WHITE PAPER 65880

An urgent need for expanded virus research

The role of mass spectrometry in understanding virus structure and function

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 several of the most prominent approaches with the intent to inform and expand upon their 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.

Human health impact of a novel pathogen

The current coronavirus pandemic is caused by the novel viral pathogen SARS-CoV-2. This novelty makes it different than other viral infections, such as seasonal influenza, and it has the potential to have a much larger human health impact. With the flu and other common viruses, most people have been exposed and have antibodies either by having been sick and recovered, or by having been intentionally immunized. When we are exposed to that specific virus again, our bodies are able to detect it and mount a rapid immune response, which limits its potential health impact.

A novel virus has not been seen by the body before, and it is not recognized by the immune system. If it infects a person, a naive host, it can replicate and become established, and there are no known drugs or therapies to treat it. We are able to treat the symptoms, but it is up to the body to mount its own defense to fight off the virus and survive the infection, after which the recovered host will have a level of immunity. This cycle will continue with each new infection until there are vaccines and therapies to prevent and treat infected individuals. Until then, our best option is to do what we can to limit new infections and reduce viral spread.

The appearance of a novel virus that can infect and spread among the human population, such as severe acute respiratory syndrome-related coronavirus (SARS, or SARS-CoV-1), Middle East respiratory syndrome (MERS) or Ebola, puts us under incredible pressure to understand it and to learn how to fight it as rapidly as possible. We will be led by insights from molecular biology and molecular biochemistry, and MS has an important role to play. This paper will describe techniques for viral detection, the detailed characterization of virus structure and function, the determination of cell-surface interactions and routes of infection, the understanding of the host immune response, and, importantly, the characterization of the mechanisms by which the virus infects and rewires human cells. All of these approaches will help guide the urgent efforts to develop vaccines and therapies to address SARS-CoV-2 and future novel viruses.

Virus structure and function

Virus architecture

The virus structure influences how it interacts with host cells and how it functions during its lifecycle. In general, viruses are composed of RNA or DNA surrounded by a capsid composed of proteins and glycoproteins. The capsid can vary depending upon the virus, but its primary goal is to protect the viral genome. Some viruses are further encompassed by membrane envelopes composed of lipids and proteins. The surface of the envelope has additional proteins, specifically glycoproteins, which are involved in the process of binding to the host cells for viral entry. For example, coronaviruses have large transmembrane proteins on their surfaces, which are referred to as spike proteins or S proteins. These proteins form the trimers that give coronaviruses their crown-like appearance, hence the term coronavirus.

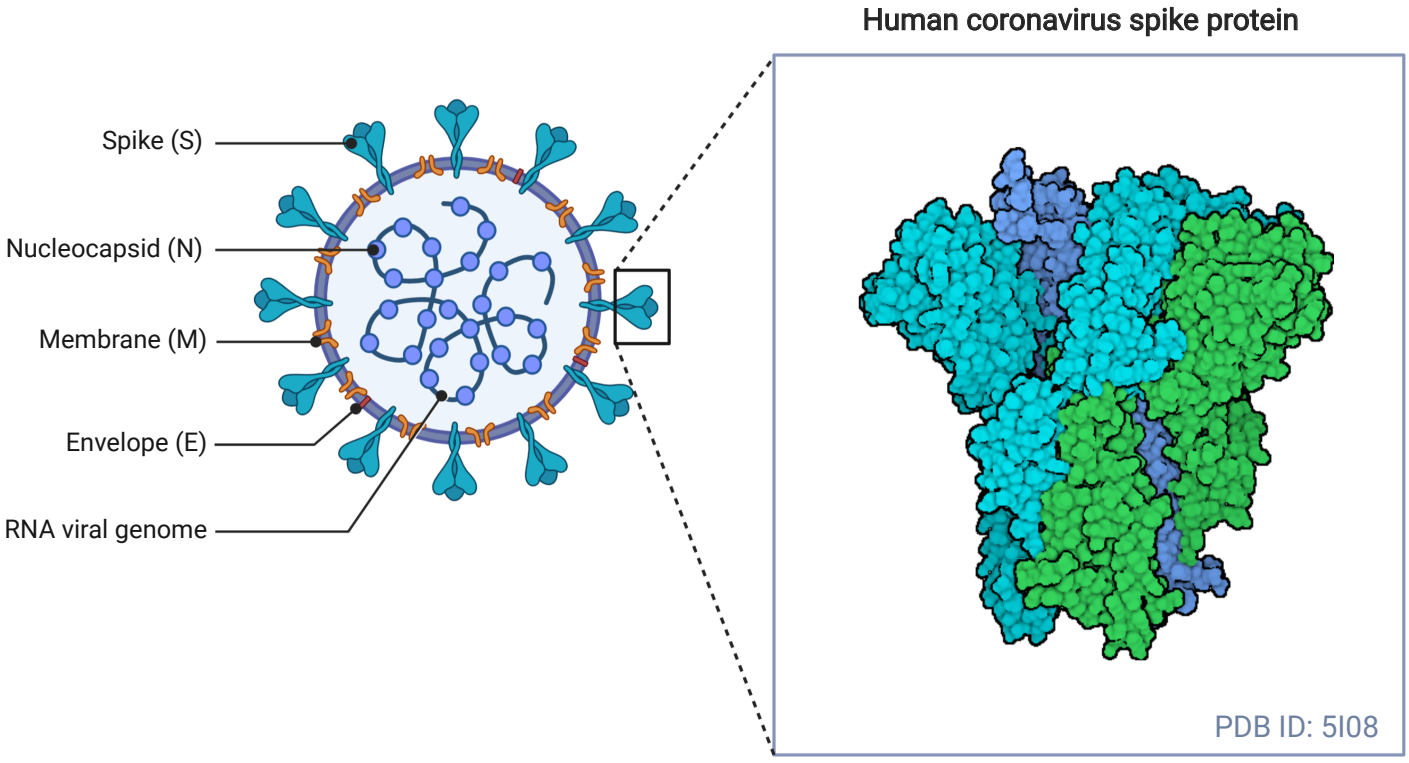

Figure 1. Structure of SARS-CoV-2.

To understand viral functions, it is essential that researchers determine the viral complex assembly and structure. Solving the structure of large dynamic complexes often requires integrating several complementary MS techniques, along with imaging techniques such as cryogenic electron microscopy (cryo-EM)—an approach known as integrative structural biology. The following section will focus on the latest proteomics MS tools and workflows that have been developed for studying virus structure.

