

A Literature Review on Quantitative Assays in Viral Immunology

Jared Galloway — Fred Hutchinson CRC

November 29, 2020

Abstract

Protection from rapidly spreading and evolving viruses is key to human health and survival. The molecular nature of infection and respective immune system defences provides us with a complex and noisy set of problems to solve if we are to combat infectious disease and understand the nature of their evolution. Most notably, we need to understand the affinity of a virus to bind to a host cell via proteins expressed on the surface of the cell's outer membrane. These proteins allow the virus to enter a host cell, then proceed to hijack the cell's own machinery to replicate and propagate an infection within the host individual. In the context of SARS-CoV-2, understanding the immune response with respect to those binding proteins is critical for prevention and prediction of disease severity. Additionally, understanding the evolution of these binding sites allows us to interpret the cross-species transmission and duration of immunity post-infection, or the potential mutational routes of binding escape. Neither of these are trivial problems to solve. Fortunately, recent advances in next generation sequencing (NGS), oligonucleotide synthesis (ONS), and PCR-induced mutagenesis have driven the development of quantitative assays and given us the ability to explore and quantify fitness of particular sequences in the context of protein interaction. These methods have laid the foundation for exploration and development of advanced vaccines providing protection against deadly pathogens from causing serious illness and even death. In this literature review, we explore the techniques which leverage *quantitative assays*, Phage Immunoprecipitation Sequencing (PhIP-Seq) and Deep Mutational Scanning (DMS), which allow us to measure and explore binding properties viruses and thus, gain insight into the nature of a virus with respect to its *pathogenesis*. Concretely, we will provide background and motivation of these we will observe the results and methods from Shrock et al. [2020] and Starr et al. [2020], two papers that focus on the binding properties of the novel betacoronavirus, SARS-CoV-2 and its potential variants. The primary motivation behind the literature review is to provide and approachable explanation of novel techniques which are being used to rapidly characterize how we are combating novel viral outbreaks.

Primer & Prologue *You are welcome to skip this section if you're well versed the history and relevance of the Central Dogma and Basic Viral Biology. For me, priming myself with the basics before diving deep into complex topics like the ones described in this paper is always beneficial.*

Every organism on earth is the aggregate of many underlying biological systems which work together to drive the complex functions which fully characterize every individual. In more detail, each biological system is driven by a defined set of proteins – each encoded and regulated by a relatively small portion the individual's genetic code. Each protein has a role to play which may be trivial by itself, but in concert with the other relevant proteins we observe incredible functionality. The advent of Next Generation Sequencing (NGS) and complimentary algorithms has provided researchers with the ability to explore entire genome sequences from almost any living organism of interest. Even more impressively, we can profile these protein coding portions of the genome by extracting messenger ribonucleic acid (mRNA) from cells to infer the set of all expressed proteins, known as the transcriptome. In contrast to living organisms the genetic code of a virus is extremely simplistic. In the case of RNA viruses such as SARS-CoV-2, genomes are consisted mostly of protein coding RNA with a severely limited set of instructions. With no means to replicate or produce energy alone, the singular function of viral proteins is to bind to complimentary proteins expressed on host cells and hijack the machinery in order to reproduce itself - a process which then destroys the cell. Unfortunately in this specific context (along with many other facets of functional biology), little is known about how the underlying peptide sequences and relates to its folding properties (tertiary structure) and most importantly, its primary function.

Introduction

Modern mammalian immune systems are constituted by the aggregate of proteins and specialized cells known as *lymphocytes* which defend against unwanted invaders (pathogens). These defences keep the pathogens from harming the delicate and complex biological systems which keep us alive and healthy. However, deadly pathogenic outbreaks which rapidly spread among humans and other species can often harm or kill a large percentage of populations [Wu et al., 2020]. In the case of viruses, replication as a function of fitness drives pathogens to evolve much in the same way we do – often meaning the most potent and infectious pathogens prevail as a product of their genome evolution [Twiddy et al., 2003, Felsenstein, 1981]. Fighting fire with fire, the adaptive immune system works through similar processes of mutation and selection, inside our own body, to evolve along-side these pathogens and confer specialized immunity - in many cases this lasts throughout the lifetime of an individual. Incredibly, the combinatorial effects of specialized (VDJ) recombination results in enough diversity to select upon that evolution of specialized cells takes place in mere days (often a week or so) when encountering a new pathogen [Jung and Alt, 2004].

