The UMN/Mayo Computational Human Immuno-Peptidome (CHIP) Project

Project Details:

1. **Provide details of your project:**

Research project name: UMN-Mayo Computational Human Immuno-Peptidome (CHIP) Project
Project expected start date:7/1/22
Proposed project duration (up to 12 months): 6/30/23
Project executive summary (3-4 sentence overview of your project): This project tackles a grand challenge in computational science: predicting how strongly peptides derived from viral proteins will bind to cell-surface molecules as part of our immune response. The set of all the peptides that bind strongly is called a person's immunopeptidome. This set is unique and determines the capacity of their immune system. The immune response to a virus such as SARS-CoV-2 hinges on whether the viral protein fragments bind into a groove in these cell-surface proteins – like a key into a lock. The problem is that simulating the binding of a single pair of molecules takes hours, or even days, on a powerful computing cluster with existing approaches. There are millions of pairs. In this research we will develop new, highly targeted algorithms to make the computation tractable. We will turn billions of days of computing into millions of minutes, and solve the problem by deploying the algorithms at scale on Oracle's cloud-computing infrastructure, completing the task in weeks.
Research public relevance, project description and goals: **The Grand Challenge** Perhaps more so than for any other discipline of science or engineering, the history of computer science has been one in which problems seem impossible until suddenly they are not. Translating languages, playing chess, recognizing objects, driving cars – all at first seemed so daunting that experts gave these as examples of tasks that computers might never be able to do as well as humans. Of course, none of these problems were solved "suddenly" with a single keystroke. In all cases, it was decades worth of concerted research, down blind alleys and with major paradigm shifts, that brought solutions. Raw computing power alone did not solve these problems; however, the tremendous increase in computing power that preceded these breakthroughs enabled them. This proposal discusses a problem that computer science currently judges to be very difficult. It is a foundational problem in computational immunology which, if solved, could inform

predictions of disease severity, enable treatments, and guide vaccine development. While such aspects of a pandemic response ultimately depend on experimental knowledge and trials, the computational results that we are proposing could play a critical role at two ends of the time spectrum:

- In the early phases, when identifying and characterizing the threat a new pathogen.
- In the long term, to develop a deep understanding of the molecular mechanisms of the infection.

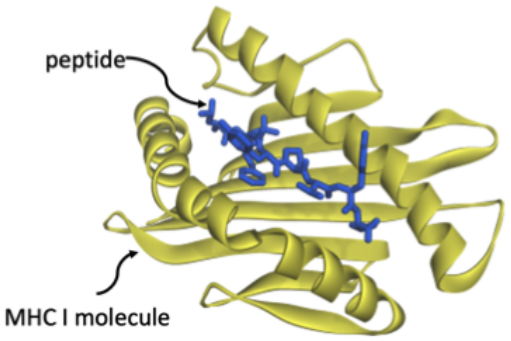


Figure 1: A peptide binds to an MHC I molecule if it fits into the cleft, like a key into a lock. There are perhaps 38,000 distinct peptides derived from proteins for a virus such as SARS-Cov-2. An individual has up to 6 variants of MHC I molecules. There are perhaps 21,000 variants of MHC I molecules in the human population.

Narrow Statement of Grand Challenge Problem

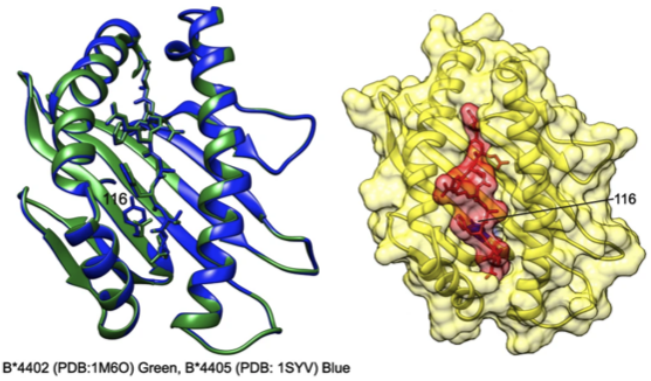
Stated succinctly, the computational problem that we will tackle is determining how strongly a given molecule binds to another. The given molecule is a peptide – a fragment of a protein derived from a pathogen, such as a virus. The other is a molecule called Major Histocompatibility Class I (MHC I) that is expressed on the surface of most of our cells. MHC I molecules have a cleft into which a peptide can bind. As illustrated in Figure 1, a peptide will only bind if it fits into the cleft like a key into a lock. The binding of a peptide to an MHC I molecule is a critical step in a critical component of the immune system, so-called cellular immunity.

