

RESEARCH ARTICLE

B-cell epitope discovery: The first protein flexibility-based algorithm–Zika virus conserved epitope demonstration

Daniel W. Biner[✉], Jason S. Grosch[✉], Peter J. Ortoleva*

Department of Chemistry, Indiana University, Bloomington, Indiana, United States of America

✉ These authors contributed equally to this work.

* ortoleva@indiana.edu



OPEN ACCESS

Citation: Biner DW, Grosch JS, Ortoleva PJ (2023) B-cell epitope discovery: The first protein flexibility-based algorithm–Zika virus conserved epitope demonstration. PLoS ONE 18(3): e0262321. <https://doi.org/10.1371/journal.pone.0262321>

Editor: Paolo Carloni, Computational Biophysics, GERMANY

Received: April 7, 2021

Accepted: December 22, 2021

Published: March 15, 2023

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0262321>

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CCO](#) public domain dedication.

Data Availability Statement: All MD simulation files are available from the IUScholarWorks Repository (DOI: [10.5967/V844-F290](https://doi.org/10.5967/V844-F290)).

Funding: PJO Grant #: UL1 TR001108 the National Institutes of Health, National Center for Advancing

Abstract

Antibody-antigen interaction—at antigenic local environments called B-cell epitopes—is a prominent mechanism for neutralization of infection. Effective mimicry, and display, of B-cell epitopes is key to vaccine design. Here, a physical approach is evaluated for the discovery of epitopes which evolve slowly over closely related pathogens (conserved epitopes). The approach is 1) protein flexibility-based and 2) demonstrated with clinically relevant enveloped viruses, simulated via molecular dynamics. The approach is validated against 1) seven structurally characterized enveloped virus epitopes which evolved the least (out of thirty-nine enveloped virus-antibody structures), 2) two structurally characterized non-enveloped virus epitopes which evolved slowly (out of eight non-enveloped virus-antibody structures), and 3) eight preexisting epitope and peptide discovery algorithms. Rationale for a new benchmarking scheme is presented. A data-driven epitope clustering algorithm is introduced. The prediction of five Zika virus epitopes (for future exploration on recombinant vaccine technologies) is demonstrated. For the first time, protein flexibility is shown to outperform solvent accessible surface area as an epitope discovery metric.

Introduction

B-cell epitopes are localities of antigens targeted by the humoral immune response, via antibodies and B-cell receptors, to protect the extracellular space (e.g. the blood plasma) [1]. Molecules which effectively mimic B-cell epitopes are vaccines [2–4]. Consequently, structure-based B-cell epitope discovery has emerged as a promising foundational step in rational vaccine design [5,6]. Despite promise, 1) ambiguity surrounding the definition of an epitope, 2) limitations of current performance benchmarking approaches, 3) a scarcity of benchmark datasets, and 4) an overabundance of solvent accessible surface area-based metrics suggest there is ample room for improvement within the field. For example, while some enveloped virus [7] epitopes (like those of the Zika virus [8] (ZIKV)) [9–11] have been studied, a comprehensive, quantitative, uniform, and structure-based epitope analysis of a clinically relevant enveloped virus (like ZIKV) has yet to be explored. Likewise, 1) “cryptic” epitopes hidden within virus structures, [12] 2) pathogen morphological diversity, [13] 3) full pathogen

Translational Sciences, Clinical and Translational Sciences Award <https://ncats.nih.gov/ctsa> NO Indiana University Grant #: CNS-0521433 National Science Foundation <https://www.nsf.gov/> NO This research was supported in part by Lilly Endowment, Inc., through its support for the Indiana University Pervasive Technology Institute. NO This work was supported in part by Shared University Research grants from IBM, Inc., to Indiana University. NO.

Competing interests: The authors have declared that no competing interests exist.

structural dynamics, [14] and 4) possible links between protein flexibility and immunogenicity [15,16] suggest an overreliance on solvent accessible surface area-based epitope discovery metrics [5] may be a major oversight within the field, especially when it comes to highly flexible, clinically relevant antigens (like ZIKV).

Although challenges still exist, recent advancements within the field of all-atom molecular dynamics simulation (MD) have opened up new opportunities to enhance our atomic level understanding of protein flexibility (within the context of the immune response) [17]. Recent MD investigations into multi-million atom VLP systems (like HIV-1 [18], satellite tobacco mosaic virus (STMV) [19], and human papillomavirus (HPV) [16,20]) show quantification of full pathogen structural dynamics 1) is possible and 2) can provide new insights into antibody-antigen interaction physics.

In addition to the limitations (within the field) which have already been noted (e.g. a scarcity of flexibility-based epitope discovery metrics), structure-based methods for the discovery of conserved epitopes (epitopes which evolve the least across closely related pathogens) have yet to be developed. Because highly conserved epitopes (by definition) generalize across many pathogens, highly conserved epitopes are of especially high value as antigenic targets. Ineffectiveness of current vaccine technologies in eliciting antibodies able to bind a diverse set of closely related pathogens has been a challenge in vaccine design [21,22]. Multivalent, live-attenuated vaccines have been introduced to solve the complications mentioned above; however, simultaneous vaccine-based display of related pathogen epitopes has resulted in unbalanced prophylactic protection—suggesting new approaches are needed [23]. One promising, new approach is the presentation (or display) of conserved epitopes on recombinant vaccines [24].

Here, previous studies of *in silico* immunology [16,20] are extended by evaluating an algorithm for conserved epitope discovery, based on protein flexibility—as measured via root-mean-square fluctuation (RMSF) of isolated protein and virus-like particle (VLP) protein residues. The method is developed with flaviviruses simulated via MD. The approach is validated against 1) seven structurally characterized flavivirus epitopes (from thirty-nine flavivirus-antibody structures) with the lowest phylogeny-based evolutionary rates (described by Ashkenazy, H., et al. (2016) [25] as the rate at which a structurally aligned residue changes over a phylogenetic tree of proteins with shared ancestry), 2) two structurally characterized human papillomavirus (HPV) epitopes (from eight HPV-antibody structures) with low phylogeny-based evolutionary rates, and 3) eight preexisting epitope and peptide discovery algorithms [5]. To enhance the epitope dataset (for the prediction of currently uncharacterized ZIKV epitopes), epitopes from seven flaviviruses are structurally aligned. Using 1) a new (data-driven) clustering algorithm, 2) a new epitope organizational model, 3) a new epitope discovery performance benchmarking scheme (which addresses bias in previous methods), and 4) a new epitope discovery benchmark dataset—all presented here for the first time—the prediction of five ZIKV epitopes (which provide starting points for future presentation on recombinant vaccine technologies) is demonstrated. Physical insights identified here 1) supply context for understanding (seemingly contradictory) previous reports on protein flexibility's role in the humoral immune response [15,16,20,26] and 2) shed new light on immunologically relevant distinctions between clinically relevant epitope subsets. Notably, for the first time [26,27], protein flexibility is shown to outperform solvent accessible surface area as an epitope discovery metric.

Results/Discussion

Epitope discovery performance benchmarking and rationale for a new method

Epitope discovery performance benchmarking on a compilation of all ZIKV-aligned epitopes and associations between isolated protein convex hull scores vs. other metrics are shown in

Table 1. Epitope discovery performance benchmarking on a compilation of epitopes.

| Metric | ROCAUC | ± | PRAUC | ± | Rho | p | Type |
|--------|--------|------|-------|------|------|--------|------|
| hvlp | 0.78 | 0.04 | 0.81 | 0.04 | 0.43 | < 0.01 | SASA |
| hmon | 0.75 | 0.04 | 0.77 | 0.04 | — | — | SASA |
| epro | 0.74 | 0.02 | 0.76 | 0.03 | 0.88 | < 0.01 | SASA |
| fmon* | 0.73 | 0.02 | 0.74 | 0.03 | 0.73 | < 0.01 | RMSF |
| dtope | 0.72 | 0.03 | 0.74 | 0.03 | 0.61 | < 0.01 | SASA |
| opia | 0.69 | 0.07 | 0.70 | 0.06 | 0.61 | < 0.01 | SASA |
| bpro | 0.66 | 0.02 | 0.71 | 0.06 | 0.51 | < 0.01 | SASA |
| fmon | 0.66 | 0.02 | 0.71 | 0.03 | 0.48 | < 0.01 | RMSF |
| tfac | 0.63 | 0.05 | 0.65 | 0.04 | 0.59 | < 0.01 | RMSF |
| fvlp | 0.61 | 0.04 | 0.66 | 0.01 | 0.45 | < 0.01 | RMSF |
| iupred | 0.57 | 0.09 | 0.63 | 0.07 | 0.08 | 0.12 | SEQ |
| ppisp | 0.56 | 0.11 | 0.56 | 0.08 | 0.11 | 0.03 | SASA |

Metrics are ordered from the top to bottom in terms of highest ROCAUC and PRAUC product. Spearman rho (r) and p-values are shown for associations between isolated protein convex hull scores (hmon) vs. VLP protein convex hull scores (hvlp), partially isolated protein RMSF (fmon*), DiscoTope scores (dtope), Epitopia scores (opia), ElliPro scores (epro), BEpro scores (bpro), isolated protein RMSF (fmon), VLP protein RMSF (fvlp), temperature factors (tfac), cons-PPISP scores (ppisp), and IUPred scores (iupred). The primary type of structural information utilized for each metric is shown under the heading titled Type (solvent accessible surface area (SASA), protein flexibility (RMSF), and sequence information (SEQ)).

<https://doi.org/10.1371/journal.pone.0262321.t001>

Table 1. The raw ZIKV-aligned epitope dataset is provided in [S1 Dataset](#). For comparison, benchmarking on individual ZIKV-aligned epitopes is shown in [S1 Table](#).

For epitope discovery performance benchmarking, the value of no discrimination [28] is equivalent to the Area Under the Curve (AUC) value associated with performance no better than chance (as discussed in the *Epitope discovery performance benchmarking and rationale for a new method* section of the *Methods*). Because the value of no discrimination assists comparison between Precision-Recall vs. Receiver Operating Characteristic analysis, it is provided for each AUC measurement discussed below.

Comparison of new epitope discovery performance indicator, Precision-Recall Area Under the Curve (PRAUC), with preexisting performance indicator, Receiver Operating Characteristic Area Under the Curve (ROCAUC), shows ROCAUC can overestimate epitope discovery performance when epitopes are analyzed individually, as expected. Over all metrics, the mean ROCAUC was 0.13 ± 0.07 (above the value of no discrimination, 0.50) [28] and the mean PRAUC was 0.07 ± 0.05 (above each value of no discrimination, a mean of 0.05 ± 0.02) when benchmarking was performed on individual epitopes 1) supporting the idea epitopes are much smaller than the entirety of an antigen (comprising only ~5% of the number of antigen residues here, 402) and 2) telling us, despite strong capacity to rank non-epitope residues below epitope residues (a high ROCAUC mean), the average metric performed poorly at pinpointing individual epitopes from other known epitopes (a low PRAUC mean) (S1 Table) [29]. For comparison, over all metrics, the mean ROCAUC was 0.17 ± 0.07 (above the value of no discrimination, 0.50) and mean PRAUC was 0.16 ± 0.07 (above the value of no discrimination, 0.54) when benchmarking was performed on a compilation of epitopes telling us 1) the proportion of epitope residues and non-epitope residues had parity when epitopes were compiled (similar ROCAUC and PRAUC values of no discrimination, 0.50 and 0.54 respectively) and 2) the average metric had a higher capacity to pinpoint a compilation of epitopes (a high PRAUC mean) than individual epitopes ($t(22) = 3.77$ $p = < 0.01$) (Table 1). These results make sense as structural metrics are associated with the entirety of a protein antigen and antigens contain multiple epitopes [30]. These results are also important for interpreting previous epitope

discovery benchmarking attempts [5]. From this point forward, all epitope discovery performance benchmarking discussion will refer to results obtained using the compilation of epitopes benchmarking scheme.

As a note going forward, because a single flavivirus gene product forms the basis for epitope discovery performance benchmarking here—the envelope protein [31], we refer to this protein as the ZIKV (or flavivirus) protein (for more detail, see the *Model preparation and molecular dynamics simulations* section of the *Methods*). Additionally, a ZIKV VLP protein refers to the envelope proteins which comprise the ZIKV VLP model examined here, a hollow protein cage, as shown in [S3 Fig](#).

