

Comprehensive profiling of polyclonal sera targeting a non-enveloped viral capsid

Beatriz Álvarez-Rodríguez¹, Javier Buceta^{1,*}, Ron Geller^{1, 2, *}

¹Institute for Integrative Systems Biology (I2SysBio), Universitat de Valencia-CSIC, Valencia, 46980, Spain.

²Lead contact

*Correspondence: javier.buceta@csic.es, ron.geller@csic.es

Abstract

Despite their fundamental role in resolving viral infections, our understanding of how polyclonal neutralizing antibody responses target non-enveloped viruses remains limited. To define these responses, we obtained the full antigenic profile of multiple human and mouse polyclonal sera targeting the capsid of a prototypical picornavirus. Our results uncover significant variation in the breadth and strength of neutralization sites targeted by individual human polyclonal responses, which contrasted with homogenous responses observed in experimentally infected mice. We further use these comprehensive antigenic profiles to define key structural and evolutionary parameters that are predictive of escape, assess epitope dominance at the population level, and reveal a need for at least two mutations to achieve significant escape from multiple sera. Overall, our data provide the first comprehensive analysis of how polyclonal sera target a non-enveloped viral capsid and help define both immune dominance and escape at the population level.

Keywords

Neutralizing antibodies; Viral capsids; Picornaviruses; Coxsackievirus B3; Deep mutational scanning; Machine learning;

Introduction

Neutralizing antibodies are key to resolving viral infections and can provide long-term protection against reinfection. These antibodies mostly target viral proteins involved in cell entry, namely membrane proteins in enveloped viruses and capsid proteins in non-enveloped viruses (Corti and Lanzavecchia 2013). Consequently, vaccination strategies frequently aim to elicit polyclonal neutralizing antibody responses utilizing these same viral proteins (D. B. Huang, Wu, and Tying 2004). In turn, viruses must overcome these immune responses for their successful spread in previously infected or immunized populations, establishing a continuous evolutionary arms race to alter immunodominant epitopes and refine antibody responses (Burton 2002).

Recent high-throughput approaches have provided new insights into how viral membrane proteins are targeted by polyclonal antibody responses for several enveloped viruses, including human immunodeficiency virus, influenza A virus, and SARS-CoV-2 (Liu et al. 2022; Greaney, Loes, et al. 2021; Greaney, Starr, Barnes, et al. 2021; Greaney, Starr, Gilchuk, et al. 2021; Dingens et al. 2021; Lee et al. 2019). These studies have revealed the breadth and relative strength of neutralization sites induced by both natural infection as well as vaccination and helped define mutations conferring escape from neutralization. However, our knowledge of how non-enveloped viruses are targeted by polyclonal sera remains limited, despite the fact that they constitute >40% of mammalian viruses (Valero-Rello and Sanjuán 2022). Moreover, fundamental differences between capsid proteins and viral envelope proteins could preclude the extrapolation of results from enveloped to non-enveloped viruses. In particular, carbohydrate modifications that alter the sensitivity of viral membrane proteins to antibody neutralization (Li et al. 2021) are absent in viral capsids. Additionally, non-enveloped viral capsids encode multiple functions not found in viral membrane proteins, including the information for assembly, genome packaging, and genome release, which could significantly constrain their ability to tolerate mutations conferring immune escape. Obtaining a deep understanding of how viral capsid proteins are targeted by, and escape, polyclonal antibody responses is therefore of key importance for understanding host-pathogen interactions and viral evolution of this large fraction of viruses.

Picornaviruses were the first human viruses to be structurally defined at the atomic level (Rossmann et al. 1985), revealing an icosahedral capsid whose symmetry is conserved across all capsids of non-enveloped viruses in vertebrates. The picornavirus capsid is comprised of 60 copies of four structural proteins, three of which are surface-exposed (VP1, VP2, and VP3) and one that lines the internal capsid surface (VP4) (Racaniello 2013). A depression in the capsid, termed the canyon, frequently harbors residues involved in receptor binding (Levy et al. 2014). Antibody neutralization in picornaviruses has been mapped to four surface-exposed structural regions using escape from monoclonal antibodies (mAbs) and structural studies: the canyon northern rim (five-fold axis), inner surface (canyon floor), and southern rim (canyon outer surface) as well as the two and three-fold plateau (see Figure 1E) (K. Y. A. Huang 2021; Martín et al. 2003; Mateu 1995; Sherry et al. 1986). The mechanisms by which mAbs neutralize picornaviruses have also been extensively studied, and include impeding receptor binding, premature induction of genome release, and virion stabilization (K. Y. A. Huang 2021; Cifuentes and Moratorio 2019). The large body of knowledge of how picornaviruses are targeted by mAbs combined

with the fact that humoral responses are essential for resolving picornavirus infections (Dotzauer and Kraemer 2012) make picornaviruses excellent models for studying antibody-capsid interactions.

In this work, we comprehensively map the antigenic profiles of multiple human and mouse neutralizing sera to a model human picornavirus, coxsackievirus B3 (CVB3). Our results reveal a surprisingly conserved antigenic profile in mice that focuses on a single common antigenic site. In contrast, we find human polyclonal responses to be diverse, targeting anywhere from a single dominant site strongly to several sites across the capsid with reduced potency. Capitalizing on the large dataset of antibody-escape mutations defined in our study, we show that a few key structural and evolutionary parameters are sufficient to predict escape with relatively high accuracy. Finally, we define the immune dominance of the identified escape sites at the population level and demonstrate that a combination of two mutations can confer significant evasion from neutralization by multiple human sera. Overall, our results provide a comprehensive understanding of how polyclonal neutralizing antibody responses target a prototypical human picornavirus and reveal epitope dominance at the population level.

Results

An experimental workflow defines mutations conferring escape from antibody neutralization across the complete CVB3 capsid

To define how picornaviruses are targeted by polyclonal antibody responses, we modified an experimental workflow that provides a quantitative evaluation of how mutations in a viral population confer escape from antibody neutralization, termed deep mutational antigenic profiling (Doud, Hensley, and Bloom 2017) (Figure 1A). Briefly, two previously characterized CVB3 populations derived from plasmid libraries encoding 96% of all possible single amino acid mutations across the capsid region were employed (Mattenberger et al. 2021). Following growth in cells, non-viable and highly deleterious mutations were purged from these viral populations, resulting in 86% of all possible single amino acid mutations being represented (~13,900 mutations; 78% and 68.6% of all possible single AA mutations in the population 1 and 2, respectively (Mattenberger et al. 2021)). To define which mutations reduce sensitivity to antibody neutralization, these populations are treated with either a monoclonal antibody (mAb) or neutralizing sera so that infectivity is reduced by 90-99% relative to mock treatment, balancing sufficient selection to observe escape without losing mutations with intermediate phenotypes (Figure 1A). Mutants escaping neutralization are then enriched by infecting cells for a single cycle. Finally, the frequencies of capsid mutations in both the neutralized and mock-neutralized populations are obtained using a high-fidelity next-generation sequencing technique (Mattenberger et al. 2021). Mutations showing positive differential selection, i.e., those whose frequency increases versus the WT amino acid at a given position, indicate the contribution of the individual mutation to escaping antibody neutralization. Similarly, the sum of all positive differential selection values at a given residue reflects its overall contribution to evading antibody neutralization (termed site positive differential selection).

To validate our ability to identify mutations conferring escape from antibody neutralization, we first treated the CVB3 populations using a commercially available mouse mAb (Figure 1B and Table S1). Experiments were performed twice with each of the two viral populations. Site positive differential selection values showed good correlations between replicates of the same viral populations (Pearson's r of 0.72-0.81; Figure S1) but were somewhat lower between biological replicates performed with different virus libraries (Pearson's r of 0.67-0.69; Figure S1), as expected due to the different mutations present in each population. The main sites of escape in the mAb profile were comprised of several residues across the VP2 EF loop and VP3 AB helix/AB loop (Figure 1F, and Table S2). These two regions on the outer surface of the canyon (Figure 1E) have been described as antigenic sites in different picornaviruses and are known as the “puff” and the “knob” (Muckelbauer et al. 1995). Mapping of the positive differential selection values to the structure of the CVB3 capsid locates these sites in a single region of escape on the outer surface of the canyon, as expected for a mAb (Figure 1F; interactive visualization available at <https://github.com/RGellerLab/CVB3-Antigenic-Profiling>).

To validate our mutational antigenic profiling approach, we compared the ability of the mAb to neutralize one mutant predicted to confer strong escape from neutralization

(K227S) and two mutations predicted to confer low (K650Y) or no (K723E) escape. As expected, the K227S mutant showed significant escape, requiring >50-fold higher serum concentrations to neutralize 50% of the virus (IC₅₀) and exceeding the assay's limit of detection (LOD: the ratio of the lowest sera dilution employed, 1:100, divided by the IC₅₀ of the WT virus; Figure 1D). On the other hand, K650Y and K723E did not alter neutralization significantly (IC₅₀ fold change of 1.9 and 1.3, respectively). Together, these results validate our ability to map escape from antibody neutralization at both the site and mutation level.

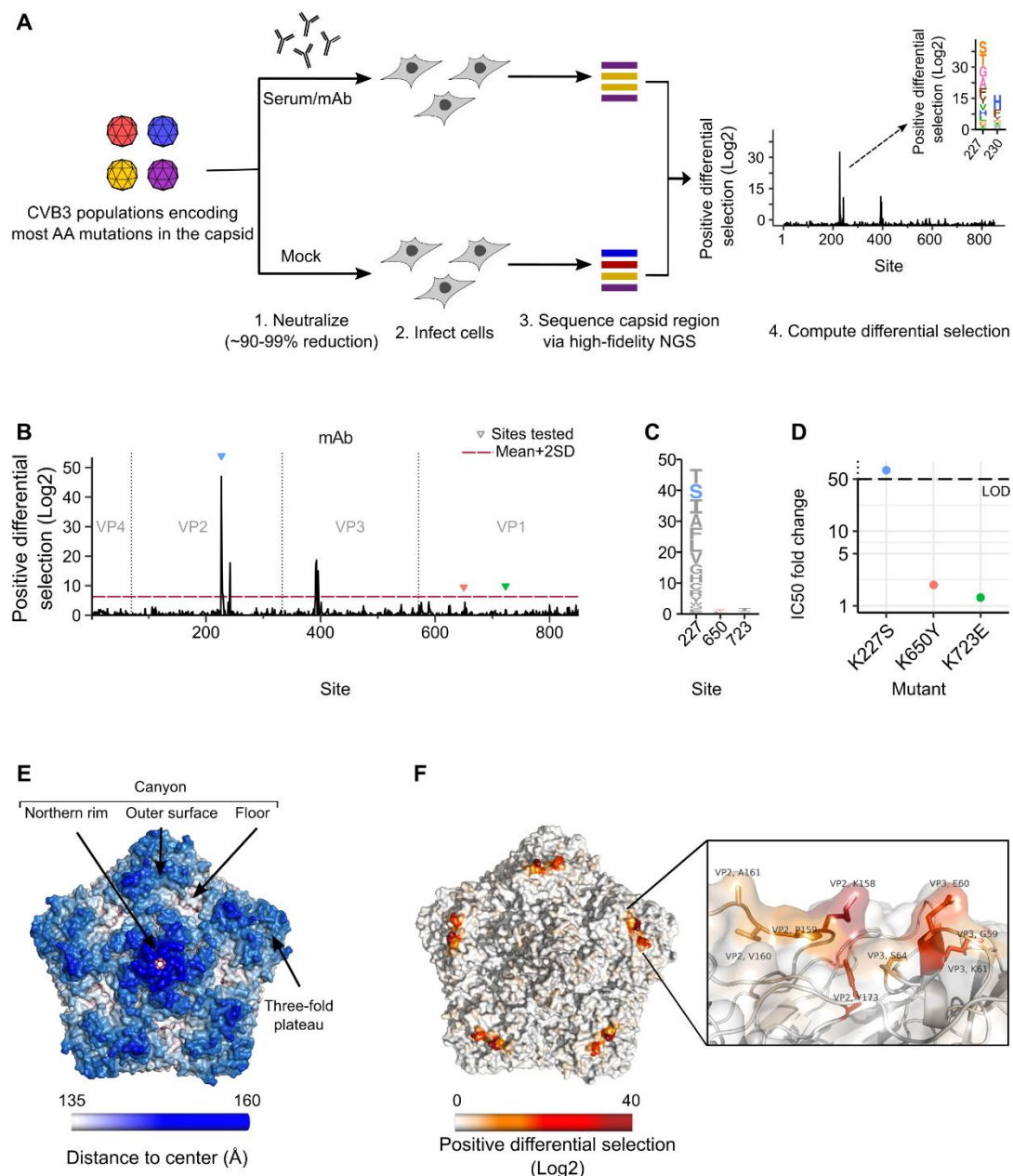


Figure 1. Mutational antigenic profiling workflow for CVB3. (A) Overview of the experimental workflow. CVB3 populations harboring high diversity in the capsid region are neutralized or mock treated and surviving viruses amplified by infection of cells. Mutation frequencies across the capsid are then obtained via high-fidelity deep sequencing. Mutations showing positive differential selection, i.e. those whose frequency relative to that of the WT has increased following neutralization versus mock-neutralized controls, define mutations conferring escape from antibody neutralization. **(B)** Mutational

