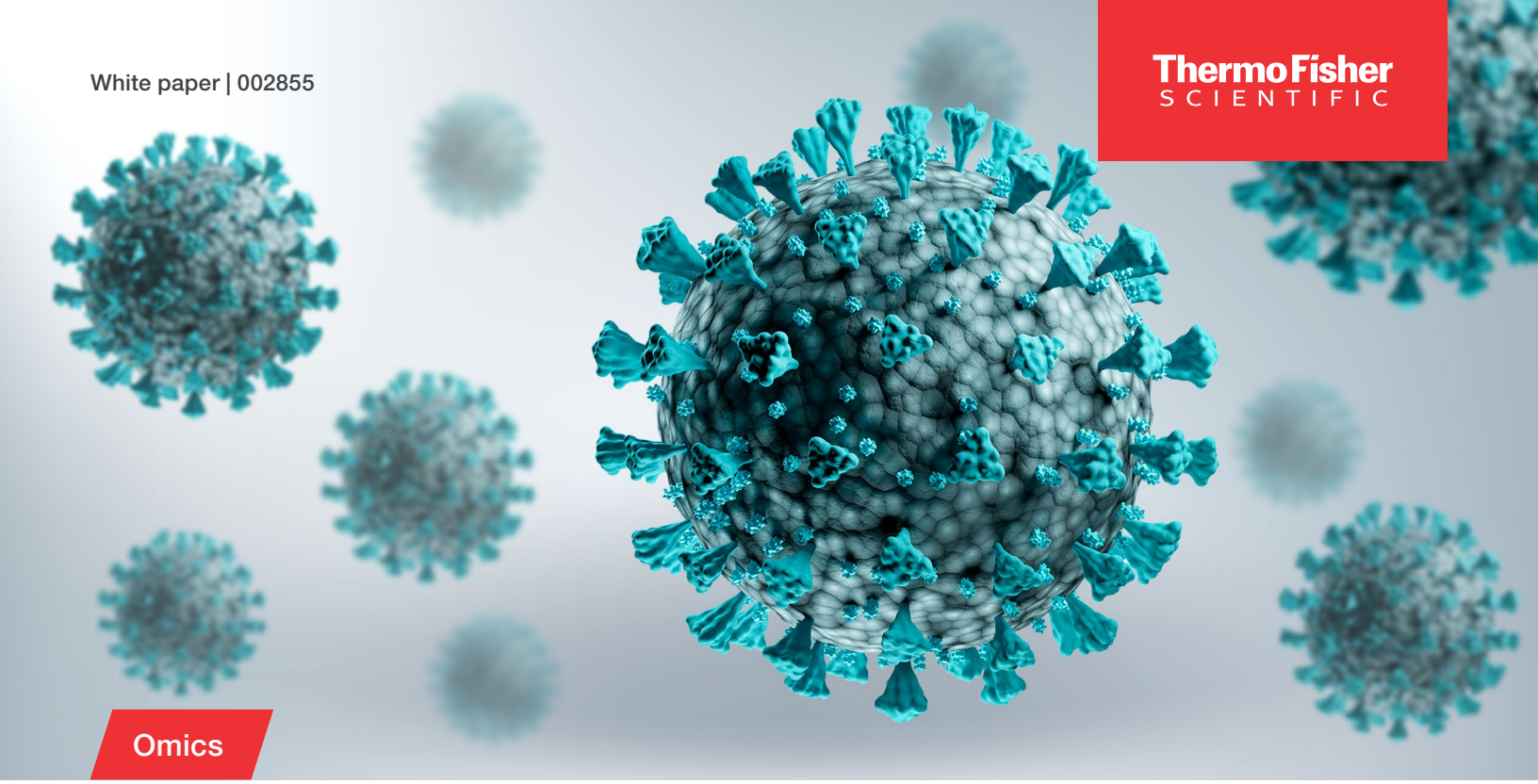
White paper | 002855



Omics

The role of immunopeptidomics in virus research

Keywords
Severe acute respiratory syndrome coronavirus 2, SARS-CoV-2, COVID-19, host-cell interactions, viral immunity, high-resolution mass spectrometry, viral epitopes, immunogenicity, Prosit algorithm, viral peptide antigens, vaccine candidates, immunotherapies

Introduction

If we needed any additional evidence that virus research is an important area to emphasize, the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic and cumulative COVID-19 human health impact have provided a global case study for the need to maintain a responsive and effective ability to rapidly understand emerging viruses. Mass spectrometry (MS) offers many approaches to gain insights into viral behavior by studying intact viral particles, their surfaces and binding characteristics, their protein compositions, and their impacts on the biochemical pathways of host cells upon infection. This information can help provide detailed insights into virus structure and function, which might enable virus detection and inform drug and vaccine programs designed to prevent infection and disease.