Explained succinctly, cellular immunity allows circulating T-cells to kill off infected cells. When a cell is infected with a virus, it hijacks the host cell's machinery, forcing it to make viral proteins. Our cells have a defense mechanism: they chop up such proteins into fragments, called peptides, and transport them to the cell surface, bound to MHC I molecules. Presented this way on the cell surface, T-cells can identify a cell as being infected and can destroy it using toxins. If this mechanism succeeds, an infection is stopped in its tracks: T-cells kill off infected cells before they can do damage. If it fails, then infected cells become factories for re- producing copies of the virus and full-blown disease results.

Success or failure depends on whether peptides derived from the viral proteins bind to the MHC I molecules and so become targets. Binding depends on the biochemical affinity between the constituent building blocks of the pair of molecules. This, in turn, largely depends on molecular shape: how well the metaphorical key fits into the metaphorical lock. There are many variants of MHC I molecules, coded for by a person's genes. These vary, sometimes subtly, in shape.

The set of all the peptides that can bind to a person's MHC I molecules is called their immunopeptidome. This set is unique and determines the capacity of their immune system.

Since the immune response of a person to, for instance, a viral infection like COVID-19 is dependent on whether their MHC I molecules present peptides derived from the virus, understanding and predicting the binding step is an important topic.



B*4402 (PDB:1M6O) Green, B*4405 (PDB: 1SYV) Blue

Figure 2: Inferring the structure of an MHC I molecule by homology and through protein-folding. **On Left**: Comparison of MHC I molecule B*4402 (in green) and MHC I molecule B*4405 (in blue). The root-mean squared distanced between these two structures is 0.3 Å. The structure of the green molecule was inferred by homology from the structure of the blue molecule.
On Right: Combined ribbon and surface representation of the green MHC I molecule, after folding, with the peptide 116 from SARS-Cov-2 bound in the cleft.

What we are proposing here appears to be a narrowly defined problem: characterizing the binding strength between specific pairs of molecules. Most aspects of the molecular biology are well understood. And yet, we argue that the problem qualifies as a grand challenge, in terms of its difficulty and in terms of the impact of a solution. The difficulty lies with the computational requirements, stemming from the combinatorial scale of the problem. The impact of the solution will stem from an ability to precisely characterize, in advance and through purely computational means, how well a person's cellular immunity will cope with a novel pathogen.

Simulating molecular interactions has been a widely studied and largely successful topic for perhaps five decades. Indeed, some of the earliest computers were applied to this problem. Sophisticated software exists to simulate molecular binding events. The conventional approach with such tools is to simulate binding from first principles, tracking the trajectories of all the atoms in all the molecules in three-dimensional space, numerically solving Newton's equations of motion. A variety of strategies are used to find low-energy configurations, including randomization, with so-called Monte Carlo methods. The problem is that simulating a single peptide-MHC I molecular binding takes hours, or even days, on a powerful computing cluster with this approach.

The SARS-CoV-2 virus, for example, has 29 distinct viral proteins. When chopped up, this translates into approximately 38,000 peptides. This is not an unmanageable number. However, the other side of the equation consists of the MHC I molecules. Every person has up to 6 variants, having inherited 3 from each parent. There are at least 21,000 variants in the human population. Indeed, the genes that code for MHC I molecules are the most diverse in our genome. Evolution has ensured this, as humans and pathogens have been co-evolving together.