In contrast to all other forms of evolution (often on ecological timescales), The process of generating specific antibodies to ward off an infection is incredibly fast. Unfortunately, the symptoms of an infection during that time frame can still make an individual very ill, or even be fatal. The ability of viruses to hijack our cell's own machinery to replicate itself in order to propagate the infection make them efficient and deadly. Luckily, the process of producing antibodies need not occur every time we encounter the same virus. Rather, once an individual has encountered a pathogen and created the necessary cells needed to fend off the virus and infected cells, the defences that were used are stored in a sort of “immuno-memory” – using another type specialized cell. Upon contact with a pathogen the individual has encountered in the past, then, the immune system has the infrastructure in place to elicit a fast and effective response, Having this cellular machinery is what's known as immunity in an individual - and is key to survival in a world filled with microbial pathogens.

One of the most impactful developments in human health has been our ability to provoke immunity to common viruses without actually infecting us with a deadly disease causing pathogen. These biologically prepared agents are known as *vaccines*, and according to the center for disease control (CDC.gov) will have prevented over 21,000,000 hospitalizations and roughly 750,000 deaths among children born within the last 20 years – in the U.S alone. While this is an extreme success, the rate at which vaccines can be produced are a function of our ability to observe the physical properties of a virus. To date, the fastest a vaccine has been successfully developed was during the mumps outbreak in 1969 and took 4 years from start to finish. Facing a more deadly pathogen, of which we are certain exists, this slow rate of development could pose an existential threat to the human race.

Commonly, a vaccine for some particular virus essentially models the virus - without any of the harmful properties. This can be thought of as giving your immune system a molecular picture of the virus so that it is prepared when the real thing is encountered. Anything that elicits an immune response is known as an *antigen* – the antigen targeted by antibodies for a particular set of pathogens is known as the *epitope*. Knowing the epitope for any virus is key in modeling it for vaccines. Inferring a particular antigen is no trivial process, the number of possible peptides chains forming a protein which constitute the epitope for any particular virus are nearly infinitesimal [Stoddard et al., 2020]. To date there is no direct way to isolate which proteins are expressed on a virus, and which constitute an antigen. To complicate further, little is known about how sequence variation affects protein function. As viruses evolve, we would like to know how possible mutations impact our immuno-defences; In the case of the novel coronavirus, SARS-CoV-2, high mutation rates have already been found in the region which binds to our cells. In order to predict or understand how long immunity will last in the face of evolution, we must explore all variants of the epitope and their respective binding affinity relative to the wild type sequence.

Fortunately, recent advances in techniques such as next generation sequencing (NGS), oligonucleotide synthesis (ONS), modern computing and more, have opened the door to a new brute force, high throughput approach to answering these questions. The conjunction of these advancements tools have provided the infrastructure necessary to create and analyze *Quantitative assays* in immunology. Using the related protocols and analysis techniques, researchers now have the ability to explore the relevant likelihoods of nearly every single possible antigen – as well as every possible mutant of a particular antigen. These large scale studies require complex and carefully executed protocols resulting in noisy data for analysis [Mohan et al., 2018]. Once the data is acquired complex modeling and computing techniques are applied to parse the signal and produce some form of likelihood surrounding a particular sequence. This likelihood then informs a great deal about the biological nature of an evolutionary tango between

pathogens and the modern immune system. Quantitative assays in Immunology, particularly in the last decade, have laid the foundation for measuring protein interactions between a virus and individual host cells at magnitude far greater than previously thought possible [Fowler and Fields, 2014, Thyagarajan and Bloom, 2014].

