



Mallari, M.<sup>1,2</sup>, Vemu, S<sup>3</sup>, Afshar, G<sup>4</sup>, Sanchez, J<sup>3</sup>, Carson, V<sup>4</sup>, Rodriguez Estrada, A.<sup>6</sup>, Wilson, C.<sup>7</sup>, Porter, S.<sup>1</sup>  
1. Digital World Biology 2. City College of San Francisco 3. Waubonsee Community College 4. Hillsborough Community College  
5. Skyline College 6. Aurora University 7. City Colleges of Chicago

## OVERVIEW

Access to undergraduate-level research is critical for the development of future researchers. The barrier for many attempting to break into a career in research is the need for more experience applying technical skills to answer scientific inquiries. Many community college programs need more resources focused on biotechnology, with many curriculum's building skills for biomanufacturing careers. Those interested in a career in biological research face many challenges in finding experiences like internships due to the significant competition from applicants attending 4-year institutions. Databases such as the Immune Epitope Database (IEDB) provide access to curated, peer-reviewed, open-source datasets that can help guide research questions. The issue for inexperienced researchers lies in understanding how to utilize these datasets. The development and implementation of guided research experiences provide users with a topic they are familiar with but not necessarily experts in, like SARS-CoV2 infections. These guides will facilitate establishing connections between these web-based tools and datasets to critical topics in immunology and protein engineering. Using tools built into databases like the IEDB allows users to learn subjects like understanding proteasomal cleavage events, mapping immunogenic regions or mutations on crystal structures, and predicting B and T-cell epitopes.

## INTRODUCTION TO THE IEDB

The Immune Epitope Database (IEDB) is a robust resource that curates and catalogs data on infectious diseases, allergies, autoimmunity, and transplantation. This data is traceable, and manuscript references are provided to understand how researchers acquired each data point. In addition, this collection of data can be used to survey some epitope characteristics. To leverage these datasets, the IEDB has also created several prediction tools and imported several third-party tools to aid in research efforts.

SARS-CoV2, specifically the Spike Glycoprotein, was used for the creation of this lesson plan.

### Creating a query from the homepage

START YOUR SEARCH HERE

**a. Epitope**: Any, Linear peptide, Exact M<sup>n</sup> (Ex: SINFEKL), Discontinuous, Non-peptide.

**b. Assay**: T Cell, B Cell (selected), MHC Ligand, Ex: neutralization, Outcome: Positive, Negative.

**c. Epitope Source**: Organism: SARS-CoV2 (ID:2697) (selected), Ex: core, capsid, myo, Host: Human, Mouse, Non-human primate, Ex: dog, camel.

**d. Host**: Any, Human, Mouse, Non-human primate, Ex: dog, camel.

**e. MHC Restriction**: Any, Class I, Class II, Non-classical, Ex: HLA-A\*02:01.

**f. B Cell Epitope Prediction**: Predict linear B cell epitopes using: Antigen Sequence Properties, Predict discontinuous B cell epitopes using: Discotope, EpiPro.

**Epitope Analysis Tools**: Analyze epitope sets: Population Coverage, Conservation Across Antigens, Clusters with Similar Sequences.

A refined query will result in more representative data, so it's essential to carefully consider your research question before sourcing data.

- a. **Epitope:** Linear vs. discontinuous
- b. **Assay:** By default, the IEDB will only show assays with **Positive Outcomes**; however, for specific tools like the **Immunoome Browser**, it is essential also to survey **Negative Outcomes** to have a more representative dataset for the target antigen (in our case, SARS-CoV2).
- c. **Epitope source:** The epitope source allows one to filter where the epitope can be found (bacteria, eukaryotes, viruses, plasmid, vectors, etc.) if a known organism is not known.
- d. **Host:** The IEDB allows one to discriminate by the host model system.
- e. **MHC Restriction and Disease:** These were intentionally left in their defaults because their MHCs will not express on SARS-CoV2, so the result will include zero data points. The 'Disease' designation can be specified, but due to SARS-CoV2 being a virus, querying for allergic or autoimmune data will also have no results.

