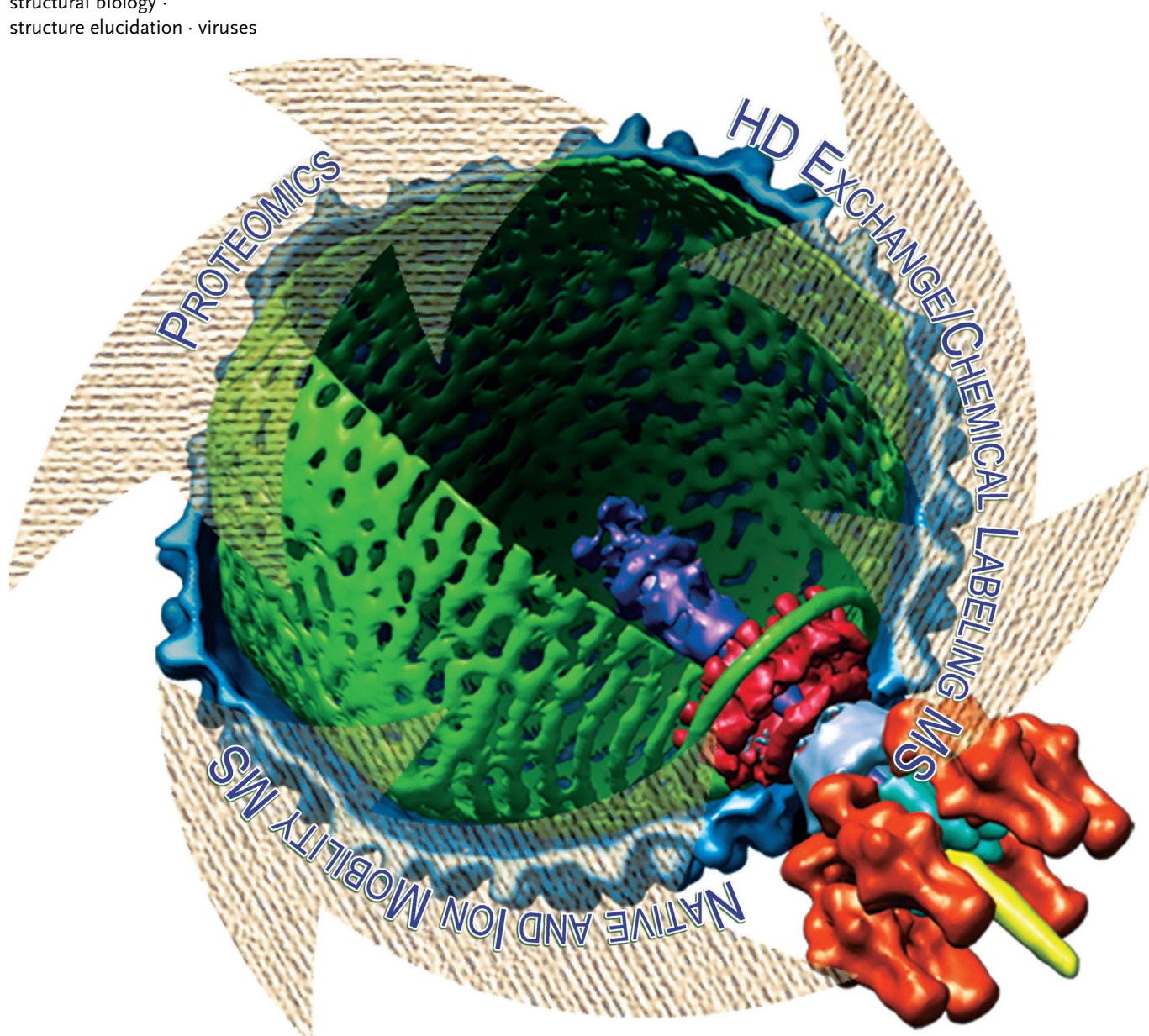# Modern Biomolecular Mass Spectrometry and its Role in Studying Virus Structure, Dynamics, and Assembly

Charlotte Uetrecht and Albert J. R. Heck*

Angewandte
Chemie

Angewandte Chemie 50 International Edition

*O*ver a century since its development, the analytical technique of mass spectrometry is blooming more than ever, and applied in nearly all aspects of the natural and life sciences. In the last two decades mass spectrometry has also become amenable to the analysis of proteins and even intact protein complexes, and thus begun to make a significant impact in the field of structural biology. In this Review, we describe the emerging role of mass spectrometry, with its different technical facets, in structural biology, focusing especially on structural virology. We describe how mass spectrometry has evolved into a tool that can provide unique structural and functional information about viral-protein and protein-complex structure, conformation, assembly, and topology, extending to the direct analysis of intact virus capsids of several million Dalton in mass. Mass spectrometry is now used to address important questions in virology ranging from how viruses assemble to how they interact with their host.

From the Contents

1. Introduction

Mass spectrometry (MS) is a very powerful analytical technique known to and used by most researchers in the natural and life sciences. The technique behind MS dates back about a hundred years, when people like Thomson and Aston were amongst the first to separate particles of different mass to charge ratios (m/z), discovering isotopes of rare gases and other elements. For half a century, MS remained primarily in the hands of physicists, and often shrouded in secrecy, as it was also used to enrich uranium for the Manhattan project.[1] Around the 1960s, the technique began to be adopted by chemists, for instance working in the petroleum industry, to investigate the chemical nature of the compounds formed in refinement processes. In the domain of chemistry, MS subsequently came to bloom particularly in the chemical analysis of unknown compounds (assisted by the coupling to separation technologies, such as gas and liquid chromatography) and as a research area by itself, termed organic MS, concerned with studying structures and fragmentation mechanisms of ions, and ion–molecule reactions.[2]

Up to the 1980s, the impossibility of transferring larger molecules as intact gaseous ions into the vacuum of the mass spectrometer represented a serious bottleneck in the analysis. This problem was gradually solved by new desorption techniques, cumulating in the introduction of matrix assisted laser desorption/ionization (MALDI)[3] and electrospray ionization (ESI) MS.[4] These new ionization methods, together with breakthrough innovations in instrumentation, really opened up applications in biology, nanotechnology, polymer science, and medicine, as well as many other fields in the natural and life sciences.

Focusing on the life sciences, MS became a key technology used for peptide sequencing,[5] through which the identity of a protein can be revealed. The speed and sensitivity in MS allows the qualitative and quantitative analysis of the protein content of a whole cell or tissue, nowadays termed proteomics.[6] Similarly, the chemical analysis of the small-molecule content of a cell or body fluid can now be performed using MS, and is termed metabolomics.[7] New desorption techniques and special dedicated mass spectrometers even allow MS-based imaging by the position-sensitive measurement of compound distributions (protein, neuropeptide, metabolite, drug molecule) in a tissue or organelle.[8] Furthermore, ambient desorption technologies have been introduced to directly sample molecular compounds of surfaces and organisms.[9]

The importance of MALDI and ESI for these revolutionary developments was recognized by awarding the Nobel Prize in Chemistry to the late John B. Fenn and to Koichi Tanaka. Fenn entitled his Nobel Lecture "Electrospray Wings for Molecular Elephants"[10] as ESI (and MALDI) expanded the mass regime attainable for MS at least 1000-fold. Soon after the introduction of ESI, it became apparent that not only the mass of intact proteins but also the tertiary and quaternary structure of these proteins could be partially retained and therefore analyzed. This potential was evidenced by the early discovery that the noncovalent complex between myoglobin and its heme cofactor could be kept intact in the gas phase.[11] Groundbreaking work of Standing et al.,[12,13] Smith, Loo, et al.,[14,15] Robinson et al.,[16–19] and our own group[20,21] in the area of MS on intact macromolecular complexes led to a new very powerful tool in structural biology, now termed native mass spectrometry.[22]

Next to the well-established peptide-sequencing approach and native MS, modern techniques have many other facets relevant for the analysis of proteins in numerous ways.

[*] Dr. C. Uetrecht,[+] Prof. Dr. A. J. R. Heck
Biomolecular Mass Spectrometry and Proteomics Group, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht
Padualaan 8, 3584 CH Utrecht (The Netherlands)
and
Netherlands Proteomics Centre (The Netherlands)
E-mail: a.j.r.heck@uu.nl

[+] Present address: Molecular Biophysics, Uppsala University
Uppsala (Sweden)

WILEY ONLINE LIBRARY

Especially, H/D (hydrogen/deuterium) exchange, and chemical labeling of solvent-accessible amino acids in combination with MS or cross-linking MS can add to our knowledge about protein structure–function relationships. In this Review, we will describe how such tools in MS can be used to study several aspects of protein structure and function, focusing in particular on the biochemical and biophysical properties of viruses and viral particles.

Viruses are ideal model systems to study the assembly of protein complexes, since the viral protein shells, capsids, often have the amazing ability to self-organize their folding and assembly even in vitro without the help of chaperones.[23] Moreover, their natural capacity of encapsulating material, that is, the viral genome, makes virus capsids an interesting target for nanotechnological applications that extend far beyond drug delivery.[24] The detailed biophysical and biochemical characterization of the virus assembly and maturation processes is crucial, as such data may potentially be used to interfere with viral infection.[25] Technically, studying virus assemblies is rather challenging as the structures formed can be very large, hampering analysis by conventional structural-biology techniques such as protein X-ray crystallography and NMR spectroscopy. Another problem is posed by the transient nature of the intermediates formed during assembly and/or maturation impeding their purification and analysis.[26] Although virus structure, dynamics, and assembly has been studied for decades, in recent years MS has entered this area of research and tackled several important questions that were less accessible by other means.[27] Proteomics approaches were early on used to study the "new" SARS virus,[28] but MS is also an emerging method to reveal the constituents of a virus, the stoichiometry of the viral structural proteins, virus assembly and the corresponding intermediates.[27] After introducing briefly a few general concepts in structural virology, we describe how modern MS can assist in virus characterization, especially focusing on structural aspects.