Epitope discovery performance benchmarking comparison shows solvent accessible surface area-based metrics (layered convex hulls, ElliPro [32], DiscoTope [33], BEpro [34], Epitopia [35], and cons-PPISP [36]) performed both the strongest and weakest of all metrics against all thirty-nine ZIKV-aligned flavivirus epitopes ([Table 1](#)). More specifically, VLP protein convex hull scores performed with one of the strongest capacities to rank epitope residues above non-epitope residues (the highest ROCAUC mean) and to pinpoint epitope residues (one of the highest PRAUC means) likely because 1) flavivirus structures mature as they are secreted from the cell into the extracellular space [37], 2) the mature ZIKV structure was used to model the VLP examined here, [8] and 3) the humoral immune response protects the extracellular space from foreign objects via B-cell antibodies [38]. For comparison, a moderate Spearman correlation and overlapping performance were observed between isolated and VLP protein convex hull scores ($r = 0.43$, $p < 0.01$, $n = 402$) perhaps because 1) the transmembrane protein region protrudes from the isolated protein, enhancing convex hull scores of residues which are also exposed on the VLP [8], or 2) the extracellular space sees a diverse ensemble of flavivirus morphologies (outside the mature cryo-EM reconstruction), as supported by recent studies [12–14]. Overall, these results 1) join growing support for the hypothesis accessibility to antibody binding is one prerequisite for an epitope’s existence [5,33,39] and 2) suggest morphological diversity may challenge our assumptions as to which regions of a pathogen are accessible to antibody binding.

Shifting focus to protein flexibility-base metrics, examination of epitope discovery performance of isolated protein flexibility and partially isolated protein flexibility (incorporating the transmembrane region and two bound M proteins; labeled fmon* in [Table 2](#)) reveals partially isolated protein flexibility performed 1) on par with ElliPro and

Table 2. Conserved epitope discovery performance benchmarking.

| Metric | ROCAUC | ± | PRAUC | ± | Rho | p | Type |
|------------------|--------|------|-------|------|------|--------|------|
| fmon-fvlp | 0.82 | 0.08 | 0.46 | 0.08 | 0.53 | < 0.01 | RMSF |
| fmon | 0.72 | 0.07 | 0.48 | 0.10 | — | — | RMSF |
| fmon* | 0.71 | 0.07 | 0.46 | 0.13 | 0.72 | < 0.01 | RMSF |
| epro | 0.69 | 0.07 | 0.45 | 0.10 | 0.53 | < 0.01 | SASA |
| opia | 0.64 | 0.03 | 0.27 | 0.05 | 0.53 | < 0.01 | SASA |
| hmon | 0.62 | 0.03 | 0.24 | 0.07 | 0.48 | < 0.01 | SASA |
| ppisp | 0.69 | 0.12 | 0.18 | 0.14 | 0.31 | < 0.01 | SASA |
| fvlp | 0.59 | 0.07 | 0.14 | 0.03 | 0.76 | < 0.01 | RMSF |
| dtope | 0.62 | 0.09 | 0.13 | 0.04 | 0.65 | < 0.01 | SASA |
| bpro | 0.58 | 0.05 | 0.13 | 0.05 | 0.58 | < 0.01 | SASA |
| tfac | 0.56 | 0.11 | 0.09 | 0.02 | 0.37 | < 0.01 | RMSF |
| iupred | 0.50 | 0.04 | 0.07 | 0.01 | 0.04 | 0.44 | SEQ |
| hvlp | 0.39 | 0.10 | 0.06 | 0.01 | 0.12 | 0.02 | SASA |

<https://doi.org/10.1371/journal.pone.0262321.t002>

DiscoTope (a PRAUC mean of 0.73 ± 0.02 and ROCAUC mean of 0.74 ± 0.03) and 2) significantly stronger than isolated protein flexibility at ranking ZIKV-aligned epitope residues above all other residues (a higher mean ROCAUC) ($t(8) = 3.88$ $p = < 0.01$). A stronger correlation was observed between partially isolated protein flexibility and isolated protein convex hull scores ($r = 0.73$, $p < 0.01$, $n = 402$) than isolated protein flexibility and isolated protein convex hull scores ($r = 0.48$, $p < 0.01$, $n = 402$) further highlighting differences between isolated and partially isolated protein flexibility. Together, these results signify 1) protein flexibility can perform with relatively high capacity to pinpoint epitopes and 2) protein flexibility metrics can have a range of similarities with solvent accessible surface area depending on protein local environment.

Unlike partially isolated protein flexibility, other flexibility metrics examined here (isolated protein flexibility, VLP protein flexibility, and temperature factors) 1) were moderately correlated with and 2) performed below most solvent accessible surface area-based metrics (against all thirty-nine ZIKV-aligned flavivirus epitopes) (Table 1). Despite relatively weak performance (in comparison with solvent accessible surface area-based metrics and partially isolated protein flexibility), these flexibility-based metrics retained some capacity to pinpoint ZIKV-aligned flavivirus epitope residues (PRAUC means above the value of no discrimination, 0.54). RMSD equilibration plots indicate, despite performance overlap, differences between isolated and VLP protein flexibility were likely associated with steric hindrance between adjacent VLP proteins, a 0.15 \AA RMSD standard deviation (on average) for all VLP proteins (analyzed individually) vs. a 0.37 \AA RMSD standard deviation for the isolated protein (over the last 20 ns; Fig 1). These results reveal 1) isolated protein flexibility was damped upon incorporation into a VLP and 2) pathogen morphological diversity may complicate our understanding of protein flexibility's role within the humoral immune response.

To ensure the ZIKV VLP was equilibrated prior to VLP protein flexibility calculation, the VLP waters exchange rates were quantified. Over the last 20 ns of the simulation, the rate of waters entering and exiting the ZIKV VLP were $122,973 \pm 1,344$ and $123,105 \pm 1,328$ per ns ($\sim 5\%$ of the total waters in the capsid), respectively. These rates are approximately two orders of magnitude higher than those observed for a simulation of the mature HIV capsid [18], highlighting the permeability of the ZIKV VLP. Relatively high permeability of the ZIKV VLP (compare to the mature HIV capsid) could be related to 1) differences between envelope proteins and capsid proteins and 2) the absence of membrane and membrane-associated protein components within the ZIKV VLP. Overall, these results show the internal pressure of the VLP was equilibrated prior to RMSF calculation.

Although protein flexibility-based metrics (which were most distinct from solvent accessible surface area-based metrics here) failed to outperform solvent accessible surface area-based metrics (in terms of epitope discovery performance against all ZIKV-aligned epitopes), 1) very strong correlations between isolated protein flexibility of seven structurally related flaviviruses (a mean Spearman r of 0.80 ± 0.04 and $p < 0.01$ over flavivirus structures; S1 Fig) and 2) seemingly contradictory conclusions from previous studies (examining flexibility's role in the humoral immune response) [15,16,20] suggest an external factor such as antibody flexibility (as hypothesized here) may influence the relationship between protein flexibility and epitopes [40,41].

To understand how consideration of protein flexibility differences across antibodies [40,41] could reveal a stronger association between protein flexibility-based metrics (which were most distinct from solvent accessible surface area-based metrics here) and epitopes, one must examine the combined influence of antibody flexibility and epitope flexibility on antibody-antigen binding affinity—as epitopes are, by definition, antigenic local environments bound by antibodies. First, one way of obtaining high antibody-antigen binding affinity is through

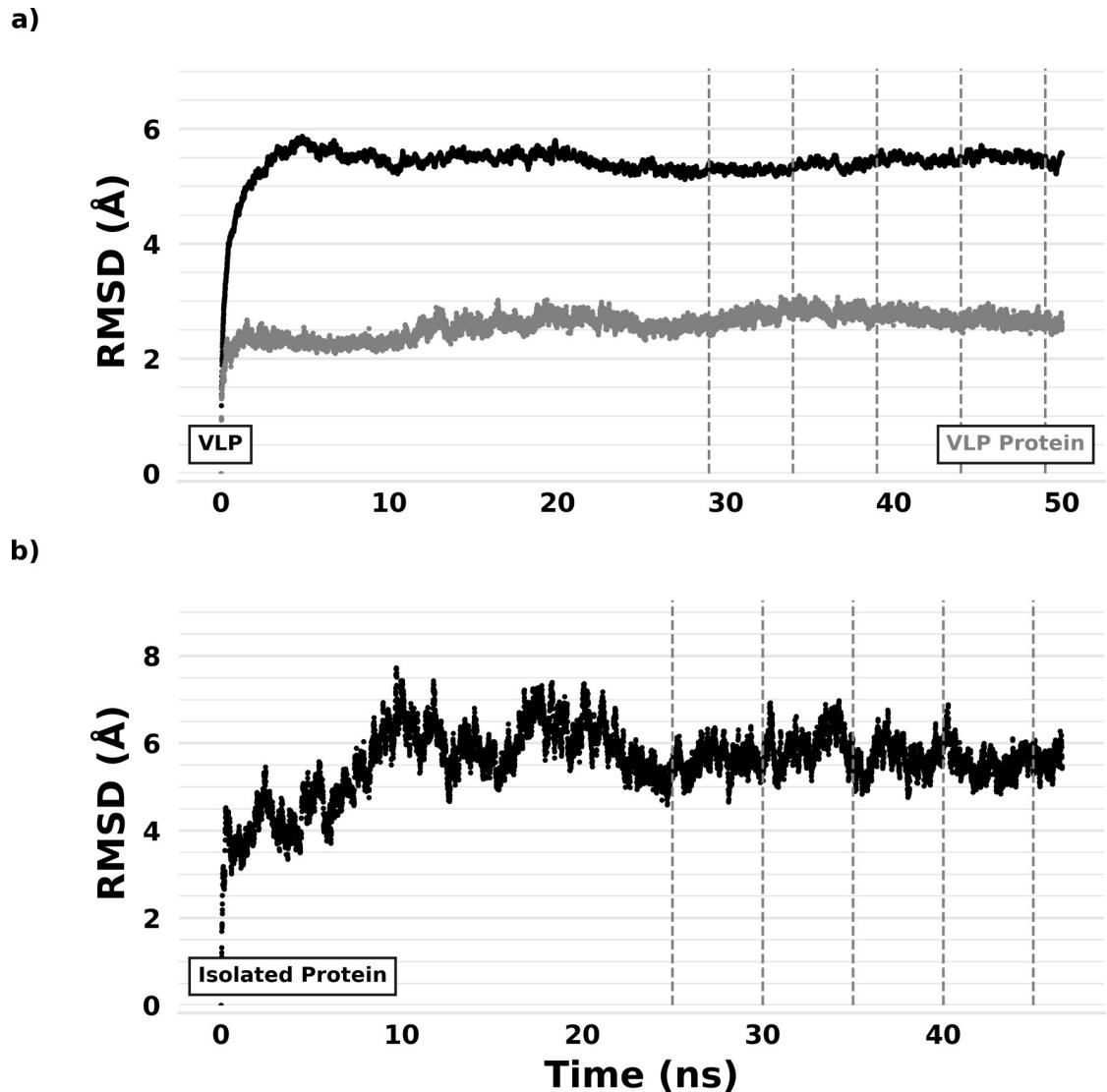


Fig 1. ZIKV structural equilibration at 300 K. a) RMSD from starting configuration for a) the ZIKV VLP (black trace), a single VLP protein (gray trace), and b) the isolated ZIKV protein are shown. Vertical lines represent trajectory bounds used for RMSF calculation.

<https://doi.org/10.1371/journal.pone.0262321.g001>

maximization of contact between an antibody and an antigen. Second, to maximize contact between an antibody and an antigen, either the antibody or the antigen (more specifically the partner epitope on the antigen) must have flexibility. For example, when an antibody and an epitope are both rigid, contact is likely limited via steric hinderance, and, when an antibody and an epitope are both flexible, contact is likely limited via a high conformational entropy barrier. Using this logic, the relationship between protein flexibility and epitopes should be enhanced for epitopes bound by antibodies with relatively high rigidity compared with all antibodies. Broadly neutralizing antibodies, for example, have recently been shown to affinity mature through a process which rigidifies the antibody, while, at the same time, enhancing antibody-antigen binding specificity [40,41] suggesting affinity matured, broad-spectrum antibodies may preferentially target flexible regions of pathogens.

Conserved epitope discovery performance benchmarking and the flexibility-based model

In light of results here and previous suggestions of flexibility's influence on broadly neutralizing antibody affinity maturation [40,41], epitope discovery performance benchmarking on a compilation of the top seven most conserved flavivirus epitopes (an epitope subset which 1) evolves slowly over structurally related pathogens and 2) is likely targeted by broad-spectrum antibodies) was examined (Table 2).

Metrics are ordered from the top to bottom in terms of highest ROCAUC and PRAUC product calculated against the top seven most conserved ZIKV-aligned epitopes. Spearman rho (r) and p-values are shown for associations between isolated protein RMSF (fmon) vs. (fmon-fvlp) the linear difference of isolated and VLP RMSF, partially isolated protein RMSF (fmon*), ElliPro scores (epro), Epitopia scores (opia), isolated protein convex hull scores (hmon), cons-PPISP scores (ppisp), BEpro scores (bpro), (hmon-hvlp) the linear difference of isolated and VLP convex hull scores, VLP protein RMSF (fvlp), DiscoTope scores (dtope), IUPred scores (iupred), temperature factors (tfac), and VLP protein convex hull scores (hvlp). The primary type of structural information utilized for each metric is shown under the heading titled Type (solvent accessible surface area (SASA), protein flexibility (RMSF), and sequence information (SEQ)).

