

**Extending the boundaries of
native mass spectrometry
to study virus structure and assembly**

Joost Snijder

ISBN 978-94-6295-272-0

The research in this thesis was performed in the Biomolecular Mass Spectrometry & Proteomics Group, Utrecht University, Utrecht, The Netherlands.

The research was financially supported by the Netherlands Proteomics Center (NPC).

MS Vision financially supported the printing of this thesis.

Thermo Fisher Scientific financially supported the printing of this thesis.

Printed by proefschriftmaken.nl || Uitgeverij BOXPress

Extending the boundaries of native mass spectrometry to study virus structure and assembly

De grenzen van native massaspectrometrie verleggen om de structuur en
assemblage van virussen te onderzoeken
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag
van de rector magnificus, prof. dr. G.J. van der Zwaan, ingevolge het besluit
van het college voor promoties in het openbaar te verdedigen op woensdag 2
september 2015 des middags te 12.45 uur

door

Joost Snijder

geboren op 17 september 1988 te Asperen

Promotor: Prof. dr. A.J.R. Heck

Table of contents

Preface:

The emerging niche of native mass spectrometry in structural biology. 7

Chapter1:

Analytical approaches to size and mass analysis of large protein complexes. 13

Chapter2:

Challenging the current upper mass limit of native MS. 47

Chapter3:

A new platform with improved resolution for native MS of virus particles. 65

Chapter4:

Mass spectrometry uncovers a novel component of mature adenovirus capsids. 81

Chapter5:

Probing the biophysical interplay between a viral genome and its capsid. 101

Chapter6:

Cargo encapsulation triggers a stability switch in the bacterial nanocompartment encapsulin. 123

Chapter7:

Future outlook, Summary, Samenvatting in het Nederlands, Acknowledgements, Curriculum vitae, List of publications. 145

Appendix:

Supplementary information. 163

Preface:

The emerging niche of native mass spectrometry in structural biology.

Some scientists study protein complexes in terms of the vast interaction networks that can be uncovered by proteomics screens, while others study them in terms of the atomic structural models derived from x-ray crystallography and NMR spectroscopy experiments. Others might focus on the interactions with other types of molecules in the cell, such as the lipid bilayer, various sorts of nucleic acids or the countless known metabolites in the cell. Or they might try to understand the post-translational modifications that proteins are susceptible to, and how they modulate protein complex formation and a protein's biological activities. Regardless of which aspect of protein complexes one focuses on, the analytical tools that are associated with a particular field offer a technically biased and limited perspective. The bigger picture may emerge by comparing notes on what protein complexes look like from these different perspectives.

I like to consider 'protein' as a blanket term, describing vastly different types of biological structures. The presence of more than a few peptide bonds is practically the only thing they all have in common. They might range from single polypeptide chains of a few kilodaltons to enormous clusters of proteins that are up to hundreds of megadaltons in mass. They might be rock solid, structured and rigid, or practically without a set structure at all. The analytical tools we use to describe them may be very well suited to describe some protein complexes, but grossly inappropriate to describe others, depending on these parameters such as size, solubility, flexibility or the presence of post-translational modification or other types of biomolecules that may also be part of the complex. Moreover, a polypeptide chain might exist in a multitude of different forms: in different oligomeric states, different structural conformations, in complex with different interaction partners and with different stoichiometries of ligands bound or side chains modified. Meanwhile, all of these different forms are pulled out the cell by the same tag or antibody recognition site in preparation for subsequent analysis. What does 'pure' really mean if the analytical tool is blind to all these different versions of the protein complex that are present in the sample? Our analytical tools limit, or conversely, open up, our perspective. How we account for our models of even the most basic biology is ultimately down to which analytical tools we used to reach our conclusions.