2. General Concepts in Structural Virology

Viruses are infectious agents that can replicate inside a host.[29] With sizes from nm to μm, most viruses are invisible under the light microscope.[30–32] Viruses are ubiquitous; they appear in archaea, bacteria, plants, and animals.[33–38] Estimates suggest that after prokaryotes, viruses account for the second largest amount of biomass on earth.[39] Amongst the viruses, those hosting in bacteria are most abundant and called phages. The viral genome encodes all viral proteins necessary for replication in the adequate host, which can range from several hundred to just a few different proteins. Owing to their inability to reproduce independently, viruses are often not regarded as a form of life.[40] However, the discussion is still ongoing and was recently fueled by the discovery of giant mimiviruses.[30,41] Furthermore, the evolutionary ancestry of viruses is still unclear: did they arise from pieces of nucleic acids replicating in cells; or reduce from cellular organisms?[42,43]

The ubiquity of viruses amongst all species already foreshadows a broad diversity complicating viral classification. Generally, viruses are distinguished based on either their host organism or morphogenetic characteristics and genome organization.[44–47] The genome can be encoded by both RNA or DNA, in single- or double-stranded (ss and ds) form as exemplified in Figure 1. The information can be located on (+) or (−) strands and in some cases requires reverse transcription for replication. Viruses can contain single or multiple pieces of nucleic acid in linear or circular form. The real evolutionary relationship between viruses is often difficult to obtain from one single feature, complicating virus taxonomy. The viral genome is typically enclosed in a protein shell termed the nucleocapsid. This capsid can be helical, icosahedral, or more complex in structure. The nucleocapsid alone can facilitate host-cell attachment and entry, but especially eukaryotic viruses are often enveloped with a lipid bilayer containing the adaptor proteins.[48]
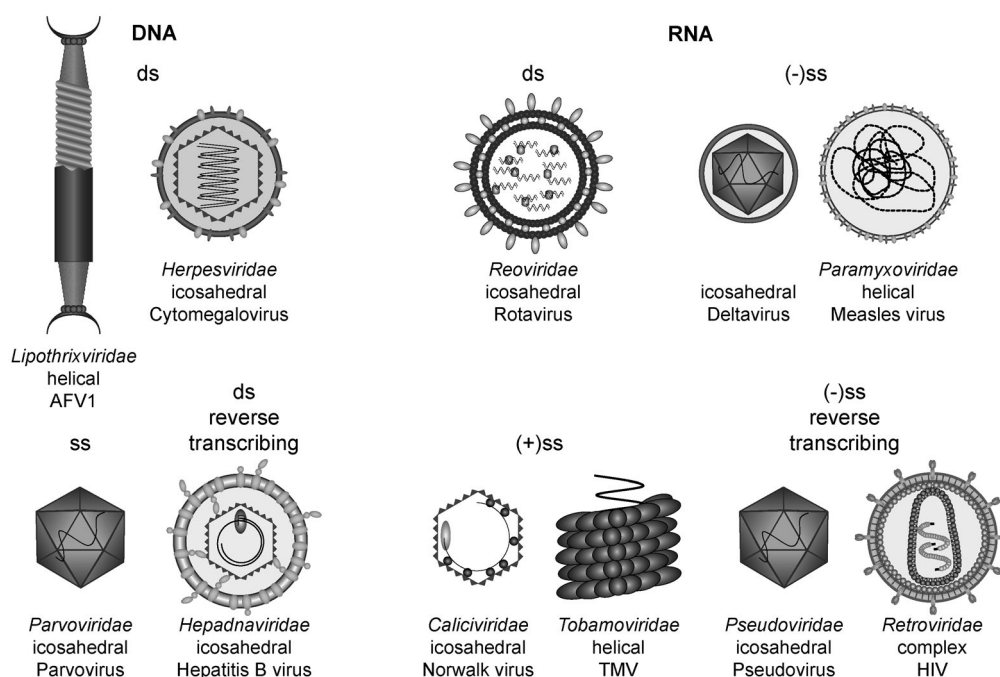
2.1. Infection and Lifecycle

To replicate, viruses first have to recognize the host and introduce their genome into the cell. Next, the protein and nucleic acid machinery of the cell is taken over to produce the viral constituents, finally comes the assembly and release of the infectious virus. After successful attachment to the host cell, the nucleocapsid can enter the cytosol by various mechanisms (Figure 2), such as, membrane fusion or phagocytosis often used by eukaryotic viruses.[37,49,50] Bacteriophages generally inject their genome directly into the cell,[51,52] whereas most other viruses release the genome from the capsid at the pore complexes of the cell nucleus.[53]

Charlotte Uetrecht, born 1982, studied biochemistry at the University of Potsdam, Germany. From 2006 to 2010, she was a PhD student at Utrecht University under supervision of Albert Heck. Her research there focused on the mass spectrometric analyses of viral assemblies. Since 2011 she is an EMBO long-term fellow at the Molecular Biophysics department, Uppsala University in Sweden, investigating viral and other large structures/particles with MS and X-ray free electron lasers in vacuum.

Albert J. R. Heck, born 1964, has held the chair of Biomolecular Mass Spectrometry & Proteomics at Utrecht University since 1998. Since 2003 he has been director of the Netherlands Proteomics Centre, and since 2006 director of the Bijvoet Centre for Biomolecular Research. His research interests comprise all aspects of protein mass spectrometry. He has received awards including the gold medal of the Dutch Royal Chemical Society, the Descartes Huygens Award from the Republique Francaise, and the Life Science Award of the German Mass Spectrometry Society.
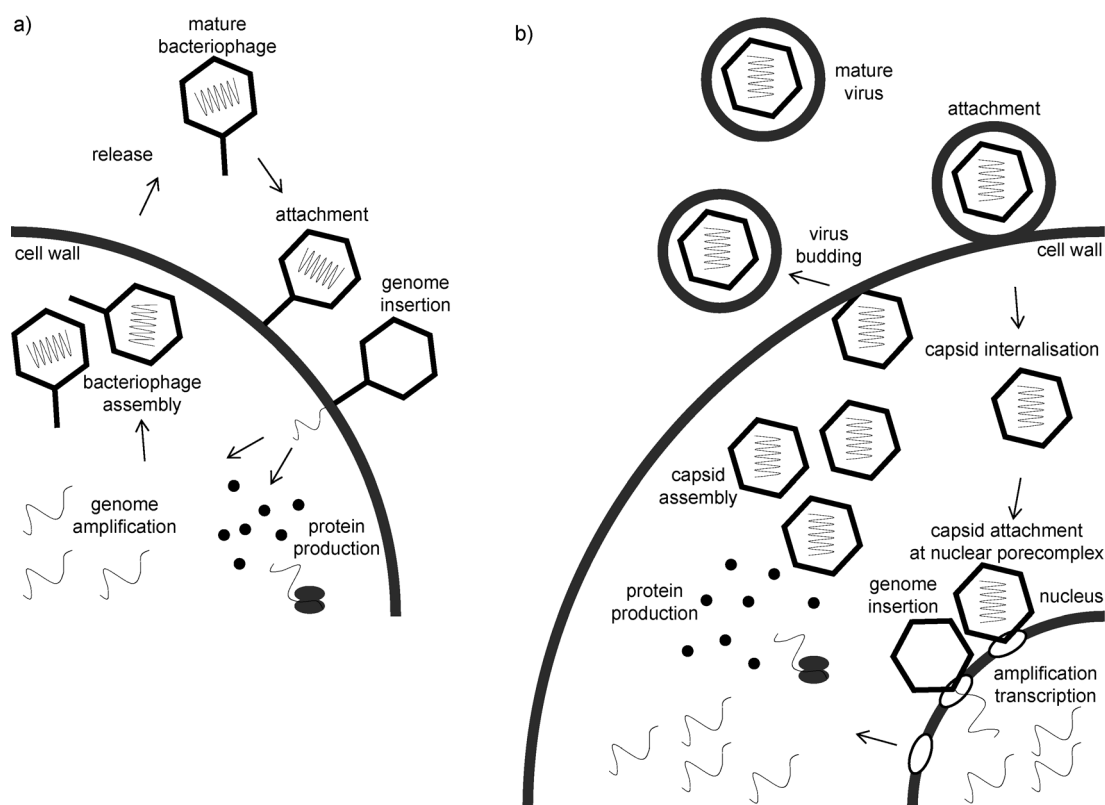
Figure 1. Virus classification is largely based on the type and organization of the nucleic acid. A selection of virus families and their corresponding morphologies are depicted.

2.2. Capsid Assembly and Structure

The viral nucleocapsid is of crucial importance in the viral lifecycle since it encapsulates and protects the viral genome. Especially their high efficiency to self-assemble, their strength, and efficiency in nucleic acid packaging mark nucleocapsids as intriguing structures. Typically, the capsid shell is formed by multiple copies of one or a few different structural proteins,[72,73] although decorations with other proteins are common in non-enveloped viruses.[74] These attached proteins can increase the capsid stability, or serve a role during infection and genome packaging. The high copy number of a limited set of small proteins beneficially decreases the length of nucleic acid needed to encode for them, requiring less space for encapsulation.[72] This illustrates some of the brilliant and efficient principles underlying virus structure and function.

Many capsid proteins (cp) can readily be produced recombinantly in high quantities and represent thus ideal model systems to study protein (self-)assembly.[23] Even though, the building blocks in capsid assembly are different for certain viruses, their formation is generally in agreement with nucleation theory as has also been proposed in amyloid assembly.[73,75,76] First, an assembly nucleus has to be formed, which may be an oligomeric assembly or just a conformationally changed cp monomer. After formation of this nucleus, further building blocks attach to it until the capsid is completed (Figure 3). Under conditions of efficient assembly, the nucleation is generally the rate-limiting step and only a small fraction of the proteins are in this intermediate state. The following elongation steps take place at a much higher rate leading to immediate propagation of the nucleus to a capsid. Therefore, the intermediate oligomeric species forming the assembly nucleus are typically only present in trace amounts under assembly conditions.[26,73,76,77] For some viruses, it is possible to change the solution conditions favoring over-nucleation, whereby intermediates become kinetically trapped. Still, even under such conditions, processes such as protein mis-folding and aggregation, can further hamper the detection of these intermediates.[67,78–80] A low nucleation rate ensures effective capsid assembly. Nucleation can be triggered by increasing concentration or posttranslational modifications to secure a sufficiently high titer of cp in the cell. Also, interactions with the nucleic acid can facilitate nucleus formation.[78,81,82]

An exception are (+)ssRNA viruses, which also uncoat in the cytoplasm, but replicate their genome outside the nucleus in close proximity to the membrane.[54] Some viruses are incorporated metastably into the host genome, for example retroviruses and lysogenic phages. Replication of such proviruses is then induced spontaneously or after a specific trigger.[55–57] Generally following infection, viruses take control of the host-cell machineries to facilitate their own replication.[58] At late stages of infection, the structural proteins either assemble around new copies of nucleic acid or assemble independently and package the nucleic acid afterwards.[59–61] Eventually, mature virions are released by budding like vesicles or the cell can rupture as a consequence of the viral load.[62,63] The virus can then spread and infect new cells, reinitiating the cycle. Also multiple ways of replication occur during the lifecycle of different viruses, an important step is usually maturation, signaling complete assembly of infectious particles that are ready for release. This process includes a variety of biochemical adaptations such as conformational changes triggered by the nucleic acid incorporated, attachment of auxiliary proteins, or protein posttranslational modifications, such as phosphorylation, glycosylation, proteolysis, and cross-linking.[64–68] Viruses represent a strong selective pressure on the host. Resistant cells have an evolutionary advantage by increased viability. As a consequence, viruses evolve by constantly altering their genome. RNA viruses, in particular, often show a high mutagenesis rate enforcing a constant struggle for survival.[69] Another mechanism which results in an increased virulence and a switch in species specificity is the recombination of partial genomes from related viral strains, as is common in influenza viruses.[70,71]

Figure 2. Viral lifecycle: a) Prokaryotic viruses (phages) attach to the host cell and directly inject their genome as shown in this case for a tailed phage. Then, the genome is amplified and proteins transcribed in the cytosol. In case of tailed phages, an empty capsid forms, then the genome is incorporated followed by maturation. The assembled phages accumulate in the cytoplasm until the cell ruptures. b) Eukaryotic viruses transfer their capsid into the cytoplasm. Internalization can occur by membrane fusion in the case of enveloped viruses. At the nucleus the capsid disintegrates and releases its genome, which is reproduced in the nucleus. Protein synthesis and assembly take place in the cytoplasm. For (+)ssRNA viruses the genome is also amplified in the cytoplasm. After assembly and possibly maturation, the virus is released from the cell through budding or destructing the cell.