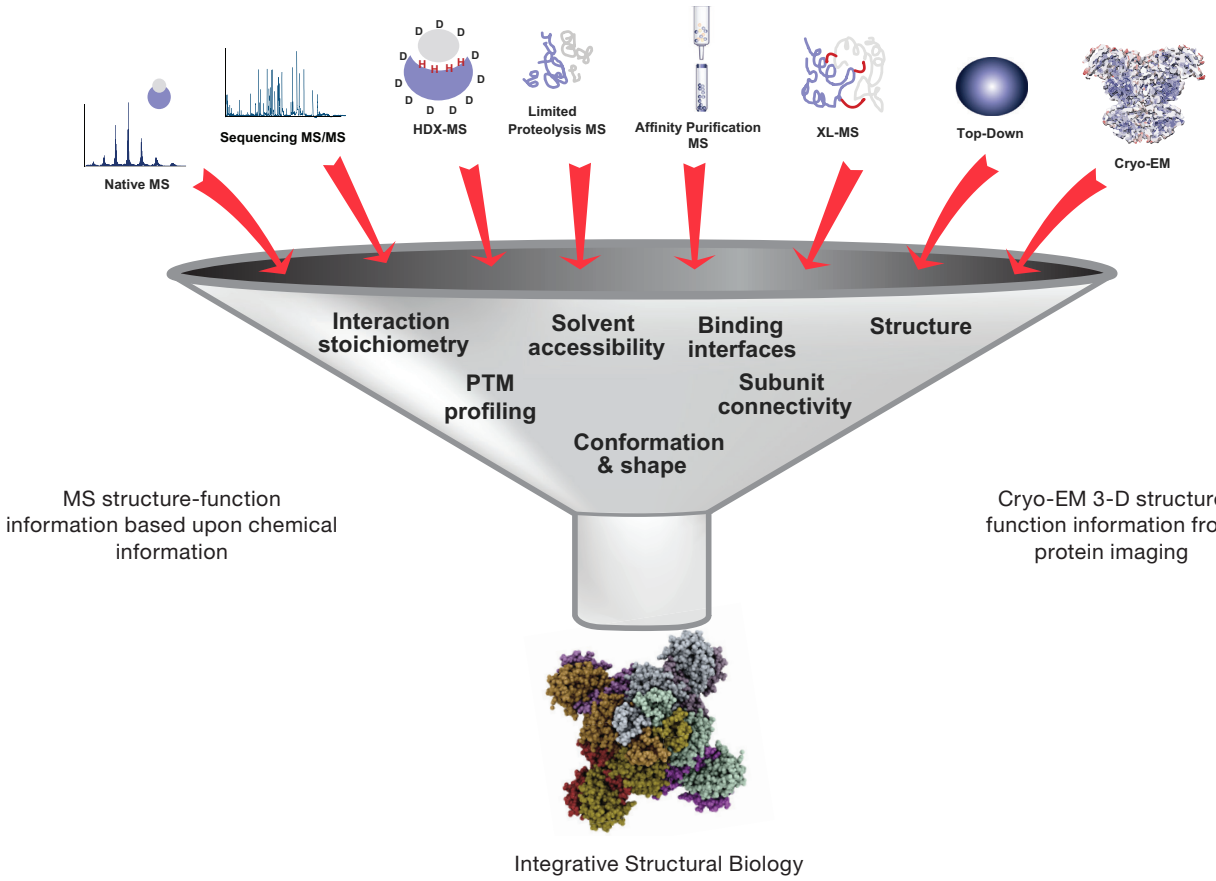

Figure 2. Tools for performing integrative structural biology.

Native mass spectrometry for studying virus structure

The viral glycoprotein is not only a key protein for host cell attachment but also a prime target for neutralizing antibodies elicited through vaccination. While RNA sequencing is extraordinarily informative with regard to viral mutation or adaptation via immune selective pressure, it cannot provide insight into the critical feature that helps enveloped viruses evade the immune system: glycosylation. Many viruses exploit the glycosylation machinery by using host glycans for immune evasion; for example, HIV-1 hides immunological epitopes behind a "glycan shield." Additionally, glycans on the viruses play a role in host cell receptor binding and associated conformational dynamics, which influence binding efficacy.[1] Measuring glycoprotein glycosylation and plasticity is crucial for the production of vaccine candidates that mimic circulating viruses and for efforts to develop drugs to block receptor attachment and subsequent infection.

A powerful tool in the MS toolbox for characterizing glycosylation on viral glycoproteins is native mass spectrometry (native MS).[2] In native MS, protein structures are kept intact and introduced into the mass spectrometer in a structural configuration similar to the one they exist in under biological conditions. Native MS provides an unaveraged picture of the solution conditions, meaning that different proteoforms (such as differential glycosylations) can be detected simultaneously, which is difficult to do using other structural biology techniques. The number and

the type of glycoform compositions that are present can be obtained using native MS. Phosphorylation is another posttranslational modification (PTM) that can be elucidated using native MS. Native MS can also provide the mass of the intact glycoprotein, which can inform researchers whether the protein of interest is actually being examined or if it's another protein.

Native MS is also a useful tool for examining virus capsids, quantifying cargo encapsulation and monitoring capsid assembly.[3-6] Combining native MS with a top-down MS allows the structural characterization of the capsid's proteins. An example of this characterization of hepatitis B virus capsid proteins is shown in Figure 3.

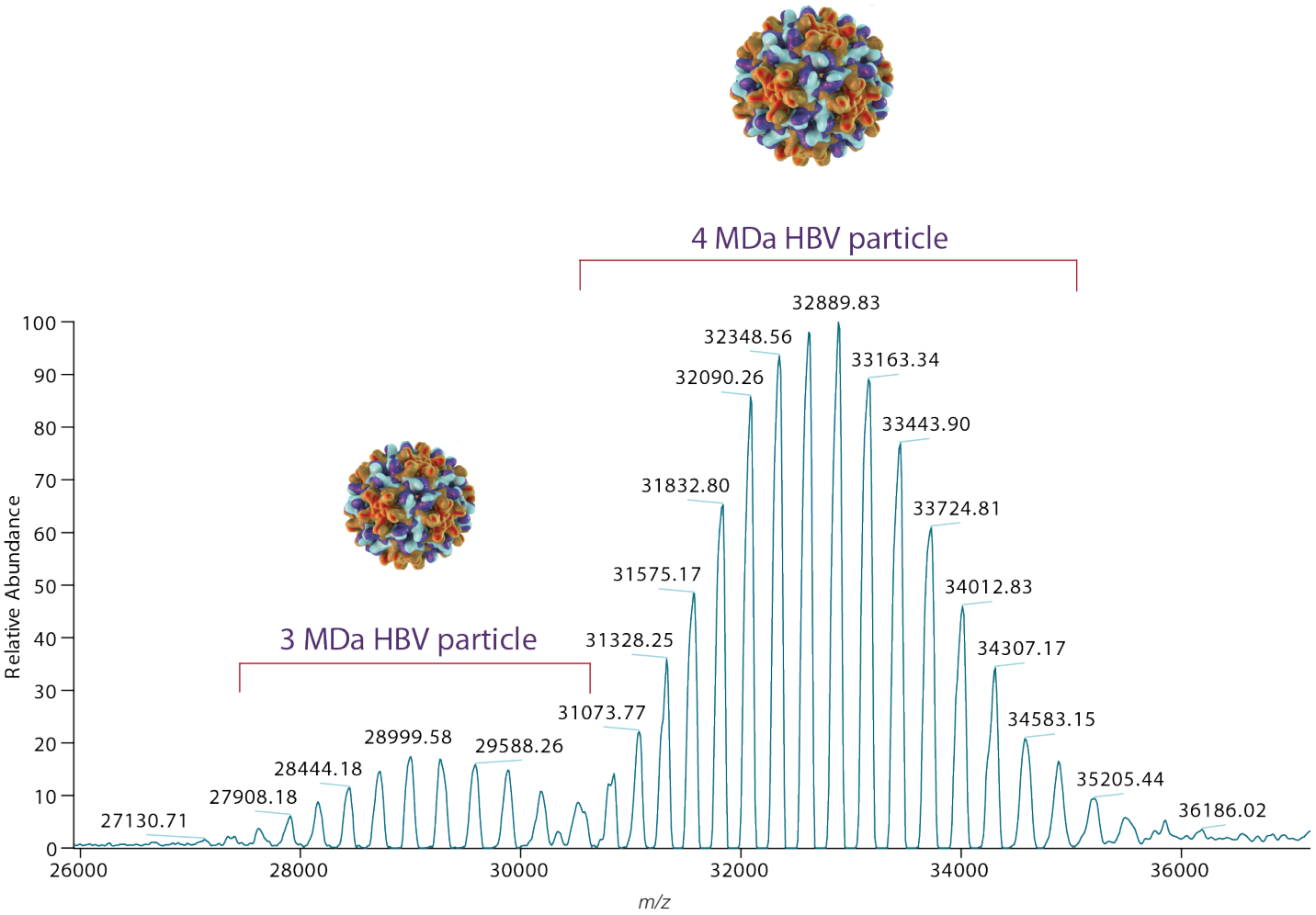

Figure 3. Native MS analysis of hepatitis B virus (HBV) capsids under charge-reducing conditions. Mixtures of 3-MDa and 4-MDa HBV capsids analyzed under charge-reducing conditions span a wide range of masses between 27,000 and 36,000 *m/z*. The spectrum contains well-resolved charge states for both particles and shows no bias for the 25% smaller 3-MDa particle.