antigenic profile of a neutralizing mAb. Triangles indicate sites that were experimentally validated in (D) and the dashed red line represents the mean+2SD of all mutations showing positive differential selection. **(C)** Logo plot representation of sites selected for validation in (D). The height of the letter is proportional to the positive differential selection value. **(D)** The IC₅₀ fold change for the mutant versus WT virus. The dashed line indicates the upper limit of detection (LOD) of the assay calculated as indicated in the text. **(E)** CVB3 capsid pentamer structure (PDB ID: 4GB3) with residues colored by distance to the center of the capsid. The four main antibody binding regions are indicated. **(F)** The CVB3 capsid pentamer structure colored by the positive differential selection values of the mAb antigenic profile, with a zoomed view in the right panel. See Figure S1 for individual replicates and the correlation between replicates, Table S1 for surviving fraction of the viral populations after neutralization and Table S2 for structural location of the main escape sites. See capsid_roadmaps for capsid roadmaps labeled by residue and dms_view for interactive visualization at <https://github.com/RGellerLab/CVB3-Antigenic-Profiling>.

Mouse polyclonal antibody responses are uniform and discrete

Laboratory mice are the most common model for investigating CVB3 in vivo (Fairweather and Rose 2007; Robinson, Wang, and Pfeiffer 2017). To understand the humoral immune response in experimentally infected mice, we challenged three mice with CVB3. The sera of all mice was neutralizing at 3 weeks post-challenge, with reciprocal IC₅₀ values of 1,863, 322, and 572 for samples m1, m2, and m3, respectively (see neutralization_ic50s at <https://github.com/RGellerLab/CVB3-Antigenic-Profiling>). We next evaluated the antigenic profile of these sera. Unexpectedly, all three sera showed a similar antigenic profile, suggesting that the response of experimentally immunized mice is relatively invariable (Figure 2A and Figure S1). The strongest escape mutations in these antigenic profiles were located in several residues across the VP1 protein, focused within a single structural area of escape in the northern rim of the canyon, involving the EF and BC loops of VP1 and residue Q264 in the C terminal loop of VP1 (Figure 2D and Table S2). Residues within these loops have been previously defined to interact with mouse neutralizing mAbs in several enteroviruses (Wang et al. 2020; Xu et al. 2017; Yin et al. 2019), suggesting a conserved response in mice.

We next evaluated whether mutations in two of the main sites of escape (K650Y and K723E, Figure 2B) affected neutralization by the mouse sera. Indeed, both mutations resulted in a nearly complete loss of neutralization at the lowest serum dilution tested (1:100), precluding the calculation of their IC₅₀ (Figure 2C). To obtain a relative quantification of how these mutations altered neutralization, we defined the IC₅₀ fold-change for each serum as greater than the limit of detection of our assays as for the mAb. For the most neutralizing mouse sera (serum m1), a >10-fold reduction in sensitivity to neutralization was observed, although the actual value could be significantly higher (Figure 2C). Overall, these results reveal the immune response of experimentally infected mice to be uniform and targeted to a delimited region of the capsid.

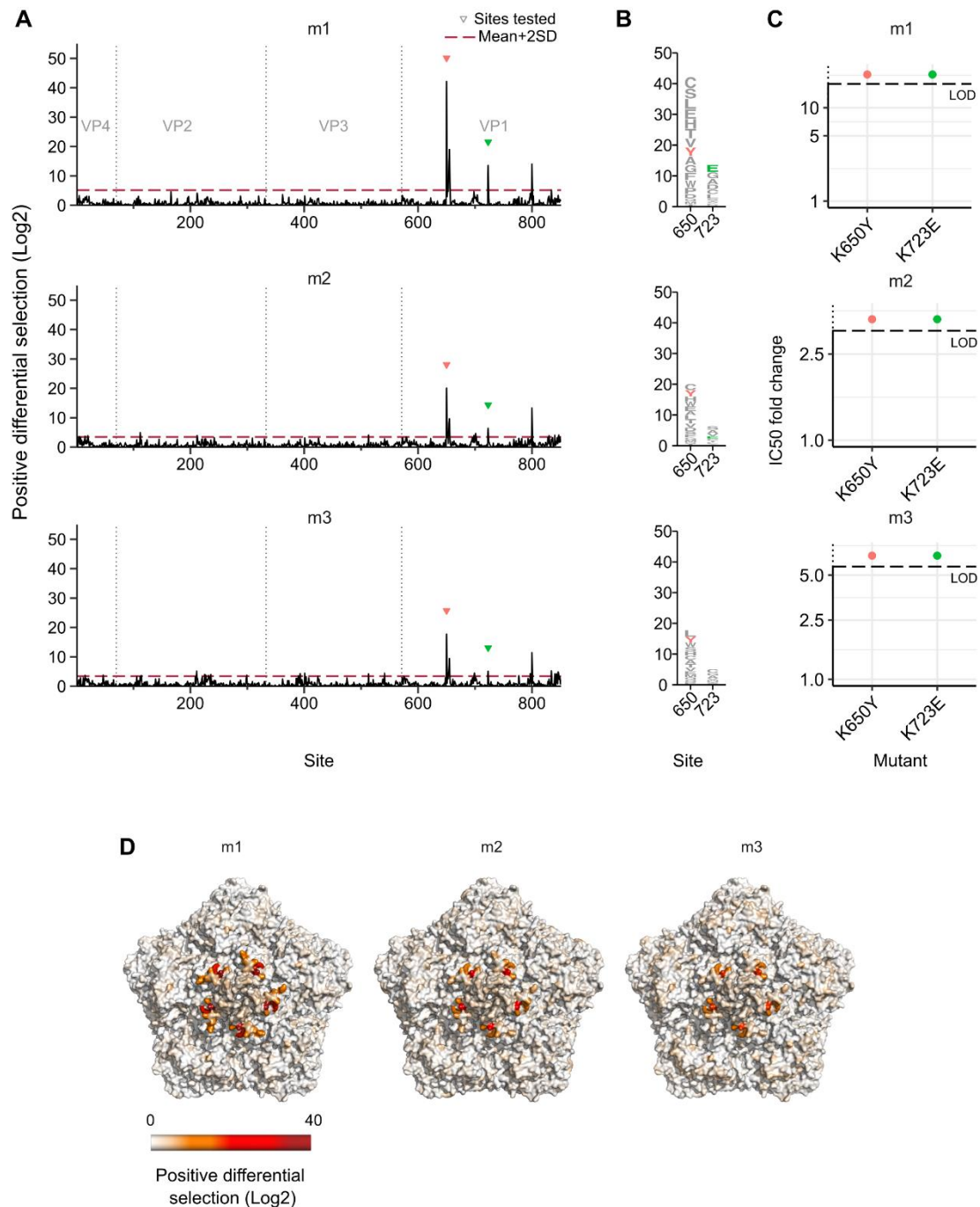


Figure 2. Antigenic profiles of sera from CVB3 infected mice. (A) Positive differential selection profiles of sera from mice infected with CVB3. Triangles indicate sites validated in (C). **(B)** Logo plot representation of sites of escape selected for validation, with mutations evaluated in (C) colored. **(C)** IC50 fold change for the mutant virus versus the WT virus. **(D)** The CVB3 capsid pentamer structure colored by positive differential selection values of the mouse antigenic profiles. See Figure S1 for individual replicates and the correlation between replicates, Table S1 for surviving fraction of the viral populations after neutralization and Table S2 for structural location of the main escape sites. See capsid_roadmaps for capsid roadmaps labeled by residue and dms_view for interactive visualization at <https://github.com/RGellerLab/CVB3-Antigenic-Profiling>.

Antigenic profiles of human sera are highly variable

Enteroviruses, such as CVB3, cause frequent infections in humans (Brouwer et al. 2021). The humoral response plays a key role in resolving these infections and confers protection from reinfection (Rachlin et al. 2022; Cifuentes and Moratorio 2019; Dotzauer and Kraemer 2012). To define how polyclonal human sera neutralizes the CVB3 capsid following natural infection, we first evaluated the ability of 140 sera from healthy adult donors to neutralize the CVB3 Nancy lab strain (see biobank_results at <https://github.com/RGellerLab/CVB3-Antigenic-Profiling>). Overall, 60% of sera was neutralizing (defined as a reciprocal IC₅₀ > 40; n=85/140, Figure S2A), with higher prevalence observed in female donors (female = 52/74 versus male = 33/66; p-value = 0.01605 by Fisher's test; Figure S2B). From these, the eight most neutralizing sera were subjected to mutational antigenic profiling. In contrast to mouse sera, human sera had heterogeneous antigenic profiles, varying in both the number and location of capsid sites conferring escape (Figures 3 and 4). Overall, two types of profiles were observed: narrow profiles, where escape sites were predominantly located in one or two regions of the capsid (Figure 3), and broad profiles, where multiple sites conferred escape to a similar extent across different regions of the capsid (Figure 4).

The narrow profile was observed in four of the eight human sera samples evaluated (Figure 3 and Figure S3). Escape mutations in these sera clustered into one or two of the four structural regions known to harbor antibody binding sites (K.-Y. A. Huang 2021) (Figure 3D). The most neutralizing serum (h8; reciprocal IC₅₀ = 9,186) showed escape in a single region of the VP2 EF loop (the “puff” region) in the outer surface of the canyon (Table S3) with multiple sites surrounding the main K227 escape residue involved (Figures 3A and 3D). The strongest escape mutation, K227S, reduced neutralization by ~19-fold (Figure 3B and 3C). Interestingly, the main site of escape in this serum (K227) is the same as that conferring the strongest escape from the mAb yet the antigenic contexts involve different additional residues (see Figures 1B and 3A). As for serum h8, serum h112 (reciprocal IC₅₀ = 2,359) presented a single prominent escape site in the three-fold plateau (site 473, VP3 EF loop, Figure 3A, 3D and Figure S3). As anticipated, mutation of this site (A473L) also resulted in strong escape (~19-fold change in the IC₅₀).

Unlike sera h8 and h112, the antigenic profile of serum h4 (reciprocal IC₅₀ = 3,348) presented escape mutations across two capsid regions: the canyon inner surface (VP1 CD loop and VP3 CD and C-terminal loops) and the northern rim (VP1 DE, EF, and HI loops; Figure 3A, 3D and Figure S3). These sites are characteristic of human and murine antibodies that bind to the inner surface of the canyon in other enteroviruses (Vogt et al. 2020; He et al. 2021). Mutations in these three dominant sites resulted in IC₅₀ fold-changes of 8.99 (A512L), 3.69 (K827V), and 2.91 (S564E, Figure 3C), in agreement with the antigenic profile. Finally, escape mutations in sample h87 (reciprocal IC₅₀ = 3,108) largely mapped to the interface between the inner surface (G2H loop of VP3) and the outer surface of the canyon (C-terminal loop of VP1 and EF loop of VP2; Figure 3A, 3D and Figure S3). In agreement with the antigenic profile, more modest escape was observed for single mutations in each region, with K222L (EF loop of VP2) and S513P (G2H loop of VP3) resulting in a fold change in the IC₅₀ of 3.42 and 4.71, respectively (Figure 3C). Overall, the neutralization profiles defined above showed prominent sites of escape in one or a few contiguous regions of the

capsid, with individual escape mutations resulting in reduced neutralization of 3 to 20-fold change.