## ACKNOWLEDGMENTS

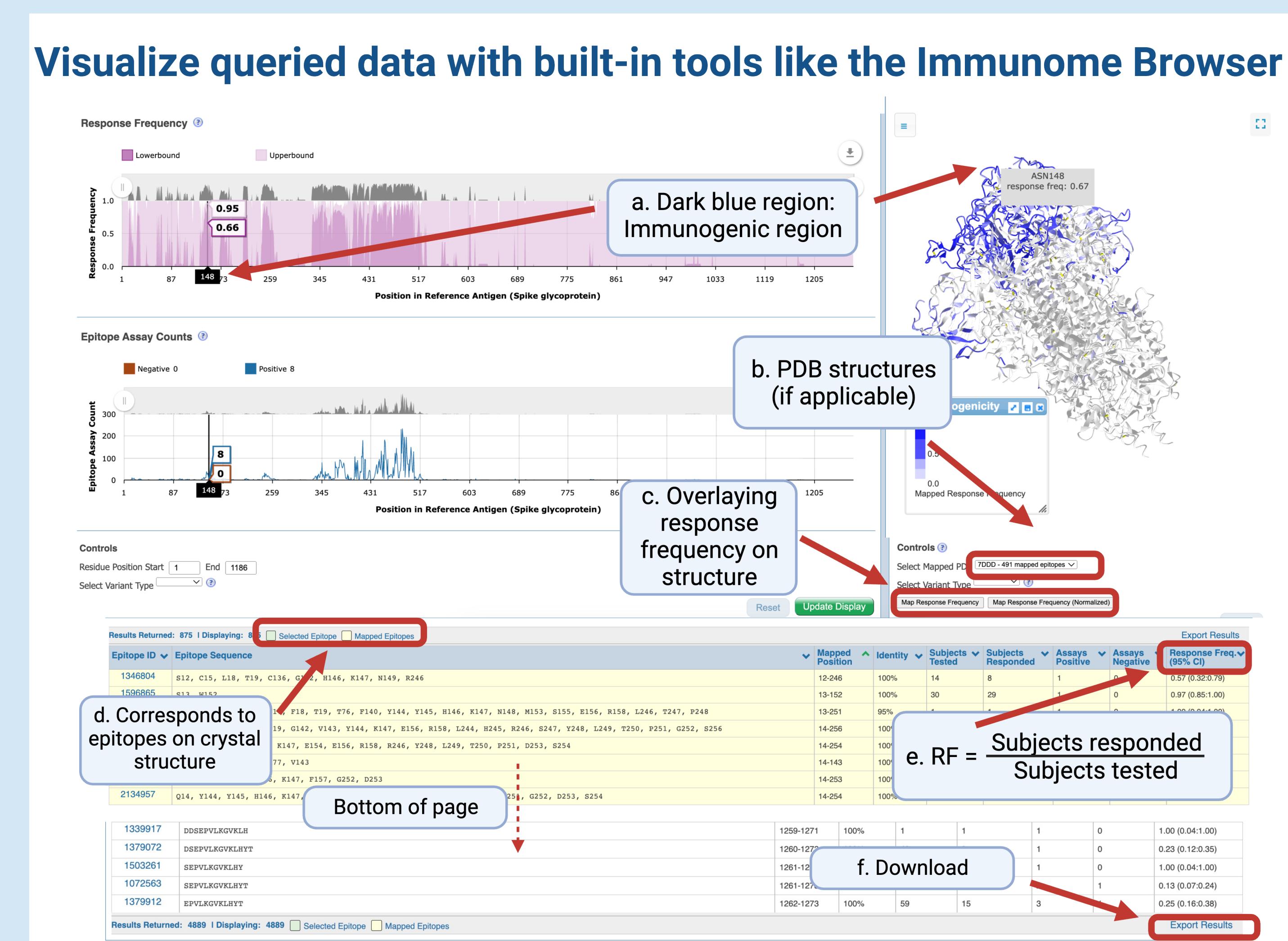
- Bethesda: Tracking Resistance and Coronavirus Evolution (TRACE). National Library of Medicine (US), National Center for Biotechnology Information
- Mallari, M., Vemu, S., Carson, V., Afshar, G., Rodriguez Estrada, A., Wilson, C., Porter, S. (2022). Using the IEDB to Predict Proteasomal Cleavage Events and T-cell Epitopes. Antibody Engineering, QUBES Educational Resources. doi:10.25334/ADRK-9ST6
- NSF: Digital World Biology
- Vita, P., Mahajan, S., Overton, J.A., Dhanda, S.K., Martini, S., Cantrell, J.R., Wheeler, D.K., Sette, A., Peters, B. The Immune Epitope Database (IEDB) update 2018. *Oct 24, 2018*. doi:10.1101/nar.gky1006. PMID: 30357391; PMCID: PMC6324067
- Wang, Y., Youkharbach, P., Zhang, D., Lanckzyk, C.J., Geer, R.C., Madej, T., Phan, L., Ward, M., Lu, S., Marchler, G.H., Wang, Y., Bryant, S. P. Web-based 3D viewer for Sharing 3D representations of Biomolecular Structures. *Bioinformatics*. 2020 Jan; 36(1):131-135. (Epub 2019 June 20, doi: 10.1093/bioinformatics/btz502). PubMed PMID: 31218344
- Wang, Y., Youkharbach, P., Marchler-Bauer, A., Lanckzyk, C., Zhang, D., Lu, S., Madej, T., Marchler, G.H., Cheng, T., Chong, L.C., Zhao, Z., Yang, K., Lin, J., Cheng, Z., Dunn, R., Makrilia, S.A., Tai, C.-H., Emoma, D., Busby, B., Johnson, N.L., Tabaro, F., Song, G., Geer, R.C., Bryant, S. P. Web-based 3D viewer to Share 3D representations of Biomolecular Structures. *Front. Mol. Biosci.* 2022;9:741. (Epub 2022 Feb 17, doi: 10.3389/fmolb.2022.91740). [PubMed PMID: 35252651]
- Zhang, G., Wang, Y., Zhu, Y. et al. Development of a dual structural basis of two-MAb cocktail for treating SARS-CoV-2 infections. *Nat Commun* 12, 264 (2021). https://doi.org/10.1038/s41467-020-20465-w