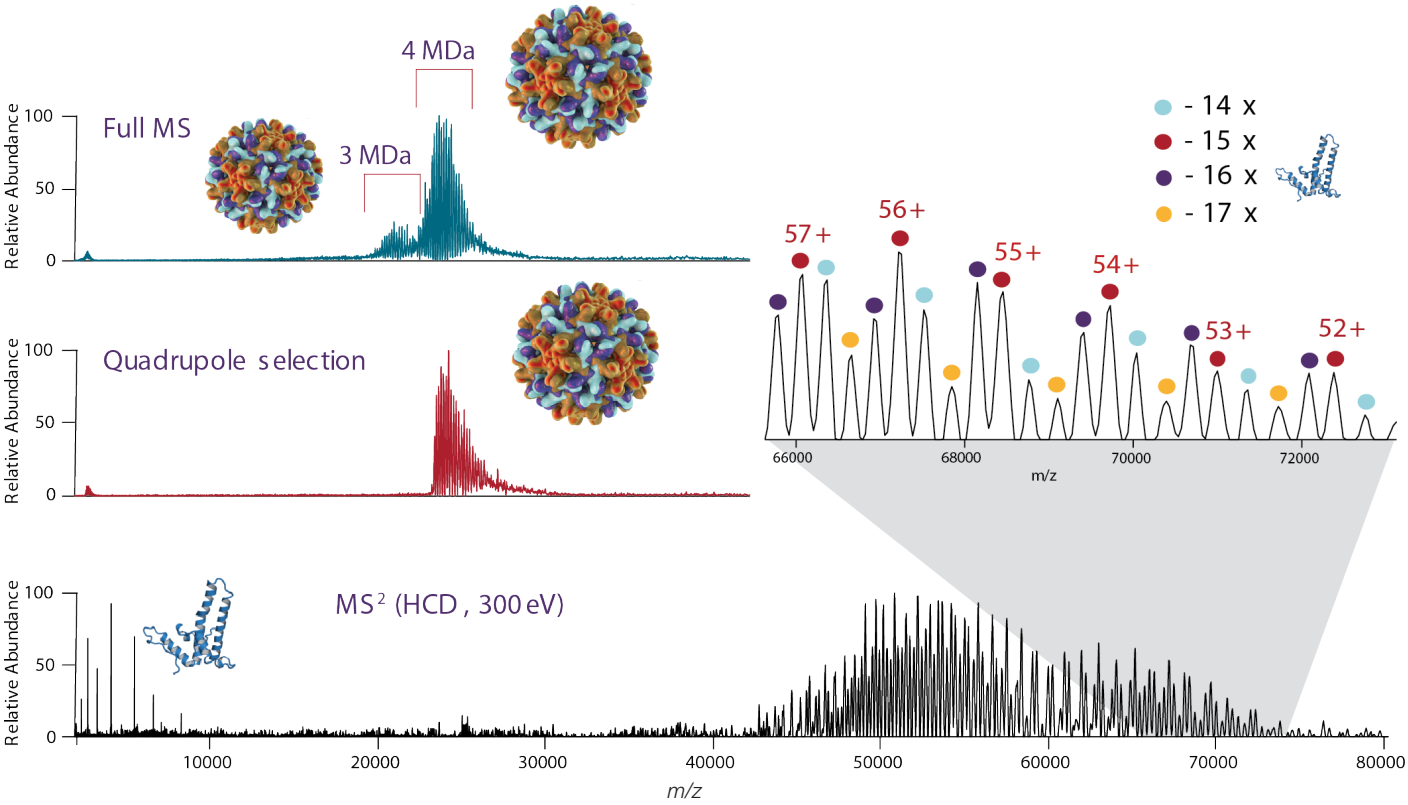

Figure 4. By combining top-down with native MS, the capsid proteins can be characterized. The larger 4-MDa HBV particle can be isolated at 25,000 *m/z* and fragmentated. Here, the tandem MS spectrum of the 4-MDa HBV capsids shows sequential ejection of up to 17 out of the 240 copies of the capsid protein. The ejected monomers appear at low *m/z* while the product ions appear at increasingly higher *m/z*. The inset shows an enlargement of the spectrum at 70,000 *m/z*, which contains baseline resolved ions of HBV capsids that have lost between 14 and 17 capsid proteins (6.5% of the original mass) and 68% of the original charge.

Glycoproteomics and glycomics to investigate virus surface proteins

In order to get more detail on the type of glycosylation and the sites where it occurs (site occupancy) on viral surface glycoproteins, a bottom-up glycoproteomics strategy can be used to supplement native MS data. A glycoproteomics strategy can reveal far more site-specific glycans on a protein than a native MS workflow can. In a bottom-up glycoproteomics analysis, the viral surface glycoproteins can be enzymatically digested, enriched specifically for glycopeptides and analyzed by MS. This strategy can provide information on the sites of glycosylation and glycan compositions. It can also reveal minor glycoforms and O-linked glycosylation that might not be seen or might be suppressed in a native MS approach. A similar strategy can be employed for other potential PTMs, such as phosphorylation. This approach has been used to examine the SARS-CoV-2 S protein. Watanabe and colleagues,[7] as well as several other labs, performed site-specific analysis of the glycan compositions with liquid chromatography – tandem mass spectrometry (LC-MS/MS). Their experiments revealed that the overall glycosylation of the SARS-CoV-2 S protein is similar to that of SARS-CoV-1, but very different from that of the HIV-1 Env protein.

If researchers require structural rather than compositional information about the glycans present on the glycoproteins, incorporating an MS-based glycomics approach would provide the desired data. Here, the glycans are released from the proteins or peptides using enzymatic or chemical means. Once the glycans are released, they can be labeled to increase ionization efficiency, separated by LC and characterized by MS. Such a strategy can provide very detailed information about the glycans, such as linkages, branching and potential structural isomers.

Probing capsid assembly, structure and dynamics with hydrogen-deuterium exchange mass spectrometry

Viruses do not exist in static states; rather, they undergo various conformational changes. The characterization of these changes can be difficult with most structural tools. A major focus of structural virology is to obtain an understanding of the construction and structural basis for stability of the protein capsids that encapsulate the viral genomic information. Hydrogen-deuterium exchange mass spectrometry (HDX-MS) has been used to measure local conformational dynamics and to gain insight into the mechanisms of assembly and capsid maturation for various viruses.

HDX-MS takes advantage of the labile nature of the protons present on protein backbone amides. When dissolved in buffer, proteins exchange these protons with hydrogen groups present in the solution. In a deuterated buffer, protons from the protein are exchanged with deuterium. Only the protons present on backbone amides are measured, as protons found on the functional groups of amino-acid side chains exchange too rapidly to measure, and the ones on the carbons are too slow to exchange.[8] The rate of hydrogen-to-deuterium exchange provides solvent accessibility data, which can be used to infer information on protein structure and conformation. MS can be used to measure the rate of deuterium uptake. For single proteins and protein complexes, HDX-MS can be used to obtain information on structure, protein-protein or protein-ligand interaction sites, allosteric effects, intrinsic disorder, and conformational changes induced by PTMs. HDX-MS has the added advantage of not being limited by the size of proteins or protein complexes, compared to traditional structural approaches. It also has the additional advantage of being highly sensitive; it can detect co-existing protein conformations and examine membrane proteins that are not amenable to traditional approaches.