Epitope discovery performance benchmarking comparison shows isolated protein flexibility performed above all other lone flexibility-based metrics (VLP protein RMSF and temperature factors) and all solvent accessible surface area-based metrics (layered convex hulls, ElliPro, DiscoTope, BEpro, Epitopia, and cons-PPISP) against the top seven most conserved flavivirus epitopes—the first report of a flexibility-based epitope discovery metric outperforming solvent accessible surface area-based metrics (Table 2) [5]. Examination of epitope discovery performance against the top two most conserved flavivirus epitopes shows isolated protein flexibility also maintained almost perfect capacity to 1) pinpoint residues from the top two conserved epitopes (a PRAUC mean of 0.98 ± 0.02 ; a value of no discrimination of 0.04) and 2) rank residues from the top two conserved epitopes above all other residues (a ROCAUC mean of 0.86 ± 0.13 ; a value of no discrimination of 0.50). Closer inspection reveals 1) isolated protein flexibility of seven structurally related flaviviruses retained comparable capacity to pinpoint residues from the top seven conserved epitopes (a mean PRAUC over flavivirus structures of 0.47 ± 0.03 ; a mean value of no discrimination of 0.08; S2 Table), 2) isolated protein flexibility from a ZIKV protein simulation without disulfides performed with a slightly lower capacity (despite overlapping capacity) to rank residues from the top seven conserved epitopes above all other residues (ROCAUC mean of 0.68 ± 0.07 ; a value of no discrimination of 0.50) than is isolated protein flexibility with disulfides, supporting separate studies suggesting disulfides play a structural role in flavivirus immunogenicity [42–46], 3) IUPred, a sequence-based measure of intrinsically disordered protein regions (which one might associate with protein flexibility) [47,48], performed no better than chance against the top seven conserved epitopes (ROCAUC and PRAUC means near their respective values of no discrimination, 0.50 and 0.08), and 4) Elli-Pro had overlapping performance with isolated protein flexibility in terms of capacity to pinpoint residues from the top seven conserved epitopes (despite a lower PRAUC mean).

Strong conserved epitope discovery performance with isolated protein flexibility comes as little surprise, as 1) conserved protein regions are necessary for successful flavivirus infection [49] and 2) protein flexibility (linked with major morphological shifts) is also necessary for successful flavivirus infection. For example, conformational dynamics are thought to facilitate flavivirus endosomal escape, and antibodies which block or alter these dynamics have been shown to neutralize infection [50].

Despite performance overlap and compared with isolated protein flexibility alone, VLP protein flexibility-based metrics (VLP protein RMSF and temperature factors) performed weakly at pinpointing conserved epitope residues (PRAUC means near the value of no discrimination, 0.08). However, when combined, the linear difference of isolated and VLP protein flexibility (using weights which minimized the root-mean-square difference between the two flexibility profiles) performed with stronger capacity to rank conserved epitope residues above all other residues (a distinctly higher ROCAUC distribution) compared with ElliPro ($t(8) = 2.43$, $p = 0.04$) (Table 2).

Examination of differences between isolated and partially isolated ZIKV protein flexibility reveals the two metrics performed similarly. However, unlike isolated protein flexibility, linear combination of partially isolated and VLP protein flexibility (using weights optimized via test and train dataset splits) resulted in conserved epitope discovery performance indistinguishable from ElliPro (a PRAUC mean of 0.49 ± 0.06 and ROCAUC mean of 0.75 ± 0.06). These results imply rotational and translation freedom of the fully isolated ZIKV protein, dampened upon assembly into the mature ZIKV virus structure, was one facet of the linearly combined, flexibility model's capacity to rank conserved epitope residues above all other residues. Like previously mentioned, broadly neutralizing flavivirus antibodies have been shown to mute or disrupt the dynamics of mature flavivirus structures, obstructing conformational motions required for endosomal escape and successful infection [50].

Visualization of the linearly combined, flexibility model on the ZIKV cryo-EM reconstruction [8] highlights a radial spiral in metric amplitude originating at the five-fold rotational icosahedral symmetry axes (Fig 2), a local environment thought to hold stress [51] of icosahedral virus curvature. Visual analysis indicates conserved epitopes may be important for 1) stabilizing the mature flavivirus morphology before endosomal escape and 2) destabilizing the mature flavivirus morphology for endosomal escape, as part of a morphological transition state. This makes sense as isolated ZIKV protein flexibility (isolated from other envelope proteins, M proteins, and the transmembrane region) is likely, at least somewhat, indicative of the flexibility sampled during flavivirus morphological transitions (e.g. immature to mature or mature to endosomal escape morphologies).

Inspection of the weights which optimized performance for the linearly combined, flexibility model using 1) test and train dataset splits and 2) root-mean-square difference minimization between the two individual flexibility metrics shows the two linear combination methods produced similar results. For example, weights which minimized the root-mean-square difference between isolated and VLP protein flexibility (0.378 and 0.622, respectively) were almost

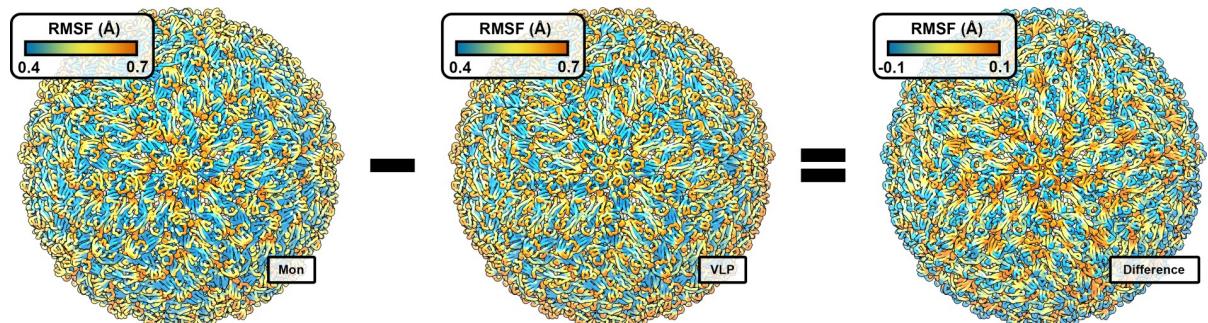


Fig 2. Visualizing the linearly combined, flexibility model via the mature ZIKV cryo-EM reconstruction. From left to right: Weighted isolated protein flexibility, weighted VLP protein flexibility, and the linearly combined, flexibility model are depicted over the mature ZIKV cryo-EM reconstruction (PDB ID: 5IRE) [8] (scales shown). VLP structures are oriented with the five-fold rotational icosahedral symmetry axis coming out of the figure plane.

<https://doi.org/10.1371/journal.pone.0262321.g002>

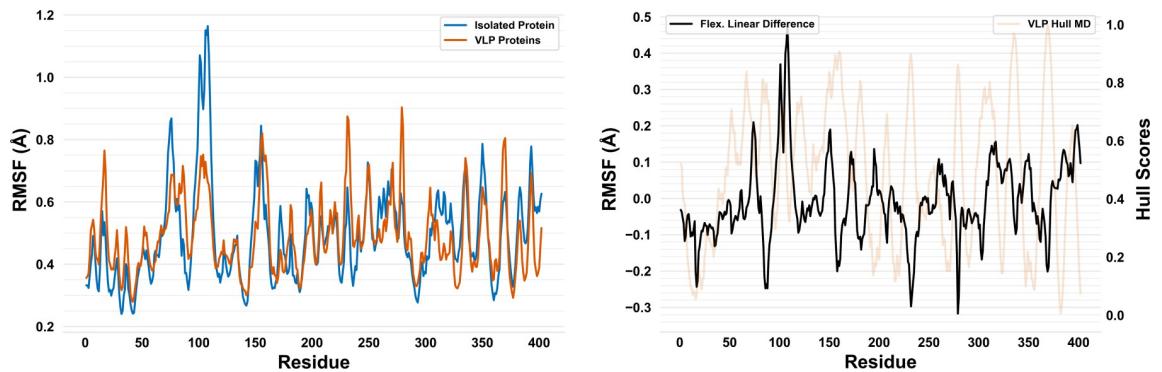


Fig 3. Visualizing the linearly combined, flexibility model via RMSF profiles. On the left, plots of the weighted isolated (blue trace) and VLP (red trace) protein flexibility metrics vs. residue number are overlaid. On the right, a plot of the linearly combined, flexibility model (black trace; using weights which minimized the root-mean-square difference between the two flexibility-based metrics) vs. residue number is shown. For comparison, on the right, a plot of VLP hull scores (red trace) is shown behind a plot of the linearly combined, flexibility model.

<https://doi.org/10.1371/journal.pone.0262321.g003>

identical to weights which optimized performance for the linear combination (0.325 and 0.675, respectively). These results suggest the linearly combined, flexibility model was representative of the relative flexibility difference between the isolated ZIKV protein and ZIKV VLP proteins after baseline flexibility differences were accounted for (Fig 3).

For comparison, solvent accessible surface area-based metrics performed with relatively weak capacity to rank conserved flavivirus epitope residues above all other residues (relatively low ROCAUC means) and to pinpoint conserved epitope residues (relatively low PRAUC means) (Table 2). In fact, although isolated ZIKV protein convex hull scores had some capacity to rank conserved epitope residues above all other residues (a ROCAUC mean ~0.10 above the 0.50 value of no discrimination), ZIKV VLP protein convex hull scores ranked all other residues above conserved flavivirus epitope residues (a ROCAUC mean ~0.10 below the 0.50 value of no discrimination). Additionally, in contrast with protein flexibility, the linear difference of ZIKV isolated and VLP protein convex hull scores (using weights optimized via test and train dataset splits) performed with 1) a higher capacity to rank conserved flavivirus epitope residues above all other ZIKV residues (a higher ROCAUC mean, 0.71 ± 0.04 ; a value of no discrimination of 0.50) but 2) a much weaker capacity to pinpoint conserved flavivirus epitope residues (a lower PRAUC mean, 0.14 ± 0.02 ; a value of no discrimination of 0.08) compared with isolated ZIKV protein convex hull scores alone. Interestingly, cons-PPISP, a metric calibrated to typical protein-protein interfaces, performed with a stronger capacity to pinpoint epitope residues against the top seven most conserved flavivirus epitopes (a PRAUC mean ~0.10 higher than the 0.08 value of no discrimination) than against all thirty-nine flavivirus epitopes (a PRAUC mean around the 0.54 value of no discrimination). These results signify 1) protein-protein interaction sites play a possibly limited role in the existence of conserved epitopes, 2) differences between ZIKV isolated and VLP protein convex hull scores were accentuated via performance benchmarking against conserved flavivirus epitopes, and, relatedly, 3) it is evolutionarily advantageous to partially or transiently conceal conserved epitopes for immune evasion [49]. Interestingly, despite potential partial or transient concealment, engineered presentation of conserved epitopes on recombinant vaccines can still induce broadly reactive humoral immune responses [24]—highlighting the potential significance of 1) conserved epitope discovery methods in vaccine design and 2) engineered presentation of conserved epitopes on vaccine, protein scaffolds.

Shifting focus to the generalizability of the protein flexibility-based epitope discovery method presented here, performance reliability shows isolated flavivirus protein flexibility pinpoints conserved flavivirus epitope residues well across seven structures (a mean PRAUC of 0.47 ± 0.03 over flavivirus structures; [S2 Table](#)). Interestingly, isolated flavivirus protein flexibility appears to be conserved across flavivirus envelope proteins (a mean Spearman r of 0.80 ± 0.04 and $p < 0.01$ over flavivirus structures; [S1 Fig](#))—a finding which may account for strong flexibility-based conserved epitope discovery performance here. The flavivirus NS3 protein was also recently shown to have a flexibility profile conserved across flaviviruses [\[52\]](#) despite a lack of sequence similarity.

Results here, along with the discovery of an ancestral protein linked with class II fusion proteins (e.g. the flavivirus envelope protein) [\[53\]](#), suggest the flexibility-based method presented here likely generalizes across pathogens with class II fusion proteins. Previous observations of protein flexibility's role in broadly neutralizing HIV antibody maturation [\[41\]](#) and the first principles nature of protein flexibility-based (as discussed in *Epitope discovery performance benchmarking and rationale for a new method* section of the *Results*) suggest the protein flexibility-based epitope discovery method could generalize beyond pathogens with class II fusion proteins.