## THE IMMUNE EPITOPE DATABASE (IEDB)

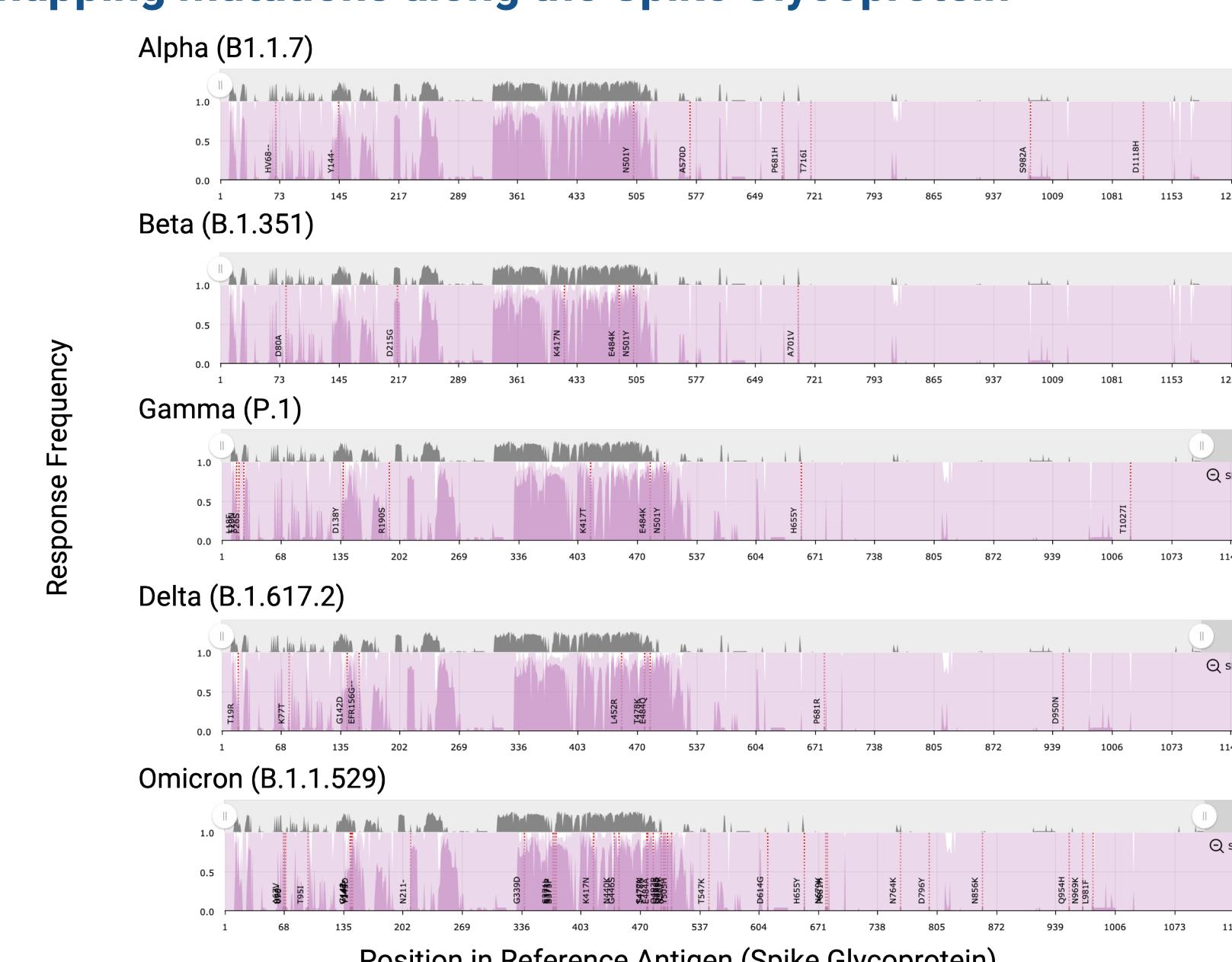
A successful query will enable the use of built-in tools. The Immunoome Browser is a powerful way to illustrate epitope data in both a linear view as well as a crystal structure powered by the iCN3D viewer. Users learn basics in protein structure, immunogenicity, as well as mutations.