HIV-1 capsid assembly and maturation, the higher-order structure of the Rous sarcoma virus SP assembly domain, and HBV core-antigen (capsid protein) dynamics have all been explored by HDX-MS.[9-13]

Virus-host interactions

The infection process

Irrespective of the type of virus, viral infections tend to follow similar processes. The initial step involves the virus binding to receptors on the host cell. Viruses infect specific cells, and this process is dictated by the receptors found on the surface of the cells, which run the gamut of molecules from proteins and lipids to glycans. The virus itself employs proteins for binding, which can be either glycoproteins or fibers. The binding of the virus to the cell surface receptor can result in conformational changes to the receptors and to the virus. These changes can aid the virus' ability to penetrate the cell or form membrane association. Sometimes, these attachments might not involve receptors, and the focus is to simply concentrate many virus particles on the surface of the host cell.

Mechanistically, there are few ways that a virus can enter the host, as it first must break through the phospholipid bilayer that covers the cell. The entry process can be summarized as follows: endocytosis, membrane fusion and viral penetration.

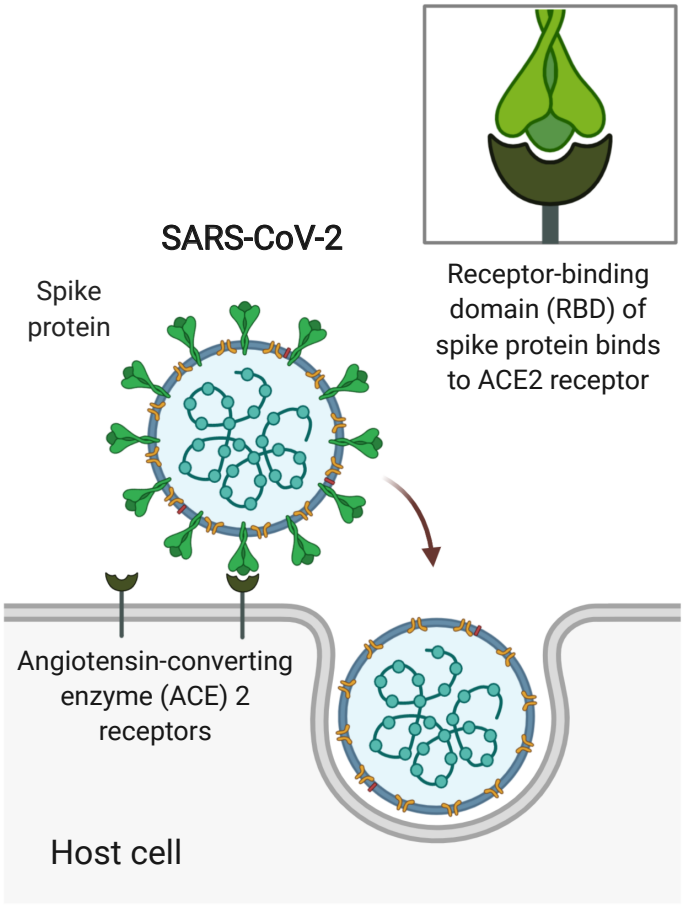

Figure 5. Viral binding to host cell receptor and entry by membrane fusion and endocytosis.

Viruses that lack viral envelopes can enter the host cell by endocytosis, although, a number of viruses with viral envelopes can also enter host cells in such a manner. After binding to host cell receptors, viruses can be internalized into the cell by various endocytic processes such as clathrin-mediated endocytosis, caveolar or lipid raft-mediated endocytosis, or macropinocytosis. Some viruses are able to use more than one pathway.

Alternatively, the virus can enter the host cell by the process of membrane fusion, where the membranes of the host cell and the virus fuse together. This process is reserved for enveloped viruses. The whole process is facilitated by viral fusion proteins that bind to the receptors on the host cell, resulting in changes in the conformation of the membrane fusion proteins. These changes make it easy for the viral envelope to insert its peptides into the cellular membrane of the host cell, fusing itself to the membrane and creating a path between the virus and the host.

Lastly, viruses can enter the host cell directly by penetration. This process is reserved for nonenveloped viruses. These viruses contain capsid proteins with hydrophobic peptides that are internal parts of the protein. These peptides allow membrane penetration through diverse approaches such as pore formation, osmolysis or induction of membrane curvature. Once the membrane is penetrated, the virus genome can be released into the host cell's cytoplasm.

Regardless of the manner of entry, once the virus has entered the host cell, the following processes occur: the viral translation of mRNA by host ribosomes, the replication of the genome of the virus, the assembly of viral particles enclosing the genome, and the release of infectious particles from the cell.

In order for a virus to be effective, it needs to hijack the host cell and replicate. Therefore, the interaction of the virus with host cell proteins is a crucial part of this strategy. This interaction not only ensures the replication of the virus, but is also a strategy for the virus to evade the defense mechanisms initiated by the host cells. Examining the virus-host cell protein interactions can provide insights into processes such as viral entry, virus genome replication within the host cell, and the ways the viral genome can spread across the cell. Additionally, examining virus-host cell protein interactions could yield valuable information and help scientists develop broadly acting antiviral therapeutics.

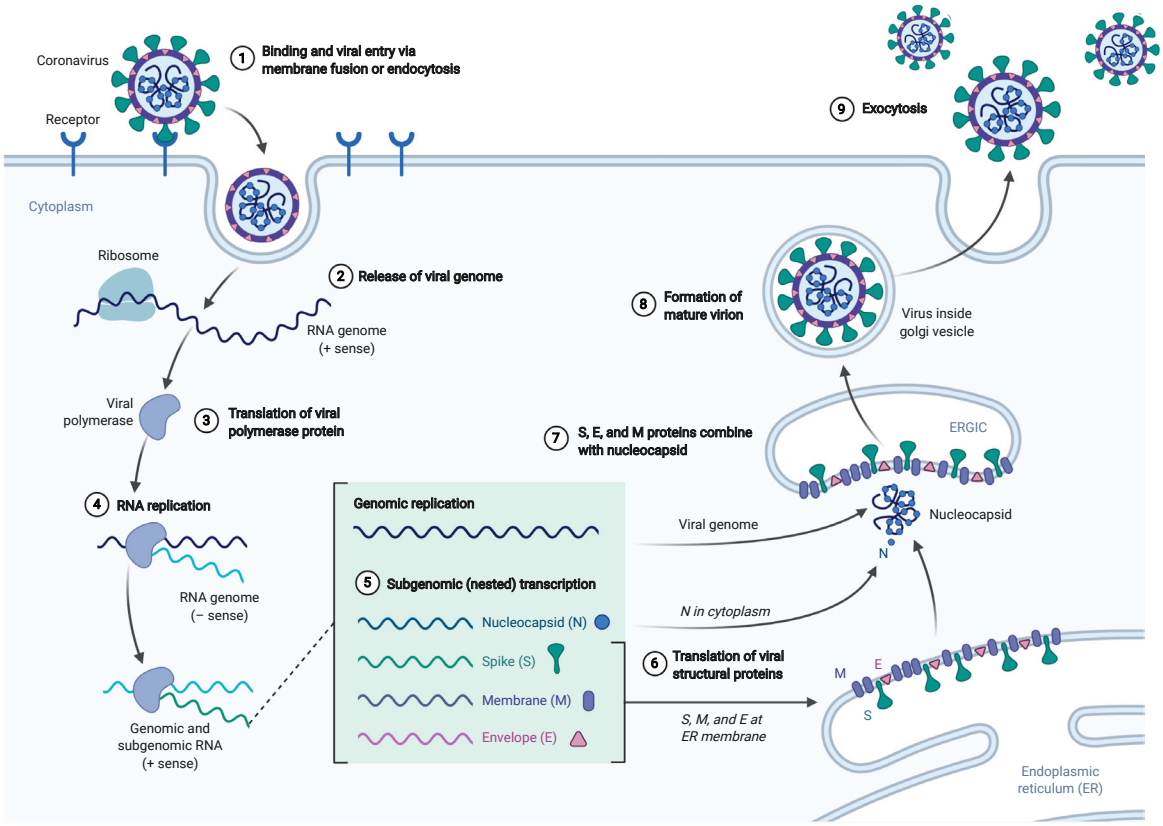

Figure 6. Schematic representation of the virus replication cycle.