It is impressive to note that results from the first MS-based studies of SARS-CoV-2's effects on human cells were published online within weeks of the start of the pandemic. Well-developed methods exist and are in routine use in specialized labs today. This document will detail one of the more prominent MS approaches with the intent to inform and expand upon its use. The current crisis will push us to further improve our methods and test their ability to contribute to our understanding of SARS-CoV-2 and potentially play a role in mitigating its impact.

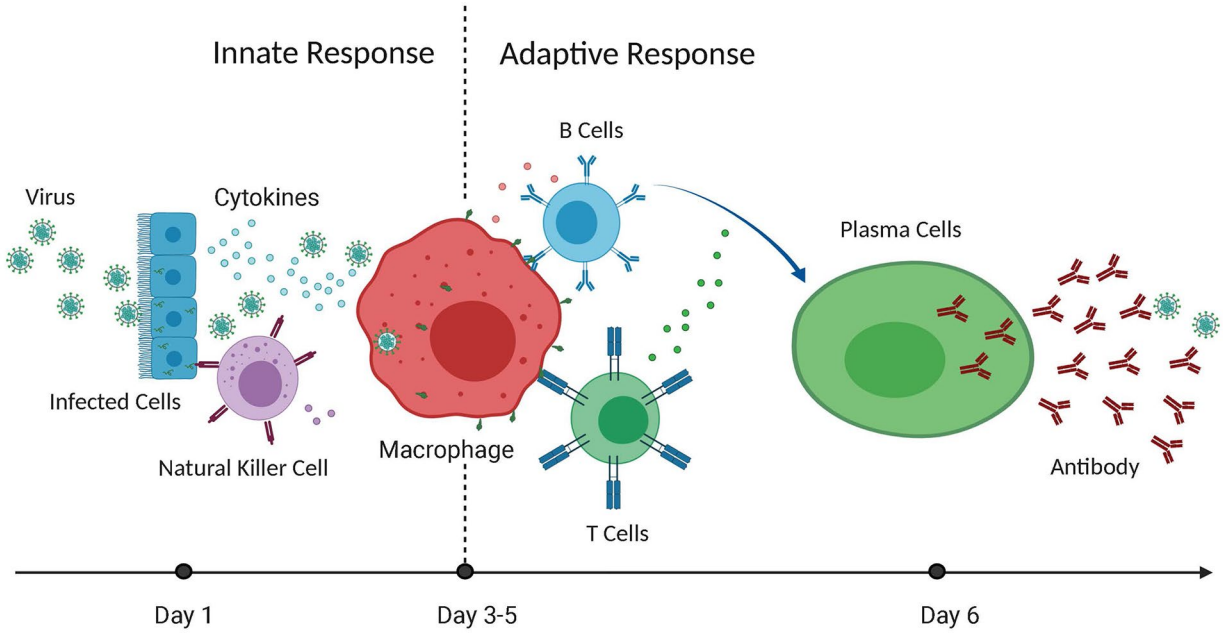


Figure 1. Lifecycle of the host immune system

Virus-host immunity and the immune system

The host immune system acts through a complex system of molecular mechanisms to defend against viral infections. Many aspects of the immune system's functions are characterized, but new aspects of host defense are being discovered through the application of new technologies such as high-resolution MS. Molecular characterization of the immune response to viruses is necessary for the development of treatments such as antiviral drugs that can be administered post-exposure to quell viral replication and propagation, as well as alleviate any symptoms of viral disease. The application of MS in understanding the mechanism of host-cell interactions for specific viral infections promises to help researchers discover the responsible viral antigens more quickly and guide the development of vaccines and other prophylactic therapies.

Immunity against viruses in mammals is achieved through two arms of the host immune system: the innate immune system and the adaptive immune system. Although varied in their modes of action, both systems function by distinguishing "self" from "non-self." Innate immunity is a non-specific defense mechanism that comes into play immediately or within hours of an infection or the appearance of an antigen in the body. While the innate immune system includes physical barriers such as the skin and chemicals in the blood, the main mode of defense is through specialized immune system cells that attack foreign cells in the body. The innate immune system provides an immediate and rapid but non-specific response to infections, whereas the adaptive immune responses are generally slower but are more specific to the virus.