Here, we dig into the benefits and limitations of two such quantitative assays, Phage Immunoprecipitation Sequencing (PhIP-Seq), and Deep Mutational Scanning (DMS) along with a hybrid technique, coined Phage-DMS. We will explore the methods, results, and analysis tools of two studies applied to proteins expressed by SARS-CoV-2. These studies will act as a template for understanding the biological insight provided by quantitative assays as a whole. First, we explore Shrock et al. [2020], a large-scale PhIP-Seq study done at Harvard university to characterize the set epitopes found in SARS-CoV-2. Next, we observe how every single mutation across one particular epitope on the novel coronavirus effects the affinity of binding to ACE2 receptor on human cells in Starr et al. [2020], where the authors used DMS. Finally, we see how these techniques are combined to offer insight into potential mutational pathways of escape from antibody recognition in Garrett et al. [2020]. Together, these studies motivate the use quantitative assays and associated analysis techniques in Immunology.

Quantitative Assays in the Context of SARS-CoV-2

The genome of SARS-CoV-2 consists of $\approx 30,000$ nucleotides which provide the landscape for 11 protein coding regions [Naqvi et al., 2020]. Of those, there exists a very small subset of less than 90 nucleotides encoding for an epitope identifying the virus and invoking an immune response able to combat the infection. Designing effective therapeutics and understanding how their effectiveness as the virus evolves over time provokes two primary questions (among many others): (1) Given the viral genome, how can we accurately infer the correct epitope which will act as the appropriate antigen and (2) how effective will the chosen antigenic sequence be in the face of inevitable genome evolution as a result of mutagenesis over time. In this section we explore two commonly used techniques which use quantitative assays in the field of viral immunology to explore these specific questions. For each of these we describe the methods and results, and follow up with future directions and limitations.

Phage-Immuno Precipitation Sequencing

Method Background Addressing the question of inferring the epitope of a particular virus, we turn to one of the more notable quantitative assay techniques, known as Phage Immunoprecipitation Sequencing (PhIP-Seq). Originally, Larman et al. [2011] introduced a this method which used oligonucleotide synthesis to create a synthetic representation of all possible epitopes. Importantly, the entire span of possible epitopes was created using a sliding window approach across the proteome and subsequently, those nucleotides were cloned into a special bacteria (T7 phage). It should be noted [Talk about size and "leap" of sliding window]. These such libraries are commonly referred to as *peptidomes* [Mohan et al., 2018]. Once cloned, the proteins of interest are then conveniently transcribed using the machinery of the phage and expressed on the exterior. This quantitative assay was then used to identify which human proteins were being mislabeled as pathogens, thus leading to auto-immune diseases such as Diabetes type I, multiple sclerosis, and rheumatoid arthritis.

Stepping back a little, the overarching concept of this technique is to create and label (barcode) a set of peptides of interest - this is the quantitative assay. Once developed, this assay is then presented (homogenized) with serum antibodies extracted from a patient of interest. Finally, the binding interactions we care about are extracted (precipitated) using specialized magnetic beads to be sequenced. When the samples are aligned with the library of oligos, we can count the number of each peptide which had a binding affinity with the antibodies. The resulting data from this process is then a counts array for which each count represents the number of binding events for each of the peptides in the library. When repeating this process with multiple samples, the arrays for each are merged to create what is most commonly referred to as the *enrichment matrix*. It then follows that large numbers in the counts matrix represent some underlying binding affinity of a particular antibody to a specific peptide. With respect to epitope detection in viruses of interest, this same technique is applied using the proteome of the virus to explore which proteins are targeted by antibodies produced during the infection of some patient or individual. Figure