### Immunoome Browser

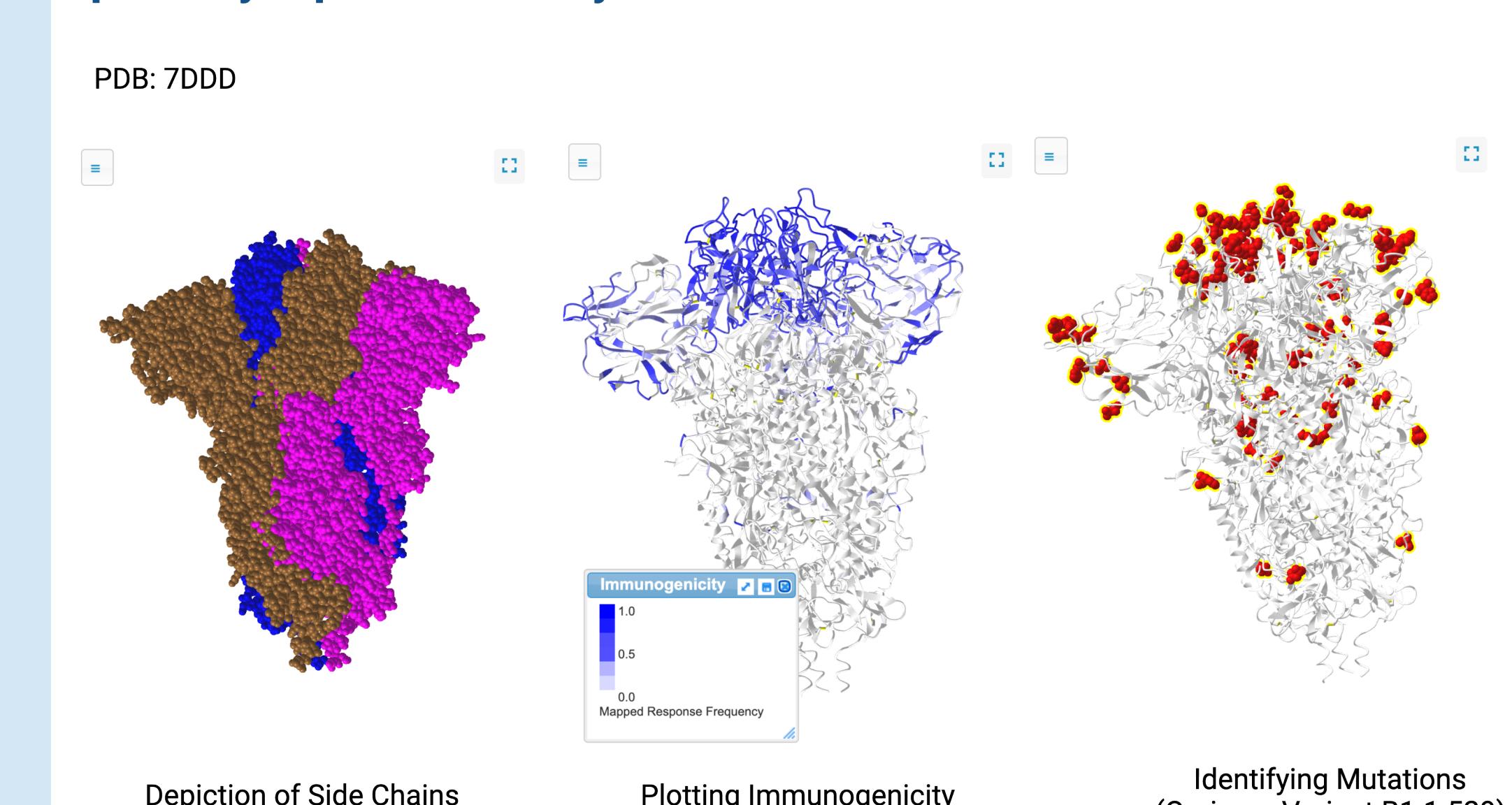
- Immunogenicity can be studied by epitopes observed to high response frequencies
- Data can be mapped onto other PDB structures if more than one is available.
- The default view does not map out immunogenic regions.
- Selecting an epitope in the table view will highlight the epitope on the structure in green.
- Subject responded refers to an assay resulting in either a positive or negative.
- Available file types: .XLSX, .CSV, .TSV, .JSON
- Variant data is also captured and can be mapped on both the plot and structure.