So, in the narrow formulation of the research problem, there are 38,000 peptides for a virus like SARS- Cov-2 each paired with 21,000 variants of MHC I molecules in the human population. This

translates to three-quarters of a billion distinct pairings. If one is using existing simulation software, which requires hours or even days of computing time per pairing, one is confronted with billions of hours, or billions of days, of computing time -- clearly an intractable proposition.

With this grant, we will develop new, highly targeted algorithms to make such computation tractable. While existing software for these sorts of atomic simulations is sophisticated, it is general-purpose, written in FORTRAN decades ago. The most widely used software packages have been written to simulate molecular interactions of nearly any type, from crystals, to proteins, to polymer chemistry. Others specifically simulate protein binding. Observing simulation trials for such general- purpose software, most of the computational time is spent moving molecules randomly in space, looking for energetically favorable states; most of the random movement is wasted for this particular problem. Only the final steps, as the peptide settles into an optimal configuration in the cleft of the MHC I molecule, matter. There is domain-specific knowledge here that can really help.

Broader Statement of Grand Challenge Problem

The grand challenge that we are positing is much broader than simply writing better algorithms and deploying the code on supercomputing clusters, although this will be a significant aspect of the effort. The problem statement is broader in two main respects:

1. Not all aspects of the biochemistry of binding are understood or have been well characterized from a computational perspective.
2. Structural models do not exist for novel peptides, nor for most variants of MHC-I molecules that one encounters in the human population.

Needless to say, the biochemistry of the immune system is a complex and vast topic of study by a large community of experimentalists. We can point to the work of Prof. Mark Davis at Stanford. His lab has been striving to understand the structural and biochemical underpinnings of peptide binding for decades.

Our focus is applied, translational, and computational. We aim to incorporate knowledge from structural and molecular biology into efficient computational models and deliver useful predictions, at scale. The goal of this grant is not to develop new, experimental or biochemical knowledge of the immune system, but rather to synthesize and apply knowledge as it evolves.

Beyond detailed and accurate bio-chemical models of binding, a significant challenge for this research is the availability of models. On the one hand, viral proteins are readily characterized. In the case of SARS-Cov-2, the virus was first identified in Dec. 2019, By February 2020, the amino acid sequence of its proteins had already been published. Viral proteins get chopped into fragments called peptides by intracellular mechanisms, each 8 to 15 amino acids in length. Given a novel pathogen, most of these peptides will be new to science, having never been encountered before. (Similarly, such peptides will be new to an individual’s immune system!)

The first challenge is to construct structural models for novel peptides. The second, more significant challenge is to construct structural models for MHC I molecules. As noted above, there is tremendous diversity in these molecules, with perhaps 21,000 variants in the human population. Only a small fraction of these have been characterized experimentally. Doing so entails significant effort, with experimental techniques such as crystallography and mass spectrometry. This has led to an unfortunate social consequence: nearly all of the variants of MHC I molecules that have been characterized experimentally are those found predominantly in people of European descent. Addressing this inequity will be a significant focus of our activities.

This project will pursue a novel strategy for constructing structural models of variants of MHC I molecules that are not available. Instead of waiting for them to be characterized experimentally, we will construct the models de novo. We will do so first by homology: beginning with a structural model for an MHC I molecule that it most closely resembles, we will construct a model for a new MHC I molecule by substituting amino acids.

This is illustrated in Figure 2. Here we will apply domain-specific expertise, provided by Prof. James Cornette from Iowa State University. (He is not a PI on this grant, but he is a close collaborator.) While all MHC I variants have different shapes, their overall structure is similar; differences are primarily in the location of side chains. These structural differences can be inferred.

Given a de novo model of an MHC I molecule that is accurate in terms of its atomic configuration, we will apply computing power to fold it into its actual shape. Here we will make use of the latest breakthrough, AlphaFold, an AI-based solution recently announced by Google's Deepmind project.

Perhaps the most ambitious aspect of this project, and the aspect that qualifies it as a Grand Challenge, is translating the computational modeling into practice. Of course, we emphasize that the computational challenge is significant; it will require the deployment of Oracle's cloud power to be realized. Characterizing the immuno-peptidome will not be a separate activity from applications. Rather, there will be tight synergy between the computing team and the practitioners.