Current techniques used to study protein-protein interactions include X-ray crystallography, cryo-EM and nuclear magnetic resonance (NMR). However, these approaches require large amounts of highly purified proteins, may not allow for the analysis of proteins in their native conditions, or may not be applicable on a proteome-wide scale. Furthermore, many proteins are simply not amenable to these types of analysis, thereby limiting their applicability. MS techniques have been used for many years to study protein structure and protein-protein interactions, albeit usually within highly specialized research groups. But in recent years, with the development of new workflows, instruments and software, MS has developed from a manual, non-standardized research tool to an emerging, powerful solution for studying protein-protein interactions. The following section will focus on the latest proteomics MS tools and workflows that have been developed to study virus-host cell interactions.

Using crosslinking MS to understand virus-host cell protein interactions

The virus-host cell protein interactions are intermolecular interactions occurring between the virus protein and the host cell receptor. Unfortunately, these interactions tend to be transient, meaning that they only occur for a brief period. Examining transient interactions is challenging due to the lack of available techniques that can measure the interactions in real time. Chemical crosslinking reagents provide a means of capturing these interactions. By combining crosslinking with MS (XL-MS), we can analyze the interacting complexes, enabling in vivo and in vitro approaches to study protein-protein interactions while maintaining the original interacting complex. The ability to visualize the interacting regions lets researchers create distance maps within the protein complexes, allowing them to generate low-resolution three-dimensional maps of the interactions or a general topology of the protein's structure. For viral research, this can include protein interaction topologies between the virus and the host protein receptor. Furthermore, XL-MS can be used to elucidate proteins or protein complex structures involved in the virus-host interactions. Protein structural information can be obtained independently or simultaneously as part of protein interaction studies. XL-MS can be used in parallel with high-resolution techniques such as cryo-EM or X-ray crystallography to obtain highly detailed structural information on protein complexes, multi-subunit complexes and protein stoichiometry. XL-MS can also be used to study protein interactions with small molecules, nucleic acids and lipids, and has recently been used to study proteome-wide interactions. Finally, XL-MS has several advantages over more traditional techniques: it works well with a small sample size, it does not require highly pure samples, it requires minimal analysis time, and it's able to examine interactions in a state similar to the physiological state of an organism.

Examples of work done with XL-MS to identify virus-host interactions include research by Frei et al.[14] The authors used XL-MS to show that the cell surface proteins AXL, M6PR, DAG1, CSPG4 and CDH13 act as binding factors on human cells for mature vaccinia viruses. Additionally, Li et al. used XL-MS to study the response of host cells to human cytomegalovirus (HCMV).[15] They specifically examined the role of interferon-inducible protein IFI16 in viral DNA sensing. The development of MS-cleavable cross-linkers significantly improved the outcomes of such studies. XL-MS can also be used to examine protein interaction topologies between a virus and the host cell proteins. For example, DeBlasio et al. used XL-MS to identify polerovirus (Potato leafroll virus) in complex with plant proteins.[16] They were able to specifically figure out the interaction sites between the virus and the host, as well as the precise residues involved in binding. This information enabled them to precisely map the functional host-pathogen protein interaction topologies.

XL-MS is not limited to protein-protein or nucleic acid–protein interaction studies; it can also be used to elucidate protein structures. Protein structural information can be obtained independently or simultaneously as part of protein interaction studies. For protein structural analysis, XL-MS is used in parallel with high-resolution techniques such as cryo-EM or X-ray crystallography to obtain structural information about protein complexes, multi-subunit complexes and protein stoichiometry. Here, XL-MS provides distance constraints between regions within the protein, resulting in low-resolution, three-dimensional structural information or a general topology of the protein's structure. In the case of viruses, XL-MS can provide extra information about spike protein structural changes upon binding to the host cell receptor.

Examining the dynamic view of virus-host cell interaction with affinity purification mass spectrometry

Enrichment strategies such as affinity purification can be coupled to qualitative and quantitative MS workflows to examine a subset of proteins or protein complexes in complex samples. This workflow, referred to as affinity purification mass spectrometry (AP-MS), uses specific binding interactions between molecules to isolate proteins or protein complexes of interest. It can be used to examine specific protein-protein interactions within protein complexes, or to look at protein complexes more globally at the interactome level. Incorporating quantitative MS with affinity purification enables researchers to examine protein-protein interactions under different conditions, thereby providing a much more dynamic view.

Quantitative techniques such as label-free quantitation (LFQ), stable isotope labeling by amino acids in cell culture (SILAC) and isobaric labeling approaches with tandem mass tags (TMT) can all be used with affinity purification. AP-MS can also be used to examine PTMs and the role they play in facilitating protein-protein interactions. The primary attractiveness of AP-MS is the large number of purification strategies that can be employed, such as various antibody approaches (antigen purification, immunoprecipitation, co-immunoprecipitation), pull-down assays, fusion tag protein purification and avidin-biotin systems.

A typical AP-MS workflow is as follows: proteins of interest are enriched from crude extracts or complex mixtures by passing the complex mixture through a particular ligand that is chemically immobilized to a solid support. The proteins with specific binding affinity to the ligand become bound. After other sample components are washed away, the bound proteins are washed from the support, resulting in their purification from the original sample. Using this strategy allows researchers to obtain information on virus-host protein interactions. By coupling this process with the quantitative MS workflows mentioned previously, researchers can gather the virus-host protein interaction information across different stages of viral infection.

A paper by Weekes et al. showed that using affinity purification coupled to a multiplexing TMT workflow allowed them to look at virus-host protein interactions during HCMV infection.[17] Using this technique, quantitative temporal viromics (QTV), the researchers were able to perform temporal quantitation of the plasma membrane and intracellular proteins. The authors infected primary human fetal foreskin fibroblasts with the HCMV strain, then enriched the plasma membrane proteins with high-affinity streptavidin agarose beads and analyzed them with TMT-based MS experiments. This process enabled the authors to look at changes in the expression of proteins from the plasma membrane across different time points. The researchers quantified 927 plasma membrane proteins and found 56% of proteins changed more than twofold and 33% more than threefold by hour 72 of infection. They also observed expected changes in 21 of 22 cellular proteins and detected six of six previously reported HCMV proteins at the plasma membrane. They determined that five proteins were virion envelope glycoproteins expressed late in infection. Taking a more in-depth look utilizing 10-plex TMT and seven time points, the team quantified 1,184 plasma proteins, getting an almost complete picture of the host and the virus proteome.

In addition to antibody-based strategies, researchers have employed proximity-dependent biotinylation assays (BioID) to examine virus-host protein interactions. Coyaud et al. used BioID to generate host cell protein-interaction profiles for each of the 10 polypeptides found in the Zika virus genome.[18] This enabled them to come up with a protein topology network that showed 1,224 human polypeptides interacting with the Zika virus.

Insights into virus-host cell protein complexes with native MS

Native MS can also be used to study virus-host cell interactions. The attractiveness of the technique comes from the fact that the protein structures and complexes are introduced into the mass spectrometer in similar structural configurations to those they exist in under biological conditions, maintaining the non-covalent interactions between the protein and its binding partners (other proteins, biomolecules or small molecules). The primary types of information obtained by native MS are listed below:

- Subunit stoichiometry—provides information on the composition of the subunits that make up the protein complex—specifically, the number of each unique subunit that is present.