To better understand the generality of flexibility-based method presented here, protein flexibility's influence on conserved human papillomavirus [\[16\]](#) (HPV) epitopes was examined. Benchmarking was performed against two highly conserved (mean residue evolutionary rate of 1.52) [\[25\]](#) HPV epitopes (PDB IDs: 7CN2 [\[54\]](#), 3J8Z [\[55\]](#)) out of eight structurally characterized HPV epitopes (evolutionary rate range of 1.18–2.47). Like flaviviruses, linear combination of HPV isolated and VLP protein flexibility (using weights which minimized the root-mean-square difference between the two flexibility profiles) performed with the strongest capacity (of all metrics examined) to pinpoint conserved HPV epitope residues (a PRAUC mean 0.21 above the value of no discrimination, 0.07) ([Table 3](#)). For comparison, ElliPro and IUPred performed with some capacity to pinpoint conserved HPV epitopes (PRAUC means ~0.10 above the value of no discrimination, 0.07), while DiscoTope, BEpro, and cons-PPISP perform no better than chance at pinpointing conserved HPV epitopes (PRAUC means around the value of no discrimination, 0.07). These results show the physics-based protein flexibility model identified here does indeed generalize beyond clinically relevant enveloped viruses to clinically relevant non-enveloped viruses.

Conserved epitope organization and epitope discovery demonstration

Of the seven highly conserved epitopes, 14% were associated with ZIKV, 0% were associated with dengue virus serotype 1, 14% were associated with dengue virus serotype 2, 14% were associated with dengue virus serotype 3, 29% were associated with dengue virus serotype 4,

Table 3. Generalizability of the flexibility-based, conserved epitope model.

| | fmon-fvlp | epro | iupred | dtope | bpro | ppisp |
|--------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| ROCAUC | 0.81 ± 0.06 | 0.82 ± 0.06 | 0.73 ± 0.06 | 0.67 ± 0.07 | 0.63 ± 0.08 | 0.44 ± 0.07 |
| PRAUC | 0.28 ± 0.06 | 0.18 ± 0.07 | 0.15 ± 0.07 | 0.09 ± 0.02 | 0.09 ± 0.02 | 0.06 ± 0.01 |
| Type | RMSF | SASA | SEQ | SASA | SASA | SASA |

ROCAUC and PRAUC values calculated against two highly conserved HPV epitopes [\[54,55\]](#) are shown for the linear difference of isolated and VLP RMSF (fmon-fvlp), ElliPro scores (epro), IUPred scores (iupred), DiscoTope scores (dtope), BEpro scores (bpro), cons-PPISP scores (ppisp). The primary type of structural information utilized for each metric is shown under the heading titled Type (solvent accessible surface area (SASA), protein flexibility (RMSF), and sequence information (SEQ)).

<https://doi.org/10.1371/journal.pone.0262321.t003>

Table 4. Flavivirus epitope mean evolutionary rates.

| PDB | ER | Flav. | PDB | ER | Flav. | PDB | ER | Flav. | PDB | ER | Flav. |
|------|-------|-------|------|-------|-------|------|------|-------|------|------|-------|
| 3I50 | -0.95 | W | 4UTB | -0.13 | D2 | 3UZV | 0.42 | D2 | 5LBS | 0.69 | Z |
| 5JHL | -0.93 | Z | 4UTA | -0.1 | D1 | 3UYP | 0.45 | D4 | 1ZTX | 0.73 | W |
| 3IXX | -0.91 | W | 5KVD | -0.04 | Z | 2R69 | 0.48 | D2 | 5VIC | 0.75 | D1 |
| 4BZ1 | -0.78 | D4 | 4UT9 | 0.03 | D2 | 5GZO | 0.52 | Z | 3IYW | 0.77 | W |
| 4AM0 | -0.7 | D4 | 3UZQ | 0.33 | D1 | 5KVE | 0.54 | Z | 5YWP | 0.78 | J |
| 4ALA | -0.63 | D3 | 4FFZ | 0.34 | D1 | 3UAJ | 0.54 | D4 | 5VIG | 0.86 | Z |
| 3IXY | -0.62 | D2 | 4FFY | 0.36 | D1 | 4L5F | 0.57 | D1 | 5Y0A | 0.86 | Z |
| 4AL8 | -0.51 | D1 | 5AAW | 0.36 | D4 | 5KVG | 0.62 | Z | 5YWO | 0.86 | J |
| 5LCV | -0.36 | Z | 3UZE | 0.38 | D3 | 5KVF | 0.63 | Z | 5GZN | 1.41 | Z |
| 4UT6 | -0.15 | D2 | 5H37 | 0.41 | Z | 5UHY | 0.64 | Z | — | — | — |

Flavivirus-antibody structures from which conserved epitopes were identified are ordered from the top to bottom and left to right in terms of mean residue evolutionary rate (ER) [25]. The PDB ID and flavivirus of the associated structures are shown (ZIKV (Z), dengue virus serotype 1–4 (D1–4), West Nile virus (W), Japanese encephalitis virus (J)).

<https://doi.org/10.1371/journal.pone.0262321.t004>

29% were associated with West Nile virus, and 0% were associated with Japanese encephalitis virus (Table 4).

Examination of ZIKV-aligned epitope organization shows conserved flavivirus epitopes fall on the lower end of the epitope distribution, in terms of epitope residue quantity and epitope discontinuity. Conserved flavivirus epitopes had a mean of 8 ± 4 total residues and 3 ± 1 subsequences with 4 ± 3 residues. All thirty-nine ZIKV-aligned flavivirus epitopes had a mean of 18 ± 9 total residues and 5 ± 3 subsequences with 4 ± 4 residues. For comparison, a previous analysis of seventy-six antibody-antigen structures [56] found epitopes had a mean of 16 ± 3 total residues and 5 ± 2 subsequences with 2 ± 2 residues. These results further support the idea partial or transient concealment of conserved epitopes is evolutionarily advantageous [49], emphasizing the potential value of 1) conserved epitopes and 2) conserved epitope discovery methods in vaccine design.

Prediction of five ZIKV epitopes with 8 ± 7 total residues and 1 ± 0 subsequence of 5 ± 4 residues was demonstrated (Table 5). Visual inspection of predicted epitopes on the ZIKV cryo-EM reconstruction shows a spiral pattern focused around the five-fold rotational icosahedral symmetry axes (Fig 4). High continuity of predicted epitopes is likely a result of 1) the structure examined, 2) the parameters set within the epitope residue clustering algorithm (described below), and 3) the residue profile of the metric used for epitope prediction. The isolated and VLP protein flexibility linear difference threshold which minimized the root-mean-square difference between known conserved epitope organization and predicted conserved epitope organization was 0.10 \AA (Fig 3). The highest performing epitope clustering residue-residue distance cutoff which minimized the root-mean-square difference between known conserved epitope organization and predicted conserved epitope organization was 6 \AA .

Table 5. Predicted conserved epitopes.

| Predicted Epitope | ER | Predicted Epitope | ER | Predicted Epitope | ER |
|--|------|--|------|---|------|
| [147, 148, 149, 150, 151, 152]; 'QHSGMT' | -.94 | [195, 196]; 'GL' | 0.52 | [72, 73, 74, 75, 76, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112]; 'SRCPTVDRGWNGCGLFGKGS' | 1.74 |
| [171, 172]; 'PN' | 0.46 | [311, 312, 313, 314, 315, 316, 317, 318, 331]; 'AFTFTKIPQ' | 0.57 | | |

Predicted ZIKV epitopes are shown and ordered according to mean residue evolutionary rate (ER) [25].

<https://doi.org/10.1371/journal.pone.0262321.t005>

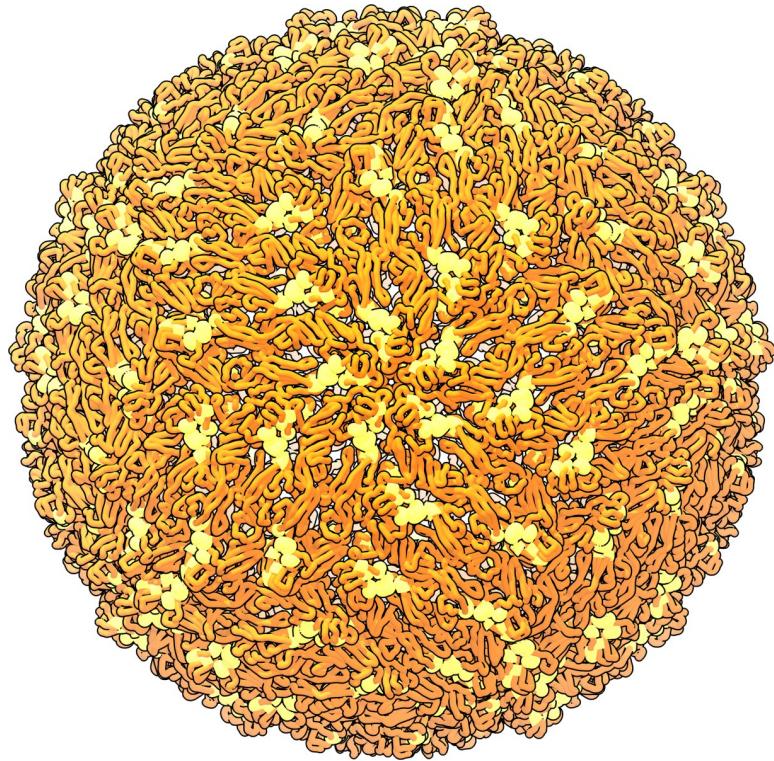


Fig 4. Conserved epitope prediction demonstration. Predicted conserved epitopes are represented in yellow on the ZIKV cryo-EM reconstruction backbone (red) (PDB ID: 5IRE) [8]. The cryo-EM reconstruction is shown with the five-fold rotational icosahedral symmetry axis coming out of the figure plane.

<https://doi.org/10.1371/journal.pone.0262321.g004>

Comparison of the similarity [57] between full, predicted epitopes (clustered from predicted epitope residues) and structurally characterized epitopes shows several epitope predictions overlap with structurally characterized epitopes which were (not necessarily within the top seven conserved epitopes (as measured via ConSurf) [25] but still) bound by cross-reactive, neutralizing antibodies. Predicted epitopes with the highest similarity to structurally characterized epitopes were: [311, 312, 313, 314, 315, 316, 317, 318, 331] and [72, 73, 74, 75, 76, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112]. Predicted epitope [311, 312, 313, 314, 315, 316, 317, 318, 331] showed 50%, 47%, and 45% similarity with epitopes bound by antibody 4E11 (PDB ID: 3UZE, 3UZQ, and 3UYP respectively). Antibody 4E11 cross-neutralizes all four dengue serotypes [58]. Predicted epitope [72, 73, 74, 75, 76, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112] showed 36% similarity with an epitope bound by antibody 2A10G6 (PDB ID: 5JHL). Antibody 2A10G6 cross-neutralizes both ZIKV and dengue virus [59]. Predicted epitope [72, 73, 74, 75, 76, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112] also overlaps with a structurally uncharacterized epitope bound by antibody 1C19 –which also cross-neutralizes all four dengue serotypes [60].

Methods

Model preparation and molecular dynamics simulations

ZIKV chains A, B, C, and E were extracted from a 3.80 Å resolution, cryogenic electron microscopy (cryo-EM) structure (Protein Databank (PDB) ID: 5IRE) [8]. Protein coordinates were run through Pulchra [61] to reconstruct missing and sterically clashing atoms. Protein

coordinates from chain A (residues 1–402) were used to construct a soluble, isolated protein simulation starting structure. An analogous process was applied to construct soluble, isolated protein simulation starting structures for six related flaviviruses (PDB: ID 4CCT (dengue virus serotype 1) [62], 3J27 (dengue 2) [63], 3J6S (dengue 3) [64], 4CBF (dengue 4) [65], 2HG0 (West Nile virus) [66], and 5WSN (Japanese encephalitis virus) [67]). For comparison, protein coordinates from ZIKV chain A (residues 1–502) and two chain Bs (touching chain A after application of the associated symmetry matrix) were used to construct a partially isolated ZIKV protein simulation starting structure (with transmembrane region and associated M proteins). Protein coordinates from ZIKV chains A, C, and E (residues 1–402) and an associated symmetry matrix (taken from PDB ID: 5IRE) were used to build a soluble ZIKV VLP simulation starting structure.

Using GROMACS 5.0.4 [68], explicit-solvent (TIP3P) [69] MD simulations were run on the flavivirus proteins and ZIKV VLP to a local structural equilibrium on the Indiana University Big Red II and III supercomputers (Fig 1). Attainment of local structural equilibrium was assessed as a steady time-average in backbone atom root-mean-square deviation (RMSD) relative to the starting configuration [20]. The CHARMM27 [70] forcefield was used with a 1 fs time step and periodic boundary conditions. The LINCS [71] algorithm was used to fix hydrogen bond lengths and the distance from protein to the edge of the cubic box was set to 1.5 nm. Steepest decent energy minimization was employed before simulation runs.

Flavivirus proteins were simulated in 142 mM NaCl (physiological) water baths. Simulation temperatures were coupled to 300 K (room temperature) using the Nosé-Hoover [72] thermostat. The coupling time constant was set to 0.5 ps, and ten Nosé-Hoover chains were used. The pressure was coupled to 1 bar using the Parrinello-Rahman [73] barostat. The coupling time constant and compressibility were set to 1 ps and 4.5E-5 bar. Short-range electrostatic cutoffs were set to 1.4 nm; long-range electrostatics were managed using the reaction-field [74] method. Epsilon-r (the relative dielectric constant) and epsilon-rf (the dielectric constant beyond a 1.4 nm reaction sphere cutoff) values were set to 1 (protein) and 78 (water).