2. **Please summarize any prior work in this research area and state how this project will build upon and/or differentiate itself from existing work:**

Prior Work and Expertise

The proposed research is targeted and focused on delivering outcomes that will aid in pandemic preparedness. However, our approach is ambitious, aiming to develop a very general capability in computational immunology. It builds upon expertise and prior research in the Vasmatzis lab at the Mayo Clinic in genomics and computational techniques for cancer immunotherapy; work in

the Block lab at the Mayo Clinic on patient immune response monitoring for SARS-CoV-2; as well as expertise and prior research by Julia Udell in graft-versus-host predictions at “Be-the-Match.”

A premise is the deployment of high-performance computing. Prof Riedel’s experience with circuit design and molecular dynamics will be brought to bear on this aspect. This project will provide the computational infrastructure to predict whether peptides derived from viral proteins will bind to allelic variants of MHC I molecules. The same computational infrastructure could be deployed for other pathogens. It could also be transformative in other contexts, for instance for treatments of cancer via immunotherapy as well as for the treatment of autoimmune diseases. Capitalizing on our past experience in modeling MHC-peptide complexes; monitoring of immune responses to peptide vaccines; our ability to develop fast computational algorithms and pipelines; and our access to clinical data, we are poised to develop transformative tools and deliver critical information for pandemic preparedness, adhering to a tight timeline.

We have assembled a strong multidisciplinary team that, combined, has both breadth and depth of expertise in computational biology, genomics/bioinformatics, immunology, graphics, computer science and high-performance computing.

Dr. Vasmatzis from the **Mayo Clinic** and **Prof. Jim Cornette** from **Iowa State University** have been involved in pioneering work related to peptide-MHC binding predictions, TCR and antibody structure prediction, and modeling statistical potentials of atomic interactions. The Vasmatzis lab at Mayo Clinic also has expertise in developing highly sophisticated genomics pipelines. For example, the group has developed a pipeline that allows the accurate determination of tumor-specific neoantigens based on tumor-specific DNA junctions that are often the source of neopeptides in tumors. Their technique, called MPseq, detects with high sensitivity, specificity and cost-effectiveness many complex rearrangement events such as chromoplexis and chromothripsis. These truncate highly expressed genes and result in altered protein sequence juxtaposed on normal truncated proteins. Their group at Mayo has used this technique for mesothelioma cases that exhibit a higher potential of rearrangements to produce neoantigens compared to single nucleotide variants. Also, the group at the Mayo Clinic has world-class expertise in genomics and sequencing, as well as access to patient data sets needed to validate the computational results.

Professor Riedel from the **University of Minnesota** has extensive experience with molecular computing that can be brought to bear on the research. Funded by seven major NSF grants, he has spear-headed the development of novel computing constructs with DNA. He currently has a DARPA grant to develop DNA storage systems. This research is predicated on algorithmic expertise in molecular simulation and design, adapted by Prof. Riedel from the realm of electronic circuit design. The circuit-design community has unique expertise that can be brought to bear on the challenging computational problems encountered in molecular simulation. Applications in molecular biology, in turn, offer a wealth of interesting problems in modeling and algorithmic development.

• Prof. Riedel and Dr. Vasmatzis co-advise **Julia Udell**, a Ph.D. student in the Bioinformatics and Computational Biology program at the University of Minnesota, who will play a significant role in this project. She has both computational and immunological expertise, having worked as a bio-statistician for Stanford's HLA typing lab prior to beginning her doctorate. She is the author of the neoantigen-ranking algorithm and will take the lead in applying this algorithm to the SARS-CoV-2 proteome.

• **Dr. Matthew Block** is an immunologist and a medical oncologist at the **Mayo Clinic** with an interest in understanding the mechanisms that influence anti-tumor immunity in patients with melanoma and ovarian cancer. His research efforts have focused on preclinical translational studies and therapeutic clinical trials testing novel cancer vaccines and immunotherapies. As part of his efforts to identify novel immunotherapy approaches to cancer, his laboratory has developed T cell and antibody-based assays to measure changes in antigen-specific immune responses using patient samples. With the onset of the COVID-19 pandemic, he has developed methods to measure immune responses to SARS-CoV-2.