As part of the innate immune system, lymphocytes recognize distinct antigens derived from viruses, and this recognition activates the adaptive immune response. The adaptive immune system is typically mobilized after the innate immune system is unsuccessful in clearing the virus. After an initial encounter with a specific virus, an immunological memory is maintained so that subsequent infections with the same virus stimulate enhanced and efficient viral eradication. The molecular basis of the adaptive immune system response can be specifically probed using modern high-resolution MS and informatics, commonly referred to as immunopeptidomics.

The role of immunopeptidomics in understanding viral immunity

Upon infection of the host cell, viruses hijack the host's cellular machinery to produce their own proteins. Just like any protein synthesized within the cell, these newly produced viral proteins are subjected to proteasomal degradation into peptides, which are transported into the endoplasmic reticulum for further processing before they bind to major histocompatibility complex (MHC) class I molecules. Proteasomal processing of these intracellular viral proteins liberates the precursors of antigenic peptides that are distinct to the virus and serve as the "canary in the coal mine" to mark this specific host cell as infected.
As part of the host cellular response, the viral peptide antigens are typically presented by MHC class I molecules on the surfaces of the infected cells, where they interact with T-cell receptors expressed on $CD8^{+}$T cells. The MHC class I molecules bind with self or endogenous peptides in healthy cells that are

8 to 11 amino acids long, but they may also present pathogen-derived peptides and form a tight MHC class I-viral peptide complex, which then may trigger an adaptive immune response. In addition to cell-mediated immunity, a second mechanism, the humoral immune response involving another class of MHC molecules (MHC class II), typically presents processed extracellular antigens to a different subset of T cell, the $CD4^+$T cells. MHC class II molecules are expressed on professional antigen-presenting cells, which include dendritic cells, macrophages, and B cells.

Besides these two well-established mechanisms for antigenic peptide generation and presentation to the host immune system, several other cellular mechanisms are being investigated as unconventional origins of antigenic peptides. These avenues include the presentation of extracellular antigens on MHC I class molecules through a process known as cross-presentation, and the presentation of antigens that were predominantly sourced from defective ribosomal viral products (DRiPs) as well as antigens derived from proteasome-spliced peptides.[1-4] The significant advancements made in understanding the processing and presentation of antigens to T cells and efforts in determining the sources of antigenic peptides using MS are generally known as immunopeptidomics. The characterization and quantitative assessment of the antigenic peptides from viral host proteins are crucial in understanding the specific and effective defenses against a particular viral infection.

The role of high-resolution MS in the identification and quantitation of viral epitopes and understanding immunogenicity

To study viral epitopes with MS, researchers use a bottom-up proteomics strategy where viral protein-expressing cells presenting MHC class I molecule complexes are isolated using monoclonal antibodies specific for MHC class I molecules, and peptides are dissociated from the MHC class I molecules and fractionated by reversed-phase high-performance liquid chromatography (LC). Subsequently, peptide-containing fractions are analyzed with LC-MS/MS. The generated mass spectra are compared to the viral proteome of interest to determine the sequences of the identified peptides. This analytical approach has been successfully applied to discover not only more peptides than had been shown using traditional overlapping peptide screening techniques, but also post-translational modifications (PTMs) and spliced peptides.[5,6] However, the distinct challenges that arise with this LC-MS-based approach in combination with database search strategies have been discussed in recent publications.[7] For example, human leukocyte antigen (HLA) peptides, with their inherently diverse C-termini (across different HLA alleles), often display poor ionization and do not always fragment predictably or yield sequencing ions compared to tryptic peptides that bear charged arginine or lysine C-termini. The resultant spectra from many HLA peptides require more substantial sequencing efforts, and often require taking into account internal fragments and reduced coverage of paired sequencing ions. Substantial improvements in both quality and quantity of spectra from HLA peptides have been achieved by using hybrid fragmentation methods or by optimizing data acquisition methods.[8,9]

In addition, the current design of database search algorithms is heavily biased toward proteolytic peptides, presenting a significant obstacle to the confident detection of heavily processed and non-tryptic sequences in immunopeptidomics. Efforts to expand the search space to include HLA peptides create additional uncertainties associated with lower confidence in peptide spectral matching scores, resulting in inflated false discovery rates.