Results from Recent Study using PhIP-Seq In Shrock et al. [2020], the immune (humoral) response of a mixed cohort of both COVID-19 positive and pre-pandemic individuals (232 and 190, respectively) was quantified by preseting antibody (*serological*) samples to a set of synthetic peptides libraries. The authors present three synthetic

peptidomes, each to provide a different scope and granularity of candidate epitopes across viral proteomes of interest. One of the libraries focused specifically on including peptides from all common human-infecting coronaviruses (HCoV's). Using this library, the authors found candidate SARS-CoV-2 epitopes which were specific to COVID-19 patients and more importantly, they found epitopes which were cross-reactive with antibodies which were likely developed in response to endemic coronaviruses from pre-pandemic samples. A "cross-reactive" epitope is defined by a binding event from a COVID-19 negative sample to a SARS-CoV-2 derived peptide. In regards to the outstanding question about why some individuals get so sick while some remain asymptomatic, cross-reactive epitopes suggest that an individual's exposure history to certain HCoV's determine the outcome of their COVID-19 infection. `¡SAY SOMETHING ABOUT THE CROSS REACTIVE EPITOPS¡`

Additionally, the authors presented a machine learning model that used a specialized z-score the peptide enrichments as input and predicted the infection status of COVID-19 with 99% sensitivity and 98% specificity. Using the XGBoost random forest approach, the model was trained using a zscore approach to normalizing the peptide enrichments across samples. `¡DEMPGRAPHICS¡ ¡MUTANT LIBRARY¡`

Deep Mutational Scanning

Method Background Understanding how underlying amino acid sequences relates to the function (*phenotype*) of proteins remains at large in many facets of biology. In particular, we are interested in the process by which a pathogen leads to diseased state in an individual - known as *pathogenesis* [Araya and Fowler, 2011, Fowler and Fields, 2014, Weile and Roth, 2018]. We know, for instance, that more than half of human disease is caused by sequence mutations, or single nucleotide polymorphisms (SNPs) [Stenson et al., 2009]. Unfortunately, the lack of ability to map a sequence to the respective phenotype poses a large problem when developing therapeutics or predicting the pathogenic outcome of mutation in a pathogen. To address the difficulty, [Araya and Fowler, 2011] popularized a method which uses mutagenesis to instead explore how sequence variation changes the binding properties of protein when compared to the wildtype protein of interest – this method is known as Deep Mutational Scanning (DMS). Taking advantage of the advancement and scaling of next generation sequencing, this method generates a quantitative assay containing every possible amino acid substitution along the protein of interest and then infers how each performs outside of a living organism (in vitro). Presented in Adams et al. [2016], the authors show how this brute force approach to mutagenesis and subsequent measuring of binding affinity of an antibody to an antigen of interest allow us to explore the function of a protein at the sequence level.

In the context of viruses, each virus infects cells by attaching to a protein on the surface of the host cell via a binding site of its own. It then follows that mutations to the binding sites on these viruses would regulate its affinity to bind, and thus, regulate the infectious properties of the virus. For example, the reason we have seasonal flu viruses. In Thyagarajan and Bloom [2014], the authors explore

Results from a recent study using DMS With the knowledge that the entry receptor for SARS-CoV-2 is the angiotensin converting enzyme 2 (ACE2) – and the Spike protein carries the most prominent candidates for the epitope – it would be very relevant to understand how mutations on the spike protein impact the mutant's binding affinity to the ACE2 receptor. Given

Phage-DMS

Future Perspectives and Conclusions

References

- Rhys M Adams, Thierry Mora, Aleksandra M Walczak, and Justin B Kinney. Measuring the sequence-affinity landscape of antibodies with massively parallel titration curves. *eLife*, 5:e23156, dec 2016. ISSN 2050-084X. doi: 10.7554/eLife.23156. URL <https://doi.org/10.7554/eLife.23156>.
- Carlos L. Araya and Douglas M. Fowler. Deep mutational scanning: assessing protein function on a massive scale. *Trends in Biotechnology*, 29(9):435 – 442, 2011. ISSN 0167-7799. doi: <https://doi.org/10.1016/j.tibtech.2011.04.003>. URL <http://www.sciencedirect.com/science/article/pii/S0167779911000692>.