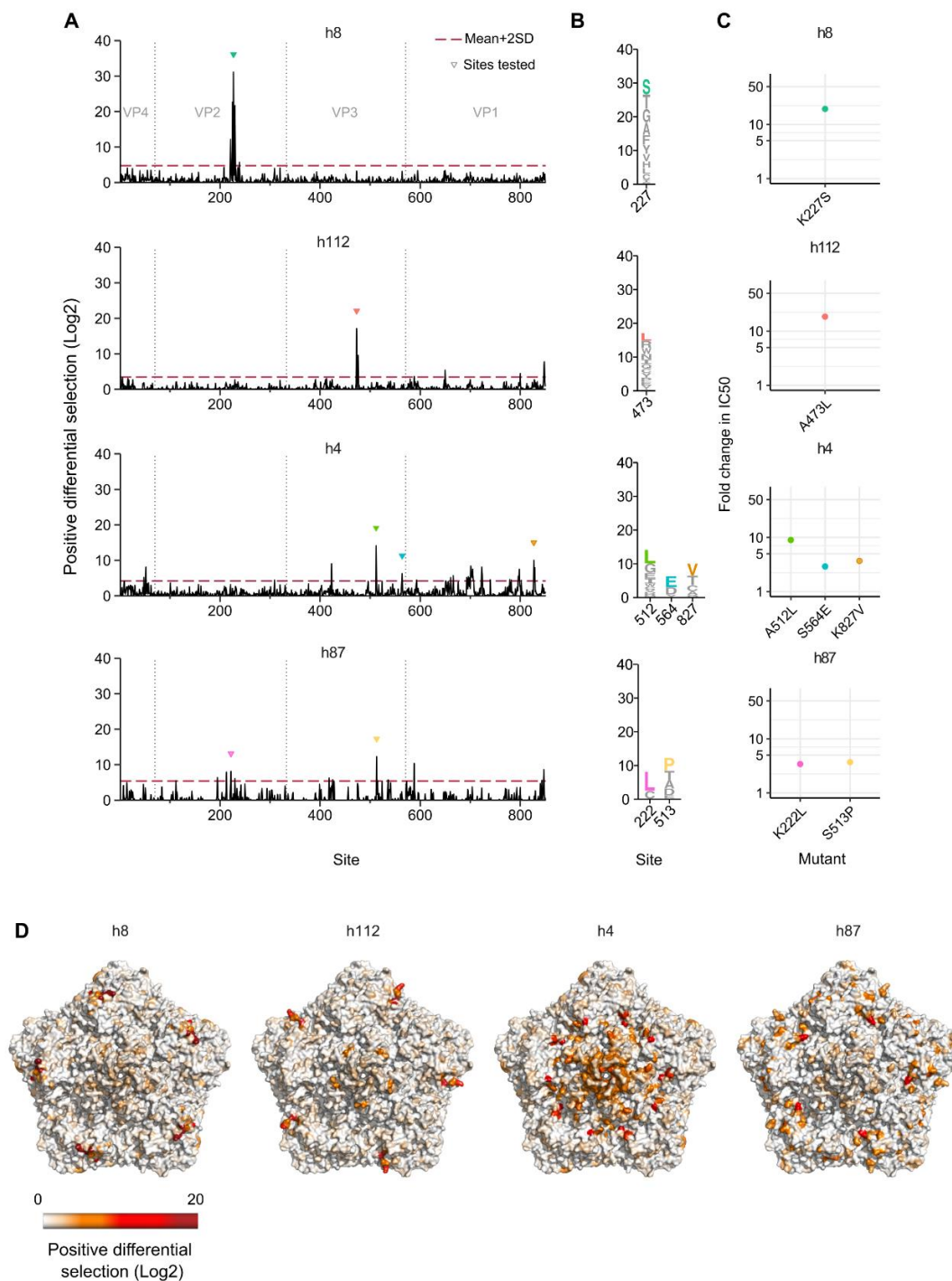


Figure 3. Narrow antigenic profiles observed in four human sera. (A) The antigenic profile of four human sera showing narrow escape profiles. Triangles indicate sites validated in (C). (B) Logo plot representation of selected sites of escape evaluated in (C). (C) The fold change in IC50 for the indicated mutants versus the WT virus. (D) The CVB3 capsid pentamer structure colored by positive differential selection values for each serum. See Figure S2 for the CVB3 neutralization titers in the evaluated donor sera, Figure S3 for individual replicates and the correlation between replicates, Table S1 for the surviving fraction of the viral populations after neutralization and Table S3 for the structural location of the main escape sites. See capsid_roadmaps for capsid roadmaps labeled by residue and dms_view for interactive visualization at <https://github.com/RGellerLab/CVB3-Antigenic-Profiling>.

The remaining four human sera presented a distinct antigenic profile, with escape sites distributing across multiple antibody neutralization regions with weaker overall differential selection (Figure 4 and Figure S4; note the change in scale versus Figure 3). In particular, sera h39 and h46 had similar neutralization capacity (reciprocal IC₅₀ = 2,520 and 2,212, respectively) and similar mutational antigenic profiles, with the principal escape sites in the northern rim of the canyon. Indeed, escape from both sera was conferred by the same Q702G mutation within the shared region of VP1 (IC₅₀ fold-change of 4.05 and 3.41, for serum h39 and h46, respectively; Figure 4C). However, less dominant sites varied between samples h39 and h46, targeting to different regions of the outer surface of the canyon and the three-fold plateau (Figure 4A,D and Supplementary Table S3). Experimental evaluation of prominent escape mutations was in agreement with the antigenic profiles, with mutations in VP1 conferring escape from serum h39 (K723E and K798H; fold change in IC₅₀ of 4.88 and 1.66, respectively) and mutations in VP2 and VP3 conferring escape from serum h46 (A221V and D535T; fold change in IC₅₀ of 1.6 and 5.56, respectively).

Escape mutations in serum h19 (reciprocal IC₅₀ = 2,052) were broadly distributed across the capsid. Some of the main escape sites mapped to the VP2 EF loop and the VP1 C terminal loop in the canyon outer surface, which are secondary structure elements also involved in the escape profile of serum h75 (reciprocal IC₅₀ = 1,414). In agreement with the profile of serum h19, mutation K227S (VP2 EF loop) and K723E (northern rim of the canyon) conferred fold changes in the IC₅₀ of 4.4 and 5.19, respectively (Figure 4C). The h75 profile presented additional sites of escape in the inner surface of the canyon (G2H loop of VP3) and the northern rim (DE loop of VP1). As expected, mutation A512L (inner surface of the canyon) and K829V (outer surface of the canyon) conferred escape from neutralization (IC₅₀ fold-change of 7.41 and 2.78, respectively; Figure 4C). Overall, escape mutations in broad antigenic profiles were distributed across the capsid and conferred smaller changes in escape from neutralization (2 to 8-fold change in the IC₅₀) compared to narrow antigenic profiles.

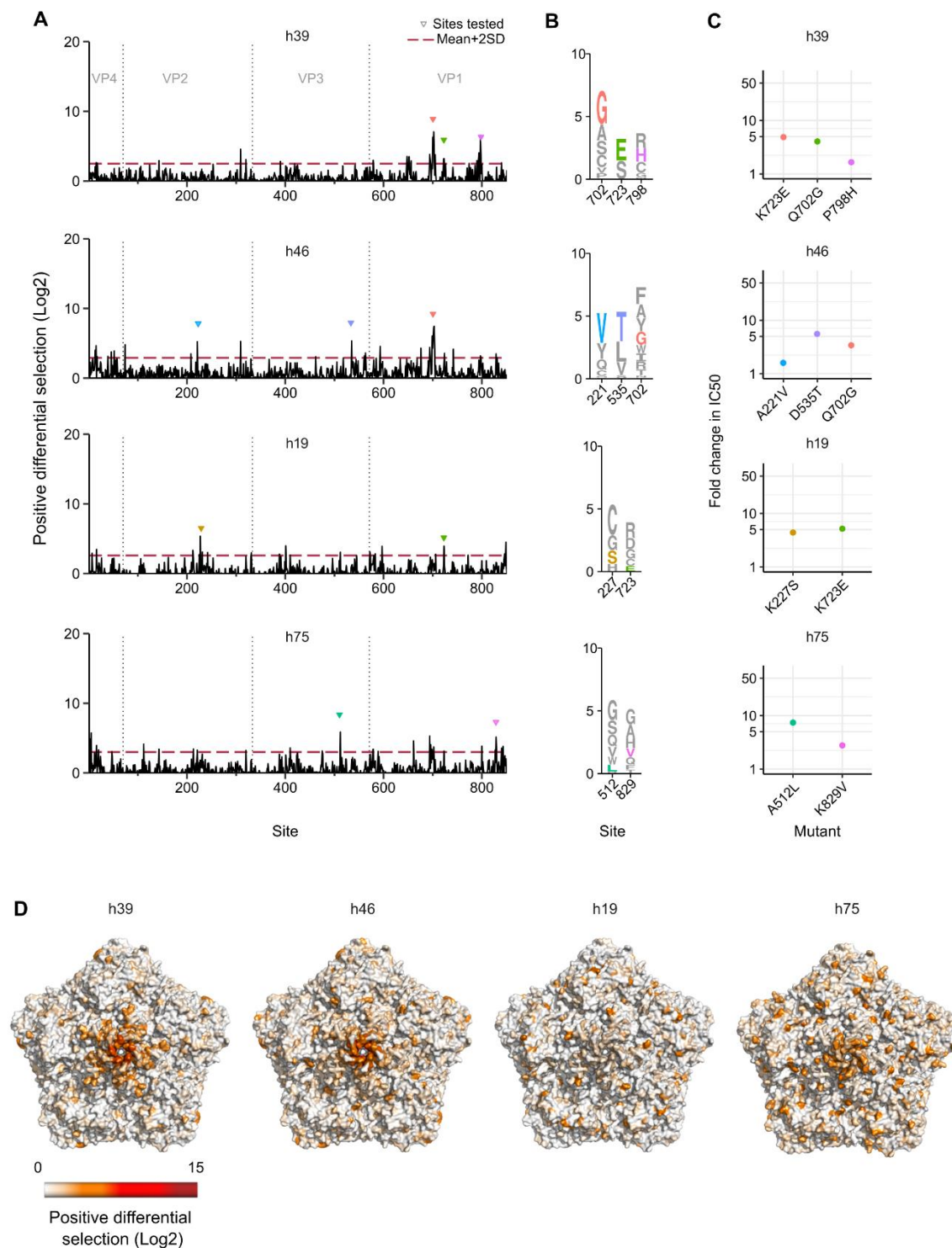


Figure 4. Broad antigenic profiles of four human sera. (A) The antigenic profile of four human sera showing broad antigenic profiles. Triangles indicate sites validated in (C). **(B)** Logo plot representation of escape sites selected for validation in (C). **(C)** The fold-change in the IC50 for the indicated mutations versus the WT virus. **(D)** The CVB3 capsid pentamer structure colored by positive differential selection values of the indicated serum. See Figure S4 for individual replicates and the correlation between replicates, Table S1 for the surviving fraction of the viral populations after neutralization and Table S2 for the structural location of the main escape sites. See capsid_roadmaps for capsid roadmaps labeled by residue and dms_view for interactive visualization at <https://github.com/RGellerLab/CVB3-Antigenic-Profiling>.