- Subunit identification—reveals information on whether the subunit is bound to the complex of interest and not simply copurified.
- Biomolecule binding—provides information on the interactions between the different subunits or ligands, whether they are happening between a protein and another protein, or with nucleic acids, lipids, glycans and small molecules.
- Protein complex topology—describes the interconnectivity of the subunits. For example, it explores the locations of the subunits within the protein complex, and whether these subunits are present within the inner or outer portion of the complex.
- Protein dynamics—the term for the conformational changes that proteins and protein complexes undergo to form new three-dimensional structures.

The capabilities of native MS can be further expanded when combined with other techniques in the structural MS toolbox. An example is shown in the work of Holmes et al., where the authors combined native MS, limited proteolysis and small-angle X-ray scattering to characterize the cap-snatching endonuclease of Crimean–Congo hemorrhagic fever virus (CCHFV).[19] Native MS was used to characterize the tryptically digested and undigested peptides from the CCHFV domain. Because the samples were being detected in their native conditions, the authors were able to infer the protein interactions that were occurring.

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, such as COVID-19, 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

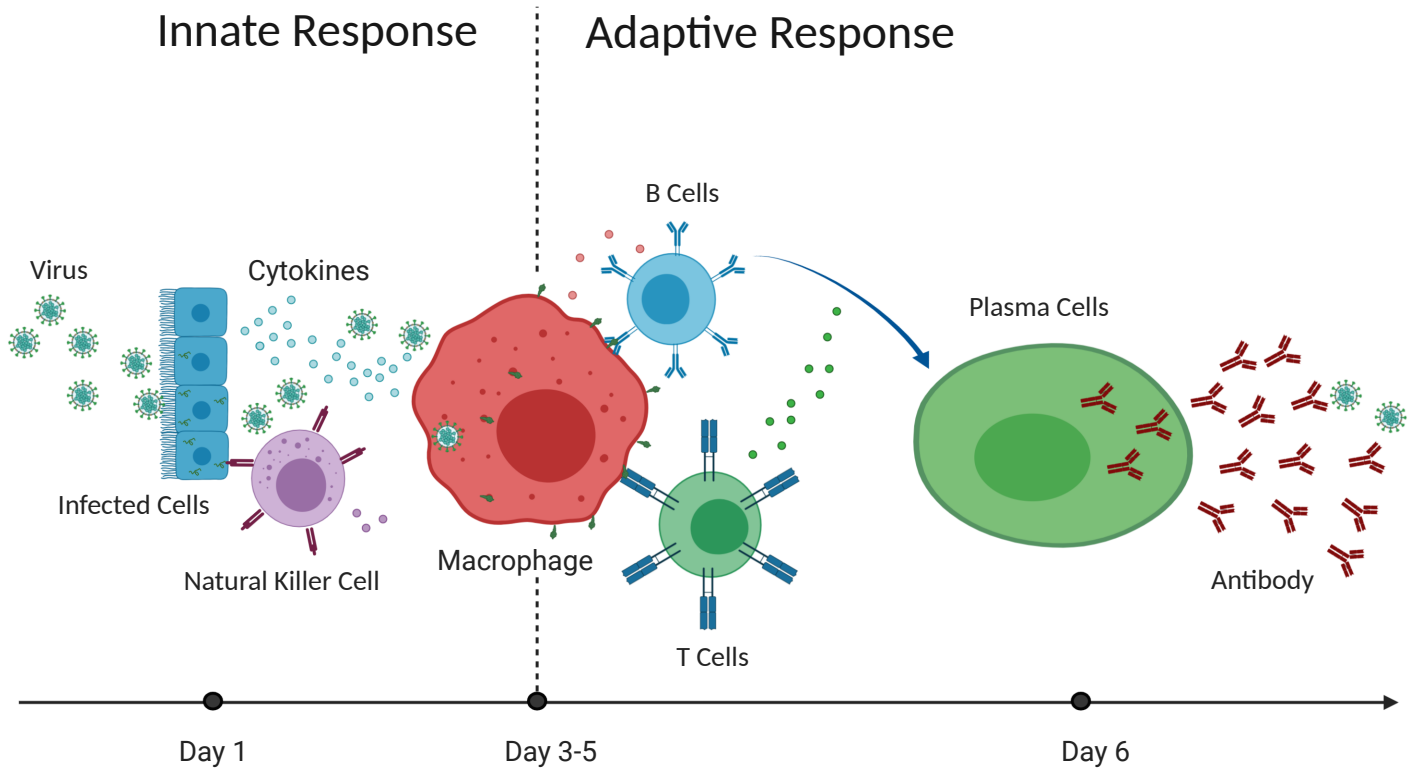

Figure 7. Lifecycle of the host immune system.

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.[20-23] 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, such as SARS-CoV-2.

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 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 PTMs and spliced peptides.[24,25] However, the distinct challenges that arise with this LC-MS-based approach in combination with database search strategies have been discussed in recent publications.[26] 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.[27,28]

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.[29] However, recent progress in machine learning–based approaches to predicting spectra, including an algorithm called PROSIT, have created now opportunities to address this challenge.[30] 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 Thermo Scientific™ Proteome Discoverer™ 2.5 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 PTMs, 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. High-resolution, accurate-mass (HRAM) 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.

The role of MS in vaccine design

The host adaptive immune system responds to viral infection by generating antibodies that neutralize virion binding to host receptors. It has been previously observed that virion composition determines virus stability, transmissibility, tropism and immunogenicity.[31] For example, in the case of influenza, viral hijacking of the host cell machinery results in error-prone replication, which leads to new virion types that are different in size and shape. Viral protein mutations caused by error-prone replication also result in antigenic drift, which allows the virus to evade neutralization by immune system molecules. Despite considerable efforts to consider these dynamic changes in virion composition, the effectiveness of the seasonal influenza vaccine remains unacceptably low, ranging from 10 to 60%.[32] Observations of these antigenic drifts and intense efforts to apply gene sequencing to efficiently characterize circulating strains and the availability of crystal coordinates for the virion envelope binding protein, the glycoprotein hemaglutinin (HA), have not led to improvements in vaccine efficacy. Under evolutionary pressure, it has been shown in influenza A infections that the amino acid residues of the HA domain mutate rapidly, thereby avoiding antibody recognition.[33] During this genetic evolution of the amino acid sequence, potential glycosylation sites also increase. In the case of influenza, the HA globular domain begins with a low degree of glycosylation, but the number of N-glycosylation sites increases as the strains circulate seasonally. More significantly, the amino acids shielded by N-glycosylation appear to mutate at a slower rate than those that are exposed to antibody binding, which emphasizes the importance of studying the glycosylation patterns of virions.[34]

Observations in influenza A indicate that glycosylation impacts the virus' antigenicity, immunogenicity and immunodominance. Antigenicity refers to the capacity of the antigen or hapten to bind specifically with T cell receptors or antibodies, while immunogenicity describes the ability of the antigen to provoke the adaptive immune response in the human body. In response to a viral challenge, the immune system acts in a hierarchical manner known as immunodominance, whereby

immunodominant antigens may suppress immune response to subdominant antigens, which typically bind to less specific but broadly protective antibodies. Those subdominant epitopes are conserved and have low variability due to the need to maintain viral function. Given the role of immunodominance in an effective immune response, the characterization of glycosylation sites and their roles in shielding immunodominant domains and affecting the adaptive immune system response are crucial to defining effective epitopes as targets for vaccine design.[35] More importantly, as the glycosylation patterns of envelope proteins appear to change over time, it is pivotal that researchers understand how glycosylation changes under immune pressure and in response to genetic drift in order to design broadly neutralizing vaccines. The application of high-resolution MS to characterize the glycosylation patterns of existing and emerging viral strains is currently the most effective method of laying the foundation for vaccine design. The glycomics and glycoproteomics strategy discussed earlier can be applied here for characterization.