Six cysteine residues pairs (([3, 30], [60, 121], [74, 105], [92, 116], [190, 291], and [308, 339]) which had sulfurs positioned 6 Å or less apart (overall mean sulfur separation distance of 3.5 ± 1.2 Å) on the ZIKV starting structures were simulated as disulfide bonds [8]. Disulfides, which aligned with those simulated for ZIKV, were also simulated for dengue virus serotype 1–4, West Nile virus, and Japanese encephalitis virus isolated protein starting structures. For comparison, the isolated ZIKV protein was also simulated without disulfides.

For simulations, the protonation state of titratable residues was chosen using a pH value of 8 (the pH associated with PDB ID: 5IRE) [8]. To identify biologically relevant pKa values for titratable flavivirus protein residues, a four-step procedure was performed as follows:

Step 1. Initial pKa guesses were estimated with PROPKA3.1 [75] using the starting model.

Step 2. Protonation states were assigned to titratable residues and a simulation was started.

Step 3. pKa values were predicted again using simulated protein coordinates averaged over an early portion of the trajectory. The simulation was continued if the pKa predictions matched pKas chosen in Step 1. Steps 2–3 were repeated with the new pKa values if predictions differed from pKas chosen in Step 1.

Step 4. pKa values were predicted again using simulated protein coordinates averaged over an equilibrated portion of the trajectory. Step 2–4 were repeated with the new pKa values if predictions differed from pKas found in Step 3.

An analogous process was performed for HPV simulation. HPV L1 protein chain A (PDB ID: 5KEP) [76] was used to construct the isolated HPV protein model. HPV L1 protein chains A-F (PDB ID: 3J6R) [77] were used to construct the HPV VLP model. No protein truncation was performed, and no disulfides were simulated.

Protein flexibility

Protein flexibility, measured as root-mean-square fluctuation (RMSF), was considered as an epitope discovery metric. Flavivirus and HPV backbone atom coordinate RMSF were calculated over structurally equilibrated 5 ns MD trajectory segments using GROMACS [68] as follows:

$$RMSF = \sqrt{\frac{1}{N} \sum_{n=1}^N \frac{1}{T} \sum_{t=1}^T |\vec{r}_n(t) - \langle \vec{r}_n \rangle|^2} \quad (1)$$

where $\vec{r}_n(t)$ is the position of backbone atom n of a residue with N backbone atoms at time t and $\langle \vec{r}_n \rangle$ is the time average position of atom n and T is the total number of discrete time steps. N, C α and C atoms were included as backbone atoms. Rotation and translation of the isolated flavivirus proteins, the soluble portion (residues 1–402) of the partially isolated ZIKV protein, and each ZIKV VLP protein were subtracted before RMSF calculation. For the VLP, mean RMSF amplitudes were calculated for each residue over all proteins to enhance statistical significance, as all proteins which comprise the VLP are quasi-equivalent [78]. An analogous process was performed to quantify HPV RMSF.

Layered convex hull depths

Solvent accessible surface area of the ZIKV isolated protein and VLP proteins were analyzed using a series of layered convex hulls. A convex hull is the smallest convex polygon which encloses all of the points within a dataset [79]. Each protein residue, (1–501) from chain A, C, and E, was reduced to its center of mass to construct three isolated ZIKV protein models for isolated protein convex hull [80] analysis (S2 Fig). Each residue (for the soluble portion (residues 1–402) of chain A, C, and E) was reduced to its center of mass and run through a symmetry matrix taken from PDB ID: 5IRE to construct a soluble ZIKV VLP model for convex hull analysis (S3 Fig).

Convex hull scores were calculated using an extension of a method described by Zheng, W., et al. (2015) [81] as follows:

Step 1. The SciPy [82] ConvexHull module was run on all model coordinates.

Step 2. Model coordinates which comprised the convex hull were given a score and then removed from the remaining model coordinates.

Step 3. Steps 1–2 were repeated until no coordinates remained.

Step 4. Residues in the first hull were given a score of one. Residues in the last hull were given a score of zero. Residues in the remaining hulls were given an evenly spaced score between zero and one based on the associated iteration of Steps 1–2. The mean of the scores was calculated for each residue.

Preexisting structural epitope discovery metrics

ElliPro [32], DiscoTope [33], BEpro [34], Epitopia [35], and cons-PPISP [36] structure-based epitope and peptide discovery algorithms were run using default settings on ZIKV chain A, C, and E individually (residues 1–501), and the mean of the results was calculated for each

residue. ElliPro, DiscoTope, BEpro, and cons-PPISP structure-based epitope and peptide discovery algorithms were run using default settings on HPV chain A (the Epitopia server was unavailable at the time of HPV benchmarking). ZIKV temperature factors of chain A, C, and E were obtained from PDB ID: 5IRE [8], the square root was taken for each value, and the mean was calculated for each residue. Sequence-based intrinsically disordered protein region prediction algorithm, IUPred2A [47,48], was run using default settings. Spearman correlations between metrics were calculated using the SciPy [82] spearmanr module. T-tests were performed using the SciPy ttest_ind module (with equal_var set to False).

Linear combinations of metric pairs

Linear combinations of select pairs of flavivirus structural metrics were calculated using two methods.

The first linear combination method is an extension of a method described by Kringelum, J. V., et al. (2012) [83] and was performed as follows:

For a pair of metrics, difference linear combinations (DLC) were calculated as

$$DLC(r, \alpha) = \alpha FS(r) - (1 - \alpha) SS(r) \quad (2)$$

where FS(r) represents the residue amplitude of the first partner of the pair, SS(r) represents the residue amplitude of the second partner of the pair, and α represents a weight from a grid of weights ranging from zero to one in increments of 0.005.

The second linear combination method was performed using weights which minimized the root-mean-square difference between the individual metrics (as discussed in the *Conserved epitope discovery performance benchmarking and the flexibility-based model* section of the *Results*).

Structurally characterized, ZIKV-aligned flavivirus epitopes and HPV epitopes

Structurally characterized flavivirus epitopes were quantified using an extension of a method described by Stave, J. W. and K. Lindpaintner (2013) [6] and aligned to the ZIKV structure as follows:

Step 1. All available ZIKV, dengue virus, West Nile virus, and Japanese encephalitis virus envelope protein-antibody structures [12,50,58,59,84–101] were compiled from the National Institute of Allergy and Infectious Diseases Immune Epitope Database [102] (IEDB) and the PDB.

Step 2. The pairwise distance between each envelope protein atom and each antibody atom was calculated for each antibody within each structure.

Step 3. Envelope protein residues which had at least one atom within 4 Å from an antibody atom were identified and compiled for each structure, providing one epitope for each structure.

Step 4. Epitope residues were renumbered according to a structural sequence alignment between ZIKV and related flaviviruses reported by Kostyuchenko, V. A., et al. (2016) [8].

No envelope protein atoms were found within 4 Å of an antibody for PDB ID: 4C2I [90].

For conserved epitope quantification: ZIKV residue evolutionary rates were estimated using the cryo-EM chain A with ConSurf [25] (default settings) and the mean amplitude was calculated over each epitope. The sets of two and seven epitopes with the lowest mean residue evolutionary rates [12,59,86,88] were compiled (Table 4).

An analogous process was performed for HPV epitope quantification. HPV epitopes were aligned to the sequence of L1 protein chain A (PDB ID: 5KEP). Two structurally characterized epitopes with low mean residue evolutionary rates (PDB IDs: 7CN2 [54], 3J8Z [55]) out of eight epitopes were compiled for benchmarking.

Epitope discovery performance benchmarking and rationale for a new method

Because of the high structural similarity between ZIKV and other flaviviruses [8], epitopes of other flaviviruses were considered as benchmark epitopes for ZIKV.

Epitope discovery performance benchmarking was extended from methods described by Ponomarenko, J., et al. (2008) [32] and Kringelum, J. V., et al. (2012) [83]. Two limitations of previous epitope discovery performance benchmarking approaches were addressed here. One limitation is the use of receiver operating characteristic curves on imbalanced datasets (a higher proportion of non-epitope residues (negatives) to epitope residues (positives)), which can overestimate how well a metric performs at identifying epitope residues (positives) [29]. This was overcome by benchmarking performance with precision-recall area under the curve (PRAUC) [103] analysis in conjunction with receiver operating characteristic area under the curve (ROCAUC) [104] analysis. A second limitation of previous studies is 1) measuring performance using individual epitopes and 2) then calculating the mean of the results. This was overcome by 1) benchmarking performance on a compilation of all epitopes using five stratified train-test data splits and then 2) calculating the mean of the results.

To understand why using of receiver operating characteristic curves to analyze epitopes individually can overestimate how well a metric performs at identifying epitope residues, one has to look at 1) the size discrepancy between an epitope and a pathogen protein and 2) the meaning of ROCAUC:

First, the average ZIKV-aligned flavivirus epitope is only ~5% of the total number (402) of soluble envelope protein residues.

Second, ROCAUC quantifies, over all thresholds, the combined probability of finding an epitope residue (positive) above a threshold over all epitope residues (positives) (true positive rates) and the probability of finding a non-epitope residue (negative) below the same threshold over all non-epitope residues (negatives) (true negative rates) (Fig 5). A 0.50 ROCAUC indicates performance no better than random chance (the value of no discrimination) [28].

Although ROCAUC is a valuable benchmark, when the proportion of non-epitope residues (negatives) is more pronounced than the proportion of epitope residues (positives), the probability of finding a non-epitope residue (negative) below a threshold over all non-epitope residues (negatives) (true negative rate) holds more weight within the combined probability than the probability of finding an epitope residue (positive) above a threshold over all epitope residues (positives) (true positive rate). In this case, ROCAUC provides more information about metric capacity to rank non-epitope residues (negatives) below epitope residues (positives) (true negative rates) than metric capacity to rank epitope residues (positives) above non-epitope residues (negatives) (true positive rates), which can mask weak true positive rates [29].

Because flavivirus epitopes are much smaller than the isolated proteins, benchmarking on individual epitopes (as done in the development of other epitope discovery algorithms) [5,32] results in datasets with a larger proportion of non-epitope residues (negatives) to epitope residues (positives)—introducing bias.

To understand why the use of precision-recall curves can more transparently reveal how well a metric performs at identifying epitope residues (positives) than the use of receiver operating characteristic curves, one must look at the meaning of PRAUC. PRAUC quantifies, over

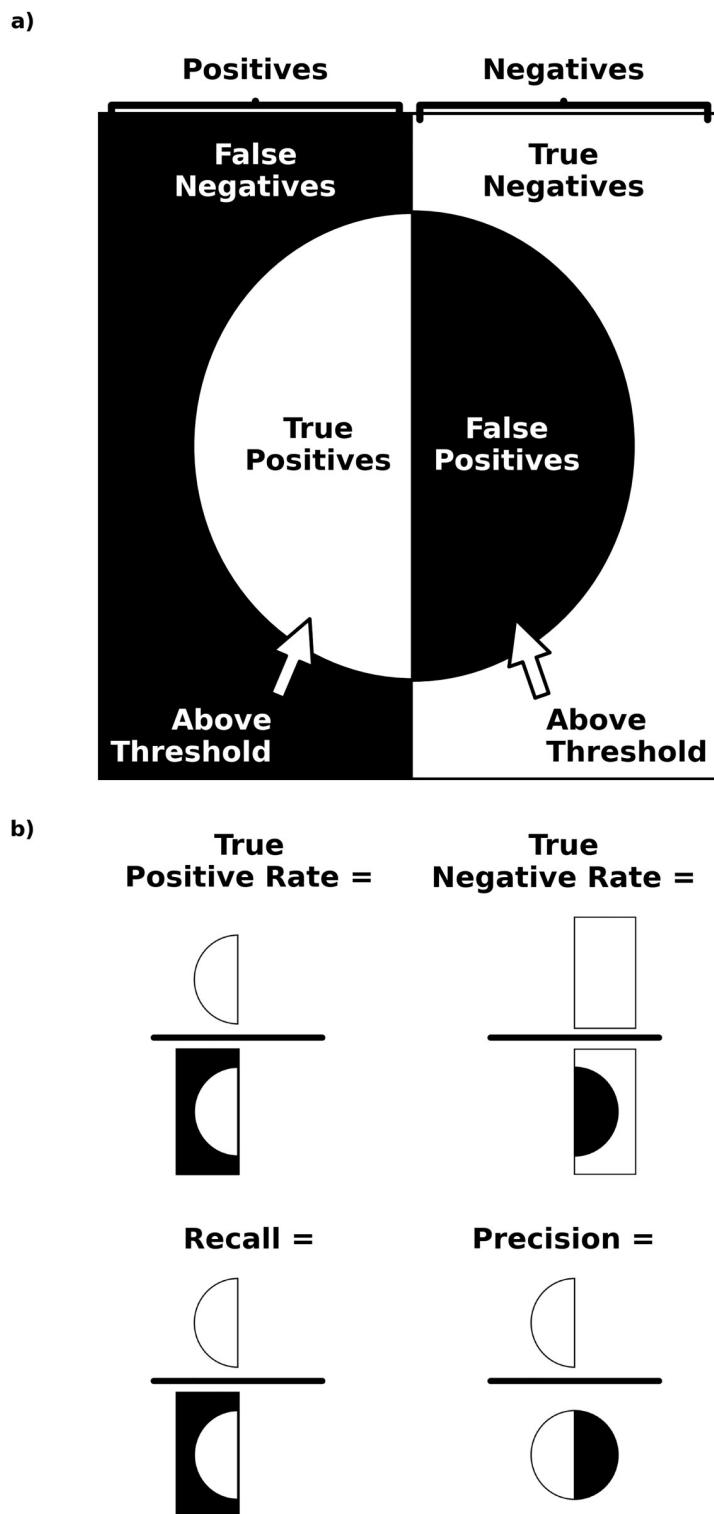


Fig 5. ROCAUC and PRAUC benchmarking. a) Possible prediction outcomes (used for ROCAUC and PRAUC benchmarking) (e.g. true positives) are depicted with black and white wedges. The vertical center line separates all positives and all negatives within the dataset. The center circle contains prediction outcomes for elements above a metric amplitude threshold. b) Rates (e.g. precision and recall) associated with possible prediction outcomes (used for ROCAUC and PRAUC benchmarking) are depicted with wedges and division bars.