Escape mutations can be predicted based on structural and evolutionary parameters

Having obtained a comprehensive dataset of mutations conferring escape from eight different human sera, we next examined whether these mutations can be predicted based on different attributes using a machine learning approach. To this end, we compiled for all possible mutations in surface-exposed capsid residues ($n=5,776$) a dataset comprising structural, physicochemical, and evolutionary parameters. In addition, we predicted the effects of all such mutations on protein stability, aggregation propensity, disorder, and fitness (19 numerical and categorical features in total; see analysis_sites_muts at <https://github.com/RGellerLab/CVB3-Antigenic-Profiling>). This data was combined with a Boolean feature indicating whether the mutation led to strong escape in the antigenic profiling experiments (“escape”) or not (“no escape”). Specifically, we define escape mutations ($n=213$) as those conferring strong escape on their own (mutational positive differential selection >2 -fold) and occurring in a residue with a strong overall contribution to escape (site differential selection of $>\text{mean}+2\text{SD}$ in their respective profile).

To implement the machine learning approaches, numerical features were first preprocessed, and the dataset was split into training and testing datasets, comprising 75% and 25% of the data, respectively (Figure 5A, see Methods for details). Different algorithms were then evaluated. Overall, the random forest algorithm (RF) (Breiman 2001) using class priors and probability recalibration produced the best results, as measured by the F_1 -score, accuracy, precision, and Receiver Operating Characteristic (ROC) output (Table S4). When tested against the training dataset, the RF classifier showed $99.77 \pm 0.05\%$ accuracy (classification precision of 100% and $99.54 \pm 0.10\%$ for predicting no escape and escape, respectively; Figure 5B and Table S5). However, while the accuracy was maintained at high values with the testing dataset ($95.7 \pm 0.5\%$) a significant drop in predictability for mutations conferring escape was observed (classification precision: no escape $97.0 \pm 0.5\%$; escape: $37 \pm 9\%$; Figure 5B and Table S5).

The reduced predictability observed in the testing dataset for mutations conferring escape could stem from their low number in the dataset ($\sim 4\%$ or $n = 53$) but could also result from incomplete identification of some mutations conferring escape in the antigenic profiling experiments. To address this possibility, we reran the RF algorithm on 1,000 different random samples of training and testing datasets and identified mutations that were consistently classified as conferring escape by the RF algorithm ($>70\%$ probability) but not in the antigenic profiling experiments (see analysis_sites_muts at <https://github.com/RGellerLab/CVB3-Antigenic-Profiling>). In total, 27 such mutations were identified. Interestingly, all of these occurred in 12 sites where strong escape was already observed with other mutations in the antigenic profiling experiments, strongly suggesting these are likely to confer escape. Reclassification of these 27 instances in the dataset increased the predictability of the RF classifier by 53% in every testing dataset following training (accuracy $96.4 \pm 0.5\%$, classification precision: no escape $97.7 \pm 0.4\%$; escape: $58 \pm 7\%$; Figure 5B and Table S5).

Taking advantage of the fact that escape mutations could be identified by the RF algorithm, we evaluated which features were most predictive of escape using both

accuracy loss after feature shuffling (ALAS) (Breiman et al. 2017) and Shapley additive explanations (SHAP) values (Christoph Molnar 2020) (Figure 5C-D). Of the top seven features, the most relevant for forecasting escape were the identity of the WT and mutant AA. We therefore examined whether mutations conferring escape were enriched in particular residues, chemical polarity, or charge that could be important for antibody binding. Indeed, sites contributing to escape (WT AA) were enriched in polar and positively charged residues, in particular K and T ($p < 0.05$ by Fisher's exact test; Figure 5E and Figure S5). In contrast, a significant depletion of non-polar residues ($p < 0.05$ by Fisher's exact test), and leucine in particular ($p < 0.01$ by Fisher's exact test), was observed at escape sites (Figure 5E). On the other hand, mutations conferring escape (mutant AA) were significantly enriched in specific AA (R, A, G, L, and T; >2 -fold enrichment; $p < 0.05$ by Fisher's exact test) and depleted in others (K and N; >2 -fold reduction; $p < 0.05$ by Fisher's exact test) but this was not generalizable to other AA with the same characteristics ($p > 0.05$ by Fisher's exact test; Figure 5F and Figure S5). The other five variables which contributed $>10\%$ to model accuracy were quantitative measures of the effect of mutations on stability ($\Delta\Delta G$), distance from the center, flexibility (B-factor value), predicted disorder, and natural variation (Shannon's entropy; Figure 5C). Specifically, the combination of thermodynamically stable mutations in flexible, disordered, and distal surface residues that are evolutionarily variable were highly predictive of escape (Figure 5C-D).

To validate the relevance of the top seven features identified above for predicting escape, we trained and tested an RF classifier using only these features on the dataset including the 27 reclassified instances from above. Notably, the predictability of the RF classifier increased appreciably when evaluated on the testing dataset using only the top seven features versus the full set of features (accuracy $96.8 \pm 0.5\%$, classification precision: no escape $97.4 \pm 0.4\%$; escape: $72 \pm 8\%$; Table S5). Together, these findings reveal mutations conferring escape can be predicted using machine learning approaches and highlight the most relevant features for such predictions.

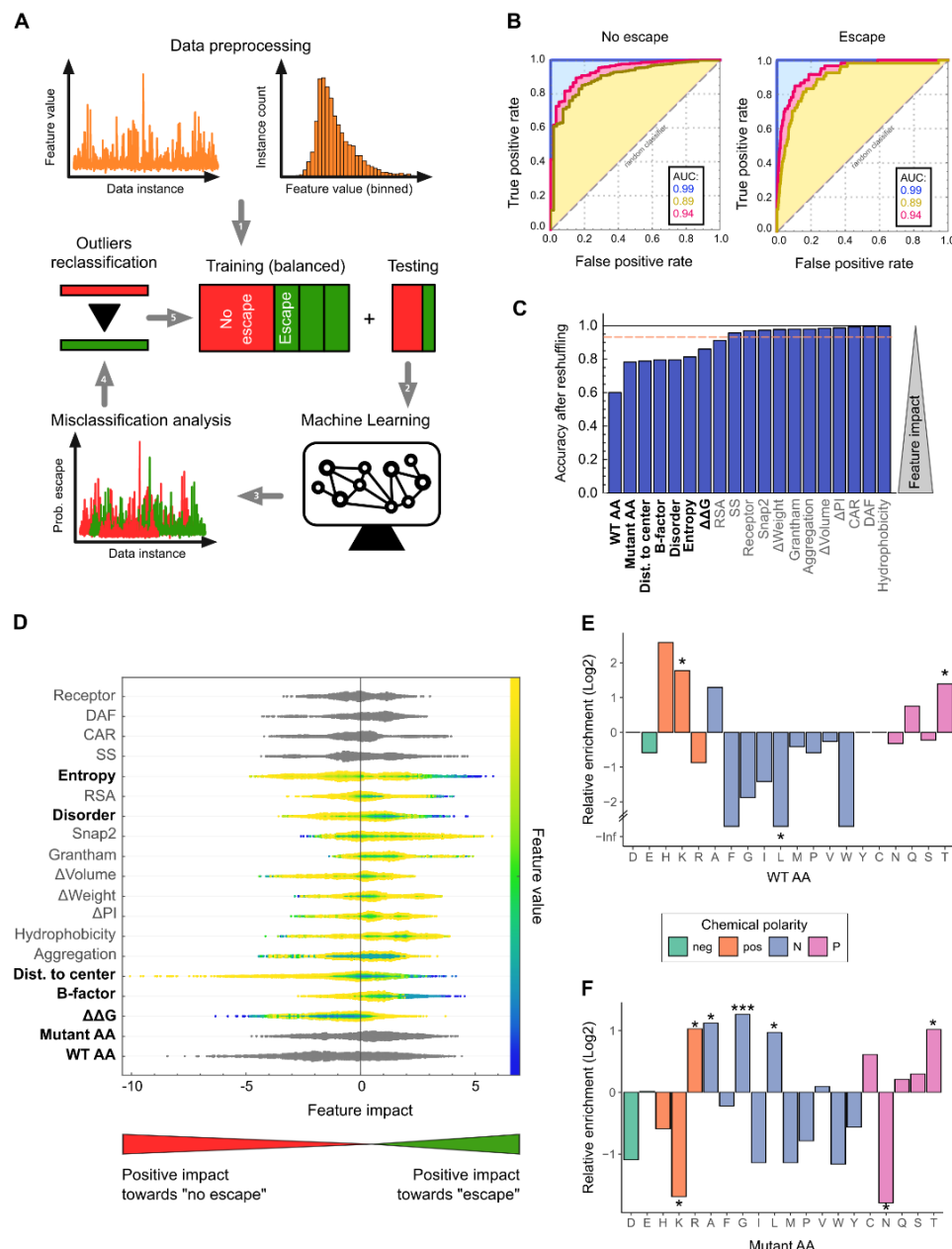


Figure 5. Escape mutations can be predicted by machine learning approaches. (A) Overview of the machine learning (ML) workflow: the dataset was preprocessed to reduce dimensionality by binning and randomly split into training (75%) and testing (25%) datasets. After 1,000 training-testing rounds, 27 outliers for which the ML classifier provided a large probability of escape (>70%) were identified. The reliability of the ML classifier was reevaluated after their reclassification (no escape to escape). (B) Receiver Operating Characteristic (ROC) curves for "no escape" (left) and "escape" (right) for training (blue), testing without reclassification (yellow), and testing with reclassification (pink). The area under the curve (AUC) is indicated in the insets. (C) Analysis of accuracy loss after feature shuffling. Features contributing to more than 10% of the model's accuracy are indicated in bold. (D) SHAP feature impact analysis. The colors indicate the relative value of numerical features and the positive or negative impact of the feature in the class is measured on the horizontal axis. Relevant predictive features (in bold) are those having a large impact, with their numeric value having directionality. (E) The relative enrichment of individual amino acids in sites conferring escape versus surface-exposed residues where no escape is observed. (F) The relative enrichment of individual amino acids in mutations conferring escape versus mutations in the same residues which do not confer escape. neg: negatively charged; pos: positively charged; N: non-polar; P: polar. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$ by Fisher's exact test. See Figure S5 for characteristics of sites and mutations of escape and Tables S3 and S4 for comparison of the different machine learning algorithms and RF classifiers tested, respectively.

Immune dominance of escape sites in the population

To better understand where polyclonal sera target the CVB3 capsid, we examined the distribution of sites conferring escape in the eight human antigenic profiles across the different capsid regions. Overall, 76 of 304 surface exposed residues were observed to confer escape (25%). Of these, 39.5% (30/76) were located on the northern rim of the canyon, 34.2% (26/76) on the outer surface of the canyon, 17.1% (13/76) on the inner surface of the canyon, and 9.2% (7/76) in the three-fold plateau (Figure 6A and Figure S6). Most samples present in common escape sites in the outer and northern rim of the canyon (Figure 6A). To assess if specific regions and/or sites were similarly targeted in a larger sample size, we screened representative mutants from each capsid region for their ability to confer escape from 18 additional human sera (reciprocal IC₅₀s of 550-1,100; Figure 6). Of the eight mutants tested, D535T (three-fold plateau) and K723E (northern rim of the canyon) were the most common mutations affecting neutralization, conferring a >2-fold reduction in sensitivity in all sera (Figure 6C). Reduced neutralization was also frequent with mutation P798H in the northern rim of the canyon (escape from 5/18 samples) and A512L in the inner surface of the canyon (escape from 4/18 samples). In contrast, the remaining mutations conferred escape from only a few (Q702G and K227S) or no (A473L and K650Y) sera despite being located in similar regions of the capsid as the most prevalent escape mutations (three-fold plateau and northern rim of the canyon, Figure 6). Interestingly, the K650Y mutation, which showed the strongest escape from mouse polyclonal sera, did not reduce neutralization by any of the human sera, highlighting potentially important differences between human and murine antibody responses. Overall, these results uncover two conserved sites of escape in the three-fold plateau and northern rim of the canyon in all sera samples, and highlight that escape is specific to a given residue rather than a region.