Some analytical tools are less appropriate for molecules of a certain size: large complexes present a special challenge for NMR spectroscopy, whereas the lower end of the mass range is particularly challenging for electron microscopy. X-ray crystallography applies well over the entire relevant mass range, but is strictly limited by the requirement for well-diffracting crystals and therefore less suited to study dynamic, flexible and heterogeneous samples. Heterogeneity of stoi-

chionometry, structure and conformation is generally speaking a very challenging aspect when analyzing protein samples. Most analytical tools provide a bulk measurement, but lack the required resolution to distinguish between coexisting states, especially as these may be present over a large dynamic range. Single-molecule techniques might provide an answer in this respect, but often lack the precision to describe aspects of a protein structure to an appreciable level of detail. Electron microscopy may provide a notable exception. However, the high resolution information from an electron microscopy experiment still requires averaging over many particles, and sorting out which single particles belong to which species becomes increasingly challenging with smaller differences in the observed densities.

In practical terms, many techniques that are used in structural biology lack chemical specificity, in the sense that the observed signals are interpreted with the expected 20 amino acids in mind. For instance, x-ray crystallography and electron microscopy both provide a detailed electron density map, but electrons on a sulfur or carbon atom don't look all that different. Especially at resolutions $>3\text{\AA}$, most of the atomic model is inferred from what is previously known about the amino acid sequence, post-translational modifications, bound metal ions and ligands: the map itself serves mostly to trace the backbone of the polypeptide chains. Therefore, when the identity of the polypeptide chains, the nature of post-translational modifications, the type of bound ligands may be brought into question, such techniques as x-ray crystallography and electron microscopy fall short, in spite of the high resolution nature of the information they provide.

Where many analytical tools thus practically lack chemical specificity for the analysis of protein complexes, the use of mass spectrometry comes into play. Determining the masses of proteins and protein complexes with high precision and accuracy provides mass spectrometry with an unmatched power to pinpoint the copy numbers of subunits and ligands assembling in a protein complex. Mass spectrometry techniques for the analysis of protein complexes are coming along with broad mass range, and sufficiently high resolving power to determine the precise identity and stoichiometry of many co-occurring components in rather heterogeneous mixtures of protein complexes. Mass spectrometry is therefore very well suited to determine the composition of protein complexes and uncovering the ensemble of quaternary structures as which the proteins exists. But, as with any technique, specific challenges and limitations still apply. Mass spectrometry works especially well for the study of quaternary structure by virtue of the high accuracy and precision of the mass determination, especially in comparison to alternative tools such as size-exclusion chromatography,

multi-angle light scattering, analytical ultracentrifugation and various types of gel electrophoresis. Mass spectrometry's ability to simultaneously detect multiple coexisting complexes in a heterogeneous mixture also requires a sufficiently high resolving power. An important limitation of mass spectrometry is that achieving this high precision, high accuracy and high resolving power in analysis of protein complexes becomes increasingly challenging with increasingly larger assemblies.

The challenge of analyzing large protein complexes has been addressed in several previous mass spectrometry studies of large protein complexes like the ribosome or proteasome, whereas the Heck lab has focused a lot on protein cages, most notably virus capsids. Virus capsids play a pivotal role in the viral replication cycle, and may therefore be important therapeutic targets. They are also being developed as vehicles for gene and vaccine delivery and as assembly scaffolds or nanoreactors/-containers in nanotechnology (1). These studies include mass spectrometry of chaperonin complexes (2-4), ferritin-like protein cages (5-7), de novo designed protein cages (8), empty virus-like particles of hepatitis b virus, norovirus and bacteriophages P22 and MS2 (9-15), as well complete, infectious viruses (16). Development of the analytical tool and studying the protein cage are two scientific aims that have gone hand-in-hand in these studies, and they continue to do so in the chapters of this thesis.