A major element of vaccine design is the examination of the interactions between the antibody and the target antigen, which can provide crucial information on binding sites, the mechanism of binding and epitope structure. HDX-MS has emerged as a powerful method for epitope/paratope mapping, with important applications in drug discovery and vaccine development. HDX-MS measures the kinetics of hydrogen-deuterium exchange at protein backbone amides. The rate of exchange for a backbone amide hydrogen is directly affected by both involvement in hydrogen-bonding and solvent exposure.[36] When a certain exposed region on an antigen protein interacts with another protein, the level of solvent accessibility decreases significantly, consequently reducing the rate of amide hydrogen exchange and the deuterium incorporation of the affected residues. Thus, decreased deuteration is indicative of protein-protein interactions, and when coupled with proteolytic digestion can provide details regarding the regions of the proteins involved in the interaction. In addition, HDX-MS can reveal the allosteric effects of protein binding in solution, as it reports on the local conformational dynamics of protein complexes in a near-native environment. Unlike X-ray crystallography, HDX-MS requires relatively low amounts and concentrations of proteins and is suitable for proteins with higher levels of polydispersity.[37] Fast, cost-effective HDX-MS studies are enabled by recently developed instrumentation that integrates complete online automated HDX-MS sample management and data collection with software for automated data analysis. While applying HDX-MS to the epitope mapping of large, highly glycosylated proteins such as S protein or antibody-antigen complexes can be challenging, there are already several studies that use this technique. For example, a group from Janssen Pharmaceuticals used this approach to develop neutralizing human monoclonal antibodies against influenza viruses that target HA, the major envelope protein of influenza A viruses.[38]

Another international group of scientists used HDX-MS to probe the local structural stability of antigen-antibody complexes, including key antigenic sites for neutralizing antibodies, for a computationally designed two-component protein nanomaterials vaccine against potent respiratory syncytial virus.[39] The ability to design new self-assembling protein complexes with atomic-level accuracy enables the production of materials with structural features tailored to specific applications.

Conclusion

Whether the focus is on studying the virus itself, virus-host interactions, host immunity, the immune system or virus cellular machinery, proteins play crucial roles. The host proteome will go through a number of processes including production, degradation and spatial reorganization. Understanding these processes will enable researchers to develop strategies to better tackle the virus under study. The vast majority of these studies involve looking at very complex structures involved in myriad interactions. Thus, the study of viruses has benefitted tremendously from the rapid advancement of MS-based structural biology tools. The integration of MS tools such as HDX-MS, XL-MS, AP-MS, native MS, and PTM characterization enables the study of these complexes and the interactions that they participate in, which promotes a better understanding of protein function and mechanism of action in biological systems. Furthermore, MS tools are complementary to traditional techniques such as cryo-EM, X-ray crystallography and NMR by either filling in or extending information that these techniques miss. Recent years have also seen the rapid advancement of quantitative MS tools. The ability to multiplex higher numbers of samples using TMT and powerful new suites of software have helped moved these studies beyond static pictures to more dynamic views of virus systems and the roles they undertake during viral infection, replication and transmission.

