive heatmap of MFE across the capsid was generated by modifying a script from a prior publication (*Starr et al., 2020*) (available at https://github.com/jbloomlab/SARS-CoV-2-RBD_DMS/blob/master/interactive_heatmap.ipynb) and can be found on this projects' GitHub page (https://github.com/RGellerLab/CVB3_Capsid_DMS).

Additional information

Author contributions

Florian Mattenberger, Conceptualization, Formal analysis, Investigation, Methodology, Writing - original draft, Writing - review and editing; Victor Latorre, Conceptualization, Investigation, Methodology, Writing - review and editing; Omer Tirosh, Software, Formal analysis, Methodology; Adi Stern, Software, Formal analysis, Supervision, Methodology, Writing - original draft, Writing - review and editing; Ron Geller, Conceptualization, Data curation, Formal analysis, Supervision, Funding acquisition, Investigation, Methodology, Writing - original draft, Writing - review and editing

Author ORCIDs

Florian Mattenberger (iD) https://orcid.org/0000-0002-2727-0284
Omer Tirosh (iD) https://orcid.org/0000-0001-8139-9866
Adi Stern (iD) http://orcid.org/0000-0002-2919-3542
Ron Geller (iD) https://orcid.org/0000-0002-7612-4611

Additional files

Supplementary files

- Supplementary file 1. Primers for next-generation sequencing.
- Supplementary file 2. Next-generation sequencing statistics.
- Supplementary file 3. Mutations observed by Sanger sequencing.
- Supplementary file 4. Mutational fitness effects of the mutagenized viral populations.
- Supplementary file 5. Results of qPCR validation of MFE.
- Supplementary file 6. Data used for random forest model and parameter explanation.
- Supplementary file 7. Differential selection results.
- Supplementary file 8. PSSMsearch results.
- Transparent reporting form

Data availability

Sequencing data have been uploaded to SRA (Bioproject PRJNA643896, SRA SRP269871, Accession SRX8663374-SRX8663384). All data used in the paper are either included as supplemental data and/or can be found at https://github.com/RGellerLab/CVB3_Capsid_DMS (copy archived at https://archive.softwareheritage.org/swh:1:rev:29dd205182f0886dc5bad3e6b4ddd6e786c58a75/).

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Mattenberger F, Geller R	2020	Deep mutational scanning of the CVB3 capsid protein	https://www.ncbi.nlm.nih.gov/bioproject/643896	NCBI BioProject, PRJNA643896

References

Acevedo A, Brodsky L, Andino R. 2014. Mutational and fitness landscapes of an RNA virus revealed through population sequencing. *Nature* **505**:686–690. DOI: https://doi.org/10.1038/nature12861, PMID: 24284629

Ashenberg O, Padmakumar J, Doud MB, Bloom JD. 2017. Deep mutational scanning identifies sites in influenza nucleoprotein that affect viral inhibition by MxA. *PLOS Pathogens* **13**:e1006288. DOI: https://doi.org/10.1371/journal.ppat.1006288, PMID: 28346537

Belnap DM, McDermott BM, Filman DJ, Cheng N, Trus BL, Zuccola HJ, Racaniello VR, Hogle JM, Steven AC. 2000. Three-dimensional structure of poliovirus receptor bound to poliovirus. *PNAS* **97**:73–78. DOI: https://doi.org/10.1073/pnas.97.1.73, PMID: 10618373

Bloom JD. 2014. An experimentally determined evolutionary model dramatically improves phylogenetic fit. *Molecular Biology and Evolution* **31**:1956–1978. DOI: https://doi.org/10.1093/molbev/msu173, PMID: 24859245

Bloom JD. 2015. Software for the analysis and visualization of deep mutational scanning data. *BMC Bioinformatics* **16**:168. DOI: https://doi.org/10.1186/s12859-015-0590-4, PMID: 25990960

Bloom JD. 2017. Identification of positive selection in genes is greatly improved by using experimentally informed site-specific models. *Biology Direct* **12**:1. DOI: https://doi.org/10.1186/s13062-016-0172-z, PMID: 28095902