Chapter 1 reviews the current state of predominant mass spectrometry technologies for the analysis of large protein complexes. Chapter 2 discusses how mass spectrometry can be applied to study an intact 18 megadalton virus capsid and discusses which properties of the samples and instrumentation ultimately limit what size of molecule is still accessible to the technique. Chapter 3 discusses how the mass range of a more recently introduced instrument may be further extended to allow high-resolution analysis of megadalton virus particles. Chapters 4 through 6 discuss specific applications of mass spectrometry to study capsid structures, thereby illustrating the full potential for the structural and biophysical characterization of large protein complexes by this technique. Finally, in Chapter 7 I will discuss some possible future directions for mass spectrometry of large protein complexes.

References.

1. Douglas, T., and M. Young. 2006. Viruses: Making friends with old foes. *Science*. 312:873-875.
2. Rostom, A. A., and C. V. Robinson. 1999. Detection of the intact GroEL chaperonin assembly by mass spectrometry. *J. Am. Chem. Soc.* 121:4718-4719.
3. van Duijn, E., P. J. Bakkes, R. M. A. Heeren, R. H. H. van den Heuvel, H. van Heerikhuizen, S. M. van der Vies, and A. J. R. Heck. 2005. Monitoring macromolecular complexes involved in the chaperonin-assisted protein folding cycle by mass spectrometry. *Nature Methods*. 2:371-376.
4. Van Duijn, E., A. Barendregt, S. Synowsky, C. Versluis, and A. J. R. Heck. 2009. Chaperonin complexes monitored by ion mobility mass spectrometry. *J. Am. Chem. Soc.* 131:1452-1459.
5. Kang, S., C. C. Jolley, L. O. Liepold, M. Young, and T. Douglas. 2009. From metal binding to nanoparticle formation: monitoring biomimetic iron oxide synthesis within protein cages using mass spectrometry. *Angewandte Chemie - International Edition*. 48:4772-4776.
6. Kang, S., J. Lucon, Z. B. Varpness, L. Liepold, M. Uchida, D. Willits, M. Young, and T. Douglas. 2008. Monitoring biomimetic platinum nanocluster formation using mass spectrometry and cluster-dependent H₂ production. *Angewandte Chemie - International Edition*. 47:7845-7848.
7. Kang, S., L. M. Oltrogge, C. C. Broomell, L. O. Liepold, P. E. Prevelige, M. Young, and T. Douglas. 2008. Controlled assembly of bifunctional chimeric protein cages and composition analysis using noncovalent mass spectrometry. *J. Am. Chem. Soc.* 130:16527-16529.
8. Lai, Y. -, E. Reading, G. L. Hura, K. -. Tsai, A. Laganowsky, F. J. Asturias, J. A. Tainer, C. V. Robinson, and T. O. Yeates. 2014. Structure of a designed protein cage that self-assembles into a highly porous cube. *Nature Chemistry*. 6:1065-1071.
9. Shoemaker, G. K., E. Van Duijn, S. E. Crawford, C. Uetrecht, M. Baclayon, W. H. Roos, G. J. L. Wuite, M. K. Estes, B. V. V. Prasad, and A. J. R. Heck. 2010. Norwalk virus assembly and stability monitored by mass spectrometry. *Molecular and Cellular Proteomics*. 9:1742-1751.
10. Uetrecht, C., I. M. Barbu, G. K. Shoemaker, E. Van Duijn, and A. J. R. Heck. 2011. Interrogating viral capsid assembly with ion mobility-mass spectrometry. *Nature Chemistry*. 3:126-132.