We have relevant expertise from the participants at University of Minnesota and the Mayo Clinic. We point to collaborative work that we have recently done on an important aspect of the binding problem, hydrophobicity, as an example. Hydrophobicity plays an important role in peptide:MHC I binding, yet has not been explicitly considered in computational models. We have shown how to incorporate it and how this improves binding prediction.

Our "grand challenge" problem is ambitious both from the standpoint of computation as well as in terms of the requisite expertise in immunology. Our team is uniquely positioned for this challenge: the researchers at Mayo have expertise in molecular modeling and immunology, as well as access to clinical data; the researchers at UMN have complementary strengths in molecular simulation, as well as experience with developing and deploying large-scale computation projects.

Existing Tools

To our knowledge, no one has attempted to simulate peptide binding at the level of physical chemistry, at scale, pairing tens of thousands of peptides with tens of thousands of variants of MHC I molecules. Rather, people have turned to neural networks.

We point to three packages that perform exactly the predictions that we are discussing: NetMHC, PickPocket, and SYFPETHI. These tools have been used to study cancer immune escape mechanisms, checkpoint blockade immunotherapy for tumors, and identifying T-cell response targets. All are efficient, returning binding predictions in a matter of seconds for queries. So, it might seem that the grand challenge that we are posting has already been solved.

Unfortunately, it has not. These tools are trained with one-dimensional data: text labels for MHC I molecules, paired with amino acid letter sequences of peptides, scored according to the experimentally observed binding strength. So, these neural networks are trained on textual data.

The network predicts how strongly a novel peptide will bind to a given MHC I molecule, according to the similarity of the amino acid sequence only. No information regarding molecular shape or binding chemistry is used.

These tools are valuable; we have used them extensively in our research. However, the predictions that they provide are coarse:

1. The neural networks are trained without any data on molecular shape, and without any reference to the underlying physical chemistry. The inferences that they provide are based on peptide amino acid sequence only. However, the peptides have three-dimensional shapes. Small differences in amino acid sequence can translate to very different shapes and very different binding affinities.

2. The neural networks are trained on experimental data that comes from a wide variety of domains. For instance, a large fraction of the data for NetMHC comes from studies of proteins derived from the HIV virus. However, peptides from a novel pathogen such as SARS-Cov-2 might bear little similarity to these. Most will be new to science. Neural networks perform statistical inference, interpolating to produce answers. Inferring from data that is too dissimilar from the target generally yields poor results.

Indeed, acknowledgement that neural networks provide spurious inferences for peptide binding strength is widely acknowledged. In particular, the tools seem to deliver many false positives. It is possible that machine learning and neural networks are the right way to attack this problem. (We are considering such techniques in our approach.) However, training on the letter sequence of amino acids simply cannot provide reliable answers to complex questions pertaining to physical chemistry, such as this one. One must incorporate molecular shape and biochemical aspects of binding into the modeling.

Novelty on Our Approach

With this grant, we will develop and apply a new mechanistic model for predicting peptide-MHC I binding. In contrast to general-purpose software for molecular simulations, such as CHARMM and Amber, ours will be specifically optimized for this problem.

The starting point is a three-dimensional molecular model of both the peptide and the MHC I molecule. We will make use of structural models of MHC I variants that have been characterized experimentally, through crystallography. When such a structural model is not available, we will infer it by homology: we will construct it starting from the structural model that it most closely resembles, substituting structural models of amino acids where it differs. Then we will make use of existing software for simulating protein folding. So, the starting point for our calculations will be optimal folded structures for the MHC I molecules. We will follow the same strategy for generating structural models for the peptides. (These are much shorter, with fairly simple molecular structures, so this is a much easier task.)

Next, in each simulation, we will position the peptide roughly aligned in the binding pockets. This is likely the most difficult step, as the optimal position of the peptide might not be known. Here, we will endeavor to incorporate as much domain-specific knowledge as possible. We anticipate that intelligent placement of the peptide is the single most important factor in reducing computation time.