Bologna G, Yvon C, Duvaud S, Veuthey AL. 2004. N-Terminal myristoylation predictions by ensembles of neural networks. *Proteomics* **4**:1626–1632. DOI: https://doi.org/10.1002/pmic.200300783, PMID: 15174132

Bou JV, Geller R, Sanjuán R. 2019. Membrane-Associated Enteroviruses undergo intercellular transmission as pools of sibling viral genomes. *Cell Reports* **29**:714–723. DOI: https://doi.org/10.1016/j.celrep.2019.09.014, PMID: 31618638

Brass AL, Huang IC, Benita Y, John SP, Krishnan MN, Feeley EM, Ryan BJ, Weyer JL, van der Weyden L, Fikrig E, Adams DJ, Xavier RJ, Farzan M, Elledge SJ. 2009. The IFITM proteins mediate cellular resistance to influenza A H1N1 virus, west nile virus, and dengue virus. *Cell* **139**:1243–1254. DOI: https://doi.org/10.1016/j.cell.2009.12.017, PMID: 20064371

Callaway A, Giesman-Cookmeyer D, Gillock ET, Sit TL, Lommel SA. 2001. The multifunctional capsid proteins of plant RNA viruses. *Annual Review of Phytopathology* **39**:419. DOI: https://doi.org/10.1146/annurev.phyto.39.1.419, PMID: 11701872

Carrasco P, Daròs JA, Agudelo-Romero P, Elena SF. 2007. A real-time RT-PCR assay for quantifying the fitness of tobacco etch virus in competition experiments. *Journal of Virological Methods* **139**:181–188. DOI: https://doi.org/10.1016/j.jviromet.2006.09.020, PMID: 17092574

Carrillo-Tripp M, Shepherd CM, Borelli IA, Venkataraman S, Lander G, Natarajan P, Johnson JE, Brooks CL, Reddy VS. 2009. VIPERdb2: an enhanced and web API enabled relational database for structural virology. *Nucleic Acids Research* **37**:D436–D442. DOI: https://doi.org/10.1093/nar/gkn840, PMID: 18981051

Chan SL, Huppertz I, Yao C, Weng L, Moresco JJ, Yates JR, Ule J, Manley JL, Shi Y. 2014. CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3' processing. *Genes & Development* **28**:2370–2380. DOI: https://doi.org/10.1101/gad.250993.114, PMID: 25301780

Chen S, Zhou Y, Chen Y, Gu J. 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**:i884–i890. DOI: https://doi.org/10.1093/bioinformatics/bty560, PMID: 30423086

Cifuente JO, Moratorio G. 2019. Evolutionary and structural overview of human Picornavirus capsid antibody evasion. *Frontiers in Cellular and Infection Microbiology* **9**:1–11. DOI: https://doi.org/10.3389/fcimb.2019.00283, PMID: 31482072

Corbic Ramljak I, Stanger J, Real-Hohn A, Dreier D, Wimmer L, Redlberger-Fritz M, Fischl W, Klingel K, Mihovilovic MD, Blaas D, Kowalski H. 2018. Cellular N-myristoyltransferases play a crucial Picornavirus genus-specific role in viral assembly, virion maturation, and infectivity. *PLOS Pathogens* **14**:e1007203. DOI: https://doi.org/10.1371/journal.ppat.1007203, PMID: 30080883

Dey KK, Xie D, Stephens M. 2018. A new sequence logo plot to highlight enrichment and depletion. *BMC Bioinformatics* **19**:473. DOI: https://doi.org/10.1186/s12859-018-2489-3, PMID: 30526486

Dingens AS, Haddox HK, Overbaugh J, Bloom JD. 2017. Comprehensive mapping of HIV-1 escape from a broadly neutralizing antibody. *Cell Host & Microbe* **21**:777–787. DOI: https://doi.org/10.1016/j.chom.2017.05.003, PMID: 28579254

Doud M, Bloom J. 2016. Accurate measurement of the effects of all Amino-Acid mutations on influenza hemagglutinin. *Viruses* **8**:155. DOI: https://doi.org/10.3390/v8060155