- J Felsenstein. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.*, 17(6): 368–376, 1981. ISSN 0022-2844. URL <http://www.ncbi.nlm.nih.gov/pubmed/7288891>.
- Douglas M. Fowler and Stanley Fields. Deep mutational scanning: a new style of protein science. *Nature Methods*, 11(8):801–807, Aug 2014. ISSN 1548-7105. doi: 10.1038/nmeth.3027. URL <https://doi.org/10.1038/nmeth.3027>.
- Meghan E. Garrett, Jared Galloway, Helen Y. Chu, Hannah L. Itell, Caitlin I. Stoddard, Caitlin R. Wolf, Jennifer K. Logue, Dylan McDonald, Frederick A. Matsen, and Julie Overbaugh. High resolution profiling of pathways of escape for sars-cov-2 spike-binding antibodies. *bioRxiv*, 2020. doi: 10.1101/2020.11.16.385278. URL <https://www.biorxiv.org/content/early/2020/11/16/2020.11.16.385278>.
- David Jung and Frederick W. Alt. Unraveling v(d)j recombination: Insights into gene regulation. *Cell*, 116(2): 299–311, Jan 2004. ISSN 0092-8674. doi: 10.1016/S0092-8674(04)00039-X. URL [https://doi.org/10.1016/S0092-8674\(04\)00039-X](https://doi.org/10.1016/S0092-8674(04)00039-X).
- H. Benjamin Larman, Zhenming Zhao, Uri Laserson, Mamie Z. Li, Alberto Ciccia, M. Angelica Martinez Gakidis, George M. Church, Santosh Kesari, Emily M. Leproust, Nicole L. Solimini, and Stephen J. Elledge. Autoantigen discovery with a synthetic human peptidome. *Nature biotechnology*, 29(6):535–541, May 2011. ISSN 1546-1696. doi: 10.1038/nbt.1856. URL <https://pubmed.ncbi.nlm.nih.gov/21602805>. 21602805[pmid].
- Divya Mohan, Daniel L. Wansley, Brandon M. Sie, Muhammad S. Noon, Alan N. Baer, Uri Laserson, and H. Benjamin Larman. Phip-seq characterization of serum antibodies using oligonucleotide-encoded peptidomes. *Nature Protocols*, 13(9):1958–1978, Sep 2018. ISSN 1750-2799. doi: 10.1038/s41596-018-0025-6. URL <https://doi.org/10.1038/s41596-018-0025-6>.
- Ahmad Abu Turab Naqvi, Kisa Fatima, Taj Mohammad, Urooj Fatima, Indrakant K. Singh, Archana Singh, Shaikh Muhammad Atif, Gururao Hariprasad, Gulam Mustafa Hasan, and Md Imtaiyaz Hassan. Insights into sars-cov-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. *Biochimica et biophysica acta. Molecular basis of disease*, 1866(10):165878–165878, Oct 2020. ISSN 1879-260X. doi: 10.1016/j.bbdis.2020.165878. URL <https://pubmed.ncbi.nlm.nih.gov/32544429>. 32544429[pmid].
- Ellen Shrock, Eric Fujimura, Tomasz Kula, Richard T. Timms, I-Hsiu Lee, Yumei Leng, Matthew L. Robinson, Brandon M. Sie, Mamie Z. Li, Yuezhou Chen, Jennifer Logue, Adam Zuiani, Denise McCulloch, Felipe J. N. Lelis, Stephanie Henson, Daniel R. Monaco, Meghan Travers, Shaghayegh Habibi, William A. Clarke, Patrizio Caturegli, Oliver Laeyendecker, Alicja Piechocka-Trocha, Jon Li, Ashok Khatrri, Helen Y. Chu, , Alexandra-Chloé Villani, Kyle Kays, Marcia B. Goldberg, Nir Hacohen, Michael R. Filbin, Xu G. Yu, Bruce D. Walker, Duane R. Wesemann, H. Benjamin Larman, James A. Lederer, and Stephen J. Elledge. Viral epitope profiling of covid-19 patients reveals cross-reactivity and correlates of severity. *Science*, 2020. ISSN 0036-8075. doi: 10.1126/science.abd4250. URL <https://science.sciencemag.org/content/early/2020/09/28/science.abd4250>.
- Tyler N. Starr, Allison J. Greaney, Sarah K. Hilton, Daniel Ellis, Katharine H.D. Crawford, Adam S. Dingens, Mary Jane Navarro, John E. Bowen, M. Alejandra Tortorici, Alexandra C. Walls, Neil P. King, David Veisler, and Jesse D. Bloom. Deep mutational scanning of sars-cov-2 receptor binding domain reveals constraints on folding and ace2 binding. *Cell*, 182(5):1295 – 1310.e20, 2020. ISSN 0092-8674. doi: <https://doi.org/10.1016/j.cell.2020.08.012>. URL <http://www.sciencedirect.com/science/article/pii/S0092867420310035>.
- Peter D. Stenson, Matthew Mort, Edward V. Ball, Katy Howells, Andrew D. Phillips, Nick ST Thomas, and David N. Cooper. The human gene mutation database: 2008 update. *Genome Medicine*, 1(1):13, Jan 2009. ISSN 1756-994X. doi: 10.1186/gm13. URL <https://doi.org/10.1186/gm13>.
- Caitlin I. Stoddard, Jared Galloway, Helen Y. Chu, Mackenzie M. Shipley, Hannah L. Itell, Caitlin R. Wolf, Jennifer K. Logue, Ariana Magedson, Kevin Sung, Meghan Garrett, Katharine H.D. Crawford, Uri Laserson, Frederick A. Matsen, and Julie Overbaugh. Epitope profiling reveals binding signatures of sars-cov-2 immune response and cross-reactivity with endemic hcovs. *bioRxiv*, 2020. doi: 10.1101/2020.10.29.360800. URL <https://www.biorxiv.org/content/early/2020/10/29/2020.10.29.360800>.
- Bargavi Thyagarajan and Jesse D Bloom. The inherent mutational tolerance and antigenic evolvability of influenza hemagglutinin. *eLife*, 3:e03300, jul 2014. ISSN 2050-084X. doi: 10.7554/eLife.03300. URL <https://doi.org/10.7554/eLife.03300>.