The corresponding viral capsids are often highly stable towards changes in environment and can resist, for example, extreme pH values, high concentrations of denaturants and organic solvents, dilution to very low concentrations that don't facilitate assembly, and even dehydration.[83–87] This effect is reflected by a strong hysteresis observed in virus dissociation experiments. Theoretical and experimental results suggest that the interaction between individual building blocks is rather weak and cannot account for the high apparent stability.[73,88] However, in the capsid the binding energies of the subunits add up, explaining the pseudo-stability of the capsids under conditions where assembly typically does not occur. Nevertheless, theory suggests that in such a pseudo-equilibrium, there are always some free building blocks in solution that could dynamically exchange with proteins in the capsid. This process has been termed "capsid breathing" and some indirect experimental evidence has been described supporting this model.[88–92] We like to note that the term "breathing" has also been used to describe the significant conformational changes that can occur in viral cp proteins, sometimes even resulting in transient external-ization of domains that, according to structural models, are on the inside.[91,93]

Most commonly, either helical or icosahedral capsid structures are observed, which both allow the formation of a regular shell with multiple copies of a single cp as a result of the high symmetry.[72] The prototype of a helical virus is the tobacco mosaic virus (TMV). In this virus, cp monomers assemble around the RNA and the length of the genome defines the nucleocapsid size.[94,95] Icosahedral structures allow the complete closure of the shell using just one type of protein.[72] Additionally, the almost spherical structure reduces the surface area relative to the enclosed volume. An icosahedron consists of 12 vertices, 20 faces, and 30 edges corresponding to the different symmetry axis (5-, 3-, and 2-fold, respectively). At least 30, generally dimeric, building blocks are required to form the smallest possible icosahedron, where all proteins are located in pentamers. Larger capsids are formed by addition of hexamers. Only certain numbers of hexamers can be inserted to produce a perfect icosahedron reflected by the triangulation number (T): $T = h^2 + hk + k^2$ where h and k can be any positive integer and $T=2$ is therefore not allowed. The number of building blocks corresponds to $30\,T$. Even though viral capsids can be built up by a single cp, the surrounding contacts between subunits in hexamers and pentamers are different.[96,97] However, the conformational changes to compensate this are often mar-ginal, yielding cp structures that are quasi-equivalent. Addi-tional hexamers may be introduced in a ring-like fashion leading to prolate capsids as in bacteriophage Phi29.[98,99]
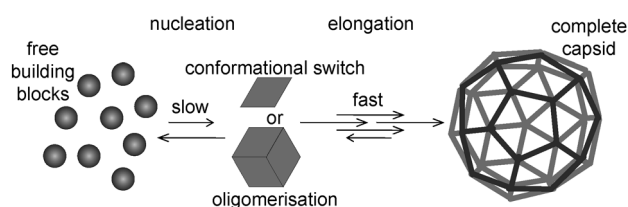
Figure 3. Model for nucleated assembly: Conformational change or oligomer formation can result in nucleation. The nucleus formation is a slow reaction, whereas the subsequent addition of building blocks proceeds fast towards capsid completion. In cases where the nucleation is a fast process, assembly intermediates accumulate because of overnucleation and few capsids are formed.[73]

Complex viruses often deviate from icosahedral symmetry, for example the HIV cp typically forms conical, but also rod-shaped capsids.[100]

3. Mass Spectrometry in Structural Virology

Next, we describe how biochemical and biophysical properties of viruses can be studied by modern MS. Before focusing on some case studies we first summarize some of the key technologies used, in four boxes: "proteomics", "native and ion mobility mass spectrometry", "H/D-exchange mass spectrometry", and "chemical-labeling approaches coupled to mass spectrometry". These refer the readers to background Reviews on these individual methods. When applied to studying virus structure, dynamics and assembly, the information obtained by these four approaches is schematically summarized in Figure 4.

Box 1: Proteomics

MS-based proteomics is currently the most powerful tool to obtain sequence information of proteins.[6] Common practice in proteome analyses is the sequencing of proteolytic peptides obtained through fragmentation by collision-induced dissociation (CID) and/or electron transfer dissociation (ETD).[101,102] These fragmentation spectra are then searched against large protein-sequence databases containing the predicted spectra, derived from continuously actualized genome/protein databases.[103] Besides identifying proteins, MS-based proteomics can also be used to identify and localize naturally occurring posttranslational modifications (such as glycosylation,[104] phosphorylation,[105] lysine and N-terminal acetylation)[106] or protein chemical modifications (induced by cross-linking reactions, surface mapping by oxidation,[107] acetylation,[108] or deuterium incorporation.[107]) For peptide/protein identification the very selective protease trypsin is the enzyme of choice, although other enzymes are becoming popular especially in conjunction with the use of ETD.[109]

Box 2: Native and Ion Mobility Mass Spectrometry

In ESI-MS, molecules are ionized by a combined process of desolvation and (de)protonation. ESI is most sensitive at low flow rates (nl min^{-1}) and optimal conditions can be obtained by bringing the analyte into a 1:1 mixture of water and organic solvent. Such solutions are typically acidified to promote protonation. Although these solvent conditions allow ultra-sensitive detection, they typically denature and unfold proteins. However, the goal of native MS[21,22,110] is to preserve higher order protein structures to enable the investigation of protein conformation, protein complex topology, and dynamics. Therefore, the samples need to be held as close as possible to physiological conditions. An ESI-MS compatible "volatile buffer" is provided by aqueous ammonium acetate, whereby salt concentrations can be varied from approximately 5 mM to 1M, retaining a neutral pH value. From numerous biophysical validation studies it has become apparent that many quaternary protein structures can be preserved under these conditions. During the ESI process the volatile buffer easily evaporates, leaving "naked" protein ions, albeit substantially less charged than when sprayed from organic ESI solvents, as the surface is more compact in these folded species. Since larger protein assemblies may attain m/z values exceeding a few thousand, dedicated/modified time of flight (ToF) mass analyzers are required for detection.[111,112] Retaining quaternary structures in the gas phase opens up ways to measure the mass of intact protein complexes and sub-complexes, from which information about stoichiometry and topology can be extracted.[19,113] To probe the structure of the protein complexes, sub-complexes may also be formed intentionally using either a low concentration of denaturant, a shift in pH value and/or ionic strength[114] or through CID.[115–117] By in vitro reconstitution of membrane protein complexes in micelles, even membrane-embedded protein complexes can be studied by native MS.[118] A further strength of native MS is its high sensitivity, allowing even the analysis of endogenously expressed protein complexes.[119,120]

The available toolbox has recently been extended by the coupling of ion mobility (IM) separation to MS (IMMS).[121] In IMMS, ions are separated not only on the basis of their m/z but also, inside a gas-filled ion-mobility chamber, according to their drift time, which depends on their overall shape or collision cross section (Ω). Typically, molecules with larger Ω values, that is, larger apparent volumes, exhibit longer drift times. Using IMMS data, the Ω value or average projected area of a protein or protein complex can be determined. Early results have revealed that solution-phase structures can be, in particular for larger protein complexes, mostly retained in the gas phase.[121,122] IMMS is nowadays used in conjunction with computational modeling to generate refined structural models for protein complexes.[114,123] For instance, by having high-resolution structures (from X-ray crystallography or NMR spectroscopy) of the protein complex constituents, the Ω value of the intact complex and/or sub-complexes can be used to predict structural models.

Box 3: H/D-Exchange Mass Spectrometry

In H/D-exchange MS, the incorporation of deuterium atoms into proteins is monitored over time.[124–128] The method is based on the exchange of solvent-accessible backbone hydrogen atoms with deuterium atoms when a protein is
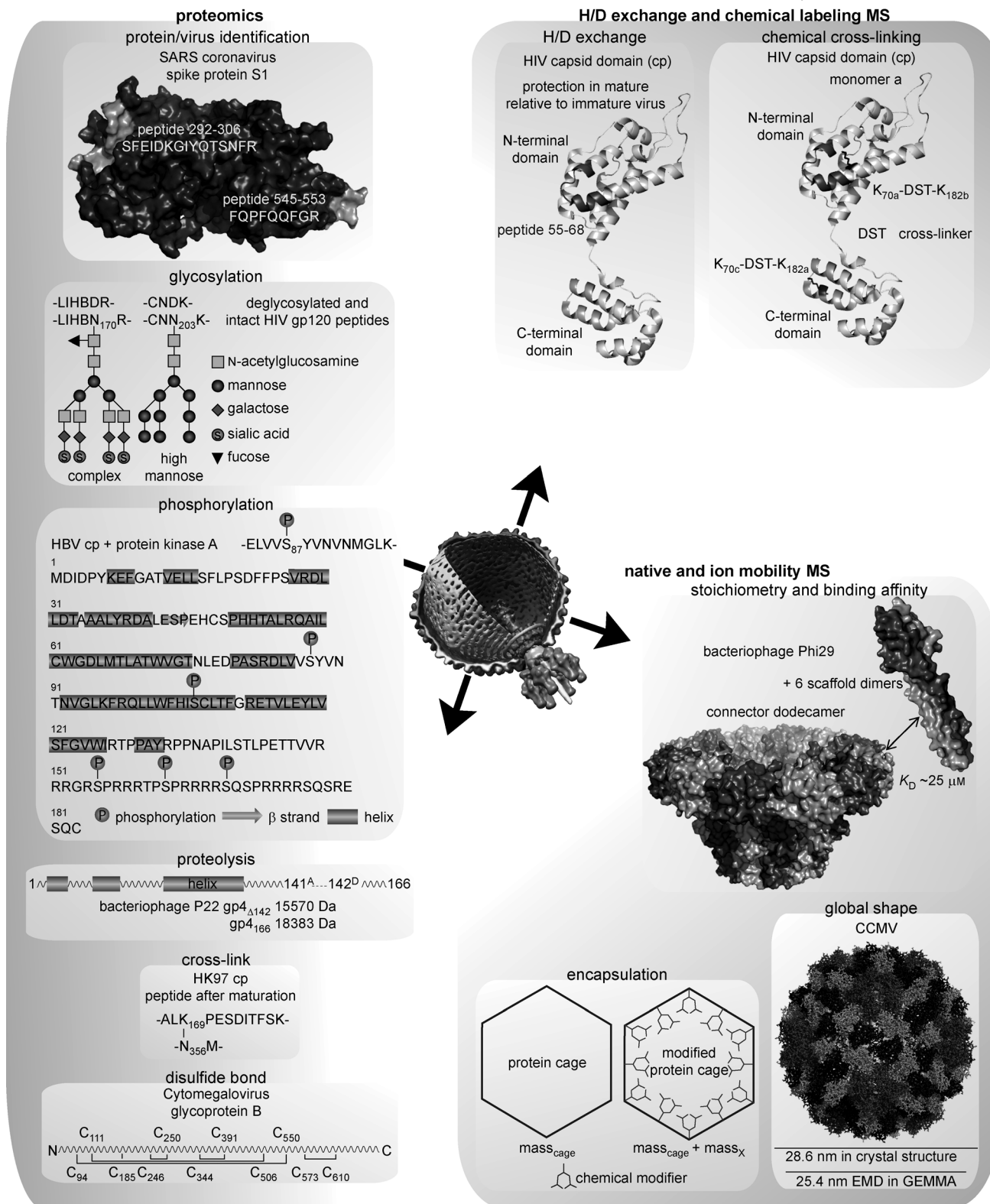
Figure 4. Applications of mass spectrometry in structural virology: Illustrative examples on how modern MS has recently contributed to structural virology. Examples include the applications of MS for virus protein identification, including their posttranslational modifications and the structural analysis of viruses using H/D-exchange MS and native ion mobility MS. Left: from top to bottom is illustrated how proteomics approaches can be used to probe the composition of a virus. Proteins of the SARS coronavirus were identified early on by MS, the detected peptides of the spike protein S1 are shown (mapped onto the structure model 1Q4Z).[143] Beneath, schematic illustrations of how proteomics was used to map the precise disposition of heterogeneous glycosylation patterns in the major HIV surface protein,[65] multiple phosphorylation sites in the HBV cp,[157] proteolytic cleavage sites in bacteriophage P22 gp4,[193] and the location of cross-links like in HK97 maturation.[67, 165] Right: How MS can be used to obtain more structural information, from top to bottom: how H/D exchange and chemical labels, for example, cross-linking were used to dissect a crucial step in HIV capsid maturation.[64] How native and ion-mobility MS provide information on the stoichiometry of an assembly, in favorable cases even the binding affinity,[160] the amount of material encapsulated,[184] and the shape of a viral protein complex.[172]