Dragovich PS, Prins TJ, Zhou R, Webber SE, Marakovits JT, Fuhrman SA, Patick AK, Matthews DA, Lee CA, Ford CE, Burke BJ, Rejto PA, Hendrickson TF, Tuntland T, Brown EL, Meador JW, Ferre RA, Harr JE, Kosa MB, Worland ST. 1999. Structure-based design, synthesis, and biological evaluation of irreversible human rhinovirus 3C protease inhibitors. 4. incorporation of P1 . *Journal of Medicinal Chemistry* **42**:1213–1224. DOI: https://doi.org/10.1021/jm9805384, PMID: 10197965

Du Y, Wu NC, Jiang L, Zhang T, Gong D, Shu S, Wu T-T, Sun R. 2016. Annotating protein functional residues by coupling High-Throughput fitness profile and Homologous-Structure analysis. *mBio* **7**:01801-16. DOI: https://doi.org/10.1128/mBio.01801-16

Fernandez-Escamilla AM, Rousseau F, Schymkowitz J, Serrano L. 2004. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nature Biotechnology* **22**:1302–1306. DOI: https://doi.org/10.1038/nbt1012, PMID: 15361882

Fields BN, Knipe DM, Howley PM. 2013. *Fields Virology*. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins.

Geller R, Vignuzzi M, Andino R, Frydman J. 2007. Evolutionary constraints on chaperone-mediated folding provide an antiviral approach refractory to development of drug resistance. *Genes & Development* **21**:195–205. DOI: https://doi.org/10.1101/gad.1505307, PMID: 17234885

Gray VE, Hause RJ, Luebeck J, Shendure J, Fowler DM. 2018. Quantitative missense variant effect prediction using Large-Scale mutagenesis data. *Cell Systems* **6**:116–124. DOI: https://doi.org/10.1016/j.cels.2017.11.003, PMID: 29226803

Haddox HK, Dingens AS, Bloom JD. 2016. Experimental estimation of the effects of all Amino-Acid mutations to HIV's Envelope Protein on Viral Replication in Cell Culture. *PLOS Pathogens* **12**:e1006114. DOI: https://doi.org/10.1371/journal.ppat.1006114, PMID: 27959955

Haddox HK, Dingens AS, Hilton SK, Overbaugh J, Bloom JD. 2018. Mapping mutational effects along the evolutionary landscape of HIV envelope. *eLife* **7**:e34420. DOI: https://doi.org/10.7554/eLife.34420, PMID: 29590010

Harrison SC. 2013. Principles of Virus Structure. In: Knipe D. M, Howley P. M (Eds). *Fields Virology*. Wolters Kluwer Health/Lippincott Williams & Wilkins. p. 52–86.

Hartman EC, Jakobson CM, Favor AH, Lobba MJ, Álvarez-Benedicto E, Francis MB, Tullman-Ercek D. 2018. Quantitative characterization of all single amino acid variants of a viral capsid-based drug delivery vehicle. *Nature Communications* **9**:1385. DOI: https://doi.org/10.1038/s41467-018-03783-y, PMID: 29643335

Hecht M, Bromberg Y, Rost B. 2015. Better prediction of functional effects for sequence variants. *BMC Genomics* **16**:S1. DOI: https://doi.org/10.1186/1471-2164-16-S8-S1

Heise MT, Virgin HW. 2013. Pathogenesis of viral infection. In: Knipe D. M, Howley P. M (Eds). *Fields Virology*. Wolters Kluwer Health/Lippincott Williams & Wilkins. p. 254–285.

Helenius A. 2013. Virus Entry and Uncoating. In: Knipe D. M, Howley P. M (Eds). *Fields Virology*. Wolters Kluwer Health/Lippincott Williams & Wilkins. p. 87–104.