Efforts to create HLA allele-tailored databases offer a suitable compromise, but may miss natural variations in sequences or mutations as well as peptides from other non-canonical reading frames.[10] However, recent progress in machine learning–based approaches to predicting spectra, including an algorithm called Prosit, have created new opportunities to address this challenge.[11] The Prosit algorithm is capable of producing spectra from any peptide sequences in silico that exceed the quality of the experimental data, resulting in chromatographic retention time and fragment ion intensity prediction capabilities that are independent of genomic information input. The Prosit predictions are generalized to non-tryptic peptides and allow the recalibration of normalized collision energy for each individual spectrum, accommodating the specific challenges associated with HLA-type peptides. The new spectral interpretation method is now available through the latest versions of the Thermo Scientific™ Proteome Discoverer™ software application, which includes a spectral prediction tool based on Prosit and enables a database-independent spectral scoring mechanism that circumvents

the HLA peptide–associated challenges of traditional search engine–based approaches. Genomically predicted or putative non-canonical viral peptide sequences can now be accurately projected through in silico prediction, increasing the HLA peptide repertoire coverage and sequence confidence of peptides investigated in immunopeptidomics. Further improvements in deep-learning approaches using additional HLA-derived and -validated sequences, including sequences with post-translational modifications, will further enhance the field of immunopeptidomics research, leading to accelerated insights in viral-associated knowledge surrounding the presentation of viral antigens by the host adaptive immune system.

The elucidation of the intricacies of antigen presentation has been immensely aided by the advent of highly accurate and sensitive mass spectrometers. MS has proven to be extremely valuable and versatile in the identification and quantitation of viral peptide antigens, and this information can be used to evaluate vaccine candidates and design effective immunotherapies.

Acknowledgement

Figure 1 was created with BioRender.com.

References

1. van de Weijer, M.L.; Luteijn, R.D.; Wiertz, E.J. Viral immune evasion: Lessons in MHC class I antigen presentation. *Semin. Immunol.* **2015,** *27*(2), 125–137.
2. Neefjes, J.; Jongsma, M.L.; Paul, P.; Bakke, O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat. Rev. Immunol.* **2011,** *11*(12), 823–836.
3. Dudek, N.L.; Croft, N.P.; Schittenhelm, R.B.; Ramarathinam, S.H.; Purcell, A.W. A systems approach to understand antigen presentation and the immune response. *Methods Mol. Biol.* **2016,** *1394,* 189–209.
4. Soethout, E.C.; Meiring, H.D.; de Jong, A.P.; van Els, C.A. Identifying the epitope-specific T cell response to virus infections. *Vaccine* **2007,** *25*(16), 3200–3203.
5. Giam, K.; Ayala-Perez, R.; Illing, P.T.; Schittenhelm, R.B.; Croft, N.P.; Purcell, A.W.; Dudek, N.L. A comprehensive analysis of peptides presented by HLA-A1. *Tissue Antigens* **2015,** *85*(6), 492–496.
6. Liepe, J.; Marino, F.; Sidney, J.; Jeko, A.; Bunting, D.E.; Sette, A.; Kloetzel, P.M.; Stumpf, M.P.; Heck, A.J.; Mishto, M. A large fraction of HLA class I ligands are proteasome-generated spliced peptides. *Science* **2016,** *354*(6310), 354–358.
7. Faridi, P.; Purcell, A.W.; Croft, N.P. In immunopeptidomics we need a sniper instead of a shotgun. *Proteomics* **2018,** *18*(12), e1700464.
8. Mommen, G.P.; Frese, C.K.; Meiring, H.D.; van Gaans-van den Brink, J.; deJong, A.P.J.M.; van Els, C.A.C.M.; Heck, A.J.R. Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD). *Proc. Natl. Acad. Sci. U. S. A.* **2014,** *111*(12), 4507–4512.
9. Sofron, A.; Ritz, D.; Neri, D.; Fugmann, T. High-resolution analysis of the murine MHC class II immunopeptidome. Eur. *J. Immunol.* **2016,** *46*(2), 319–328.
10. Shao, W.; Pedrioli, P.G.A.; Wolski, W.; et al. The SysteMHC Atlas project. *Nucleic Acids Res.* **2018,** *46*(D1), D1237–D1247.
11. Gessulat, S.; Schmidt, T.; Zolg, D.P.; et al. Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning. *Nat. Methods* **2019,** *16*(6), 509–518.

Learn more at **thermofisher.com/immunopeptidomics**