For binding strength, we will implement energy calculations based on a variety of biochemical factors: electrostatic interaction, acidic/basic pH, hydrogen bonds, van der Waals forces, shape complementarity, hydrophobicity, p-interaction, steric effects, and solvation energy.

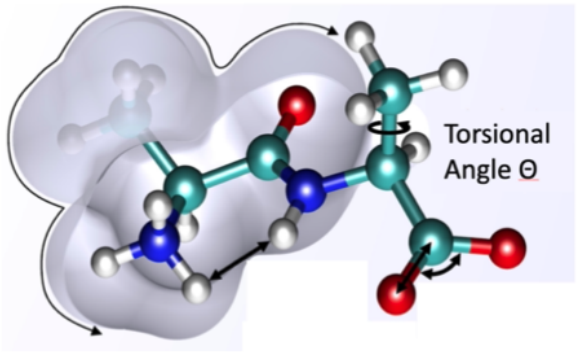


Figure 3: We will explore moves in the torsional space to find minimum-energy configurations. Such moves are much more efficient than moves in the general cartesian space.

Finally, we will perform a rigorous search for a minimum-energy binding configuration. Compared to existing methods in general-purpose software, we will reduce the dimensional space by fixing bond lengths and amino acids sidechains in advance. Thus, the only variables that we will manipulate will be the dihedral angles along the backbones, as well as the dihedral angles of the amino acid sidechains. If these moves are not sufficient, we will also flex the backbone of the MHC I molecule.

Instead of carrying out the search in a space with cartesian coordinates, we will do most of the molecular maneuvering in the torsional space, as shown in Figure 3. That is to say, we will rotate sidechains instead of randomly displacing and flexing them. Moving in the cartesian space necessitates tracking 3 x 19 = 57 variables per amino acid; however, moving in the torsional space necessitates tracking only 3 dihedral angles per amino acid. Furthermore, we can restrict moves to just the residues present in the binding pockets of the MHC I molecule, reducing the 275 amino acids per MHC I molecule to no more than 70. We can only justify this claim by developing and deploying the code, but we anticipate that by devising efficient, custom algorithms, we can turn one billion days of simulation time into one million minutes for our grand challenge problem, where we are looking at SARS-Cov-2 binding predictions, with 38,000 peptides paired with 21,000 MHC I variants.

Applying the Computational Toolset

As a follow up, using the data and algorithms discussed in the prior sections, we will identify commonly occurring haplotypes in the U.S. population that may make individuals vulnerable to COVID-19. We will apply the algorithms that we have developed to predict viral peptide binding to variants of MHC I molecules across populations. We will reduce the combinatorial space by concentrating on the low-mutational rate regions of the viral proteins as well as the most frequent MHC I alleles found in a given population. If we set cutoff thresholds, the numbers

become much more manageable. For example, there are only 374 variants of MHC I molecules that occur with a frequency above one-in-one-thousand in the U.S. population; only 77 are above 2% frequency in the U.S. population. These are considerably smaller sets than the 21,000 possible variants discussed above.

Using the algorithms developed, we will identify the variants of MHC I molecules that commonly occur in U.S. populations, as reported by. We will then identify those that may confer more, and those that may confer less, protection that average against COVID-19. According to our hypothesis, individuals with variants of MHC I molecules to which some viral peptides bind strongly are likely to mount an effective immune response. Those with variants to which all viral peptides bind only weakly are not. With this data, we can make predictions across sample populations. These will be validated against any clinical data that is available. (No new clinical data will be generated.)