Hilton SK, Doud MB, Bloom JD. 2017. Phydms: software for phylogenetic analyses informed by deep mutational scanning. *PeerJ* **5**:e3657. DOI: https://doi.org/10.7717/peerj.3657, PMID: 28785526

Hom N, Gentles L, Bloom JD, Lee KK. 2019. Deep mutational scan of the highly conserved influenza A virus M1 matrix protein reveals substantial intrinsic mutational tolerance. *Journal of Virology* **93**:1–16. DOI: https://doi.org/10.1128/JVI.00161-19

Hunter E. 2013. Virus Assembly. In: Knipe D. M, Howley P. M (Eds). *Fields Virology*. Wolters Kluwer Health/Lippincott Williams & Wilkins. p. 127–152.

Jagdeo JM, Dufour A, Klein T, Solis N, Kleifeld O, Kizhakkedathu J, Luo H, Overall CM, Jan E. 2018. N-Terminomics TAILS identifies host cell substrates of poliovirus and coxsackievirus B3 3C proteinases that modulate virus infection. *Journal of Virology* **92**:e02211-17. DOI: https://doi.org/10.1128/JVI.02211-17, PMID: 29437971

Jiang P, Liu Y, Ma HC, Paul AV, Wimmer E. 2014. Picornavirus morphogenesis. *Microbiology and Molecular Biology Reviews* **78**:418–437. DOI: https://doi.org/10.1128/MMBR.00012-14, PMID: 25184560

Kennedy SR, Schmitt MW, Fox EJ, Kohrn BF, Salk JJ, Ahn EH, Prindle MJ, Kuong KJ, Shen JC, Risques RA, Loeb LA. 2014. Detecting ultralow-frequency mutations by duplex sequencing. *Nature Protocols* **9**:2586–2606. DOI: https://doi.org/10.1038/nprot.2014.170, PMID: 25299156

Kodigepalli KM, Bowers K, Sharp A, Nanjundan M. 2015. Roles and regulation of phospholipid scramblases. *FEBS Letters* **589**:3–14. DOI: https://doi.org/10.1016/j.febslet.2014.11.036

Kristensen T, Belsham GJ. 2019. Identification of a short, highly conserved, motif required for picornavirus capsid precursor processing at distal sites. *PLOS Pathogens* **15**:e1007509. DOI: https://doi.org/10.1371/journal.ppat.1007509

Krystkowiak I, Manguy J, Davey NE. 2018. PSSMSearch: a server for modeling, visualization, proteome-wide discovery and annotation of protein motif specificity determinants. *Nucleic Acids Research* **46**:W235–W241. DOI: https://doi.org/10.1093/nar/gky426

Laitinen OH, Svedin E, Kapell S, Nurminen A, Hytönen VP, Flodström-Tullberg M. 2016. Enteroviral proteases: structure, host interactions and pathogenicity. *Reviews in Medical Virology* **26**:251–267. DOI: https://doi.org/10.1002/rmv.1883

Lee JM, Eguia R, Zost SJ, Choudhary S, Wilson PC, Bedford T, Stevens-Ayers T, Boeckh M, Hurt AC, Lakdawala SS, Hensley SE, Bloom JD. 2019. Mapping person-to-person variation in viral mutations that escape polyclonal serum targeting influenza hemagglutinin. *eLife* **8**:e49324. DOI: https://doi.org/10.7554/eLife.49324, PMID: 31452511

Livesey BJ, Marsh JA. 2020. Using deep mutational scanning to benchmark variant effect predictors and identify disease mutations. *Molecular Systems Biology* **16**:1–12. DOI: https://doi.org/10.15252/msb.20199380

Macejak DG, Sarnow P. 1992. Association of heat shock protein 70 with enterovirus capsid precursor P1 in infected human cells. *Journal of Virology* **66**:1520–1527. DOI: https://doi.org/10.1128/JVI.66.3.1520-1527.1992

Mattenberger F. 2021. CVB3_Capsid_DMS. *Software Heritage.* swh:1:rev:29dd205182f0886dc5bad3e6b4ddd6e786c58a75.https://archive.softwareheritage.org/swh:1:dir:40e0b58a90057980be80c0bbf4e39a10450c31c7;origin=https://github.com/RGellerLab/CVB3_Capsid_DMS;visit=swh:1:snp:d162eac2e7160d4ff881f6fc792188a5875294a9;anchor=swh:1:rev:29dd205182f0886dc5bad3e6b4ddd6e786c58a75/