Strong escape is conferred by multiple mutations but comes at a cost to fitness

When single mutations conferred escape, their effect on neutralization were relatively moderate (average IC₅₀ fold-change of 5.1 ± 4.0). As viruses can accumulate multiple mutations during adaptation, we assessed the ability of viral mutants encoding two escape mutations to evade neutralization by the top 26 most neutralizing sera. Surprisingly, a double mutant encoding the two most prevalent escape mutations at the population level, D535T and K723E, did not reduce neutralization compared to the single mutations alone and even reduced the degree of escape in some cases (figure 6C, average fold-change in IC₅₀ of: 3.8 ± 1.9). In contrast, combining the first and the third most prevalent escape mutations (K723E + P798H), both of which are located in the northern rim of the canyon, conferred a similar or stronger escape from neutralization than the individual mutations in all but a single serum (h19; Figure 6C). Indeed, the average reduction in sensitivity across all sera was increased by nearly two-fold when combining these mutations (average IC₅₀ fold-change of 11.2 ± 8.2 , 6.24 ± 4.22 , and 2.36 ± 0.39 for K723E + P798H, K723E, and K798H, respectively; $p=0.01$ and $p=1.08 \times 10^{-5}$ vs. K723E and K798H, respectively, by a two-tail t-test). Moreover, in nearly 20% of the sera (5/26 sera), the K723E + P798H double mutant

reduced neutralization by >20 fold versus the WT virus. Such escape was not observed for any of the single mutants, suggesting strong escape from antibody neutralization requires the combination of at least two mutations.

The ability of a virus to successfully infect a host is dependent on viral fitness. As some of the mutations examined above conferred escape from nearly all sera, we sought to determine whether variation at these particular sites is not observed due to strong effects on viral fitness. We therefore directly assayed the effect of all single and double escape mutations on viral fitness in the absence of antibody selection. Of the eight single mutants tested, five incurred a fitness cost compared to the WT virus. In particular, the two mutations conferring escape in all sera (D535T and K723E), reduced viral fitness to 20% and 30% relative to the WT virus, respectively ($p=0.01$ and $p=0.003$ by two-tail t-test for D535T and K723E, respectively; Figure 6C). Combining these mutations further reduced fitness to 10% relative to the WT virus, significantly lower than each mutant alone ($p\text{-value}=0.0007$ and $p=0.0021$ versus D535T and K723E by two-tail t-test, respectively). The fitness of the mutant showing the strongest escape (K723E + P798H) was 50% lower than the WT ($p=0.011$ by two-tail t-test) but was not significantly different from the fitness of the individual mutants ($p\text{-value} > 0.5$ by two-tail t-test for both). Overall, mutants conferring escape from multiple sera (>20% of the sera tested) had strong fitness costs (average fitness of 0.28 ± 0.11), while those conferring escape from only a few sera (<20% of the sera tested) were less deleterious (average fitness of 0.73 ± 0.35), although this difference did not reach statistical significance ($p=0.08$ by a two-tail t-test). Together, these results suggest that escape from existing immunity may come at a significant cost to viral fitness, in particular for mutations in the most common sites.

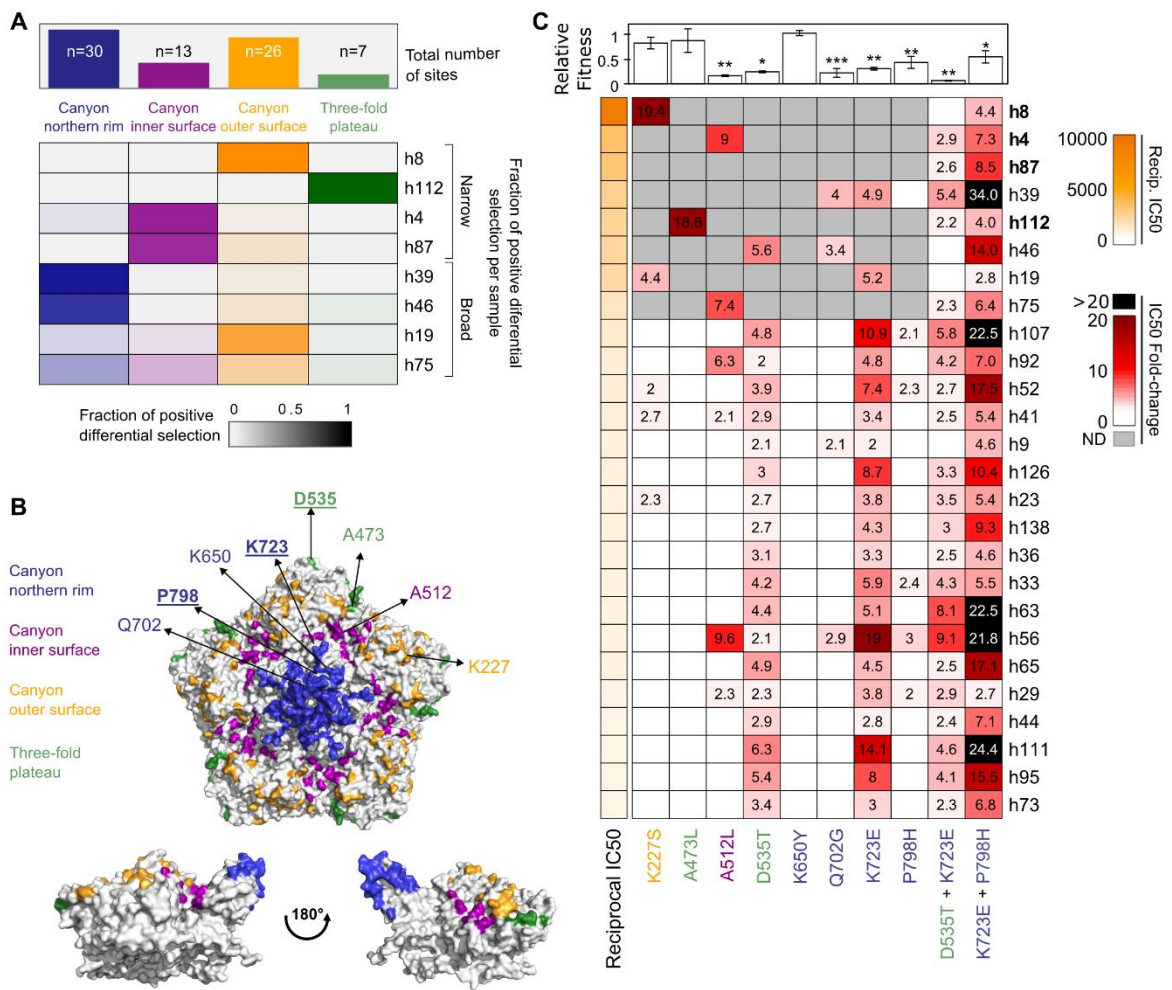


Figure 6. Epitope dominance for neutralizing antibodies targeting CVB3. (A) Heatmap representation of the total (top) or per serum (bottom) positive differential selection by capsid region. **(B)** The distribution of the main escape sites identified in the human antigenic profiles mapped onto the CVB3 capsid pentamer, colored according to their structural location. The mutations evaluated in (c) are indicated, with sites combined in the double mutants represented in bold and underlined. **(C)** Heatmap representation of the change in neutralization sensitivity conferred by single and double mutations. Samples with narrow antigenic profiles are highlighted in bold. Numbers indicate fold change in the IC50 of the mutant versus the WT virus when exceeding 2-fold. ND: not determined. The upper bar plot indicates the relative fitness of each mutant versus the WT virus. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$ by two-tailed t-test. See Figure S6 for a heatmap representation of positive differential selection per secondary structure element in the capsid.

Discussion

Neutralizing antibodies play fundamental roles in resolving viral infections and providing long-term protection against reinfection. Nevertheless, we have limited knowledge of how these responses target the capsids of non-enveloped viruses in polyclonal sera. Herein, we address this by comprehensively mapping the neutralizing antibody responses of both humans and mice to a prototypical picornavirus. Analysis of polyclonal sera from infected mice revealed a surprisingly homogenous antigenic profile (Figure 2). Indeed, all mouse sera targeted a single discrete region of the viral capsid at the northern rim of the canyon, which has been previously shown to be a target of neutralizing mouse mAbs (Wang et al. 2020; Xu et al. 2017; Yin et al. 2019). In contrast, the antigenic profiles from human sera were heterogeneous, targeting epitopes across all major antigenic regions in picornavirus capsids, and showed two distinct types of antigenic profiles (Figures 3 and 4). Four of the sera had a narrow antigenic profile, with neutralizing responses focused on one dominant antigenic site that, unlike the mouse sera, varied between individuals (Figure 3). In contrast, the remaining four sera had broader antigenic profiles, with multiple regions being targeted co-dominantly (Figure 4). These different profiles appear to have functional consequences: narrow profiles show an overall stronger neutralization capacity, comprising 4 of the 5 top neutralizing sera (labeled in bold in Figure 6), but are more susceptible to immune evasion, with single mutations conferring a stronger degree of escape than in broad-antigenic profiles (Figures 3 and 4).

Analyses of polyclonal immune responses to the membrane proteins of enveloped viruses have also revealed both narrow and broad neutralization profiles. Narrow profiles were reported for neutralizing antibodies targeting the influenza A hemagglutinin (HA) protein, which were focused on a single region and could be efficiently evaded by individual mutations (Lee et al. 2019a). In contrast, polyclonal responses to the H glycoprotein of measles virus were shown to co-dominantly target multiple sites (Muñoz-Alía et al. 2021), with individual mutations conferring weak escape from neutralization, as observed for broad antigenic profiles targeting CVB3. Interestingly, antibodies targeting influenza A virus generally do not confer significant protection against reinfection with related strains, while those targeting the measles virus H protein do (Bouche, Ertl, and Muller 2002), raising the possibility that individuals with broad antigenic profiles are more resistant to reinfection with related viruses. It is currently unclear what drives the evolution of broad or narrow antigenic profiles in CVB3. One possibility is that broad profiles represent the consequence of multiple infections with related strains. Unfortunately, we are unable to address this question in the current study as the number of times the donors in our study were infected with CVB3 and the infecting strain(s) are unknown. In this sense, analysis of the antigenic profiles from children may be informative to distinguish these possibilities as, unlike adults, they are likely to have been exposed to a single CVB3 variant.

Having obtained a comprehensive dataset of mutations conferring escape in eight human sera, we sought to understand whether such mutations could be predicted using machine learning approaches and to define the most relevant attributes for such prediction. We found that a random forest algorithm could predict escape to relatively high accuracy based on only seven features, indicating that mutations conferring escape have unique, typifying attributes (Figure 5). Specifically, we found escape mutants to be enriched in particular amino acid combinations, to be less destabilizing,

and to occur in residues that are flexible, distal from the core, and evolutionarily variable (Figure 5). As these parameters can be obtained for any virus for which a high-resolution capsid structure is available, this machine learning algorithm could potentially be used to identify escape mutations in related picornaviruses without the need to perform antigenic profiling.

The relative dominance of epitopes targeted by polyclonal human responses following natural infection with CVB3 are unknown. To address this, we assessed the ability of individual escape mutants from different capsid regions to alter neutralization by an additional 18 sera (Figure 6). Several observations were made from this analysis. First, the two mutants conferring the most significant escape in the antigenic profiles, K277S and A473L (>18-fold reduction in IC₅₀; Figure 6), which were derived from sera showing narrow antigenic profiles, did not confer significant escape in the remaining sera (maximum reduction in IC₅₀ of 4.4-fold). Second, the most frequent mutations conferring escape were observed in broad antigenic profiles. Specifically, the D535T and K723E mutations conferred some degree of escape in all sera tested, highlighting an unexpected conservation in human polyclonal responses to natural CVB3 infection. Finally, the K650Y mutation that conferred strong escape from all mouse sera had no appreciable effect in any of the human sera, revealing significant differences in human and murine humoral responses.