3. **Provide expected milestones (milestones should outline expected progress at regular intervals throughout the project):**

Milestone #1 summary: Predict disease severity for a new pathogen for different individuals.
Estimated completion date: 9/1/22
Description:
Applications: Early Response & Triaging. Translating the computing results to practice will generally entail a focus on the individual. As explained above, every person inherits up to 6 distinct variants of MHC I molecules, three from each parent. A form of genetic testing called human leukocyte antigen (HLA) typing can be performed to establish which variants a person has. This type of testing is convenient, widely available, and inexpensive, as it is used for paternity testing. Performing such tests on a large group, say everyone at risk in a pandemic, is feasible. With population-wide typing, our computational tools could predict which individuals are most likely to mount a strong antiviral immune response to a novel pathogen, given their MHC I variants; these individuals would be at the lowest risk for severe disease. Conversely, our tools could predict which individuals are least likely to mount a strong antiviral response; these individuals would be at the highest risk. Consider the implications for early response and biodefense. With the computational ability that we will deliver, when faced with a novel biological threat, an early-response team could predict which of its personnel are likely to have immunity and which might be most vulnerable. This could be assessed based on prior HLA typing of the personnel. With the computing ability described above, all that would be required is a proteome profile of the novel virus or bacteria. Such profiles are usually easy to obtain, often available within weeks when a new pathogen is identified.

Milestone #2 summary: Predict disease severity for different variants of a virus for different individuals.
Estimated completion date: 12/31/22
Description: Applications: Resource Allocation & Population Monitoring. As we have seen with SARS-Cov-2, viral mutations are perhaps the single greatest confounding factor to a pandemic response. The more widespread a pandemic, the more hosts a virus infects. With more hosts, there are more opportunities for it to mutate. Given the extent of the COVID19 pandemic, some virologists have hypothesized that nearly all mutations favorable to its spread will be discovered by the virus before the pandemic abates. Mutations confound a response because vaccines and treatments may be less effective against new variants. Here our prediction tools could transform both planning and resource allocation. Once the protein sequence of a new variant is identified, the differences from the original strain can be analyzed. Differences in the proteins expressed will translate to a different set of peptides. With population-wide HLA typing, a distribution of MHC I variants can be constructed for sub-groups -- perhaps different demographic groups in different geographic regions, or perhaps even a fine-grained map tagging all individuals in the group with their specific MHC I variants. These MHC variants can be paired up with the novel peptides from the viral variant to assess the risk of severe disease for the individual. If the analysis is done at the level of a group, then a statistical analysis of the risk can be performed against the distribution of MHC I molecules in the group.

Milestone #3 summary: Predict effectiveness of different vaccines for different variants of a virus for different individuals.
Estimated completion date: 4/1/23
Description: Applications: Tailoring Vaccines to Individuals. It is likely that future historians will point to this pandemic as an inflection point for society, not due to the damage that was inflicted, as great as this has been, but due to the progress made in science as a consequence of it. The development of mRNA vaccines, in particular, is a startling success story. They have been deployed in record time, and on an unprecedented scale. Significantly, mRNA vaccines are readily customizable. Once the infrastructure is developed, different mRNA vaccines could be administered to different groups at different times, with little extra production cost; all that is required is swapping out the RNA sequence in the vaccine doses. This flexibility offers the possibility to administer mRNA vaccines that elicit the best immune response for each individual, in response to the specific viral variants that pose a threat at that time. Recall that mRNA codes for proteins, such as the infamous spike protein of the SARS-Cov-2

virus. Dividing proteins into units 8 to 15 amino acids long yields the requisite peptides to target in mRNA vaccine production. So, our computational tools will allow screening of the peptides of viral variants, matching of those against MHC I molecular variants, and then choosing which peptides to target in the vaccine production.

Milestone #4 summary: Launch at-scale simulations for SARS-Cov-2 Peptides. Publish results. Launch a full website.
Estimated completion date: 6/30/23
Description: This research will be applied and translational: it will explore the solution to the grand challenge discussed above and will translate these into aspects of pandemic preparedness such as predictions of disease severity; resource allocation; and vaccine development. By the conclusion of the 12-month project, the team will deploy a risk assessment of COVID-19 severity given an individual's HLA typing. This will be implemented on a web-based platform and will be available to the public. Users will be able to enter -- or link if available electronically – their results of HLA typing. The website will provide a summary and detailed analysis of their risk for COVID-19, based on the binding strength of SARS-Cov-2 derived peptides to their allelic variants of MHC-I molecules.

(Add additional milestones as needed)