Moratorio G, Henningsson R, Barbezange C, Carrau L, Bordería AV, Blanc H, Beaucourt S, Poirier EZ, Vallet T, Boussier J, Mounce BC, Fontes M, Vignuzzi M. 2017. Attenuation of RNA viruses by redirecting their evolution in sequence space. *Nature Microbiology* **2**:17088. DOI: https://doi.org/10.1038/nmicrobiol.2017.88

Muckelbauer JK, Kremer M, Minor I, Diana G, Dutko FJ, Groarke J, Pevear DC, Rossmann MG. 1995. The structure of coxsackievirus B3 at 3.5 å resolution. *Structure* **3**:653–667. DOI: https://doi.org/10.1016/S0969-2126(01)00201-5

Ogden PJ, Kelsic ED, Sinai S, Church GM. 2019. Comprehensive AAV capsid fitness landscape reveals a viral gene and enables machine-guided design. *Science* **366**:1139–1143. DOI: https://doi.org/10.1126/science.aaw2900

Perlmutter JD, Hagan MF. 2015. Mechanisms of virus assembly. *Annual Review of Physical Chemistry* **66**:217–239. DOI: https://doi.org/10.1146/annurev-physchem-040214-121637, PMID: 25532951

Pickett BE, Sadat EL, Zhang Y, Noronha JM, Squires RB, Hunt V, Liu M, Kumar S, Zaremba S, Gu Z, Zhou L, Larson CN, Dietrich J, Klem EB, Scheuermann RH. 2012. ViPR: an open bioinformatics database and analysis resource for virology research. *Nucleic Acids Research* **40**:D593–D598. DOI: https://doi.org/10.1093/nar/gkr859

Qing J, Wang Y, Sun Y, Huang J, Yan W, Wang J, Su D, Ni C, Li J, Rao Z, Liu L, Lou Z. 2014. Cyclophilin A associates with enterovirus-71 virus capsid and plays an essential role in viral infection as an uncoating regulator. *PLOS Pathogens* **10**:e1004422. DOI: https://doi.org/10.1371/journal.ppat.1004422, PMID: 25275585

Racaniello VR. 2013. Picornaviridae: The Viruses and Their Replication. In: Knipe M. D, Howley M. P (Eds). *Fields Virology*. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins, c2013. p. 453–489.

Reeb J, Wirth T, Rost B. 2020. Variant effect predictions capture some aspects of deep mutational scanning experiments. *BMC Bioinformatics* **21**:107. DOI: https://doi.org/10.1186/s12859-020-3439-4, PMID: 32183714

Rossmann MG, He Y, Kuhn RJ. 2002. Picornavirus-receptor interactions. *Trends in Microbiology* **10**:324–331. DOI: https://doi.org/10.1016/S0966-842X(02)02383-1, PMID: 12110211

Schmitt MW, Kennedy SR, Salk JJ, Fox EJ, Hiatt JB, Loeb LA. 2012. Detection of ultra-rare mutations by next-generation sequencing. *PNAS* **109**:14508–14513. DOI: https://doi.org/10.1073/pnas.1208715109

Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. 2005. The FoldX web server: an online force field. *Nucleic Acids Research* **33**:W382–W388. DOI: https://doi.org/10.1093/nar/gki387

Shami Shah A, Batrouni AG, Kim D, Punyala A, Cao W, Han C, Goldberg ML, Smolka MB, Baskin JM. 2019. PLEKHA4/kramer attenuates dishevelled ubiquitination to modulate wnt and planar cell polarity signaling. *Cell Reports* **27**:2157–2170. DOI: https://doi.org/10.1016/j.celrep.2019.04.060, PMID: 31091453

Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KHD, Dingens AS, Navarro MJ, Bowen JE, Tortorici MA, Walls AC, King NP, Veesler D, Bloom JD. 2020. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell* **182**:1295–1310. DOI: https://doi.org/10.1016/j.cell.2020.08.012, PMID: 32841599