11. Uetrecht, C., C. Versluis, N. R. Watts, W. H. Roos, G. J. L. Wuite, P. T. Wingfield, A. C. Steven, and A. J. R. Heck. 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.* 105:9216-9220.
12. Uetrecht, C., C. Versluis, N. R. Watts, P. T. Wingfield, A. C. Steven, and A. J. R. Heck. 2008. Stability and shape of hepatitis B virus capsids in vacuo. *Angewandte Chemie - International Edition.* 47:6247-6251.
13. Keifer, D. Z., E. E. Pierson, J. A. Hogan, G. J. Bedwell, P. E. Prevelige, and M. F. Jarrold. 2014. Charge detection mass spectrometry of bacteriophage P22 procapsid distributions above 20 MDa. *Rapid Communications in Mass Spectrometry.* 28:483-488. doi: 10.1002/rcm.6809.
14. Pierson, E. E., D. Z. Keifer, L. Selzer, L. S. Lee, N. C. Contino, J. C. -. Wang, A. Zlotnick, and M. F. Jarrold. 2014. Detection of Late Intermediates in Virus Capsid Assembly by Charge Detection Mass Spectrometry. *J. Am. Chem. Soc.* 136:3536-3541.
15. Tito, M. A., K. Tars, K. Valegard, J. Hajdu, and C. V. Robinson. 2000. Electrospray time-of-flight mass spectrometry of the intact MS2 virus capsid [15]. *J. Am. Chem. Soc.* 122:3550-3551.
16. Fuerstenau, S. D., W. H. Benner, J. J. Thomas, C. Brugidou, B. Bothner, and G. Siuzdak. 2001. Mass spectrometry of an intact virus. *Angewandte Chemie - International Edition.* 40:542-544.

Chapter 1:

Analytical approaches to size and mass analysis of large protein complexes.

Joost Snijder^{1,2} and Albert J.R. Heck^{1,2}

¹ Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH, Utrecht, the Netherlands.

² Netherlands Proteomics Centre, Padualaan 8, 3584 CH, Utrecht, The Netherlands.

Based on:

Snijder, J. & Heck, A. J. R. Analytical approaches for size and mass analysis of large protein assemblies. *Annual Review of Analytical Chemistry* 7, 43-64 (2014).

Abstract.

The size and mass analysis of newly discovered protein complexes may be a first important step towards determining the constituents, the stoichiometry and possibly the shape and topology of these assemblies. Moreover, methods for size and mass analysis may be used to study assembly and dissociation behavior of these complexes as a function of time or under the influence of changing environmental conditions. Here, emerging analytical methods to characterize protein complexes are reviewed. The focus lies on native mass spectrometry (native MS) in particular, but a range of other electrospray ionization based methods are also covered. A selected number of applications that are illustrative of the current state of these analytical approaches for the size and mass analysis of protein assemblies are highlighted.

Whereas it has become relatively straightforward to obtain a list of protein interaction partners with affinity-purification and MS-based identification protocols (AP-MS), more precise structural analysis of a protein complex presents a much bigger analytical challenge (1-4). AP-MS methods do not generally yield information on the size of a protein complex or the stoichiometry of the co-assembled subunits, and also often do not disclose important information on co-interacting DNA, RNA and small molecules (nucleotides, lipids, etc.). It is also not usually possible to distinguish between different coexisting states of higher order oligomers. To obtain a more complete picture of a supramolecular protein assembly and its composition, it is thus necessary to analyze the intact assembly. Moreover, if the analytical approach would allow one to further dissect the purified protein complex from the top-down, additional information on protein subunit arrangement and shape can be obtained.

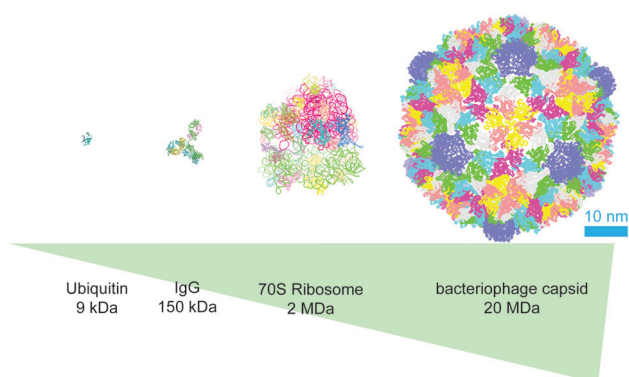


Figure 1. Structural models of Ubiquitin, IgG, the ribosome, and a virus capsid to illustrate the size and mass range of some biologically relevant protein complexes.