placed in deuterated water (D$_2$O). The subsequent increase in protein mass over time is measured with MS. Using intact or native MS global exchange in a protein or protein complex can be monitored providing information on major conformational changes, for example upon ligand binding.[129] The more detailed location of the deuterium incorporation can be determined by monitoring the mass shift in peptic fragments that are produced after the H/D-exchange reaction. Therefore, the samples are incubated in D$_2$O over different times and then typically diluted to acidic solution conditions (pH ≈ 2.5) at low temperatures (0 °C) to slow down back-exchange processes. However, few proteases can efficiently digest proteins at such pH values. Pepsin has an optimum activity at low pH values and is the preferred enzyme for H/D-exchange applications. Even though pepsin is regarded an unspecific protease, the detected peptides are reproducible.[130] After digestion, the peptides obtained are subjected to MS analysis and identified by exact mass and fragmentation pattern. The mass shift owing to deuterium incorporation can then be monitored over time, elucidating which peptides are engaged in structural changes occurring upon protein complex formation or, for instance virus maturation.[66,131] H/D exchange coupled to MS has become a valuable analytical tool for the study of protein dynamics. By combining this information with classical functional data, a more thorough understanding of protein function can be obtained. The H/D-exchange MS approach is comparable to, and has been used in conjunction with, NMR spectroscopic experiments in which the H/D exchange is monitored over time. Although H/D exchange analysis was, for a long time, somewhat limited to small proteins or protein domains, improved resolution and sensitivity in mass analyzers combined with better software for data interpretation now also allow large proteins, such as whole antibodies, to be investigated.[132]

Box 4: Chemical-Labeling Approaches Coupled to Mass Spectrometry

Next to H/D exchange, there are a couple of alternative chemical approaches to probe the surface accessibility and interconnectivity of proteins in protein complexes which are frequently used in combination with MS.[133] In these approaches specific amino acids are rapidly and efficiently chemically labeled under pseudo-physiological conditions. MS is then used to probe the chemically induced mass shifts in the peptides/amino acids affected by the label. The idea is that only the accessible amino acids will be modified, allowing conformational changes in proteins to be monitored, for instance upon ligand binding or protein-complex formation. Most popular in the field of structural biology are oxidative labeling by hydroxyl radicals[134–136] or labeling of free accessible amine groups, for instance by acetylation.[108,133,137] Chemical labeling using molecules with at least two reactive groups can also be used for cross-linking specific amino acids that are in close proximity to each other. Using such cross-linking approaches, intra- and intermolecular interactions can be identified in a protein complex.[138–140] The most commonly used bifunctional chemical cross-linkers target lysine residues, whereby the linker rigidity and spacing in between the two

reactive groups defines the range of interactions that can be probed, providing distance restraints for computational modeling. After modification of the proteins, residual reagent needs to be removed or inactivated. Following proteolysis, peptides originating from specific cross-linked regions of the proteins need to be filtered out of the background of unmodified peptides, for which dedicated software is usually a prerequisite.[139,141] For instance, chemical cross-linking and MS were applied to probe subunit–subunit interactions in the bacteriophage P22 procapsid.[138]

3.1. Identification of Viral Proteins and Their Posttranslational Modifications

New viral infectious pathogens are most efficiently characterized by genotyping.[142] However, such experiments do not directly provide information about the expressed viral proteins. Using standard proteomics experiments (Box 1), viral proteins can easily be identified even from complex samples such as host-cell lysates. Such data provides direct information about the composition of the virion and may reveal its structural and accompanied proteins.[143] An illustrative example is provided by the human coronavirus causing SARS, which emerged around 2003 in Asia. After the outbreak, all the predicted structural proteins were soon identified by MS (Figure 4), and also several glycosylation and phosphorylation sites could be mapped.[28,143,144] With a time delay of a few years, some of these proteins could be structurally analyzed using high-resolution techniques, such as X-ray crystallography.[145]

Quantitative proteomics experiments have been used to study dynamic temporal changes in the host proteome upon infection by the pathogen, with the aim to identify the host–pathogen interactome.[146,147] Therefore, often some form of isotope labeling is used enabling the identification of proteins whose expression level is most affected by the infection. For example, isotope-labeling strategies in combination with proteomics provided information on the cellular changes upon SARS infection and allowed the identification of host-cell factors putatively involved in virus replication.[148,149]

Surface-exposed viral proteins or protein domains usually mediate the attachment of the virus to the host. For enveloped viruses, such as HIV, those proteins are located in the lipid bilayer and are typically highly glycosylated. These glycans are likely involved in antigenicity, shielding the virus from the immune system. The sites and in particular types of glycosylation are heavily affected by mutations, hampering vaccine development against HIV.[150] Proteomics methods have been used to reveal major changes in the glycosylation patterns between different viral strains.[151–153] After peptide digestion, the mass discrepancy between glycosylated and enzymatically deglycosylated samples determines the size of the carbohydrate. Using tandem MS, the modified amino acid and the composition of the glycans can be determined.[65,151,152] Viral protein decoration by glycans may be very complex, as attached carbohydrates typically vary in length and type, whereby all three structural classes, high mannose, complex, and hybrid, occur and have been detected

on the HIV gp120 protein (Figure 4). Presumably because of the high structural flexibility and heterogeneity of the glycan layer, only a deglycosylated variant of gp120 could be crystallized to date. Binding of CD4 and chemokine receptors on human T-cells to gp120 is apparently influenced by the glycosylation pattern, making this sort of analysis important.

Although virus capsids may self-assemble, the assembly process in vivo is much more complex, and the host cellular machinery can regulate the assembly process. For example, phosphorylation of HBV cp enhances the formation of DNA from RNA in the assembling capsid.[154] This is likely connected to the enhanced capsid assembly and encapsulation of RNA observed after phosphorylation of multiple HBV cp amino acid residues.[155–157] Proteomics and mutagenesis studies have revealed that the host kinases PKA and PKC each phosphorylate one of the two central serine residues on HBV cp (Figure 4).[156,157]

3.2. Monitoring Virus Maturation

Proteins involved in viral assembly and maturation processes are less prone to mutational changes than the surface proteins, and therefore offer potential targets for interference. MS has been used extensively to elucidate not only virus structure, but also to monitor dynamic changes therein throughout the viral lifecycle. Such modifications and rearrangements, which can take place during virus assembly

and maturation, and also upon infection, have traditionally been mapped using fluorescent labels or globally monitored by spectroscopic techniques.[158,159] Moreover, mutagenesis studies, such as alanine scanning, have been very valuable to characterize and localize posttranslational modifications occurring during maturation.[67,156,160,161]

In an elegant example, several MS-based technologies have been used and combined with electron microscopy (EM) and X-ray crystallography to investigate capsid formation and maturation of the icosahedral bacteriophage HK97. This lambdoid phage stores its dsDNA under high pressure and thus demands a stable capsid structure and irreversible assembly to ensure protection of the genome.[68,162] Initially, 420 cp subunits organize into pentamers and hexamers. These capsomers then build the spherical, thick-walled prohead I (Figure 5). The N-terminal Δ-domain functions as a scaffold and is cleaved off during maturation leading to prohead II. Such examples of proteolytic cleavage have been observed in many viruses and the exact location of cleavage can be identified using proteomics based methods (Figure 4).[163] A critical step in HK97 maturation is the conformational change leading to a thin-walled icosahedral capsid or phagehead. This expansion is usually accompanied by DNA packaging, but can be triggered in vitro in the absence of DNA and of the packaging machinery. Major structural changes associated with this transition were recognized in EM and crystallography studies. However, the resolution achieved was limited for some intermediate structures. H/D-exchange MS (Box 3)
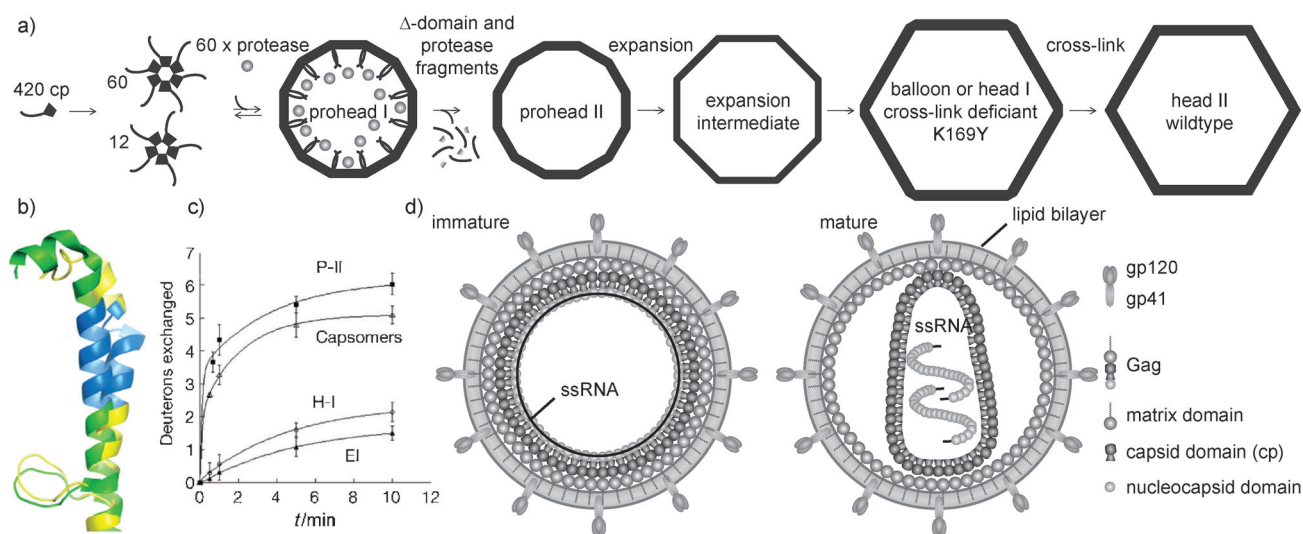
Figure 5. HK97 and HIV virus-maturation studied by mass spectrometry. a) The cp of bacteriophage HK97 preassembles into hexamers and pentamers. These then form the prohead I together with the viral protease. Assembly is facilitated by the cp Δ-domain, which functions as a scaffold. The protease cleaves the Δ-domain and itself leading to prohead II marking the first irreversible step in capsid formation. Under certain conditions expansion occurs followed by the second irreversible step, auto-catalytic cross-linking of cp subunits. The head I of a cross-link deficient mutant closely resembles the mature head II.[66] b) The spine helix from the crystal structures of one subunit in prohead II (yellow) was aligned with the straightened helix in head II of the same subunit. A peptide analyzed by H/D exchange (c) is shown in blue. As revealed by the exchange data in (c), the straightening of the helix occurs upon transition from the prohead II (P-II) to the expansion intermediate (EI). Also shown are the deuterons incorporated in the corresponding peptide fragments for the free capsomers and head I (H-I). Reprinted by permission from Macmillan Publishers Ltd: Nature,[66] 2009. d) In HIV, the immature capsid of the Gag polyprotein encloses the (+)ssRNA and becomes enveloped. After release from the cell, the viral protease cleaves the Gag into three major proteins and some small peptides. The matrix domain stays bound to the lipid envelope through a myristoyl residue and the nucleocapsid domain is associated with the ssRNA. A conformational change in cp leads to a collapse of the spherical towards a conical capsid. At this stage the virus is infective. Other proteins, including the protease and peptide fragments, are omitted for clarity.[64]