Sun D, Chen S, Cheng A, Wang M. 2016. Roles of the picornaviral 3C proteinase in the viral life cycle and host cells. *Viruses* **8**:82. DOI: https://doi.org/10.3390/v8030082, PMID: 26999188

Thibaut HJ, van der Linden L, Jiang P, Thys B, Canela MD, Aguado L, Rombaut B, Wimmer E, Paul A, Pérez-Pérez MJ, van Kuppeveld FJ, Neyts J. 2014. Binding of glutathione to Enterovirus capsids is essential for virion morphogenesis. *PLOS Pathogens* **10**:e1004039. DOI: https://doi.org/10.1371/journal.ppat.1004039, PMID: 24722756

Thyagarajan B, Bloom JD. 2014. The inherent mutational tolerance and antigenic evolvability of influenza hemagglutinin. *eLife* **3**:e03300. DOI: https://doi.org/10.7554/eLife.03300

Verbist BMP, Thys K, Reumers J, Wetzels Y, Van der Borght K, Talloen W, Aerssens J, Clement L, Thas O. 2015. VirVarSeq: a low-frequency virus variant detection pipeline for Illumina sequencing using adaptive base-calling accuracy filtering. *Bioinformatics* **31**:94–101. DOI: https://doi.org/10.1093/bioinformatics/btu587

Wala J, Zhang C-Z, Meyerson M, Beroukhim R. 2016. VariantBam: filtering and profiling of next-generational sequencing data using region-specific rules. *Bioinformatics* **32**:2029–2031. DOI: https://doi.org/10.1093/bioinformatics/btw111

Wu NC, Olson CA, Du Y, Le S, Tran K, Remenyi R, Gong D, Al-Mawsawi LQ, Qi H, Wu TT, Sun R. 2015. Functional constraint profiling of a viral protein reveals discordance of evolutionary conservation and functionality. *PLOS Genetics* **11**:e1005310. DOI: https://doi.org/10.1371/journal.pgen.1005310, PMID: 26132554

Xing L, Tjarnlund K, Lindqvist B, Kaplan GG, Feigelstock D, Cheng RH, Casasnovas JM. 2000. Distinct cellular receptor interactions in poliovirus and rhinoviruses. *The EMBO Journal* **19**:1207–1216. DOI: https://doi.org/10.1093/emboj/19.6.1207

Xu L, Zheng Q, Li S, He M, Wu Y, Li Y, Zhu R, Yu H, Hong Q, Jiang J, Li Z, Li S, Zhao H, Yang L, Hou W, Wang W, Ye X, Zhang J, Baker TS, Cheng T, et al. 2017. Atomic structures of coxsackievirus A6 and its complex with a neutralizing antibody. *Nature Communications* **8**:505. DOI: https://doi.org/10.1038/s41467-017-00477-9, PMID: 28894095

Yoder JD, Cifuente JO, Pan J, Bergelson JM, Hafenstein S. 2012. The crystal structure of a coxsackievirus B3-RD variant and a refined 9-angstrom cryo-electron microscopy reconstruction of the virus complexed with decay-accelerating factor (DAF) provide a new footprint of DAF on the virus surface. *Journal of Virology* **86**:12571–12581. DOI: https://doi.org/10.1128/JVI.01592-12, PMID: 22973031

Ypma-Wong MF, Dewalt PG, Johnson VH, Lamb JG, Semler BL. 1988. Protein 3CD is the major poliovirus proteinase responsible for cleavage of the p1 capsid precursor. *Virology* **166**:265–270. DOI: https://doi.org/10.1016/0042-6822(88)90172-9

Yun T, Park A, Hill TE, Pernet O, Beaty SM, Juelich TL, Smith JK, Zhang L, Wang YE, Vigant F, Gao J, Wu P, Lee B, Freiberg AN. 2015. Efficient reverse genetics reveals genetic determinants of budding and fusogenic differences between nipah and Hendra viruses and enables real-time monitoring of viral spread in small animal models of Henipavirus infection. *Journal of Virology* **89**:1242–1253. DOI: https://doi.org/10.1128/JVI.02583-14, PMID: 25392218