<https://doi.org/10.1371/journal.pone.0262321.g005>

all thresholds, the combined probability of finding an epitope residue (positive) above a threshold over all epitope residues (positives) (recalls or true positive rates) and the probability of finding an epitope residue (positive) above the same threshold over all residues (positives and negatives) also above the same threshold (precisions) (Fig 5). The proportion of epitope residues (positives) in the dataset indicates performance no better than random chance (the value of no discrimination) [28]. PRAUC is less biased than ROCAUC when there is a higher proportion of non-epitope residues (negatives) to epitope residues (positives) because, unlike ROCAUC, the probability of identifying non-epitope residues (negatives) below a threshold over all non-epitope residues (negatives) (true negative rate) is given no weight [29].

Epitope discovery benchmarking was performed as follows:

Step 1. In the first case, positives were defined as all structurally characterized, ZIKV-aligned flavivirus epitope residues and negatives were defined as all remaining flavivirus protein residues (from ZIKV envelope protein residues 1–402) (Table 1). In the second case, positives were defined as residues from the two and seven most conserved ZIKV-aligned flavivirus epitopes (Table 2).

Step 2. The epitope dataset was split into test (20%) and training (80%) datasets five times using the scikit-learn [105] StratifiedKFold module (n_splits = 5 and shuffle = True).

Step 3. For lone metrics, ROCAUC and PRAUC were calculated on the test datasets using the scikit-learn [105] roc_curve, precision_recall_curve, and auc modules and the means were taken.

Step 4. For linear combinations of metric pairs (as described in the *Linear Combination Metric Pairs* section above), ROCAUC and PRAUC were first calculated on the training datasets.

Step 5. The mean linear combination weights which produced the highest mean PRAUC and then the highest mean ROCAUC over the training datasets in Step 4 were identified.

Step 6. Using the optimized weights obtained in Step 5, ROCAUC and PRAUC were calculated against the associated test datasets and the means were taken.

An analogous process was used for HPV epitopes. Positives were defined as residues from two highly conserved HPV epitopes (PDB IDs: 7CN2 [54], 3J8Z [55]) out of eight epitopes. Negatives were defined as all other HPV residues in L1 protein chain A (PDB ID: 5KEP) (Table 3).

Epitope organization and epitope discovery demonstration

After epitopic residues have been identified they must be connected to create whole epitopes—in the same way a collection of notes must be connected to form a song. Epitopes can be thought of as collections of subsequences along protein chains e.g. a continuous epitope has one subsequence and a discontinuous epitope with two parts has two subsequences. From this, each epitope was considered to have three major organizational properties: 1) a number of distinct subsequences, 2) a distribution of the number of residues in each subsequence, and 3) a number of total residues.

Experimentally characterized epitope subsequences were defined here through an extension of a definition described by Rubinstein, N. D., et al. (2008) and Sivalingam, G. N. and A. J. Shepherd (2012) [106,107] as one or more positives (including internal stretches of two or less negatives) bordered by stretches of three or more negatives on either side of the protein chain (when all epitopic residues were assumed to reside on an isolated protein). Predicted epitope subsequences were defined here as one or more residues above the prediction threshold

(including internal stretches of two or less residues below the prediction threshold) bordered by stretches of three or more residues below the prediction threshold on either side of the protein chain.

Epitope discovery was extended from methods described by Ponomarenko, J., et al. (2008) [32] and Kringelum, J. V., et al. (2012) [83] and performed as follows:

Step 1. For the top performing metric, a grid of metric amplitude thresholds (determined via the scikit-learn [105] precision_recall_curve module) and residue-residue distance cutoffs (1–10 Å in increments of 1 Å) was constructed.

Step 2. An amplitude threshold and residue-residue distance cutoff were chosen from the grid.

Step 3. Residues above the threshold chosen in Step 2 were clustered using the chosen residue-residue distance cutoff and a clustering method described by Ponomarenko, J., et al. (2008) [32].

Step 4. The mean number of distinct subsequences, the mean number of residues in each subsequence, and the mean number of total residues were calculated for residue clusters identified in Step 3.

Step 5. The mean number of distinct subsequences, the mean number of residues in each subsequence, and the mean number of total residues were calculated for experimentally characterized epitopes.

Step 6. The root-mean-square difference between quantities obtained in Step 4 and 5 was taken.

Step 7. Steps 2–6 were repeated until all grid points were exhausted.

Step 8. Residue clusters associated with the grid point which 1) minimized the root-mean-square difference calculated in Step 6 (sans those which contained residues 401–402 because the transmembrane region was not included in soluble protein simulations) and 2) contained five or more clusters were predicted as epitopes.

To compare the similarity of predicted epitopes with characterized epitopes, the Jaccard Similarity index [57] was calculated using residue numbering along the protein chain.

Supporting information

S1 Fig. Flavivirus isolated protein flexibility correlation plots. Correlation plots for ZIKV isolated protein RMSF vs. dengue serotype 1 (DENV-1) isolated protein RMSF, dengue serotype 2 (DENV-2) isolated protein RMSF, dengue serotype 3 (DENV-3) isolated protein RMSF, dengue serotype 4 (DENV-4) isolated protein RMSF, West Nile virus (WNV) isolated protein RMSF, and Japanese encephalitis virus (JEV) isolated protein RMSF are shown. Isolated protein refers to a flavivirus envelope protein (with a transmembrane region truncation). Spearman rho (r) and p-values are also shown.
(TIF)

S2 Fig. ZIKV isolated protein convex hull scoring. The a) lowest scoring (most inner) and b) median scoring (middle) and c) highest scoring (most outer) isolated protein residue center of masses obtained from convex hull analysis are shown. Isolated protein refers to the ZIKV envelope protein.

(TIF)

S3 Fig. ZIKV VLP protein convex hull scoring. The a) lowest scoring (most inner) and b) median scoring (middle) and c) highest scoring (most outer) VLP protein residue center of masses obtained from convex hull analysis are shown. VLP protein refers to a ZIKV envelope protein (with a transmembrane region truncation) which comprises the VLP (a hollow protein cage).

(TIF)

S1 Table. Epitope discovery performance benchmarking on individual epitopes. Metrics are ordered from the top to bottom in terms of highest ROCAUC and PRAUC product. Spearman rho (r) and p-values are shown for associations between isolated protein convex hull scores (hmon) vs. Epitopia scores (opia), ElliPro scores (epro), isolated protein RMSF (fmon), VLP protein convex hull scores (hvlp), temperature factors (tfac), DiscoTope scores (dtopo), VLP protein RMSF (fvlp), BEPro scores (bpro), cons-PPISP scores (ppisp), partially isolated protein RMSF (fmon*), and IUPred scores (iupred). The primary type of structural information utilized for each metric is shown under the heading titled Type (solvent accessible surface area (SASA), protein flexibility (RMSF), and sequence information (SEQ)).

(PDF)

S2 Table. Flavivirus isolated protein flexibility: Conserved epitope discovery performance benchmarking. Isolated protein flexibility of seven flavivirus structures is examined for epitope discovery performance against the top seven ZIKV-aligned, conserved flavivirus epitopes. Metrics are ordered from the top to bottom in terms of highest ROCAUC and PRAUC product. Spearman rho (r) and p-values are shown for associations between ZIKV isolated protein RMSF (zikv) vs. Japanese encephalitis virus isolated protein RMSF (jev), dengue serotype 2 isolated protein RMSF (denv2), West Nile virus isolated protein RMSF (wnv), dengue serotype 4 isolated protein RMSF (denv4), dengue serotype 3 isolated protein RMSF (denv3), and dengue serotype 1 isolated protein RMSF (denv1).

(PDF)

S1 Dataset.

(CSV)

Author Contributions

Conceptualization: Daniel W. Biner, Jason S. Grosch, Peter J. Ortoleva.

Data curation: Daniel W. Biner.

Formal analysis: Daniel W. Biner.

Investigation: Daniel W. Biner.

Methodology: Daniel W. Biner, Jason S. Grosch, Peter J. Ortoleva.

Validation: Daniel W. Biner.

Visualization: Daniel W. Biner.

Writing – original draft: Daniel W. Biner, Jason S. Grosch, Peter J. Ortoleva.

Writing – review & editing: Daniel W. Biner.

References

1. Davies DR, Sheriff S, Padlan EA. Antibody-antigen complexes. *Journal of Biological Chemistry*. 1988; 263(22):10541–4. PMID: [2455717](https://doi.org/10.1074/jbc.263.22.10541)