© 2011 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

on the different intermediates in the head assembly, from capsomers to the nearly mature head I, could precisely demonstrate which amino acids were involved in the changes accompanying maturation (Figure 5).[66, 131, 164] The next step in maturation is the autocatalytic cross-linking of the cp subunits locking the capsid in the expanded conformation. The specific peptide carrying a rather unusual lysine–asparagine cross-link could be identified by MS after proteolysis by trypsin (Figure 4).[67]

More typical inter- and intramolecular cross-links are provided by disulfide bonds, which are commonly observed in mature viruses. For instance, stabilization of the Cytomegalovirus glycoprotein B involves extensive disulfide bonds, which have largely been mapped using MS and such disulfide linkages are apparently common amongst *Herpesviridae* (Figure 4).[165]

Another approach commonly used to identify and map structural changes or binding surfaces between viral proteins is chemical cross-linking coupled to MS (Box 4).[141, 160, 166] H/D-exchange and chemical cross-linking experiments[64, 166–168] have been combined to analyze the pleiomorphic HIV capsids. Their variable appearance obstructs classical high-resolution structural analysis, since this generally relies on averaging multiple particles.[100]

In HIV, the Gag polyprotein assembles around the (+)ssRNA at the plasma membrane. After budding of the enveloped virus, the viral protease residing in the spherical immature capsid cleaves the Gag protein. The released domains reorganize, leading to mature virions. The nucleocapsid domain is associated with the RNA in the now collapsed and conical core formed by the capsid domain (from here on cp), whereas the matrix domain stays bound to the envelope by an N-terminal myristoylation (Figure 5). This myristoyl residue was also detected by MS in the matrix domain in virus-like particles demonstrating their similarity to the fully infectious virus.[167] Alanine scanning had already identified certain sites in the N- and C-terminal domain of the cp involved in intersubunit binding between these homotypic domains. However, only by using H/D exchange, a previously unknown site in the N-terminal region of cp was found to be highly protected for in-vitro assembled and mature particles, whereas immature capsids showed behavior similar to the free cp subunit.[167, 168] In general, the cp arrangement is similar for different HIV capsid appearances. Moreover, chemical cross-linking elucidated an unknown interaction between the N- and C-terminal domains of adjacent monomers (Figure 4). This contact probably drives maturation and is probably protected or inhibited in the Gag polyprotein.

Another example focusing on virus assembly used H/D-exchange MS to locate regions in the MS2 bacteriophage cp dimer that exhibit conformational changes upon binding to a stem-loop from genomic ssRNA known to initiate assembly.[169] The data revealed specific areas within the cp dimer that altered their exchange kinetics in the presence of the RNA, including the known RNA-binding sites.

3.3. Mass Analysis of Intact Viral Assemblies

ESI-MS has expanded the attainable mass range for the analysis of biomolecules tremendously, and viruses and their capsids have provided showcase benchmarks. The successful transfer of the TMV into the gas phase using ESI was demonstrated as early as 1996.[83] In these pioneering studies, the exact mass determination was precluded by the limitations of the mass analyzer employed. However, collection of the electrosprayed TMV particles and subsequent EM analysis disclosed that the viral structures had been largely retained. The TMV harvested after MS was even infective. Since then, multiple instrumental setups have been applied to estimate the mass of virus particles. A combination of m/z and charge detection in a ToF analyzer resulted in an estimated mass of 40 MDa for TMV, with a substantial uncertainty of 15%.[170] ESI can thus be readily used to ionize viruses and viral particles, but accurate mass analysis is not straightforward when these particles become too big or heterogeneous. ESI has been combined with gas-phase electrophoretic mobility molecular analyzers (GEMMA).[171, 172] The high charge of particles produced by ESI is reduced to obtain singly charged ions, which are subsequently separated and sized by their electrophoretic mobility. GEMMA analysis was successfully applied on the 4.6 MDa cowpea chlorotic mottle virus (CCMV). Although the mass resolution of such an instrument is still too low to enable accurate mass measurement, GEMMA does provide in parallel information about the electrophoretic mobility diameter of the analyzed particle. Such analysis indicated that the gas-phase CCMV particle had largely retained its quaternary structure (Figure 4).[171, 172]
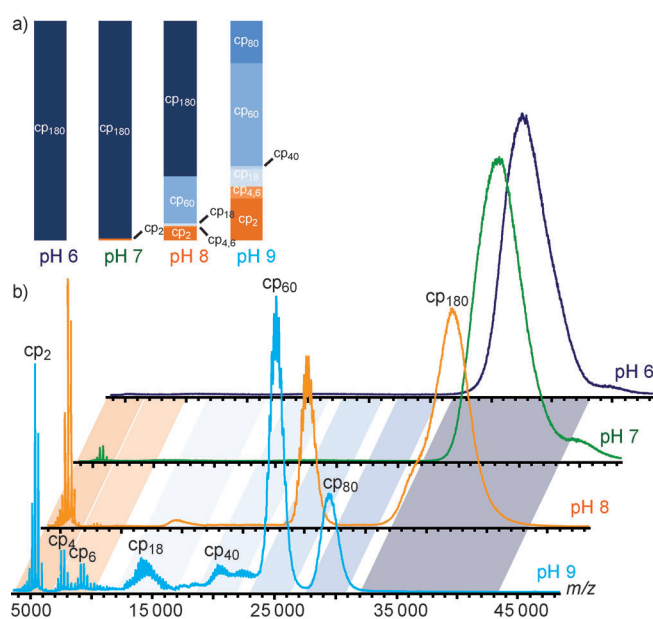
A first more-accurate mass assessment was performed on intact MS2 particles, whereby ESI-ToF analysis allowed the identification of partly resolved charge states.[173] Unprecedented high-resolution data on intact viral capsids of HBV were obtained using a modified Q-ToF instrument.[111, 174] HBV capsids are rather unique in exhibiting two distinct icosahedral morphologies even in vivo, composed of 90 and 120 dimers with masses of approximately 3 and 4 MDa, respectively. The mass spectra displayed well-separated charge-state distributions (Figure 6) for both capsids enabling a mass assignment within 0.1%, revealing that both lattices were complete. The HBV capsids were surprisingly stable during transfer into the gas phase and through the vacuum of the mass spectrometer. Measuring the Ω values of the capsids by IMMS allowed an estimation of the capsid radii in good agreement with the dimensions of the particles in EM, verifying a largely retained capsid morphology in the gas phase.[175] Although the HBV capsid turned out to be very stable, models have suggested that dimers in the particles are exchanging with very low abundant "free" cp dimers in solution. Native MS combined with CID on the intact HBV capsids was used to monitor the incorporation of isotopically labeled cp dimers into preassembled unlabeled HBV capsids. Slow exchange could be observed over a timeframe of months, albeit only for the 3 MDa particles and exclusively at low temperatures[176] providing experimental evidence for the theoretically predicted "capsid breathing".[88]
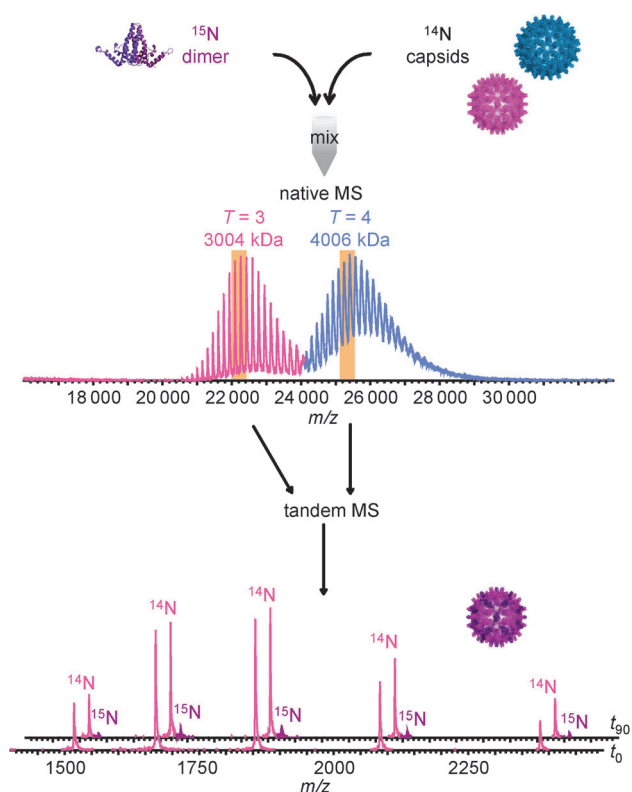
Figure 6. Capsid breathing in HBV monitored by tandem MS: ^{15}N-labeled cp dimers of HBV were incubated with preassembled unlabeled capsids. At certain times native MS spectra were recorded and either $T=3$ or $T=4$ capsids of HBV were selected for CID. Subunit exchange was detected in the $T=3$ capsids after prolonged times as judged by the growth of a signal assigned to labeled monomers, which were ejected from the capsids in tandem MS. Reproduced by permission of the PCCP Owner Societies from Ref. [176].

Figure 7. Capsid stability and morphology of norovirus depend on pH value: ESI-MS analysis of norovirus capsids incubated at pH 6–9. The abundance of the different cp oligomers are depicted as bar graph (a). The graphs and mass spectra (b) indicate that the capsid (cp$_{180}$) is instable at alkaline pH values. The dissociated dimers form larger oligomers (up to 80 subunits) at alkaline pH values and high ionic strength. Adapted from Ref. [177] the American Society for Biochemistry and Molecular Biology.

The even larger (ca. 10 MDa) norovirus capsid proved much less stable and prone to dissociation when altering pH value and ionic strength as monitored by native MS.[177] At pH 6 in solution only the intact $T=3$ capsid was detected, although with insufficient resolution to allow an accurate mass assignment. More acidic pH values resulted in chiefly cp dimers. Intriguingly at basic pH values, higher order oligomers were formed preferentially at high ionic strength. Remarkably, the transition between $T=3$ particles and other oligomers was fully reversible. The larger oligomers most likely arose from initial dissociation of the $T=3$ capsids into dimers and subsequent reassembly. The main products of this pathway were identified as 60- and 80-mers, although also smaller oligomers were present under certain conditions (Figure 7). Atomic force microscopy (AFM)[177] and earlier EM studies[178] confirmed that $T=1$ norovirus capsids consisting of 60 cp subunits can be formed at basic pH values. Ionic strength and pH value could thus drive the norovirus capsid into various morphologies rendering this a particularly ideal model system to study capsid (dis)assembly.

Owing to their low abundance and transient nature, it is extremely hard to probe potential oligomeric intermediates occurring during virus assembly in between the stage of the usually dimeric building block and the intact capsid. The nucleating intermediate is thought to form rather slowly, whereas the subsequent addition of building blocks proceeds fast towards capsid completion (Figure 3). The high sensitivity of native IMMS was used to detect and structurally characterize a wide variety of intermediate oligomers of the norovirus and HBV coexisting with the intact capsid forms.[179] In combination with computational modeling, the shape of these intermediates could be assessed which revealed that they all exhibited extended sheet-like structures as would be expected for on-pathway products. Moreover, the anticipated assembly nuclei for norovirus (decamer) and HBV capsids (hexamer) could be confirmed from the MS data. An assembly pathway common for both viruses was proposed.[179]

In case of the MS2 bacteriophage, the assembly pathways are restricted by the bound RNA. Introducing longer RNA stretches favors assembly along the threefold axis (C_3 axis). Importantly not only the polyanionic character, but also the sequence of the RNA affects the assembly efficiency.[180] The abundance of two major intermediates was monitored by MS over time in presence of various RNAs and interpreted by kinetic modeling.[181] In conclusion, the common protein-centric view of capsid assembly seems rather simplified. Furthermore, the structure of the two intermediates was deduced by IMMS of the intact and CID fragmented species.[182] Both exhibit an extended, ring-like topology as found in the norovirus and HBV oligomers.