- S. Susanna Twiddy, Edward C. Holmes, and Andrew Rambaut. Inferring the Rate and Time-Scale of Dengue Virus Evolution. *Molecular Biology and Evolution*, 20(1):122–129, 01 2003. ISSN 0737-4038. doi: 10.1093/molbev/msg010. URL <https://doi.org/10.1093/molbev/msg010>.
- Jochen Weile and Frederick P. Roth. Multiplexed assays of variant effects contribute to a growing genotype–phenotype atlas. *Human Genetics*, 137(9):665–678, Sep 2018. ISSN 1432-1203. doi: 10.1007/s00439-018-1916-x. URL <https://doi.org/10.1007/s00439-018-1916-x>.
- Fan Wu, Su Zhao, Bin Yu, Yan-Mei Chen, Wen Wang, Zhi-Gang Song, Yi Hu, Zhao-Wu Tao, Jun-Hua Tian, Yuan-Yuan Pei, Ming-Li Yuan, Yu-Ling Zhang, Fa-Hui Dai, Yi Liu, Qi-Min Wang, Jiao-Jiao Zheng, Lin Xu, Edward C. Holmes, and Yong-Zhen Zhang. A new coronavirus associated with human respiratory disease in china. *Nature*, 579(7798):265–269, Mar 2020. ISSN 1476-4687. doi: 10.1038/s41586-020-2008-3. URL <https://doi.org/10.1038/s41586-020-2008-3>.

Supplementary Figures