Having established the epitope dominance of individual escape mutants, we next assessed the degree of escape conferred by combining two individual escape mutations. Somewhat unexpectedly, a virus encoding the two mutations conferring reduced neutralization by all sera tested (D535T and K723E) did not lead to increased escape compared to the individual mutations (Figure 6). In contrast, combining the two most frequent mutations in the northern rim of the canyon (K723E and P798H) resulted in enhanced escape from multiple sera (20 to 34-fold reduction in IC₅₀; Figure 6). Despite being in the same capsid region, these mutations are separated by a distance of 20Å, exceeding the distance of typical epitope footprints (Ramaraj et al. 2012). Together, these findings suggests that neutralizing antibody responses targeting CVB3 are focused on multiple epitopes in the northern rim of the canyon and that more than one mutation is required to confer strong escape from neutralization by polyclonal sera. Whether these patterns are conserved in other picornaviruses and/or non-enveloped viruses remains to be elucidated. In this sense, the application of mutational antigenic profiling to additional non-enveloped viruses can help shed light on variation in polyclonal responses targeting viral capsids and may provide clues as to why some viruses harbor only a limited number of serotypes (e.g. three for poliovirus) while others can reach much higher antigenic diversity (>100 for rhinovirus; (Racaniello 2013)).

Acknowledgments. We want to particularly acknowledge the donors of the sera samples and the IBSP-CV Biobank (PT17/0015/0017) integrated in the Spanish National Biobanks Network for their collaboration. We would like to thank Luciana Rusu and Joao Zulaica for technical support.

Funding. This work was supported by grants to R.G. from the Spanish Ministerio de Ciencia y Innovacion (grants BFU2017-86094-R and PID2021-125063NB-I00), the Generalitat Valenciana (grant AICO/2020/216), and the European Commission NextGenerationEU fund through Consejo Superior de Investigaciones Científicas (CSIC's) Global Health Platform (Plataforma Temática Interdisciplinar, PTI Salud Global). B.A-R. holds a Generalitat Valenciana postdoctoral fellowship (APOSTD/2021/017). J.B. acknowledges support from the Spanish Ministry of Science and Innovation through grant PID2019- 105566GB-I00 and from the LifeHUB Research Network through grant PIE-202120E047- Conexiones-Life (CSIC).

Author contributions. B.A-R., J.B., and R.G. analyzed the data, wrote the manuscript, and acquired funding. R.G. conceived the project. J.B. designed and implemented the machine learning analysis. B.A-R. performed all experiments.

Methods

Cells, viruses, and reagents

HeLa H1 (CRL-1958; RRID: CVCL_3334) cells were obtained from ATCC and were periodically validated to be free of mycoplasma. Cells were cultured in culture media (Dulbecco's modified Eagle's medium, Pen-Strep, and l-glutamine) supplemented with 10% or 2% heat-inactivated fetal bovine serum for culturing or infection, respectively. The anti-Coxsackievirus B3 monoclonal antibody was obtained from Merck (clone 280-5F-4E-5E, MAB948). The human codon-optimized T7 polymerase plasmid was obtained from Addgene (#65974) and the CVB3 infectious clones encoding mCherry (CVB3-mCherry) or eGFP (CVB3-eGFP) were previously described (Bou, Geller, and Sanjuán 2019). The titer of these reporter viruses was obtained by infecting Hela H1 cells with serial dilutions of each virus in 96-well plates and counting the number of fluorescent cells at 8 hours post-infection using an Incucyte SX5 Live-Cell Analysis System (Sartorius). The CVB3 populations harboring high diversity in the capsid region have been previously described (Mattenberger et al. 2021). Briefly, a reverse genetics plasmid encoding CVB3 genome was mutagenized at the codon level to generate viral populations containing >96% of all possible single amino acid mutations in the capsid region. The virus populations were then passaged at low multiplicity of infection (MOI) to create a passage 1 population in which there is a genotype-phenotype link between the capsid proteins of each virus and the genome it carries. The viral titers in these libraries were determined by plaque assay in Hela H1 cells.

Neutralizing sera

The 140 human sera used in this study were provided by the IBSP-CV Biobank (PT17/0015/0017), integrated in the Spanish National Biobanks Network and in the Valencian Biobanking Network and they were processed following standard operating procedures with the appropriate approval of the Ethics and Scientific Committees. All samples were collected from healthy donors between 2010-2021 and included 74 females and 66 males. No information was available on the CVB3 infection record of these individuals. For mouse sera (ethics protocol approval 2019/VSC/PEA/0151), five-week-old male Balb/C mice were infected via intraperitoneal injection with 10^4 (m1) or 10^5 (m2 and m3) PFUs of CVB3 and sera was collected three weeks post-infection. All sera were heat inactivated for 30 minutes at 55°C before use.

Antigenic profiling

For mutational antigenic profiling, we used the passage 1 mutant virus populations derived from libraries 1 and 2 (L1 and L2) described in (Mattenberger et al. 2021). For each serum or antibody, we performed two biological replicates (i.e., L1 and L2), with one or two replicates of each of the virus libraries (technical replicates). Briefly, 10^6 PFU of each virus library were mixed with sera or antibodies at the indicated dilutions to have 1% to 10% of the virus library survive the neutralization treatment, as judged by qPCR (see below). Mock neutralized controls were included for each neutralized library replicate. After a 1h incubation at 37°C, Hela H1 cells plated the day before in 6-well plates were infected with the virus-antibody mix or virus alone. At 8hpi, cells were subjected to three freeze-thaw cycles, cell debris were removed by centrifugation at 500xg, and the supernatants were collected. Finally, 200 µL of each sample were treated with 2 µL of RNase-Free DNaseI (ThermoFisher) for 30 min at 37°C, and viral RNA was extracted using the Quick-RNA Viral Kit (Zymo Research), eluting in 20 µL.

To estimate the fraction of virus surviving the neutralization treatment, viral RNA was quantified by RT-qPCR using the GoTaq Probe 1-Step RT-qPCR System (Promega) on a QuantStudio3 (ThermoFisher Scientific). Ten-fold serial dilutions of RNA extracted from a known titer virus stock were used to generate a standard curve. RT-qPCR was performed using 2 μ L of template RNA of neutralized and control samples (forward primer: RTqPCR_F (GAT CGC ATA TGG TGA TGA TGT GA), reverse primer: RTqPCR_R (AGC TTC AGC GAG TAA AGA TGC A), and TaqManProbe: CVB3_probe (6FAM-CGC ATC GTA CCC ATG G-TAMRA). RT-qPCR Ct values were interpolated to the standard curve to quantify viral RNA in each sample and the fraction of virus surviving neutralization was determined versus the untreated controls (see Table S1 for details on the number of replicates performed, amount of antibody used and surviving fraction after neutralization).

Duplex sequencing

For sequencing, viral RNA extracted as indicated above was reverse transcribed with the high-fidelity OneScript Plus Reverse Transcriptase (Applied Biological Materials) using 8 μ L of RNA and the primer CVB3_RT_3450 (GTG CTG TGG TCG TGC TCA CTA A). To ensure sufficient diversity was present in the cDNA, the number of capsid copies in the cDNA was quantified via qPCR using PowerUp SYBR Green Master Mix (ThermoFisher Scientific) in a 10 μ L total reaction, using 1 μ L of the template and the primers CVB3_qPCR_F (CCC TGA ATG CGG CTA ATC C) and CVB3_qPCR_R (AAA CAC GGA CAC CCA AAG TAG TC). A standard curve was generated using the original reverse genetic plasmid encoding the CVB3 genome. The full capsid region was then amplified from the cDNA using KOD polymerase Mastermix (Novagen) and the primers CVB3_P1_seq_F (CCC TTT GTT GGG TTT ATA CCA CTT AG) and CVB3_P1_seq_R (CCT GTA GTT CCC CAC ATA CAC TG). Duplex sequencing libraries were prepared as previously described (Mattenberger et al. 2021) and sequenced on an Illumina Novaseq 6000 sequencer. The resulting files were analyzed as previously described (Mattenberger et al. 2021) to obtain the counts of each codon at each position (codon tables, available at https://github.com/RGellerLab/CVB3-Antigenic-Profiling/tree/main/1_codon_tables). As before, all single mutations in codons were omitted from the analysis to increase the signal-to-noise ratio (Mattenberger et al. 2021).

Calculation and visualization of positive differential selection

Codon tables were analyzed for differential selection versus mock neutralized controls of each library using the DMS_tools2 package (Bloom 2015) with the default settings. Two batches of experiments were performed with a different number of replicates. The first batch included samples mAb, m1, m2, m3, h4, h8, h39, and h46, and three non-neutralized controls for each library. Four replicates were performed for each sample (2 technical replicates of 2 biological replicates from libraries 1 and 2). Differential selection of each sample was calculated versus the mean of the three controls. In the second batch, 2 biological replicates (libraries 1 and 2) were performed for samples h19, h75, h87, and h112 together with an individual non-neutralized control for each library. One of the replicates of h87 (library 1) had to be discarded due to the low mutation rate observed in this sample after sequencing (30% smaller than the average of the other selected libraries), and positive differential selection was calculated versus the individual control of each library. All positive differential selection plots in the main figures represent the mean differential selection for each mutation at each particular

site across replicates. Line plots represent the mean positive differential selection per site and logoplots represent the mean positive differential selection for each amino acid mutation at selected sites. Differential selection tables (https://github.com/RGellerLab/CVB3-Antigenic-Profiling/tree/main/2_meansitediffsel, https://github.com/RGellerLab/CVB3-Antigenic-Profiling/tree/main/3_meanmutdiffsel) and positive differential selection logoplots of the full capsid region (https://github.com/RGellerLab/CVB3-Antigenic-Profiling/tree/main/4_Logoplots) are available on Github. Interactive visualization of the antigenic profiles using the dms-view software (Hilton et al. 2020) is available on Github (https://github.com/RGellerLab/CVB3-Antigenic-Profiling/tree/main/6_dms_view).

Neutralization assays

For the neutralization assays, 2.5×10^3 fluorescent virus units of WT or mutant CVB3 mCherry virus were mixed in duplicate with serial dilutions of each antibody or sera sample and incubated for 1h at 37°C. Virus-only samples without antibodies were used as a control to measure maximal fluorescence in the absence of neutralization. Virus-antibody mixes were then transferred to 96-well plates containing Hela H1 cells plated the day before. After an 8-hour incubation at 37°C, mCherry-derived red fluorescent signal was measured using an Incucyte SX5 Live-Cell Analysis System (Sartorius) to determine fluorescent virus units per well. The antibody concentration resulting in a 50% reduction of virus signal (IC₅₀) was obtained using the two-parameter log-logistic function LL2 from the R drc package (version 3.0.1). All IC₅₀ values represent the mean and standard error of the curve fitting of two replicate curves run in the same 96-well plate. Results of the human samples neutralization screening (https://github.com/RGellerLab/CVB3-Antigenic-Profiling/tree/main/7_biobank_results) and the neutralization assays performed with different CVB3 mutants (https://github.com/RGellerLab/CVB3-Antigenic-Profiling/tree/main/8_neutralization_ic50s) are available on GitHub.

Generation of CVB3 capsid mutants

CVB3 mutants were generated by site-directed mutagenesis using the mCherry-CVB3 fluorescent infectious clone mentioned above (Bou, Geller, and Sanjuán 2019). For each mutant, non-overlapping primers containing the desired mutation in one of the primers were used to introduce the mutation with Q5 polymerase, followed by DpnI (Thermo Scientific) treatment, phosphorylation, ligation, and transformation of chemically competent bacteria (NZY5α Competent Cells, NZY Tech). Successful mutagenesis was verified by Sanger sequencing. Subsequently, plasmids were linearized with Sall-HF (ThermoFisher), and 600 ng of plasmid were transfected into 5×10^4 HEK293T cells, together with 600 ng of a plasmid encoding the T7 polymerase (Addgene 65974) using Lipofectamine 2000 (ThermoFisher) according to manufacturer's instructions of use. Cells were then incubated until cytopathic effect (CPE) was observed and passage 0 virus was collected. Infectious virus titer was determined using an Incucyte system (Sartorius) to quantify the red fluorescence from the mCherry reporter signal as detailed above. If needed, mutants with low titers were amplified for an additional passage. The capsid region of the mutant virus populations was reverse-transcribed, PCR-amplified, and sequenced to ensure no compensatory mutations or reversions arose during replication.