For applications in nanotechnology the integrity of the capsids is important, but also their dynamic properties, such as reversible assembly. Knowledge about such abilities is important for particle modification and to support reactions

of encapsulated materials. Limited proteolysis in combination with MS on intact capsids revealed the dynamic nature of the cp in flock-house virus and CCMV in which cp regions located at the inner face become transiently exposed to the exterior.[91,183] The incorporation of material into the caspids and modification of the capsids and other protein cages utilizes such dynamic properties. Mass shifts induced by functionalizing virus particles or creation of mixed assemblies can be readily measured with MS providing a tool to monitor the reaction and control the product quality (Figure 4).[184–186] Dendrimers covalently attached to the inner face of a protein cage can be nurtured to subsequent generations. Thereby, the particle may increase its stability enabling applications in more extreme environments. The possibility to specifically modify the enclosed dendrimers further increases the utility in imaging.[184]

4. Outlook

MS has evolved into a technology that contributes meaningful insights into structural and functional aspects of proteins and protein complexes, such as the viral capsids and viruses described herein. While focusing herein on structural virology, the application areas are not limited, and elegant work has also been described studying ribosomes,[17] RNA polymerases,[115] and proteasomes.[187,188] Although further improvement in mass-spectrometric analyzers and ionization techniques will develop the field, most progress in the near future will come from the linking of MS with other technologies, such as computational modeling, microscopy, and spectroscopy.

Structural modeling of protein complexes is still largely underdeveloped partly because of the size of the particles to be modeled. However, advances made in recent years, for instance in modeling of the yeast nuclear pore complex,[189] may be suitable for wider application and benefit from Ω value and topology data of protein (sub-)complexes obtained by native IMMS.

Coupling of MS with EM or AFM can be envisaged to be bi-directional. The mass analyzer can be used as sample purification step, using soft-landing approaches to select particles of interest that may be further structurally characterized by EM or AFM.[190] In this way, maybe low abundant viral-assembly intermediates can be trapped and studied by microscopic techniques. Conversely, EM- and AFM-based methods can be used to "grap" selective organelles or even protein complexes out of their natural cellular environment, and subsequently bring those to the mass spectrometer for identification and/or structural analysis.[191]

Another exciting development that may become available in the near future uses ultra-short and intense X-ray laser pulses to obtain high-resolution structures of individual protein complexes in the gas phase.[192] Such a method would be highly complementary to the herein described native MS and IMMS approaches and could provide higher resolution structural information, which may be directly compared with solution-phase structural data obtained by more conventional structural-biology techniques, such as NMR spectroscopy, X-ray crystallography and EM. These are exciting times in biomolecular mass spectrometry.

Abbreviations

CID	collision induced dissociation
cp	capsid protein
ds	double-stranded
EM	electron microscopy
ESI	electrospray ionization
HBV	Hepatitis B virus
HIV	Human Immunodeficiency Virus
TMV	Tobacco Mosaic Virus
CCMV	Cowpea Chlorotic Mottle Virus
IM	ion mobility
IMMS	ion mobility mass spectrometry
MALDI	matrix assisted laser desorption/ionization
MS	mass spectrometry
m/z	mass to charge ratio
NVLP	norovirus-like particle
Q	quadrupole
ss	single-stranded
ToF	time of flight

[1] A. O. Nier, J. Chem. Educ. 1989, 66, 385.
[2] H. Budzikiewicz, C. Djerassi, D. H. Williams, Interpretation of mass spectra of organic compounds., Holden-Day, San Francisco, 1964.
[3] M. Karas, F. Hillenkamp, Anal. Chem. 1988, 60, 2299.
[4] J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, C. M. Whitehouse, Science 1989, 246, 64.
[5] S. K. Chowdhury, V. Katta, B. T. Chait, Biochem. Biophys. Res. Commun. 1990, 167, 686.
[6] R. Aebersold, M. Mann, Nature 2003, 422, 198.
[7] K. Dettmer, P. A. Aronov, B. D. Hammock, Mass Spectrom. Rev. 2007, 26, 51.
[8] K. Chughtai, R. M. Heeren, Chem. Rev. 2010, 110, 3237.
[9] Z. Takats, J. M. Wiseman, B. Gologan, R. G. Cooks, Science 2004, 306, 471.

[10] J. B. Fenn, *Angew. Chem.* **2003**, *115*, 3999; *Angew. Chem. Int. Ed.* **2003**, *42*, 3871.

[11] V. Katta, B. T. Chait, *J. Am. Chem. Soc.* **1991**, *113*, 8534.

[12] M. C. Fitzgerald, I. Chernushevich, K. G. Standing, C. P. Whitman, S. B. Kent, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 6851.

[13] X. J. Tang, C. F. Brewer, S. Saha, I. Chernushevich, W. Ens, K. G. Standing, *Rapid Commun. Mass Spectrom.* **1994**, *8*, 750.

[14] K. J. Light-Wahl, J. A. Loo, C. G. Edmonds, R. D. Smith, H. E. Witkowska, C. H. Shackleton, C. S. Wu, *Biol. Mass Spectrom.* **1993**, *22*, 112.

[15] J. A. Loo, *Mass Spectrom. Rev.* **1997**, *16*, 1.

[16] A. A. Rostom, P. Fucini, D. R. Benjamin, R. Juenemann, K. H. Nierhaus, F. U. Hartl, C. M. Dobson, C. V. Robinson, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 5185.

[17] D. R. Benjamin, C. V. Robinson, J. P. Hendrick, F. U. Hartl, C. M. Dobson, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 7391.

[18] J. L. Benesch, C. V. Robinson, *Curr. Opin. Struct. Biol.* **2006**, *16*, 245.

[19] M. Sharon, C. V. Robinson, *Annu. Rev. Biochem.* **2007**, *76*, 167.

[20] W. J. van Berkel, R. H. van den Heuvel, C. Versluis, A. J. Heck, *Protein Sci.* **2000**, *9*, 435.

[21] A. J. Heck, R. H. Van den Heuvel, *Mass Spectrom. Rev.* **2004**, *23*, 368.

[22] R. H. van den Heuvel, A. J. Heck, *Curr. Opin. Chem. Biol.* **2004**, *8*, 519.

[23] N. F. Steinmetz, T. Lin, G. P. Lomonossoff, J. E. Johnson, *Curr. Top. Microbiol. Immunol.* **2009**, *327*, 23.

[24] P. Singh, M. J. Gonzalez, M. Manchester, *Drug Dev. Res.* **2006**, *67*, 23.

[25] S. Mitragotri, J. Lahann, *Nat. Mater.* **2009**, *8*, 15.

[26] T. Dokland, *Structure* **2000**, *8*, R157.

[27] V. L. Morton, P. G. Stockley, N. J. Stonehouse, A. E. Ashcroft, *Mass Spectrom. Rev.* **2008**, *27*, 575.

[28] O. Krokhin, Y. Li, A. Andonov, H. Feldmann, R. Flick, S. Jones, U. Stroeher, N. Bastien, K. V. Dasuri, K. Cheng, J. N. Simonsen, H. Perreault, J. Wilkins, W. Ens, F. Plummer, K. G. Standing, *Mol. Cell. Proteomics* **2003**, *2*, 346.

[29] T. Pe'ery, M. B. Mathews in *Fields Virology, Vol. 1*, 5th ed. (Eds.: D. M. Knipe, P. M. Howley, D. E. Griffin, R. A. Lamb, M. A. Martin, B. Roizman, S. E. Straus), Lippincott Williams & Wilkins, Philadelphia, **2007**.

[30] J. M. Claverie, C. Abergel, *Annu. Rev. Genet.* **2009**, *43*, 49.

[31] A. C. Steven, J. F. Conway, N. Cheng, N. R. Watts, D. M. Belnap, A. Harris, S. J. Stahl, P. T. Wingfield, *Adv. Virus Res.* **2005**, *64*, 125.

[32] M. W. Beijerinck, *Verh. Akad. Wet. Amsterdam Afd. Natuurkd. Sect. 1* **1898**, *6*, 3.

[33] C. M. Lawrence, S. Menon, B. J. Eilers, B. Bothner, R. Khayat, T. Douglas, M. J. Young, *J. Biol. Chem.* **2009**, *284*, 12599.

[34] M. N. Pearson, R. E. Beever, B. Boine, K. Arthur, *Mol. Plant Pathol.* **2009**, *10*, 115.

[35] H. W. Ackermann, A. M. Kropinski, *Res. Microbiol.* **2007**, *158*, 555.

[36] R. S. Nelson, V. Citovsky, *Plant Physiol.* **2005**, *138*, 1809.

[37] J. Mercer, M. Schelhaas, A. Helenius, *Annu. Rev. Biochem.* **2010**, *79*, 803.

[38] T. Yamada, H. Onimatsu, J. L. Van Etten, *Adv. Virus Res.* **2006**, *66*, 293.

[39] C. A. Suttle, *Nature* **2005**, *437*, 356.

[40] D. Moreira, P. Lopez-Garcia, *Nat. Rev. Microbiol.* **2009**, *7*, 306.

[41] B. La Scola, C. Desnues, I. Pagnier, C. Robert, L. Barrassi, G. Fournous, M. Merchat, M. Suzan-Monti, P. Forterre, E. Koonin, D. Raoult, *Nature* **2008**, *455*, 100.

[42] E. V. Koonin, *Ann. N. Y. Acad. Sci.* **2009**, *1178*, 47.

[43] P. Forterre, D. Prangishvili, *Res. Microbiol.* **2009**, *160*, 466.

[44] R. C. Condit in *Fields Virology, Vol. 1*, 5th ed. (Eds.: D. M. Knipe, P. M. Howley, D. E. Griffin, R. A. Lamb, M. A. Martin, B. Roizman, S. E. Straus), Lippincott Williams & Wilkins, Philadelphia, **2007**.

[45] C. M. Fauquet, D. Fargette, *Virol. J.* **2005**, *2*, 64.

[46] F. Rohwer, R. Edwards, *J. Bacteriol.* **2002**, *184*, 4529.

[47] M. H. van Regenmortel, B. W. Mahy, *Emerging Infect. Dis.* **2004**, *10*, 8.

[48] S. C. Harrison in *Fields Virology, Vol. 1*, 5th ed. (Eds.: D. M. Knipe, P. M. Howley, D. E. Griffin, R. A. Lamb, M. A. Martin, B. Roizman, S. E. Straus), Lippincott Williams & Wilkins, Philadelphia, **2007**.

[49] J. M. White, S. E. Delos, M. Brecher, K. Schornberg, *Crit. Rev. Biochem. Mol. Biol.* **2008**, *43*, 189.

[50] J. T. Huiskonen, S. J. Butcher, *Curr. Opin. Struct. Biol.* **2007**, *17*, 229.

[51] D. A. Marvin, *Curr. Opin. Struct. Biol.* **1998**, *8*, 150.

[52] P. Grayson, I. J. Molineux, *Curr. Opin. Microbiol.* **2007**, *10*, 401.

[53] G. R. Whittaker, M. Kann, A. Helenius, *Annu. Rev. Cell Dev. Biol.* **2000**, *16*, 627.

[54] M. R. Denison, *PLoS Biol.* **2008**, *6*, e270.