There are several challenges associated with a comprehensive analysis of protein assemblies. For instance, protein complexes can range from a few kilodaltons to several tens of megadaltons in molecular weight (See Figure 1), which ideally would be covered with the same analytical approach. Whereas the core-assembly of a protein complex is often formed through stable high-affinity interactions, accessory components may be weakly bound and transiently associated. The transient nature of the assembly requires that the principal of analytical separation and detection is relatively fast; so as to prevent ‘smearing’ as often observed in gel electrophoresis or size-exclusion chromatography. In other words, when analytically separating complexes of different size/mass, the mass information must be extracted before dissociation or re-association can take place.

As a result, many commonly used techniques for determining protein size and mass, such as light-scattering based techniques, are unable to efficiently distinguish between multiple coexisting states due to ensemble averaging or a lack of resolution. Many proteins within protein complexes are also often decorated by post-translational modifications, and can bind small molecule substrates or effectors such as nucleotides, lipids or sugars, yielding very subtle variations in mass that can only be uncovered with high mass resolution and precision. Generally speaking, the shorter the time-scale of analytical separation and detection and the higher the precision and accuracy of the mass determination, the more information can be gained on the protein complex under study.

Electrospray ionization as a source of gas-phase ionized protein assemblies.

Because most mass analyzers are operated under high vacuum, protein complexes need to be ionized and carefully transferred into the gas phase. In most approaches, electrospray ionization (ESI) is the method of choice (5-8). In ESI, an analyte solution is transferred into a conductive capillary. A potential is applied between the capillary and the inlet of the mass spectrometer. This potential difference leads to the formation of a so-called Taylor-cone at the tip of the capillary (9). Small charged droplets, that contain the analytes, are formed at the tip of the Taylor cone by jet-fission. Solvent evaporates from the droplets, which increases their charge density. Once the charge density reaches a certain limit, i.e., the Rayleigh limit, the droplets may undergo a Coulombic-driven explosion and jet fission. This process repeats until the droplets are so small that only a single analyte is contained within them. As the remainder of the solvent evaporates, part of the excess charge on the droplet surface is carried over to the analyte, forming multiply charged (protonated) ions (according to the charged residue model of ESI) (7, 9, 10).

Whereas the solvent-free analysis of protein complexes does offer some advantage (i.e., interpretation of the obtained masses is not complicated by additional mass that originates from hydration layers, counter-ions, etc.), special care needs to be taken so as not to disrupt the native contacts of a protein complex or introduce gas-phase related artifacts to its structure. Several experiments and computational studies have pointed out that whereas some structural rearrangements do take place, much of the backbone solution-phase structure can be retained following complete desolvation (under carefully controlled conditions with minimal ion heating) (11-13). ESI is generally considered to be one of the most gentle ionization methods available and is the preferred ionization method in most MS studies of protein complexes. By using lower flow rates through the

capillary, the initially formed droplets are smaller and fewer fission events take place before effective ionization of the analyte, which makes ionization even gentler (9, 14). Scaling down to flow rates of only a few tens of nanoliters per minute is known as nanoESI and is the preferred ionization method in most native MS studies (see below). It offers the additional advantage of enhanced sensitivity.

The use of ESI in MS does pose a few restrictions on experimental design. It is for instance not feasible to analyze the protein complexes in buffers that are typically used in biochemistry and molecular biology (i.e., solutions containing high salt concentrations, Tris or HEPES etc.). These buffer components ionize very efficiently, while present at millimolar concentrations, thereby dominating the spectrum and even suppressing ionization of the protein analyte (typically present at micromolar concentrations). Moreover, these buffer components may form adducts with the protein complex under study, which complicate the interpretation of the obtained ion signals and distorts spectral quality. One is therefore restricted to the use of volatile buffer components such as ammonium acetate, ammonium bicarbonate, or the recently introduced ethylenediamine diacetate (EDDA). Of these examples, ammonium acetate is by far the most commonly used for the analysis of intact protein complexes, even though it has a very limited buffer capacity. Ammonium bicarbonate has a higher buffer capacity, but may cause protein unfolding at the gas-water interface introduced by bubble formation of the CO₂ gas (15). EDDA was recently introduced and shown to be particularly suitable for the analysis of the ATP-binding GroEL chaperone, resulting in very good desolvation and high mass resolving power (16).