2. He L, Cheng Y, Kong L, Azadnia P, Giang E, Kim J, et al. Approaching rational epitope vaccine design for hepatitis C virus with meta-server and multivalent scaffolding. *Scientific reports*. 2015; 5. <https://doi.org/10.1038/srep12501> PMID: 26238798
3. Schellenbacher C, Roden R, Kirnbauer R. Chimeric L1-L2 virus-like particles as potential broad-spectrum human papillomavirus vaccines. *Journal of virology*. 2009; 83(19):10085–95. <https://doi.org/10.1128/JVI.01088-09> PMID: 19640991
4. Steichen JM, Kulp DW, Tokatlian T, Escolano A, Dosenovic P, Stanfield RL, et al. HIV vaccine design to target germline precursors of glycan-dependent broadly neutralizing antibodies. *Immunity*. 2016; 45(3):483–96. <https://doi.org/10.1016/j.jimmuni.2016.08.016> PMID: 27617678
5. Mukonyora M. A review of important discontinuous B-cell epitope prediction tools. *Journal of Clinical & Cellular Immunology*. 2015; 2015.
6. Stave JW, Lindpaintner K. Antibody and antigen contact residues define epitope and paratope size and structure. *The Journal of Immunology*. 2013; 191(3):1428–35. <https://doi.org/10.4049/jimmunol.1203198> PMID: 23797669
7. Kielian M, Jungerwirth S. Mechanisms of enveloped virus entry into cells. *Molecular Biology and Medicine*. 1990; 7(1):17–31. PMID: 2182968
8. Kostyuchenko VA, Lim EX, Zhang S, Fibriansah G, Ng T-S, Ooi JS, et al. Structure of the thermally stable Zika virus. *Nature*. 2016. <https://doi.org/10.1038/nature17994> PMID: 27093288
9. Mirza MU, Rafique S, Ali A, Munir M, Ikram N, Manan A, et al. Towards peptide vaccines against Zika virus: Immunoinformatics combined with molecular dynamics simulations to predict antigenic epitopes of Zika viral proteins. *Scientific reports*. 2016; 6:37313. <https://doi.org/10.1038/srep37313> PMID: 27934901
10. Xu X, Vaughan K, Weiskopf D, Grifoni A, Diamond MS, Sette A, et al. Identifying candidate targets of immune responses in Zika virus based on homology to epitopes in other flavivirus species. *PLoS currents*. 2016; 8. <https://doi.org/10.1371/currents.outbreaks.9aa2e1fb61b0f632f58a098773008c4b> PMID: 28018746
11. Alam A, Ali S, Ahamed S, Malik MZ, Ishrat R. From ZikV genome to vaccine: in silico approach for the epitope-based peptide vaccine against Zika virus envelope glycoprotein. *Immunology*. 2016; 149(4):386–99. <https://doi.org/10.1111/imm.12656> PMID: 27485738
12. Li J, Watterson D, Chang C-W, Che X-Y, Li X-Q, Ericsson DJ, et al. Structural and functional characterization of a cross-reactive dengue virus neutralizing antibody that recognizes a cryptic epitope. *Structure*. 2018; 26(1):51–9. e4. <https://doi.org/10.1016/j.str.2017.11.017> PMID: 29249606
13. Morrone SR, Chew VS, Lim X-N, Ng T-S, Kostyuchenko VA, Zhang S, et al. High flavivirus structural plasticity demonstrated by a non-spherical morphological variant. *Science*. 2020; 11(1):1–10. <https://doi.org/10.1126/science.1970254> PMID: 32561757
14. Sharma KK, Lim X-X, Tantirimudalige SN, Gupta A, Marzinek JK, Holdbrook D, et al. Infectivity of dengue virus serotypes 1 and 2 is correlated with E-protein intrinsic dynamics but not to envelope conformations. *Science*. 2019; 27(4):618–30. e4. <https://doi.org/10.1126/science.1970254> PMID: 30686666
15. Westhof E, Altschuh D, Moras D, Bloomer A, Mondragon A, Klug A, et al. Correlation between segmental mobility and the location of antigenic determinants in proteins. *Nature*. 1984; 311(5982):123–6. <https://doi.org/10.1038/311123a0> PMID: 6206398
16. Joshi H, Cheluvaraja S, Somogyi E, Brown DR, Ortoleva P. A molecular dynamics study of loop fluctuation in human papillomavirus type 16 virus-like particles: a possible indicator of immunogenicity. *Vaccine*. 2011; 29(51):9423–30. <https://doi.org/10.1016/j.vaccine.2011.10.039> PMID: 22027487
17. Perilla JR, Goh BC, Cassidy CK, Liu B, Bernardi RC, Rudack T, et al. Molecular dynamics simulations of large macromolecular complexes. *Science*. 2015; 31(64–74). <https://doi.org/10.1126/science.1250007> PMID: 25845770
18. Perilla JR, Schulten KJNc. Physical properties of the HIV-1 capsid from all-atom molecular dynamics simulations. *Science*. 2017; 8(1):1–10. <https://doi.org/10.1126/science.1250007> PMID: 28722007
19. Freddolino PL, Arkhipov AS, Larson SB, McPherson A, Schulten KJS. Molecular dynamics simulations of the complete satellite tobacco mosaic virus. *Science*. 2006; 14(3):437–49. <https://doi.org/10.1126/science.1111014> PMID: 16531228
20. Grosch JS, Yang J, Shen A, Sereda YV, Ortoleva PJ. Broad spectrum assessment of the epitope fluctuation—immunogenicity hypothesis. *Vaccine*. 2015; 33(44):5945–9. <https://doi.org/10.1016/j.vaccine.2015.06.111> PMID: 26187254
21. Escolano A, Steichen JM, Dosenovic P, Kulp DW, Golijanin J, Sok D, et al. Sequential immunization elicits broadly neutralizing anti-HIV-1 antibodies in Ig knockin mice. *Cell*. 2016; 166(6):1445–58. e12. <https://doi.org/10.1016/j.cell.2016.07.030> PMID: 27610569

22. Dejnirattisai W, Supasa P, Wongwiwat W, Rouvinski A, Barba-Spaeth G, Duangchinda T, et al. Dengue virus sero-cross-reactivity drives antibody-dependent enhancement of infection with zika virus. *Nature immunology*. 2016; <https://doi.org/10.1038/ni.3515> PMID: 27339099
23. Moi ML, Takasaki T, Kurane I. Efficacy of tetravalent dengue vaccine in Thai schoolchildren. *The Lancet*. 2013; 381(9872):1094. [https://doi.org/10.1016/S0140-6736\(13\)60754-0](https://doi.org/10.1016/S0140-6736(13)60754-0) PMID: 23540846
24. Vujadinovic M, Khan S, Oosterhuis K, Uil TG, Wunderlich K, Damman S, et al. Adenovirus based HPV L2 vaccine induces broad cross-reactive humoral immune responses. *Vaccine*. 2018; 36(30):4462–70. <https://doi.org/10.1016/j.vaccine.2018.06.024> PMID: 29914845
25. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic acids research*. 2016; 44(W1):W344–W50. <https://doi.org/10.1093/nar/gkw408> PMID: 27166375
26. Thornton J, Edwards M, Taylor W, Barlow D. Location of 'continuous' antigenic determinants in the protruding regions of proteins. *The EMBO journal*. 1986; 5(2):409. <https://doi.org/10.1002/j.1460-2075.1986.tb04226.x> PMID: 2423325
27. Novotný Jí, Handschumacher M, Haber E, Brucolieri RE, Carlson WB, Fanning DW, et al. Antigenic determinants in proteins coincide with surface regions accessible to large probes (antibody domains). *Proceedings of the National Academy of Sciences*. 1986; 83(2):226–30. <https://doi.org/10.1073/pnas.83.2.226> PMID: 2417241
28. Mandrekar JNJJoTO. Receiver operating characteristic curve in diagnostic test assessment. 2010; 5(9):1315–6. <https://doi.org/10.1097/JTO.0b013e3181ec173d> PMID: 20736804
29. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS one*. 2015; 10(3):e0118432. <https://doi.org/10.1371/journal.pone.0118432> PMID: 25738806
30. Lok S-M. The interplay of dengue virus morphological diversity and human antibodies. *Trends in microbiology*. 2016; 24(4):284–93. <https://doi.org/10.1016/j.tim.2015.12.004> PMID: 26747581
31. Mandl CW, Guirakhoo F, Holzmann H, Heinz FX, Kunz CJJov. Antigenic structure of the flavivirus envelope protein E at the molecular level, using tick-borne encephalitis virus as a model. 1989; 63(2):564–71.
32. Ponomarenko J, Bui H-H, Li W, Fuseder N, Bourne PE, Sette A, et al. ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC bioinformatics*. 2008; 9(1):1. <https://doi.org/10.1186/1471-2105-9-514> PMID: 19055730
33. Yao B, Zheng D, Liang S, Zhang C. Conformational B-cell epitope prediction on antigen protein structures: a review of current algorithms and comparison with common binding site prediction methods. *PLoS One*. 2013; 8(4):e62249. <https://doi.org/10.1371/journal.pone.0062249> PMID: 23620816
34. Sweredoski MJ, Baldi P. PEPITO: improved discontinuous B-cell epitope prediction using multiple distance thresholds and half sphere exposure. *Bioinformatics*. 2008; 24(12):1459–60. <https://doi.org/10.1093/bioinformatics/btn199> PMID: 18443018
35. Rubinstein ND, Mayrose I, Martz E, Pupko T. Epitopia: a web-server for predicting B-cell epitopes. *BMC bioinformatics*. 2009; 10(1):287. <https://doi.org/10.1186/1471-2105-10-287> PMID: 19751513
36. Qin S, Zhou H-X. meta-PPISP: a meta web server for protein-protein interaction site prediction. *Bioinformatics*. 2007; 23(24):3386–7. <https://doi.org/10.1093/bioinformatics/btm434> PMID: 17895276
37. Sager G, Gabaglio S, Sztul E, Belov GA. Role of Host Cell Secretory Machinery in Zika Virus Life Cycle. *Viruses*. 2018; 10(10):559. <https://doi.org/10.3390/v10100559> PMID: 30326556
38. Hoffman W, Lakkis FG, Chalasani G. B cells, antibodies, and more. *Clinical Journal of the American Society of Nephrology*. 2016; 11(1):137–54. <https://doi.org/10.2215/CJN.09430915> PMID: 26700440
39. Rao KV. Selection in a T-dependent primary humoral response: new insight from polypeptide models. *APMIS*. 1999; 107(7-12):807–18.
40. Ovchinnikov V, Louveau JE, Barton JP, Karplus M, Chakraborty AK. Role of framework mutations and antibody flexibility in the evolution of broadly neutralizing antibodies. *Elife*. 2018; 7:e33038. <https://doi.org/10.7554/elife.33038> PMID: 29442996
41. Manivel V, Sahoo NC, Salunke DM, Rao KV. Maturation of an antibody response is governed by modulations in flexibility of the antigen-combining site. *Immunity*. 2000; 13(5):611–20. [https://doi.org/10.1016/s1074-7613\(00\)00061-3](https://doi.org/10.1016/s1074-7613(00)00061-3) PMID: 11114374
42. Winkler G, Heinz F, Kunz C. Characterization of a disulphide bridge-stabilized antigenic domain of tick-borne encephalitis virus structural glycoprotein. *Journal of general virology*. 1987; 68(8):2239–44. <https://doi.org/10.1099/0022-1317-68-8-2239> PMID: 2440984
43. Lin B, Parrish CR, Murray JM, Wright PJ. Localization of a neutralizing epitope on the envelope protein of dengue virus type 2. *Virology*. 1994; 202(2):885–90. <https://doi.org/10.1006/viro.1994.1410> PMID: 7518164

44. Mason PW, Zügel MU, Semproni AR, Fournier MJ, Mason TL. The antigenic structure of dengue type 1 virus envelope and NS1 proteins expressed in *Escherichia coli*. *Journal of General Virology*. 1990; 71(9):2107–14. <https://doi.org/10.1099/0022-1317-71-9-2107> PMID: 1698924
45. Sourisseau M, Lawrence DJ, Schwarz MC, Storrs CH, Veit EC, Bloom JD, et al. Deep mutational scanning comprehensively maps how Zika envelope protein mutations affect viral growth and antibody escape. *Journal of virology*. 2019; 93(23). <https://doi.org/10.1128/JVI.01291-19> PMID: 31511387
46. Wengler G, Wengler G. An analysis of the antibody response against West Nile virus E protein purified by SDS-PAGE indicates that this protein does not contain sequential epitopes for efficient induction of neutralizing antibodies. *Journal of general virology*. 1989; 70(4):987–92.
47. MacRaid CA, Richards JS, Anders RF, Norton RS. Antibody recognition of disordered antigens. *Structure*. 2016; 24(1):148–57. <https://doi.org/10.1016/j.str.2015.10.028> PMID: 26712277
48. Erdős G, Dosztányi Z. Analyzing Protein Disorder with IUPred2A. *Current Protocols in Bioinformatics*. 2020; 70(1):e99. <https://doi.org/10.1002/cbpi.99> PMID: 32237272
49. Carrat F, Flahault A. Influenza vaccine: the challenge of antigenic drift. *Vaccine*. 2007; 25(39–40):6852–62. <https://doi.org/10.1016/j.vaccine.2007.07.027> PMID: 17719149
50. Zhang S, Kostyuchenko VA, Ng T-S, Lim X-N, Ooi JS, Lambert S, et al. Neutralization mechanism of a highly potent antibody against Zika virus. *Nature communications*. 2016; 7:13679. <https://doi.org/10.1038/ncomms13679> PMID: 27882950
51. Zandi R, Reguera D. Mechanical properties of viral capsids. *Physical Review E*. 2005; 72(2):021917. <https://doi.org/10.1103/PhysRevE.72.021917> PMID: 16196614
52. Palanisamy N, Akaberi D, Lennerstrand JMp, evolution. Protein backbone flexibility pattern is evolutionarily conserved in the Flaviviridae family: a case of NS3 protease in Flavivirus and Hepacivirus. 2018; 118:58–63.
53. Pinello JF, Lai AL, Millet JK, Cassidy-Hanley D, Freed JH, Clark TG. Structure-function studies link class II viral fusogens with the ancestral gamete fusion protein HAP2. *Current Biology*. 2017; 27(5):651–60. <https://doi.org/10.1016/j.cub.2017.01.049> PMID: 28238660
54. Huang W, He M, Ning T, Nie J, Zhang F, Zheng Q, et al. Structural characterization of a neutralizing mAb H16. 001, a potent candidate for a common potency assay for various HPV16 VLPs. 2020; 5(1):1–9. <https://doi.org/10.1038/s41541-020-00236-w> PMID: 33042588
55. Guan J, Bywaters SM, Brendle SA, Lee H, Ashley RE, Makhov AM, et al. Structural comparison of four different antibodies interacting with human papillomavirus 16 and mechanisms of neutralization. 2015; 483:253–63. <https://doi.org/10.1016/j.virol.2015.04.016> PMID: 25996608
56. Haste Andersen P, Nielsen M, Lund O. Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. *Protein Science*. 2006; 15(11):2558–67. <https://doi.org/10.1110/ps.062405906> PMID: 17001032
57. Niwattanakul S, Singthongchai J, Naenudorn E, Wanapu S, editors. Using of Jaccard coefficient for keywords similarity. *Proceedings of the international multiconference of engineers and computer scientists*; 2013.
58. Cockburn JJ, Sanchez MEN, Fretes N, Urvoas A, Staropoli I, Kikuti CM, et al. Mechanism of dengue virus broad cross-neutralization by a monoclonal antibody. *Structure*. 2012; 20(2):303–14. <https://doi.org/10.1016/j.str.2012.01.001> PMID: 22285214
59. Dai L, Song J, Lu X, Deng Y-Q, Musyoki AM, Cheng H, et al. Structures of the Zika Virus Envelope Protein and Its Complex with a Flavivirus Broadly Protective Antibody. *Cell host & microbe*. 2016; 19(5):696–704. <https://doi.org/10.1016/j.chom.2016.04.013> PMID: 27158114
60. Smith SA, de Alwis AR, Kose N, Harris E, Ibarra KD, Kahle KM, et al. The potent and broadly neutralizing human dengue virus-specific monoclonal antibody 1C19 reveals a unique cross-reactive epitope on the bc loop of domain II of the envelope protein. *MBio*. 2013; 4(6):e00873–13. <https://doi.org/10.1128/mBio.00873-13> PMID: 24255124
61. Rotkiewicz P, Skolnick J. Fast procedure for reconstruction of full-atom protein models from reduced representations. *Journal of computational chemistry*. 2008; 29(9):1460–5. <https://doi.org/10.1002/jcc.20906> PMID: 18196502
62. Kostyuchenko VA, Zhang Q, Tan JL, Ng T-S, Lok S-M. Immature and mature dengue serotype 1 virus structures provide insight into the maturation process. *Journal of virology*. 2013; 87(13):7700–7. <https://doi.org/10.1128/JVI.00197-13> PMID: 23637416
63. Zhang X, Ge P, Yu X, Brannan JM, Bi G, Zhang Q, et al. Cryo-EM structure of the mature dengue virus at 3.5-Å resolution. *Nature structural & molecular biology*. 2013; 20(1):105.
64. Fibriansah G, Tan J, Smith S, de Alwis R, Ng T, Kostyuchenko V, et al. A highly potent human antibody neutralizes dengue virus serotype 3 by binding across three surface proteins. *Nat Commun*. 2015; 6:6341. <https://doi.org/10.1038/ncomms7341> PMID: 25698059