[55] H. Brussow, C. Canchaya, W. D. Hardt, *Microbiol. Mol. Biol. Rev.* **2004**, *68*, 560.

[56] A. B. Oppenheim, O. Kobiler, J. Stavans, D. L. Court, S. Adhya, *Annu. Rev. Genet.* **2005**, *39*, 409.

[57] F. Bushman, M. Lewinski, A. Ciuffi, S. Barr, J. Leipzig, S. Hannenhalli, C. Hoffmann, *Nat. Rev. Microbiol.* **2005**, *3*, 848.

[58] C. Netherton, K. Moffat, E. Brooks, T. Wileman, *Adv. Virus Res.* **2007**, *70*, 101.

[59] S. Sun, V. B. Rao, M. G. Rossmann, *Curr. Opin. Struct. Biol.* **2010**, *20*, 114.

[60] J. T. Patton, R. Vasquez-Del Carpio, M. A. Tortorici, Z. F. Taraporewala, *Adv. Virus Res.* **2007**, *69*, 167.

[61] R. Patient, C. Hourioux, P. Roingeard, *Cell. Microbiol.* **2009**, *11*, 1561.

[62] B. J. Chen, R. A. Lamb, *Virology* **2008**, *372*, 221.

[63] I. N. Wang, D. L. Smith, R. Young, *Annu. Rev. Microbiol.* **2000**, *54*, 799.

[64] J. Lanman, P. E. Prevelige, Jr., *Adv. Virus Res.* **2005**, *64*, 285.

[65] J. M. Cutalo, L. J. Deterding, K. B. Tomer, *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 1545.

[66] I. Gertsman, L. Gan, M. Guttman, K. Lee, J. A. Speir, R. L. Duda, R. W. Hendrix, E. A. Komives, J. E. Johnson, *Nature* **2009**, *458*, 646.

[67] R. L. Duda, J. Hempel, H. Michel, J. Shabanowitz, D. Hunt, R. W. Hendrix, *J. Mol. Biol.* **1995**, *247*, 618.

[68] A. C. Steven, J. B. Heymann, N. Cheng, B. L. Trus, J. F. Conway, *Curr. Opin. Struct. Biol.* **2005**, *15*, 227.

[69] M. K. Estes, B. V. Prasad, R. L. Atmar, *Curr. Opin. Infect. Dis.* **2006**, *19*, 467.

[70] S. R. Casjens, *Curr. Opin. Microbiol.* **2005**, *8*, 451.

[71] H. L. Yen, R. G. Webster, *Curr. Top. Microbiol. Immunol.* **2009**, *333*, 3.

[72] D. L. Caspar, A. Klug, *Cold Spring Harbor Symp. Quant. Biol.* **1962**, *27*, 1.

[73] A. Zlotnick, *J. Mol. Recognit.* **2005**, *18*, 479.

[74] J. E. Johnson, W. Chiu, *Curr. Opin. Struct. Biol.* **2007**, *17*, 237.

[75] R. Wetzel, *Acc. Chem. Res.* **2006**, *39*, 671.

[76] R. Zandi, P. van der Schoot, D. Reguera, W. Kegel, H. Reiss, *Biophys. J.* **2006**, *90*, 1939.

[77] M. F. Hagan, O. M. Elrad, *Biophys. J.* **2010**, *98*, 1065.

[78] P. G. Stockley, O. Rolfsson, G. S. Thompson, G. Basnak, S. Francese, N. J. Stonehouse, S. W. Homans, A. E. Ashcroft, *J. Mol. Biol.* **2007**, *369*, 541.

[79] A. Zlotnick, *J. Mol. Biol.* **2007**, *366*, 14.

[80] A. Zlotnick, P. Ceres, S. Singh, J. M. Johnson, *J. Virol.* **2002**, *76*, 4848.

[81] A. Kivenson, M. F. Hagan, *Biophys. J.* **2010**, *99*, 619.

[82] M. F. Hagan, *J. Chem. Phys.* **2009**, *130*, 114902.

8260   www.angewandte.org

© 2011 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Angew. Chem. Int. Ed. **2011**, *50*, 8248–8262

[83] G. Siuzdak, B. Bothner, M. Yeager, C. Brugidou, C. M. Fauquet, K. Hoey, C. M. Chang, *Chem. Biol.* **1996**, *3*, 45.

[84] M. L. Flenniken, M. Uchida, L. O. Liepold, S. Kang, M. J. Young, T. Douglas, *Curr. Top. Microbiol. Immunol.* **2009**, *327*, 71.

[85] M. Newman, F. M. Suk, M. Cajimat, P. K. Chua, C. Shih, *J. Virol.* **2003**, *77*, 12950.

[86] S. Singh, A. Zlotnick, *J. Biol. Chem.* **2003**, *278*, 18249.

[87] W. H. Roos, G. L. Wuite, *Adv. Mater.* **2009**, *21*, 1187.

[88] P. Ceres, A. Zlotnick, *Biochemistry* **2002**, *41*, 11525.

[89] J. K. Hilmer, A. Zlotnick, B. Bothner, *J. Mol. Biol.* **2008**, *375*, 581.

[90] N. Reisdorph, J. J. Thomas, U. Katpally, E. Chase, K. Harris, G. Siuzdak, T. J. Smith, *Virology* **2003**, *314*, 34.

[91] B. Bothner, X. F. Dong, L. Bibbs, J. E. Johnson, G. Siuzdak, *J. Biol. Chem.* **1998**, *273*, 673.

[92] K. N. Parent, M. M. Suhanovsky, C. M. Teschke, *J. Mol. Biol.* **2007**, *365*, 513.

[93] B. Bothner, A. Schneemann, D. Marshall, V. Reddy, J. E. Johnson, G. Siuzdak, *Nat. Struct. Biol.* **1999**, *6*, 114.

[94] A. Klug, *Philos. Trans. R. Soc. London Ser. B* **1999**, *354*, 531.

[95] W. K. Kegel, P. van der Schoot, *Biophys. J.* **2006**, *91*, 1501.

[96] X. Zhang, L. Jin, Q. Fang, W. H. Hui, Z. H. Zhou, *Cell* **2010**, *141*, 472.

[97] J. Vellinga, S. Van der Heijdt, R. C. Hoeben, *J. Gen. Virol.* **2005**, *86*, 1581.

[98] K. H. Choi, M. C. Morais, D. L. Anderson, M. G. Rossmann, *Structure* **2006**, *14*, 1723.

[99] M. F. Moody, *J. Mol. Biol.* **1999**, *293*, 401.

[100] B. K. Ganser-Pornillos, U. K. von Schwedler, K. M. Stray, C. Aiken, W. I. Sundquist, *J. Virol.* **2004**, *78*, 2545.

[101] R. A. Zubarev, D. M. Horn, E. K. Fridriksson, N. L. Kelleher, N. A. Kruger, M. A. Lewis, B. K. Carpenter, F. W. McLafferty, *Anal. Chem.* **2000**, *72*, 563.

[102] J. E. Syka, J. J. Coon, M. J. Schroeder, J. Shabanowitz, D. F. Hunt, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 9528.

[103] R. G. Sadygov, D. Cociorva, J. R. Yates, 3rd, *Nat. Methods* **2004**, *1*, 195.

[104] B. Tissot, S. J. North, A. Ceroni, P. C. Pang, M. Panico, F. Rosati, A. Capone, S. M. Haslam, A. Dell, H. R. Morris, *FEBS Lett.* **2009**, *583*, 1728.

[105] P. A. Grimsrud, D. L. Swaney, C. D. Wenger, N. A. Beauchene, J. J. Coon, *ACS Chem. Biol.* **2010**, *5*, 105.

[106] N. Mischerikow, A. J. Heck, *Proteomics* **2011**, *11*, 571.

[107] L. Konermann, D. A. Simmons, *Mass Spectrom. Rev.* **2003**, *22*, 1.

[108] A. Scholten, N. F. Visser, R. H. van den Heuvel, A. J. Heck, *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 983.

[109] N. Taouatas, M. M. Drugan, A. J. Heck, S. Mohammed, *Nat. Methods* **2008**, *5*, 405.

[110] A. J. Heck, *Nat. Methods* **2008**, *5*, 927.

[111] R. H. van den Heuvel, E. van Duijn, H. Mazon, S. A. Synowsky, K. Lorenzen, C. Versluis, S. J. Brouns, D. Langridge, J. van der Oost, J. Hoyes, A. J. Heck, *Anal. Chem.* **2006**, *78*, 7473.

[112] J. L. Benesch, B. T. Ruotolo, F. Sobott, J. Wildgoose, A. Gilbert, R. Bateman, C. V. Robinson, *Anal. Chem.* **2009**, *81*, 1270.

[113] T. Taverner, H. Hernandez, M. Sharon, B. T. Ruotolo, D. Matak-Vinkovic, D. Devos, R. B. Russell, C. V. Robinson, *Acc. Chem. Res.* **2008**, *41*, 617.

[114] S. R. Geiger, K. Lorenzen, A. Schreieck, P. Hanecker, D. Kostrewa, A. J. Heck, P. Cramer, *Mol. Cell* **2010**, *39*, 583.

[115] K. Lorenzen, A. Vannini, P. Cramer, A. J. Heck, *Structure* **2007**, *15*, 1237.

[116] E. van Duijn, D. A. Simmons, R. H. van den Heuvel, P. J. Bakkes, H. van Heerikhuizen, R. M. Heeren, C. V. Robinson, S. M. van der Vies, A. J. Heck, *J. Am. Chem. Soc.* **2006**, *128*, 4694.

[117] J. L. Benesch, J. A. Aquilina, B. T. Ruotolo, F. Sobott, C. V. Robinson, *Chem. Biol.* **2006**, *13*, 597.

[118] N. P. Barrera, N. Di Bartolo, P. J. Booth, C. V. Robinson, *Science* **2008**, *321*, 243.

[119] S. A. Synowsky, R. H. van den Heuvel, S. Mohammed, P. W. Pijnappel, A. J. Heck, *Mol. Cell. Proteomics* **2006**, *5*, 1581.

[120] S. A. Synowsky, M. van Wijk, R. Raijmakers, A. J. Heck, *J. Mol. Biol.* **2009**, *385*, 1300.

[121] C. Uetrecht, R. J. Rose, E. van Duijn, K. Lorenzen, A. J. Heck, *Chem. Soc. Rev.* **2010**, *39*, 1633.

[122] B. T. Ruotolo, K. Giles, I. Campuzano, A. M. Sandercock, R. H. Bateman, C. V. Robinson, *Science* **2005**, *310*, 1658.

[123] T. L. Pukala, B. T. Ruotolo, M. Zhou, A. Politis, R. Stefanescu, J. A. Leary, C. V. Robinson, *Structure* **2009**, *17*, 1235.

[124] L. Wang, L. C. Lane, D. L. Smith, *Protein Sci.* **2001**, *10*, 1234.

[125] L. Wang, D. L. Smith, *Protein Sci.* **2005**, *14*, 1661.

[126] C. S. Maier, M. L. Deinzer, *Methods Enzymol.* **2005**, *402*, 312.

[127] J. R. Engen, D. L. Smith, *Anal. Chem.* **2001**, *73*, 256 A.

[128] T. E. Wales, J. R. Engen, *Mass Spectrom. Rev.* **2006**, *25*, 158.

[129] V. Alverdi, H. Mazon, C. Versluis, W. Hemrika, G. Esposito, R. van den Heuvel, A. Scholten, A. J. Heck, *J. Mol. Biol.* **2008**, *375*, 1380.

[130] A. Sinz, *ChemMedChem* **2007**, *2*, 425.