Competition assays to assess the fitness of virus mutants

A 1:1 mixture of CVB3-eGFP and either CVB3-mCherry viruses encoding the WT or mutant capsid was used to infect Hela H1 cells at an MOI of 0.001 in a 24-well plate. Automated real-time quantitative fluorescence microscopy was used to track the spread of each virus in an Incucyte SX5 Live-Cell Analysis System (Sartorius). The ratio of the mCherry signal (derived from the WT or mutant virus) to the eGFP signal of the reference virus at 20hpi (Ratio_{20hpi}) provides a measure of the relative success of both the WT and mutant virus following several rounds of infection. The ratio of mCherry to eGFP signal at 8hpi (Ratio_{8hpi}) reflects the initial relative abundance of each virus before competition. To calculate the fitness of each mutant versus that of the WT virus, the formula $(\text{Ratio}_{20\text{hpi}}^{\text{Mutant}} / \text{Ratio}_{8\text{hpi}}^{\text{Mutant}}) / (\text{Ratio}_{20\text{hpi}}^{\text{WT}} / \text{Ratio}_{8\text{hpi}}^{\text{WT}})$ was used. The results of the competition assay are available on Github (https://github.com/RGellerLab/CVB3-Antigenic-Profiling/tree/main/9_competition_assay).

Bioinformatic analyses

For bioinformatics analysis, R version 4.2.0 was used with packages tidyverse, reshape2, drc (IC50 curve fitting), ggplot2 and ggh4x (plotting), ggseqlogo (logoplots), ComplexHeatmap (heatmaps), viridis (color scale), ggcorrplot (correlation matrices), igraph (bipartite network analysis), DECIPHER and Biostrings. R (Team 2021), PyMOL (Schrödinger, n.d.), and Inkscape (Inkscape, n.d.) were used to generate the figures. CVB3 Nancy capsid roadmaps were generated using RIVEM (Xiao and Rossmann 2007) with surface exposed residues colored according to their positive differential selection values and labeled according to the position of each residue in the polyprotein, available at https://github.com/RGellerLab/CVB3-Antigenic-Profiling/tree/main/5_capsid_roadmaps.

All structural features were calculated based on a CVB3 capsid pentamer (PDB ID: 4GB3) mutated to encode the Nancy strain amino acid sequence with FoldX (Schymkowitz et al. 2005) as previously reported (Mattenberger et al. 2021). Outer surface residues were defined as residues with a distance to the center of the capsid >130 Å and a surface exposure >2.5 Å² based on the FindSurfaceResidues script in PyMol (Schrödinger, n.d.). The relative surface area (RSA) of residues in the pentamer was calculated using PSAIA (Mihel et al. 2008). For entropy calculation, all available enterovirus A, B, and C full polyprotein sequences in BV-BRC (bv-brc.org) were downloaded on May 14th, 2022. Duplicate sequences were removed, and the alignments were randomly subsampled to harbor the same number of sequences using the Biostrings package in R. Sequences were then combined and aligned using the DECIPHER package in R. All gaps versus the CVB3 Nancy capsid sequence were removed, and Shannon's entropy was calculated using a custom R script. Both the alignment file and the Shannon entropy can be found on GitHub (https://github.com/RGellerLab/CVB3-Antigenic-Profiling/tree/main/11_Sequence_alignments). Finally, B factors were obtained from the PDB file, distance to center was calculated in PyMol using the command distance (cmd.dist) and disorder was calculated on the capsid sequence using IUpred3 (Erdős, Pajkos, and Dosztányi 2021). Tables used for the analysis of the characteristics of the main escape sites and mutations are available on GitHub (https://github.com/RGellerLab/CVB3-Antigenic-Profiling/tree/main/10_analysis_sites_muts).

Machine Learning

The initial dataset used for the machine learning analyses was obtained by merging the structural, physicochemical, and evolutionary features for all surface exposed capsid residues (see Bioinformatic analyses section for details). Sites of escape were defined as surface exposed residues with a positive differential selection value above the mean+2SD value per sample. Within these sites, escape mutations for each sample were defined as mutations with a >2-fold positive differential selection value (https://github.com/RGellerLab/CVB3-Antigenic-Profiling/tree/main/10_analysis_sites_muts). To implement the analysis, the data was preprocessed. Specifically, some of the numerical features in the original dataset had nearly as many different values as the number of instances ($n=5,776$). This complexity (dimensionality) was reduced to increase the predictive capabilities of ML by implementing a binning approach that assumes an underlying smooth distribution of values and assigning all the instances falling within a bin the same value (the bin mid value). For example, $\Delta\Delta G$, which had 5,772 values originally, was reduced to 93 values following binning.

Next, the dataset was randomly sampled and split into training (75%) and testing (25%) datasets for each iteration of training and testing. The training dataset was balanced in terms of "escape" and "no escape" entries by repeating the "escape" entries. Following training, we evaluated the performance of three algorithms: neural network, support vector machine, and random forest (Table S4). Given that our data was clearly biased towards the "no escape" class, we assessed the possible improvement when using prior probabilities and a probability recalibration approach. Overall, the random forest algorithm combined with prior probabilities and probability recalibration showed the best performance.

Finally, to identify mutations that were consistently misclassified in the ML compared to the antigenic profiling dataset, we estimated the expected number of times that the same instance is wrongly classified in N different testing samples of a dataset by a random classifier using the formula $Nf_{test}/2$, where f_{test} is the fraction of data for testing. We then sampled $N = 1,000$ different training/testing sets and, for each sample, we identified during the testing step the wrongly classified instances (with respect to the antigenic profiling data) having a probability >70% to belong to the "escape" class. We then selected among those the ones that appear more than expected ($Nf_{test}/2 = 125$ times). This procedure yields the list of 27 instances that we reclassified (https://github.com/RGellerLab/CVB3-Antigenic-Profiling/tree/main/10_analysis_sites_muts). All machine learning analyses were performed using Mathematica 13.0 (Wolfram Research 2022) and the code is available on GitHub (https://github.com/RGellerLab/CVB3-Antigenic-Profiling/tree/main/12_ML_code).

Statistical analyses

Wilcoxon tests, Fisher's exact tests, and t-tests were performed in R (Team 2021). ns: $p>0.05$, *: $p<0.05$, **: $p<0.01$, ***: $p<0.005$, ****: $p<0.001$. All tests were two-tailed.

Data availability

Sequencing data have been uploaded to SRA (BioProject PRJNA779606, SRA: SRP345591). The following supplementary files are available on GitHub (<https://github.com/RGellerLab/CVB3-Antigenic-Profiling>):

Codon_tables. Codon files for all samples.

https://github.com/RGellerLab/CVB3-Antigenic-Profiling/tree/main/1_codon_tables

Meansitediffsel. Mean positive site differential selection per site calculated for each sample. A subfolder is included with the positive differential selection of each replicate.
https://github.com/RGellerLab/CVB3-Antigenic-Profiling/tree/main/2_meansitediffsel

Meanmutdiffsel. Mean positive site differential selection per mutation calculated for each sample. A subfolder is included with the positive differential selection of each replicate.
https://github.com/RGellerLab/CVB3-Antigenic-Profiling/tree/main/3_meanmutdiffsel

Logoplots. Logoplots representing the mean positive differential selection values of positively selected mutations for each sample.
https://github.com/RGellerLab/CVB3-Antigenic-Profiling/tree/main/4_Logoplots

Capsid_roadmaps. Capsid roadmap with surface exposed residues colored according to the positive differential selection values obtained and labeled according to their position in the polyprotein.
https://github.com/RGellerLab/CVB3-Antigenic-Profiling/tree/main/5_capsid_roadmaps

Dms_view. Files for the interactive visualization of the antigenic profiles in dms-view (dms-view.github.io).
https://github.com/RGellerLab/CVB3-Antigenic-Profiling/tree/main/6_dms_view

Biobank_results. Results of the neutralization screening performed for 140 human samples. Anonymized ID, sex and age of donors are indicated, together with collection date, ic50 obtained and sd error in ic50 calculation.
https://github.com/RGellerLab/CVB3-Antigenic-Profiling/tree/main/7_biobank_results

Neutralization_ic50s. IC50 results of the neutralization assays for the human and mouse sera samples. Sample and virus tested are indicated, together with the ic50 obtained for the mutant and for the WT virus, and the fold change in the IC50.
https://github.com/RGellerLab/CVB3-Antigenic-Profiling/tree/main/8_neutralization_ic50s

Competition_assay. Data from the competition assay. The ratio between fluorescence signal obtained at 20hpi versus 8hpi for each mutant or wildtype virus is indicated.
https://github.com/RGellerLab/CVB3-Antigenic-Profiling/tree/main/9_competition_assay

Analysis_sites_muts. Tables used for the structural analyses and prediction of characteristics of the main escape sites and mutations.
https://github.com/RGellerLab/CVB3-Antigenic-Profiling/tree/main/10_analysis_sites_muts

Sequence_alignments. Alignment file of all available enterovirus A, B, and C full polyprotein sequences in BV-BRC (bv-brc.org) and the calculated Shannon entropy.

https://github.com/RGellerLab/CVB3-Antigenic-Profiling/tree/main/11_Sequence_alignments

ML_code. Custom code used for the machine learning analyses in Mathematica.
https://github.com/RGellerLab/CVB3-Antigenic-Profiling/tree/main/12_ML_code

References

- Bloom, Jesse D. 2015. "Software for the Analysis and Visualization of Deep Mutational Scanning Data." *BMC Bioinformatics* 16 (1): 168. <https://doi.org/10.1186/s12859-015-0590-4>.
- Bou, Juan Vicente, Ron Geller, and Rafael Sanjuán. 2019. "Membrane-Associated Enteroviruses Undergo Intercellular Transmission as Pools of Sibling Viral Genomes." *Cell Reports* 29 (3): 714-723.e4. <https://doi.org/10.1016/j.celrep.2019.09.014>.
- Bouche, Fabienne B., Oliver T. Ertl, and Claude P. Muller. 2002. "Neutralizing B Cell Response in Measles." *Viral Immunology* 15 (3): 451–71. <https://doi.org/10.1089/088282402760312331>.
- Breiman, Leo. 2001. "Random Forests." In *Machine Learning*, 45:5–32. <https://doi.org/10.1023/A:1010933404324>.
- Breiman, Leo, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone. 2017. *Classification And Regression Trees*. Routledge. <https://doi.org/10.1201/9781315139470>.
- Brouwer, Lieke, Giulia Moreni, Katja C. Wolthers, and Dasja Pajkrt. 2021. "World-Wide Prevalence and Genotype Distribution of Enteroviruses." *Viruses* 13 (3): 434. <https://doi.org/10.3390/v13030434>.
- Burton, Dennis R. 2002. "Antibodies, Viruses and Vaccines." *Nature Reviews Immunology* 2 (9): 706–13. <https://doi.org/10.1038/nri891>.
- Christoph Molnar. 2020. *Interpretable Machine Learning*. Edited by Lulu.com.
- Cifuentes, Javier Orlando, and Gonzalo Moratorio. 2019. "Evolutionary and Structural Overview of Human Picornavirus Capsid Antibody Evasion." *Frontiers in Cellular and Infection Microbiology* 9 (August): 1–11. <https://doi.org/10.3389/fcimb.2019.00283>.
- Corti, Davide, and Antonio Lanzavecchia. 2013. "Broadly Neutralizing Antiviral Antibodies." *Annual Review of Immunology* 31 (1): 705–42. <https://doi.org/10.1146/annurev-immunol-032712-095916>.
- Dingens, Adam S., Payal Pratap, Keara Malone, Sarah K. Hilton, Thomas Ketas, Christopher A. Cottrell, Julie Overbaugh, et al. 2021. "High-Resolution Mapping of the Neutralizing and Binding Specificities of Polyclonal Sera Post Hiv Env Trimer Vaccination." *ELife* 10: 1–32. <https://doi.org/10.7554/ELIFE.64281>.
- Dotzauer, Andreas, and Leena Kraemer. 2012. "Innate and Adaptive Immune Responses against Picornaviruses and Their Counteractions: An Overview." *World Journal of Virology* 1 (3): 91–107. <https://doi.org/10.5501/wjv.v1.i3.91>.
- Doud, Michael B., Scott E. Hensley, and Jesse D. Bloom. 2017. "Complete Mapping of Viral Escape from Neutralizing Antibodies." *PLoS Pathogens* 13 (3): 1–20. <https://doi.org/10.1371/journal.ppat.1006271>.
- Erdős, Gábor, Mátyás Pajkos, and Zsuzsanna Dosztányi. 2021. "IUPred3: Prediction of Protein Disorder Enhanced with Unambiguous Experimental Annotation and Visualization of Evolutionary Conservation." *Nucleic Acids Research* 49 (W1): W297–303. <https://doi.org/10.1093/nar/gkab408>.
- Fairweather, DeLisa, and Noel R. Rose. 2007. "Coxsackievirus-Induced Myocarditis in Mice: A Model of Autoimmune Disease for Studying Immunotoxicity." *Methods* 41 (1): 118–22. <https://doi.org/10.1016/j.ymeth.2006.07.009>.
- Greaney, Allison J., Andrea N. Loes, Lauren E. Gentles, Katharine H.D. Crawford, Tyler N. Starr, Keara D. Malone, Helen Y. Chu, and Jesse D. Bloom. 2021. "Antibodies Elicited by mRNA-1273 Vaccination Bind More Broadly to the Receptor Binding Domain than Do Those from SARS-CoV-2 Infection." *Science*