There are also concentration limits for the analysis, especially when using nanoESI. The lower limit depends on the sensitivity of the mass analyzer (it is typically below 100 nM in native MS studies employing nanoESI), whereas at higher concentrations (more than several tens of μM) the crowded solution starts to restrict flow through the ESI capillary. When studying protein-ligand interactions, high concentrations of ligands in ESI will generate protein ions with non-specifically bound ligands (the higher the concentration, the more likely a free ligand will share a droplet with a free protein and adsorb onto the protein as desolvation proceeds). Similarly, non-specific clustering of the protein assembly will also occur.

ESI generally yields a distribution of multiply charged ions. The number of charges that is carried over to the protein complex depends to some extent on the conditions of the spray solution but is dominated by the solvent-accessible area of the analyte (17). The charge state signature of a protein ion can thus be used

to monitor conformational changes (that result in large changes in solvent-accessibility) and is sometimes used to monitor protein folding (18, 19). Several supercharging and charge-reducing agents have been described that allow one to manipulate the average number of charges that is acquired by the analyte (20-24). As most MS approaches yield information on mass-to-charge ratio, charge state assignment is crucial for calculating accurate masses and is rather straight forward for highly resolved spectra of relatively small protein complexes (charge states can be calculated from matched peaks) (25). Specialized software has been described to aid in the interpretation of more complicated spectra and a more elaborate strategy for charge state assignment is often required in the cases of complicated spectra and lesser-resolved peaks (26-29).

Alternative ionization methods.

Although ESI is widely reported as the most suitable ionization technique for the analysis of non-covalent protein complexes, alternative ionization techniques have also been successfully applied; mentioned here briefly. In particular, large non-covalent assemblies have been studied using matrix assisted laser desorption/ionization (MALDI) and laser-induced liquid bead ion desorption (LILBID) (30-32). In the case of MALDI, the analyte is embedded in a dry matrix that, upon laser-excitation, ionizes, desorbs, and transfers charge to the analyte. In LILBID, micro-droplets are excited at the absorption maximum of water, which, beyond a threshold laser intensity, disrupts the droplet such that the analyte is transferred into the gas phase. Most notably, the use of MALDI and LILBID has facilitated the analysis of intact IgG antibody, RNA polymerase, detergent solubilized membrane proteins, amyloid assemblies, and virus particles.

Native mass spectrometry.

Native MS is a term coined to describe the mass analysis of non-covalent protein complexes under non-denaturing conditions from buffered aqueous solution, as close to physiological conditions as is still compatible with ESI (33, 34). The information that native MS offers on protein complexes is outlined in Figure 2a and discussed in detail in this section. Most reported native MS studies are carried out with time-of-flight (TOF) mass analyzers because of their superior resolution and sensitivity for high mass, or m/z , ions. In TOF analyzers, the mass-to-charge ratio of an ion is determined by measuring its flight-time along a defined path in the analyzer. These instruments are often slightly modified versions of commercially available hybrid Quadrupole-TOF (QTOF) instruments, where most modifications are aimed at optimizing transmission of high mass ions (35-37). A schematic of a QTOF for native MS is presented in Figure 2b.