[131] I. Gertsman, C. Y. Fu, R. Huang, E. Komives, J. E. Johnson, *Mol. Cell. Proteomics* **2010**, *9*, 1752.

[132] D. Houde, Y. Peng, S. A. Berkowitz, J. R. Engen, *Mol. Cell. Proteomics* **2010**, *9*, 1716.

[133] V. L. Mendoza, R. W. Vachet, *Mass Spectrom. Rev.* **2009**, *28*, 785.

[134] G. Xu, M. R. Chance, *Chem. Rev.* **2007**, *107*, 3514.

[135] L. Konermann, B. B. Stocks, Y. Pan, X. Tong, *Mass Spectrom. Rev.* **2010**, *29*, 651.

[136] B. C. Gau, J. S. Sharp, D. L. Rempel, M. L. Gross, *Anal. Chem.* **2009**, *81*, 6563.

[137] D. Suckau, M. Mak, M. Przybylski, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 5630.

[138] S. Kang, A. M. Hawkridge, K. L. Johnson, D. C. Muddiman, P. E. Prevelige, Jr., *J. Proteome Res.* **2006**, *5*, 370.

[139] A. Leitner, T. Walzthoeni, A. Kahraman, F. Herzog, O. Rinner, M. Beck, R. Aebersold, *Mol. Cell. Proteomics* **2010**, *9*, 1634.

[140] A. Sinz, *Mass Spectrom. Rev.* **2006**, *25*, 663.

[141] S. Kang, L. Mou, J. Lanman, S. Velu, W. J. Brouillette, P. E. Prevelige, Jr., *Rapid Commun. Mass Spectrom.* **2009**, *23*, 1719.

[142] S. Belák, P. Thoren, N. LeBlanc, G. Viljoen, *Expert Rev. Mol. Diagn.* **2009**, *9*, 367.

[143] R. Zeng, H. Q. Ruan, X. S. Jiang, H. Zhou, L. Shi, L. Zhang, Q. H. Sheng, Q. Tu, Q. C. Xia, J. R. Wu, *J. Proteome Res.* **2004**, *3*, 549.

[144] W. Ying, Y. Hao, Y. Zhang, W. Peng, E. Qin, Y. Cai, K. Wei, J. Wang, G. Chang, W. Sun, S. Dai, X. Li, Y. Zhu, J. Li, S. Wu, L. Guo, J. Dai, P. Wan, T. Chen, C. Du, D. Li, J. Wan, X. Kuai, W. Li, R. Shi, H. Wei, C. Cao, M. Yu, H. Liu, F. Dong, D. Wang, X. Zhang, X. Qian, Q. Zhu, F. He, *Proteomics* **2004**, *4*, 492.

[145] M. Bartlam, Y. Xu, Z. Rao, *J. Struct. Funct. Genomics* **2007**, *8*, 85.

[146] K. Radtke, D. Kieneke, A. Wolfstein, K. Michael, W. Steffen, T. Scholz, A. Karger, B. Sodeik, *PLoS Pathog.* **2010**, *6*, e1000991.

[147] D. Munday, E. Emmott, R. Surtees, C. H. Lardeau, W. Wu, W. P. Duprex, B. K. Dove, J. N. Barr, J. A. Hiscox, *Mol. Cell. Proteomics* **2010**, *9*, 2438.

[148] X. S. Jiang, L. Y. Tang, J. Dai, H. Zhou, S. J. Li, Q. C. Xia, J. R. Wu, R. Zeng, *Mol. Cell. Proteomics* **2005**, *4*, 902.

[149] L. Zhang, Z. P. Zhang, X. E. Zhang, F. S. Lin, F. Ge, *J. Virol.* **2010**, *84*, 6050.

[150] D. J. Vigerust, V. L. Shepherd, *Trends Microbiol.* **2007**, *15*, 211.

[151] X. Zhu, C. Borchers, R. J. Bienstock, K. B. Tomer, *Biochemistry* **2000**, *39*, 11194.

[152] S. Liedtke, R. Geyer, H. Geyer, *Glycoconjugate J.* **1997**, *14*, 785.

[153] E. P. Go, Q. Chang, H. X. Liao, L. L. Sutherland, S. M. Alam, B. F. Haynes, H. Desaire, *J. Proteome Res.* **2009**, *8*, 4231.

[154] M. Melegari, S. K. Wolf, R. J. Schneider, *J. Virol.* **2005**, *79*, 9810.

[155] E. V. Gazina, J. E. Fielding, B. Lin, D. A. Anderson, *J. Virol.* **2000**, *74*, 4721.

[156] H. Kang, J. Yu, G. Jung, *Biochem. J.* **2008**, *416*, 47.

[157] H. Y. Kang, S. Lee, S. G. Park, J. Yu, Y. Kim, G. Jung, *Biochem. J.* **2006**, *398*, 311.

[158] A. Masi, R. Cicchi, A. Carloni, F. S. Pavone, A. Arcangeli, *Adv. Exp. Med. Biol.* **2010**, *674*, 33.

[159] J. R. Lakowicz, *J. Biochem. Biophys. Methods* **1980**, *2*, 91.

[160] C. Fu, C. Uetrecht, S. Kang, M. Morais, A. J. Heck, M. R. Walter, P. E. Prevelige, *Mol. Cell. Proteomics* **2010**, *9*, 1764.

[161] I. S. Moreira, P. A. Fernandes, M. J. Ramos, *Proteins Struct. Funct. Genet.* **2007**, *68*, 803.

[162] R. L. Duda, P. D. Ross, N. Cheng, B. A. Firek, R. W. Hendrix, J. F. Conway, A. C. Steven, *J. Mol. Biol.* **2009**, *391*, 471.

[163] R. B. Pepinsky, I. A. Papayannopoulos, S. Campbell, V. M. Vogt, *J. Virol.* **1996**, *70*, 3313.

[164] I. Gertsman, E. A. Komives, J. E. Johnson, *J. Mol. Biol.* **2010**, *397*, 560.

[165] M. Lopper, T. Compton, *J. Virol.* **2002**, *76*, 6073.

[166] J. Lanman, T. T. Lam, S. Barnes, M. Sakalian, M. R. Emmett, A. G. Marshall, P. E. Prevelige, Jr., *J. Mol. Biol.* **2003**, *325*, 759.

[167] J. Lanman, T. T. Lam, M. R. Emmett, A. G. Marshall, M. Sakalian, P. E. Prevelige, Jr., *Nat. Struct. Mol. Biol.* **2004**, *11*, 676.

[168] T. T. Lam, J. K. Lanman, M. R. Emmett, C. L. Hendrickson, A. G. Marshall, P. E. Prevelige, *J. Chromatogr. A* **2002**, *982*, 85.

[169] V. L. Morton, W. Burkitt, G. O'Connor, N. J. Stonehouse, P. G. Stockley, A. E. Ashcroft, *Phys. Chem. Chem. Phys.* **2010**, *12*, 13468.

[170] S. D. Fuerstenau, W. H. Benner, J. J. Thomas, C. Brugidou, B. Bothner, G. Siuzdak, *Angew. Chem.* **2001**, *113*, 559; *Angew. Chem. Int. Ed.* **2001**, *40*, 541.

[171] J. J. Thomas, B. Bothner, J. Traina, W. H. Benner, G. Siuzdak, *Spectrosc. Int. J.* **2004**, *18*, 31.

[172] C. S. Kaddis, S. H. Lomeli, S. Yin, B. Berhane, M. I. Apostol, V. A. Kickhoefer, L. H. Rome, J. A. Loo, *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 1206.

[173] M. A. Tito, K. Tars, K. Valegard, J. Hajdu, C. V. Robinson, *J. Am. Chem. Soc.* **2000**, *122*, 3550.

[174] C. Uetrecht, C. Versluis, N. R. Watts, W. H. Roos, G. J. Wuite, P. T. Wingfield, A. C. Steven, A. J. Heck, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 9216.

[175] C. Uetrecht, C. Versluis, N. R. Watts, P. T. Wingfield, A. C. Steven, A. J. Heck, *Angew. Chem.* **2008**, *120*, 6343; *Angew. Chem. Int. Ed.* **2008**, *47*, 6247.

[176] C. Uetrecht, N. R. Watts, S. J. Stahl, P. T. Wingfield, A. C. Steven, A. J. Heck, *Phys. Chem. Chem. Phys.* **2010**, *12*, 13368.

[177] G. K. Shoemaker, E. van Duijn, S. E. Crawford, C. Uetrecht, M. Baclayon, W. H. Roos, G. J. Wuite, M. K. Estes, B. V. Prasad, A. J. Heck, *Mol. Cell. Proteomics* **2010**, *9*, 1742.

[178] L. J. White, M. E. Hardy, M. K. Estes, *J. Virol.* **1997**, *71*, 8066.

[179] C. Uetrecht, I. M. Barbu, G. K. Shoemaker, E. van Duijn, A. J. Heck, *Nat. Chem.* **2011**, *3*, 126.

[180] G. Basnak, V. L. Morton, O. Rolfsson, N. J. Stonehouse, A. E. Ashcroft, P. G. Stockley, *J. Mol. Biol.* **2010**, *395*, 924.

[181] V. L. Morton, E. C. Dykeman, N. J. Stonehouse, A. E. Ashcroft, R. Twarock, P. G. Stockley, *J. Mol. Biol.* **2010**, *401*, 298.

[182] T. W. Knapman, V. L. Morton, N. J. Stonehouse, P. G. Stockley, A. E. Ashcroft, *Rapid Commun. Mass Spectrom.* **2010**, *24*, 3033.

[183] L. O. Liepold, J. Revis, M. Allen, L. Oltrogge, M. Young, T. Douglas, *Phys. Biol.* **2005**, *2*, S166.

[184] M. J. Abedin, L. Liepold, P. Suci, M. Young, T. Douglas, *J. Am. Chem. Soc.* **2009**, *131*, 4346.

[185] S. Kang, L. M. Oltrogge, C. C. Broomell, L. O. Liepold, P. E. Prevelige, M. Young, T. Douglas, *J. Am. Chem. Soc.* **2008**, *130*, 16527.

[186] S. Kang, C. C. Jolley, L. O. Liepold, M. Young, T. Douglas, *Angew. Chem.* **2009**, *121*, 4866; *Angew. Chem. Int. Ed.* **2009**, *48*, 4772.

[187] M. Sharon, S. Witt, E. Glasmacher, W. Baumeister, C. V. Robinson, *J. Biol. Chem.* **2007**, *282*, 18448.

[188] J. A. Loo, B. Berhane, C. S. Kaddis, K. M. Wooding, Y. Xie, S. L. Kaufman, I. V. Chernushevich, *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 998.

[189] F. Alber, S. Dokudovskaya, L. M. Veenhoff, W. Zhang, J. Kipper, D. Devos, A. Suprapto, O. Karni-Schmidt, R. Williams, B. T. Chait, A. Sali, M. P. Rout, *Nature* **2007**, *450*, 695.

[190] J. L. Benesch, B. T. Ruotolo, D. A. Simmons, N. P. Barrera, N. Morgner, L. Wang, H. R. Saibil, C. V. Robinson, *J. Struct. Biol.* **2010**, *172*, 161.

[191] S. Nickell, C. Kofler, A. P. Leis, W. Baumeister, *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 225.

[192] R. Neutze, R. Wouts, D. van der Spoel, E. Weckert, J. Hajdu, *Nature* **2000**, *406*, 752.

[193] K. Lorenzen, A. S. Olia, C. Uetrecht, G. Cingolani, A. J. Heck, *J. Mol. Biol.* **2008**, *379*, 385.