References

1. Vankadari, N., and Wilce, J.A. (2020) "Emerging WuHan (COVID-19) coronavirus: Glycan shield and structure prediction of spike glycoprotein and its interaction with human CD26," Emerg Microbes Infect., 9(1), pp. 601–604. doi:10.1080/22221751.2020.1739565
2. Thermo Fisher Scientific, "Native MS for structural biology research," White Paper 73306. https://assets.thermofisher.com/TFS-Assets/CMD/Reference-Materials/wp-73306-ms-lc-native-ms-for-structural-biology-wp73306-en.pdf
3. Snijder, J., Rose, R.J., Veesler, D., Johnson, J.E., and Heck, A.J. (2013) "Studying 18 MDa virus assemblies with native mass spectrometry," Angew Chem Int Ed Engl., 52(14), pp. 4020-4023.
4. Uetrecht, C., Versluis, C., Watts, N.R., Roos, W.H., Wuite, G.J.L., Wingfield, P.T., Steven, A.C., and Heck, A.J.R. (2008) "High-resolution mass spectrometry of viral assemblies: Molecular composition and stability of dimorphic hepatitis B virus capsids," Proc. Natl. Acad. Sci. U.S.A., 105, pp. 9216–9220.
5. Shoemaker, G.K., Van Duijn, E., Crawford, S.E., Uetrecht, C., Baclayon, M., Roos, W.H., Wuite, G.J.L, Estes, M.K., Prasad, B.V.V., and Heck, A.J.R. (2010) "Norwalk virus assembly and stability monitored by mass spectrometry," Molecular and Cellular Proteomics, 9, pp. 1742–1751.
6. Uetrecht, C., Barbu, I.M., Shoemaker, G.K., Van Duijn, E., Heck, A.J.R. (2011) "Interrogating viral capsid assembly with ion mobility–mass spectrometry," Nature Chemistry, 3, pp. 126–132.
7. Watanabe, Y., Allen, J.D., Wrapp, D., McLellan, J.S., and Crispin, M. (2020) "Site-specific glycan analysis of the SARS-CoV-2 spike," Science, May 4: eabb9983.
8. Huang, R.Y., and Chen, G. (2014) "Higher order structure characterization of protein therapeutics by hydrogen/deuterium exchange mass spectrometry," Anal Bioanal Chem., 406(26), pp. 6541–6558.
9. Lanman, J., Lam, T.T., Barnes, S., Sakalian, M., Emmett, M.R., Marshall, A.G., and Prevelige Jr., P.E. (2003) "Identification of novel interactions in HIV-1 capsid protein assembly by high-resolution mass spectrometry," J Mol Biol., 325(4), pp. 759–772.
10. Lanman, J., Lam, T.T., Emmett, M.R., Marshall, A.G., Sakalian, M., and Prevelige Jr., P.E. (2004) "Key interactions in HIV-1 maturation identified by hydrogen-deuterium exchange," Nat Struct Mol Biol., 11(7), pp. 676–677.
11. Monroe, E.B., Kang, S., Kyere, S.K., Li, R., and Prevelige Jr., P.E. (2010) "Hydrogen/deuterium exchange analysis of HIV-1 capsid assembly and maturation," Structure, 18(11), pp. 1483–1491.
12. Bush, D.L., Monroe, E.B., Bedwell, G.J., Prevelige Jr., P.E., Phillips, J.M., and Vogt, V.M. (2014) "Higher-order structure of the Rous sarcoma virus SP assembly domain," J Virol., 88(10), pp. 5617–5629.
13. Bereszczak, J.Z., Watts, N.R., Wingfield, P.T., Steven, A.C., and Heck A.J. (2014) "Assessment of differences in the conformational flexibility of hepatitis B virus core-antigen and e-antigen by hydrogen deuterium exchange-mass spectrometry," Protein Science, 23(7), pp. 884–896.
14. Frei, A.P., Jeon, O-Y., Kilcher, S., Moest, H., Henning, L.M., Jost, C., Plückthun, A., Mercer, J., Aebersold, R., Carreira, E.M., and Wollscheid, B. (2012) "Direct identification of ligand-receptor interactions on living cells and tissues," Nat Biotechnol., 30(10), pp. 997–1001.
15. Li, T., Chen, J., and Cristea, I.M. (2013) "Human cytomegalovirus tegument protein pUL83 inhibits IFI16-mediated DNA sensing for immune evasion," Cell Host Microbe., 14(5), pp. 591–599.
16. DeBlasio, S.L., Chavez, J.D., Alexander, M.M., Ramsey, J., Eng, J.K., Mahoney, J., Gray, S.M., Bruce, J.E., and Cilia, M. (2016) "Visualization of host-polero virus interaction topologies using protein interaction reporter technology," J Virol., 90(4), pp. 1973–1987.
17. Weekes, M.P., Tomasec, P., Huttlin, E.L., Fielding, C.A., Nusinow, D., Stanton, R.J., Wang, E.C.Y., Aicheler, R., Murrell, I., Wilkinson, G.W.G., Lehner, P.J., and Gygi, S.P. (2014) "Quantitative temporal viromics: An approach to investigate host-pathogen interaction," Cell 157(6), pp. 1460–1472.
18. Coyaud, E., Ranadheera, C., Cheng, D., Gonçalves, J., Dyakov, B.J.A., Laurent, E.M.N., St-Germain, J., Pelletier, L., Gingras, A-C., Brumell, J.H., Kim, P.K., Safronetz, D., and Raught, B. (2018) "Global interactomics uncovers extensive organellar targeting by zika virus," Mol Cell Proteomics., 17(11), pp. 2242–2255.
19. Holm, T., Kopicki, J-D., Busch, C., Olschewski, S., Rosenthal, M., Uetrecht, C., Günther, S., and Reindl, S. (2018) "Biochemical and structural studies reveal differences and commonalities among cap-snatching endonucleases from segmented negative-strand RNA viruses," J Biol Chem., 293(51), pp. 19686–19698.
20. van de Weijer, M.L., Luteijn, R.D., and Wiertz, E.J. (2015) "Viral immune evasion: Lessons in MHC class I antigen presentation," Semin. Immunol., 27(2), pp. 125–137.
21. Neefjes, J., Jongsma, M.L., Paul, P., and Bakke, O. (2011) "Towards a systems understanding of MHC class I and MHC class II antigen presentation," Nat. Rev. Immunol., 11(12), pp. 823–836.
22. Dudek, N.L., Croft, N.P., Schittenhelm, R.B., Ramarathinam, S.H., and Purcell, A.W. (2016) "A systems approach to understand antigen presentation and the immune response," Methods Mol. Biol., 1394, pp. 189–209.
23. Soethout, E.C., Meiring, H.D., de Jong, A.P., and van Els, C.A. (2007) "Identifying the epitope-specific T cell response to virus infections," Vaccine, 25(16), pp. 3200–3203.
24. Giam, K., Ayala-Perez, R., Illing, P.T., Schittenhelm, R.B., Croft, N.P., Purcell, A.W., and Dudek, N.L. (2015) "A comprehensive analysis of peptides presented by HLA-A1," Tissue Antigens, 85(6), pp. 492–496.
25. Liepe, J., Marino, F., Sidney, J., Jeko, A., Bunting, D.E., Sette, A., Kloetzel, P.M., Stumpf, M.P., Heck, A.J., and Mishto, M. (2016) "A large fraction of HLA class I ligands are proteasome-generated spliced peptides," Science, 354(6310), pp. 354–358.
26. Faridi, P., Purcell, A.W., and Croft, N.P. (2018) "In immunopeptidomics we need a sniper instead of a shotgun," Proteomics, 18(12): e1700464.
27. 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., and Heck, A.J.R. (2014) "Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD)," Proc.Natl. Acad. Sci. U. S. A., 111(12), pp. 4507–4512.
28. Sofron, A., Ritz, D., Neri, D., and Fugmann, T. (2016) "High-resolution analysis of the murine MHC class II immunopeptidome," Eur J Immunol. 46(2), pp. 319–328.
29. Shao, W., Pedrioli, P.G.A., Wolski, W., Scurtescu, C., Schmid, E., Vizcaino, Courcelles, M., Schuster, H., Kowalewski, D., Marino, F., Arlehamn, C.S.L., Vaughan, K., Peters, B., Sette, A., Ottenhoff, T.H.M., Meijgaarden, K.E., Nieuwenhuizen, N., Kaufmann, S.H.E., Schlapbach, R., Castle, J.C., Nesvizhskii, A.I., Nielsen, M., Deutsch, E.W., Campbell, D.S., Moritz, R.L., Zubarev, R.A., Ytterberg, A.J., Purcell, A.W., Marcilla, M., Paradela, A., Wang, Q., Costello, C.E., Ternette, N., van Veelen, P.A., van Els, C.A.C.M., Heck, A.J.R., de Souza, G.A., Sollid, L.M., Admon, A., Stevanovic, S., Rammensee, H.G., Thibault, P., Perreault, C., Bassani-Sternberg, M., Aebersold, R., and Caron, E. (2018) "The SysteMHC atlas project," Nucleic Acids Res., 46(D1), pp. D1237–D1247.
30. Gessulat, S., Schmidt, T., Zolg, D.P., Samaras, P., Schnatbaum, K., Zerweck, J., Knaute, T., Rechenberger, J., Delanghe, B., Huhmer, A., Reimer, U., Ehrlich, H.C., Aiche, S., Kuster, B., and Wilhelm, M. (2019) "Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning," Nat Methods., 16(6), pp. 509–518.
31. Hutchinson, E.C., Charles, P.D., Hester, S.S., Thomas, B., Trudgian, D., Martinez-Alonso, M., and Fodor, E. (2014) "Conserved and host-specific features of influenza virion architecture," Nat. Commun. 5, p. 5816.
32. Centers for Disease Control and Prevention. "Seasonal influenza vaccine effectiveness, 2005–2018." https://www.cdc.gov/flu/professionals/vaccination/effectiveness-studies.htm.
33. Hensley, S.E., Das, S.R., Bailey, A.L., Schmidt, L.M., Hickman, H.D., Jayaraman, A., Viswanathan, K., Raman, R., Sasisekharan, R., Bennink, J.R., and Yewdell, J.W. (2009) "Hemagglutinin receptor binding avidity drives influenza A virus antigenic drift," Science 326(5953), pp. 734–736.
34. Skehel, J.J., Stevens, D.J., Daniels, R.S., Douglas, A.R., Knossow, M., Wilson, I.A., and Wiley, D.C. (1984) "A carbohydrate side chain on hemagglutinins of Hong Kong influenza viruses inhibits recognition by a monoclonal antibody," Proc. Natl. Acad. Sci. U.S.A. 81(6), pp. 1779–1783.
35. Chang, D., and Zaia, J. (2019) "Why glycosylation matters in building a better flu vaccine," Mol Cell Proteomics., 18(12), pp. 2348–2358.
36. Engen, J.R., and Wales, T.E. (2015) "Analytical aspects of hydrogen exchange mass spectrometry," Annu Rev Anal Chem., 8, pp. 127–148.
37. Sevy, A.M., Healey, J.F., Deng, W., Spiegel, P.C., Meeks, S.L., and Li, R. (2013) "Epitope mapping of inhibitory antibodies targeting the C2 domain of coagulation factor VIII by hydrogen-deuterium exchange mass spectrometry," 11(12), pp. 2128–2136.

38. Puchades, C., Kűkrer, B., Diefenbach, O., Sneekes-Vriese, E., Juraszek, J., Koudstaal, W., and Apetri, A. (2019) “Epitope mapping of diverse influenza Hemagglutinin drug candidates using HDX-MS,” Sci Rep., 9(1), p. 4735.

39. Marcandalli, J., Fiala, B., Ols, S., Perotti, M., de van der Schueren, W., Snijder, J., Hodge, E., Benhaim, M., Ravichandran, R., Carter, L., Sheffler, W., Brunner, L., Lawrenz, M., Dubois, P., Lanzavecchia, A., Sallusto, F., Lee, K.K., Veesler, D., Correnti, C.E., Stewart, L.J., Baker, D., Loré, K., Perez, L., and King, N.P. (2019) “Induction of potent neutralizing antibody responses by a designed protein nanoparticle vaccine for respiratory syncytial virus,” Cell, 176(6), pp. 1420–1431.

thermo scientific

Find out more at **thermofisher.com/proteomics**

ThermoFisher SCIENTIFIC

Acknowledgement: Figures 1,5,6 and 7 were created with BioRender.com. **WP65880-EN 0620S**