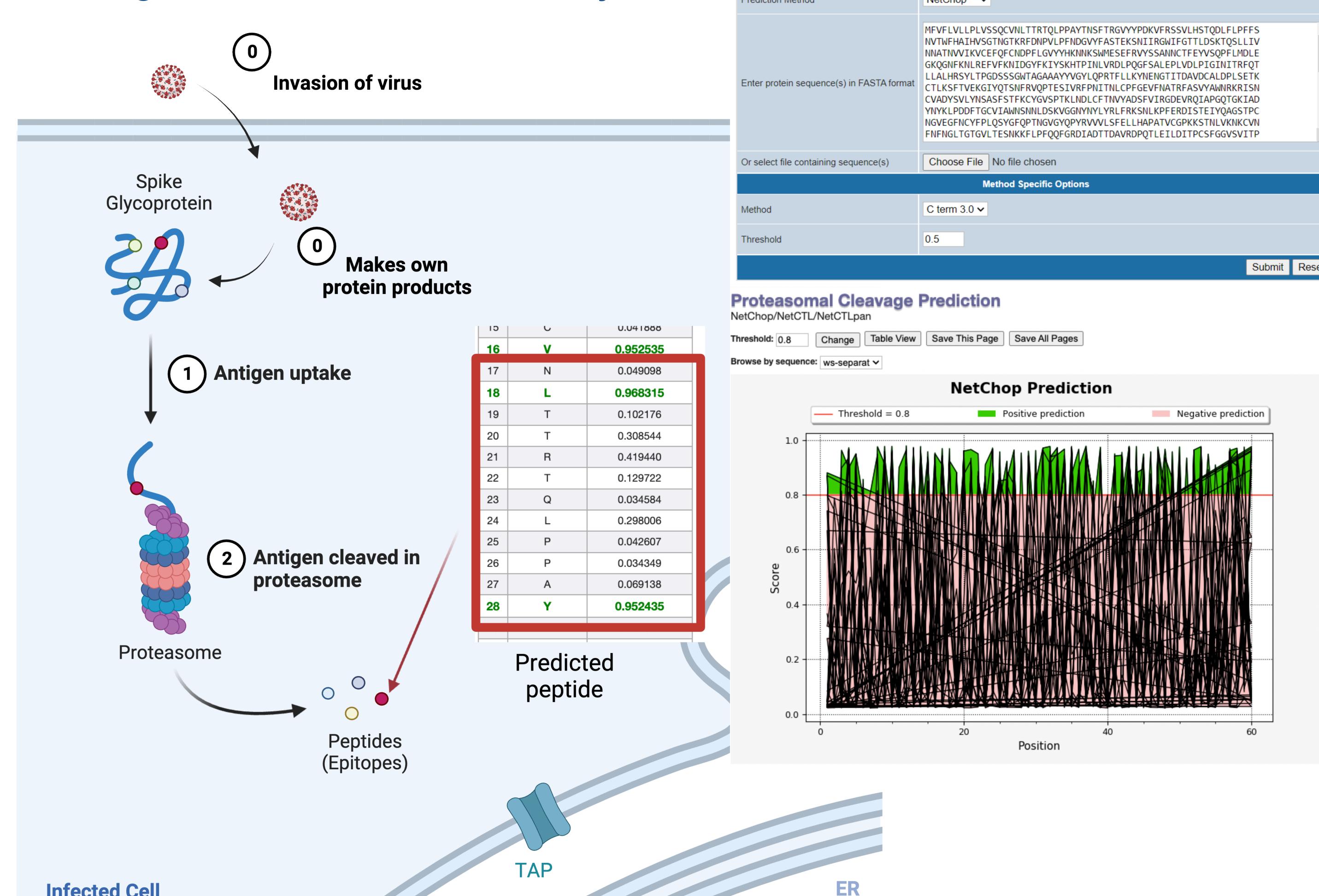
### Mapping mutations along the Spike Glycoprotein



### Visualization of immunogenicity and mutations of SARS-CoV2 Spike Glycoprotein on Crystal Structures



### Leveraging Data to predict proteasomal cleavages in MHC Class I Pathway



Aside from mapping out available data on epitopes, the IEDB also allows users to learn about more significant topics in immunology, like MHC I antigen presentation.

### Epitope Analysis Resource

Utilizing the epitope sequences collected on the database, the Epitope Analysis Resource has tools in T-cell Epitope Prediction, B-Cell Epitope Prediction, and other analysis tools.

NetChop, as seen in the MHC I pathway example, allows users to make predictions of proteasomal cleavage events, enabling users to identify peptides with the potential for antigen presentation. These guided lessons will help connect the tools to immunological and protein engineering topics.

## FUTURE DIRECTIONS

The development of digestible guides will enable users to learn how to carefully select data from databases and manipulate them with either web-based tools or even export the data to be analyzed offline. Epitope data is used in many fields beyond virology, and we hope users can apply these learnings in their studies to produce original research. In addition, we hope to create research opportunities accessible to those willing to learn.