65. Kostyuchenko VA, Chew PL, Ng T-S, Lok S-M. Near-atomic resolution cryo-electron microscopic structure of dengue serotype 4 virus. *Journal of virology*. 2014; 88(1):477–82. <https://doi.org/10.1128/JVI.02641-13> PMID: 24155405
66. Nybakken GE, Nelson CA, Chen BR, Diamond MS, Fremont DH. Crystal structure of the West Nile virus envelope glycoprotein. *Journal of virology*. 2006; 80(23):11467–74. <https://doi.org/10.1128/JVI.01125-06> PMID: 16987985
67. Wang X, Li S-H, Zhu L, Nian Q-G, Yuan S, Gao Q, et al. Near-atomic structure of Japanese encephalitis virus reveals critical determinants of virulence and stability. *Nature communications*. 2017; 8(1):1–9.
68. Berendsen HJ, van der Spoel D, van Drunen R. GROMACS: a message-passing parallel molecular dynamics implementation. *Computer Physics Communications*. 1995; 91(1):43–56.
69. Mark P, Nilsson L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *The Journal of Physical Chemistry A*. 2001; 105(43):9954–60.
70. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *Journal of computational chemistry*. 1983; 4(2):187–217.
71. Hess B, Bekker H, Berendsen HJ, Fraaije JG. LINCS: a linear constraint solver for molecular simulations. *Journal of computational chemistry*. 1997; 18(12):1463–72.
72. Martyna GJ, Klein ML, Tuckerman M. Nosé–Hoover chains: the canonical ensemble via continuous dynamics. *The Journal of chemical physics*. 1992; 97(4):2635–43.
73. Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics*. 1981; 52(12):7182–90.
74. Tironi IG, Sperb R, Smith PE, van Gunsteren WF. A generalized reaction field method for molecular dynamics simulations. *The Journal of chemical physics*. 1995; 102(13):5451–9.
75. Søndergaard CR, Olsson MH, Rostkowski M, Jensen JH. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of p K a values. *Journal of chemical theory and computation*. 2011; 7(7):2284–95. <https://doi.org/10.1021/ct200133y> PMID: 26606496
76. Guan J, Bywaters SM, Brendle SA, Ashley RE, Makhov AM, Conway JF, et al. Cryoelectron microscopy maps of human papillomavirus 16 reveal L2 densities and heparin binding site. 2017; 25(2):253–63. <https://doi.org/10.1016/j.str.2016.12.001> PMID: 28065506
77. Cardone G, Moyer AL, Cheng N, Thompson CD, Dvoretzky I, Lowy DR, et al. Maturation of the human papillomavirus 16 capsid. 2014; 5(4). <https://doi.org/10.1128/mBio.01104-14> PMID: 25096873
78. Caspar DL, Klug A, editors. Physical principles in the construction of regular viruses. Cold Spring Harbor symposia on quantitative biology; 1962: Cold Spring Harbor Laboratory Press.
79. Jarvis RA. On the identification of the convex hull of a finite set of points in the plane. *Information processing letters*. 1973; 2(1):18–21.
80. Barber CB, Dobkin DP, Dobkin DP, Huhdanpaa H. The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software (TOMS)*. 1996; 22(4):469–83.
81. Zheng W, Ruan J, Hu G, Wang K, Hanlon M, Gao J. Analysis of conformational B-cell epitopes in the antibody-antigen complex using the depth function and the convex hull. *PloS one*. 2015; 10(8): e0134835. <https://doi.org/10.1371/journal.pone.0134835> PMID: 26244562
82. Jones E, Oliphant T, Peterson P. {SciPy}: open source scientific tools for {Python}. 2014.
83. Kringelum JV, Lundsgaard C, Lund O, Nielsen M. Reliable B cell epitope predictions: impacts of method development and improved benchmarking. *PLoS computational biology*. 2012; 8(12). <https://doi.org/10.1371/journal.pcbi.1002829> PMID: 23300419
84. Zhang S, Vogt MR, Oliphant T, Engle M, Bovshik EI, Diamond MS, et al. Development of resistance to passive therapy with a potently neutralizing humanized monoclonal antibody against West Nile virus. *Journal of Infectious Diseases*. 2009; 200(2):202–5. <https://doi.org/10.1086/599794> PMID: 19527169
85. Lok S-M, Kostyuchenko V, Nybakken GE, Holdaway HA, Battisti AJ, Sukupolvi-Petty S, et al. Binding of a neutralizing antibody to dengue virus alters the arrangement of surface glycoproteins. *Nature structural & molecular biology*. 2008; 15(3):312–7. <https://doi.org/10.1038/nsmb.1382> PMID: 18264114
86. Cherrier MV, Kaufmann B, Nybakken GE, Lok SM, Warren JT, Chen BR, et al. Structural basis for the preferential recognition of immature flaviviruses by a fusion-loop antibody. *The EMBO journal*. 2009; 28(20):3269–76. <https://doi.org/10.1038/emboj.2009.245> PMID: 19713934
87. Kaufmann B, Vogt MR, Goudsmit J, Holdaway HA, Aksyuk AA, Chipman PR, et al. Neutralization of West Nile virus by cross-linking of its surface proteins with Fab fragments of the human monoclonal antibody CR4354. *Proceedings of the National Academy of Sciences*. 2010; 107(44):18950–5. <https://doi.org/10.1073/pnas.1011036107> PMID: 20956322

88. Midgley CM, Flanagan A, Tran HB, Dejnirattisai W, Chawansuntati K, Jumnainsong A, et al. Structural analysis of a dengue cross-reactive antibody complexed with envelope domain III reveals the molecular basis of cross-reactivity. *The Journal of Immunology*. 2012; 188(10):4971–9. <https://doi.org/10.4049/jimmunol.1200227> PMID: 22491255
89. Austin SK, Dowd KA, Shrestha B, Nelson CA, Edeling MA, Johnson S, et al. Structural basis of differential neutralization of DENV-1 genotypes by an antibody that recognizes a cryptic epitope. *PLoS Pathog*. 2012; 8(10):e1002930. <https://doi.org/10.1371/journal.ppat.1002930> PMID: 23055922
90. Fibriansah G, Tan JL, Smith SA, de Alwis AR, Ng TS, Kostyuchenko VA, et al. A potent anti-dengue human antibody preferentially recognizes the conformation of E protein monomers assembled on the virus surface. *EMBO molecular medicine*. 2014;e201303404. <https://doi.org/10.1002/emmm.201303404> PMID: 24421336
91. Edeling MA, Austin SK, Shrestha B, Dowd KA, Mukherjee S, Nelson CA, et al. Potent dengue virus neutralization by a therapeutic antibody with low monovalent affinity requires bivalent engagement. *PLoS Pathog*. 2014; 10(4):e1004072. <https://doi.org/10.1371/journal.ppat.1004072> PMID: 24743696
92. Rouvinski A, Guardado-Calvo P, Barba-Spaeth G, Duquerroy S, Vaney M-C, Kikuti CM, et al. Recognition determinants of broadly neutralizing human antibodies against dengue viruses. *Nature*. 2015; 520(7545):109–13. <https://doi.org/10.1038/nature14130> PMID: 25581790
93. Swanstrom J, Plante J, Plante K, Young E, McGowan E, Gallichotte E, et al. Dengue virus envelope dimer epitope monoclonal antibodies isolated from dengue patients are protective against Zika virus. *MBio*. 2016; 7(4):e01123–16. <https://doi.org/10.1128/mBio.01123-16> PMID: 27435464
94. Robinson LN, Tharakaraman K, Rowley KJ, Costa VV, Chan KR, Wong YH, et al. Structure-guided design of an anti-dengue antibody directed to a non-immunodominant epitope. *Cell*. 2015; 162(3):493–504. <https://doi.org/10.1016/j.cell.2015.06.057> PMID: 26189681
95. Barba-Spaeth G, Dejnirattisai W, Rouvinski A, Vaney M-C, Medits I, Sharma A, et al. Structural basis of potent Zika–dengue virus antibody cross-neutralization. *Nature*. 2016; 536(7614):48–53. <https://doi.org/10.1038/nature18938> PMID: 27338953
96. Zhao H, Fernandez E, Dowd KA, Speer SD, Platt DJ, Gorman MJ, et al. Structural basis of Zika virus-specific antibody protection. *Cell*. 2016; 166(4):1016–27. <https://doi.org/10.1016/j.cell.2016.07.020> PMID: 27475895
97. Wang Q, Yang H, Liu X, Dai L, Ma T, Qi J, et al. Molecular determinants of human neutralizing antibodies isolated from a patient infected with Zika virus. *Science Translational Medicine*. 2016; 8(369):369ra179–369ra179. <https://doi.org/10.1126/scitranslmed.aai8336> PMID: 27974667
98. Hasan SS, Miller A, Sapparapu G, Fernandez E, Klose T, Long F, et al. A human antibody against Zika virus crosslinks the E protein to prevent infection. *Nature communications*. 2017; 8:14722. <https://doi.org/10.1038/ncomms14722> PMID: 28300075
99. Robbiani DF, Bozzacco L, Keeffe JR, Khouri R, Olsen PC, Gazumyan A, et al. Recurrent potent human neutralizing antibodies to Zika virus in Brazil and Mexico. *Cell*. 2017; 169(4):597–609. e11. <https://doi.org/10.1016/j.cell.2017.04.024> PMID: 28475892
100. Wang J, Bardelli M, Espinosa DA, Pedotti M, Ng T-S, Bianchi S, et al. A human bi-specific antibody against Zika virus with high therapeutic potential. *Cell*. 2017; 171(1):229–41. e15. <https://doi.org/10.1016/j.cell.2017.09.002> PMID: 28938115
101. Qiu X, Lei Y, Yang P, Gao Q, Wang N, Cao L, et al. Structural basis for neutralization of Japanese encephalitis virus by two potent therapeutic antibodies. *Nature microbiology*. 2018; 3(3):287. <https://doi.org/10.1038/s41564-017-0099-x> PMID: 29379207
102. Vita R, Overton JA, Greenbaum JA, Ponomarenko J, Clark JD, Cantrell JR, et al. The immune epitope database (IEDB) 3.0. *Nucleic acids research*. 2014; 43(D1):D405–D12. <https://doi.org/10.1093/nar/gku938> PMID: 25300482
103. Buckland M, Gey F. The relationship between recall and precision. *Journal of the American society for information science*. 1994; 45(1):12–9.
104. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical chemistry*. 1993; 39(4):561–77. PMID: 8472349
105. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*. 2011; 12(Oct):2825–30.
106. Rubinstein ND, Mayrose I, Halperin D, Yekutieli D, Gershoni JM, Pupko T. Computational characterization of B-cell epitopes. *Molecular immunology*. 2008; 45(12):3477–89. <https://doi.org/10.1016/j.molimm.2007.10.016> PMID: 18023478
107. Sivalingam GN, Shepherd AJ. An analysis of B-cell epitope discontinuity. *Molecular immunology*. 2012; 51(3–4):304–9. <https://doi.org/10.1016/j.molimm.2012.03.030> PMID: 22520973