- Translational Medicine* 13 (600). <https://doi.org/10.1126/scitranslmed.abi9915>.
- Greaney, Allison J., Tyler N. Starr, Christopher O. Barnes, Yiska Weisblum, Fabian Schmidt, Marina Caskey, Christian Gaebler, et al. 2021. "Mapping Mutations to the SARS-CoV-2 RBD That Escape Binding by Different Classes of Antibodies." *Nature Communications* 12 (1): 4196. <https://doi.org/10.1038/s41467-021-24435-8>.
- Greaney, Allison J., Tyler N. Starr, Pavlo Gilchuk, Seth J. Zost, Elad Binshtein, Andrea N. Loes, Sarah K. Hilton, et al. 2021. "Complete Mapping of Mutations to the SARS-CoV-2 Spike Receptor-Binding Domain That Escape Antibody Recognition." *Cell Host and Microbe* 29 (1): 44-57.e9. <https://doi.org/10.1016/j.chom.2020.11.007>.
- He, Maozhou, Longfa Xu, Qingbing Zheng, Rui Zhu, Zhichao Yin, Yu Lin, Lisheng Yang, et al. 2021. "Identification of Antibodies with Non-Overlapping Neutralization Sites That Target Cocksackievirus A16" 27 (2): 249–61. <https://doi.org/10.1016/j.chom.2020.01.003>.Identification.
- Hilton, Sarah, John Huddleston\$, Allison Black, Khrystyna North, Adam Dingens, Trevor Bedford, and Jesse Bloom. 2020. "Dms-View: Interactive Visualization Tool for Deep Mutational Scanning Data." *Journal of Open Source Software* 5 (52): 2353. <https://doi.org/10.21105/joss.02353>.
- Huang, David B., Jashin J. Wu, and Stephen K. Tying. 2004. "A Review of Licensed Viral Vaccines, Some of Their Safety Concerns, and the Advances in the Development of Investigational Viral Vaccines." *Journal of Infection* 49 (3): 179–209. <https://doi.org/10.1016/j.jinf.2004.05.018>.
- Huang, Kuan-Ying A. 2021. "Structural Basis for Neutralization of Enterovirus." *Current Opinion in Virology* 51 (December): 199–206. <https://doi.org/10.1016/j.coviro.2021.10.006>.
- Huang, Kuan Ying A. 2021. "Structural Basis for Neutralization of Enterovirus." *Current Opinion in Virology* 51: 199–206. <https://doi.org/10.1016/j.coviro.2021.10.006>.
- Inkscape. n.d. "Inkscape Project." <https://inkscape.org>.
- Lee, Juhye M., Rachel Eguia, Seth J. Zost, Saket Choudhary, Patrick C. Wilson, Trevor Bedford, Terry Stevens-Ayers, et al. 2019. "Mapping Person-to-Person Variation in Viral Mutations That Escape Polyclonal Serum Targeting Influenza Hemagglutinin." *ELife* 8: 1–28. <https://doi.org/10.7554/eLife.49324>.
- Levy, Hazel, Mihnea Bostina, David J. Filman, and James M. Hogle. 2014. "Cell Entry: A Biochemical and Structural Perspective." In *The Picornaviruses*, 87–104. Washington, DC, USA: ASM Press. <https://doi.org/10.1128/9781555816698.ch6>.
- Li, Yuqing, Dongqi Liu, Yating Wang, Wenquan Su, Gang Liu, and Weijie Dong. 2021. "The Importance of Glycans of Viral and Host Proteins in Enveloped Virus Infection." *Frontiers in Immunology* 12 (April). <https://doi.org/10.3389/fimmu.2021.638573>.
- Liu, Tongyu, Yiquan Wang, Timothy J.C. Tan, Nicholas C. Wu, and Christopher B. Brooke. 2022. "The Evolutionary Potential of Influenza A Virus Hemagglutinin Is Highly Constrained by Epistatic Interactions with Neuraminidase." *Cell Host & Microbe* 30 (10): 1363-1369.e4. <https://doi.org/10.1016/j.chom.2022.09.003>.
- Martín, Javier, Graham Crossland, David J. Wood, and Philip D. Minor. 2003. "Characterization of Formaldehyde-Inactivated Poliovirus Preparations Made from Live-Attenuated Strains." *Journal of General Virology* 84 (7): 1781–88. <https://doi.org/10.1099/vir.0.19088-0>.

- Mateu, Mauricio G. 1995. "Antibody Recognition of Picornaviruses and Escape from Neutralization: A Structural View." *Virus Research* 38 (1): 1–24. [https://doi.org/10.1016/0168-1702\(95\)00048-U](https://doi.org/10.1016/0168-1702(95)00048-U).
- Mattenberger, Florian, Victor Latorre, Omer Tirosh, Adi Stern, and Ron Geller. 2021. "Globally Defining the Effects of Mutations in a Picornavirus Capsid." *ELife* 10: 1–56. <https://doi.org/10.7554/ELIFE.64256>.
- Mihel, Josip, Mile Šikić, Sanja Tomić, Branko Jeren, and Kristian Vlahoviček. 2008. "PSAIA – Protein Structure and Interaction Analyzer." *BMC Structural Biology* 8 (1): 21. <https://doi.org/10.1186/1472-6807-8-21>.
- Muckelbauer, Jodi K, Marcia Kremer, Iwona Minor, Guy Diana, Frank J Dutko, James Groarke, Daniel C Pevear, and Michael G Rossmann. 1995. "The Structure of Coxsackievirus B3 at 3.5 Å Resolution." *Structure* 3 (7): 653–67. [https://doi.org/10.1016/S0969-2126\(01\)00201-5](https://doi.org/10.1016/S0969-2126(01)00201-5).
- Muñoz-Alía, Miguel Ángel, Rebecca A. Nace, Lianwen Zhang, and Stephen J. Russell. 2021. "Serotypic Evolution of Measles Virus Is Constrained by Multiple Co-Dominant B Cell Epitopes on Its Surface Glycoproteins." *Cell Reports Medicine* 2 (4): 100225. <https://doi.org/10.1016/j.xcrm.2021.100225>.
- Racaniello, V. R. 2013. "Picornaviridae: The Viruses and Their Replication." *Fields Virology*, 1, 453–489.
- Rachlin, Audrey, Jaymin C. Patel, Cara C. Burns, Jaume Jorba, Graham Tallis, Aidan O'Leary, Steven G.F. Wassilak, and John F. Vertefeuille. 2022. "Progress Toward Polio Eradication — Worldwide, January 2020–April 2022." *MMWR. Morbidity and Mortality Weekly Report* 71 (19): 650–55. <https://doi.org/10.15585/mmwr.mm7119a2>.
- Ramaraj, Thiruvarangan, Thomas Angel, Edward A. Dratz, Algirdas J. Jesaitis, and Brendan Mumey. 2012. "Antigen–Antibody Interface Properties: Composition, Residue Interactions, and Features of 53 Non-Redundant Structures." *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1824 (3): 520–32. <https://doi.org/10.1016/j.bbapap.2011.12.007>.
- Robinson, Christopher M., Yao Wang, and Julie K. Pfeiffer. 2017. "Sex-Dependent Intestinal Replication of an Enteric Virus." Edited by Susana López. *Journal of Virology* 91 (7). <https://doi.org/10.1128/JVI.02101-16>.
- Rossmann, Michael G., Edward Arnold, John W. Erickson, Elizabeth A. Frankenberger, James P. Griffith, Hans-Jürgen Hecht, John E. Johnson, et al. 1985. "Structure of a Human Common Cold Virus and Functional Relationship to Other Picornaviruses." *Nature* 317 (6033): 145–53. <https://doi.org/10.1038/317145a0>.
- Schrödinger, LLC. n.d. "The PyMOL Molecular Graphics System, Version 2.0."
- Schymkowitz, Joost, Jesper Borg, Francois Stricher, Robby Nys, Frederic Rousseau, and Luis Serrano. 2005. "The FoldX Web Server: An Online Force Field." *Nucleic Acids Research* 33 (Web Server issue): W382-8. <https://doi.org/10.1093/nar/gki387>.
- Sherry, B, A G Mosser, R J Colonno, and R R Rueckert. 1986. "Use of Monoclonal Antibodies to Identify Four Neutralization Immunogens on a Common Cold Picornavirus, Human Rhinovirus 14." *Journal of Virology* 57 (1): 246–57. <https://doi.org/10.1128/JVI.57.1.246-257.1986>.
- Team, R Core. 2021. "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing, Vienna, Austria.
- Valero-Rello, Ana, and Rafael Sanjuán. 2022. "Enveloped Viruses Show Increased Propensity to Cross-Species Transmission and Zoonosis." *Proceedings of the*

- National Academy of Sciences* 119 (50).
<https://doi.org/10.1073/pnas.2215600119>.
- Vogt, Matthew R, Jianing Fu, Nurgun Kose, Lauren E Williamson, Robin Bombardi, Ian Setliff, Ivelin S Georgiev, et al. 2020. "Human Antibodies Neutralize Enterovirus D68 and Protect against Infection and Paralytic Disease." *Science Immunology* 5 (49). <https://doi.org/10.1126/sciimmunol.aba4902>.
- Wang, Kang, Binyang Zheng, Li Zhang, Lunbiao Cui, Xuan Su, Qian Zhang, Zhenxi Guo, et al. 2020. "Serotype Specific Epitopes Identified by Neutralizing Antibodies Underpin Immunogenic Differences in Enterovirus B." *Nature Communications* 11 (1): 1–12. <https://doi.org/10.1038/s41467-020-18250-w>.
- Wolfram Research, Inc. 2022. "Mathematica." Champaign, Illinois: Wolfram Research, Inc. <https://www.wolfram.com/mathematica>.
- Xiao, Chuan, and Michael G. Rossmann. 2007. "Interpretation of Electron Density with Stereographic Roadmap Projections." *Journal of Structural Biology* 158 (2): 182–87. <https://doi.org/10.1016/j.jsb.2006.10.013>.
- Xu, Longfa, Qingbing Zheng, Shaowei Li, Maozhou He, Yangtao Wu, Yongchao Li, Rui Zhu, et al. 2017. "Atomic Structures of Coxsackievirus A6 and Its Complex with a Neutralizing Antibody." *Nature Communications* 8 (1): 1–12. <https://doi.org/10.1038/s41467-017-00477-9>.
- Yin, Zhichao, Yangtao Wu, Yongchao Li, Lisheng Yang, Wangheng Hou, Z Hong Zhou, Barney S Graham, Tong Cheng, and Shaowei Li. 2019. "Atomic Structures of Enterovirus D68 in Complex with Two Monoclonal Antibodies Define Distinct Mechanisms of Viral Neutralization Qingbing" 4 (1): 124–33. <https://doi.org/10.1038/s41564-018-0275-7>.Atomic.