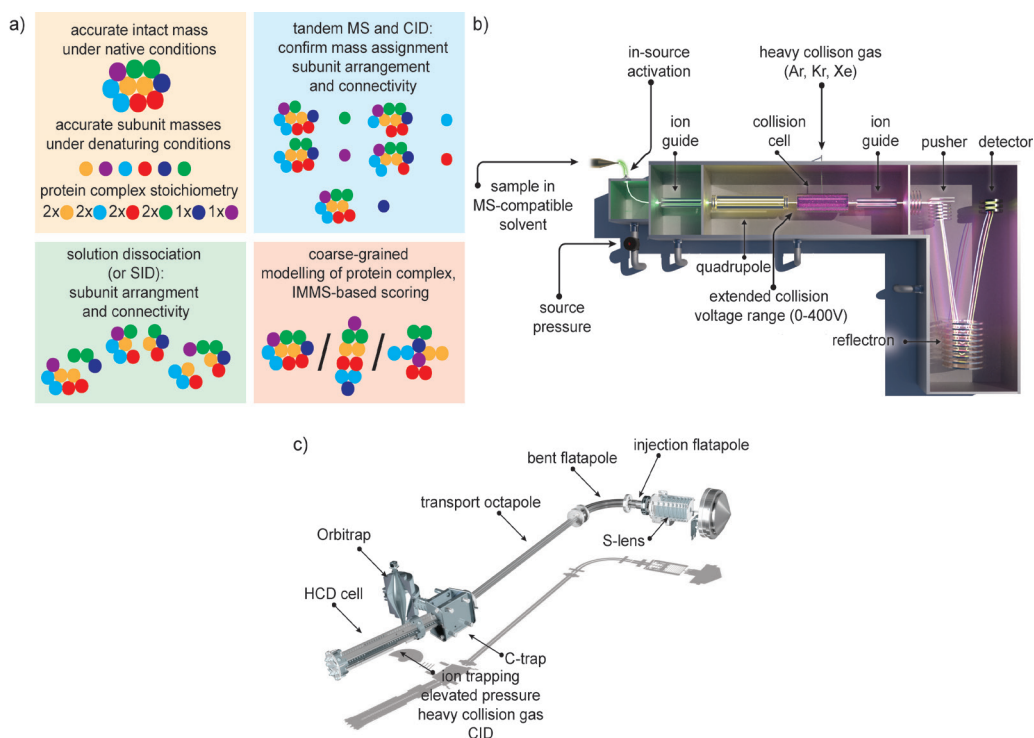


Figure 2. Workflow and Instrumentation used in Native Mass Spectrometry. A) Illustration of different experiments used in native MS. B) Schematic of a modified QTOF instrument. Adapted from reference 61, with permission. C) Schematic of a native MS Orbitrap instrument.

Whereas the standard commercial instruments are operated at a typical pressure of 1 millibar in the source region, the transmission of large ions is greatly improved by increasing the pressure to approximately 10 millibar (38, 39). The increased pressure focuses large ions by collisional cooling and a similar approach is often also employed in the multipole ion guides of high-mass QTOF instruments (40). This can be achieved by fitting a flow-restrictive sleeve on the ion guide, such that more gas from the previous pumping stage leaks into the next region, or by feeding a gas-line into the pumping stage. The use of heavier buffer gas (such as argon or xenon, compared to helium or nitrogen) further improves the transmission of high mass ions. In addition to the source and transfer stages of the instrument, the collision cell is usually also operated at elevated pressures using a heavier inert collision gas (xenon, krypton) to further improve transmission. The heavy collision gas also aids in collisional activation of large ions as more energy is transferred per individual collision. This activation in the collision cell can also be used to improve spectral quality by stripping ions and/or buffer adducts from the analyte of interest (41). Sufficiently high activation leads

to gas-phase dissociation, whereby the formed fragment ions can help confirm mass assignments and provide structural constraints on subunit arrangement (42-46). The ability for collisional activation has been further expanded in some instruments by extending the applicable DC collision potential to the collision cell (47).

Most TOF analyzers that are used in native MS have a nominal resolution ($M/\Delta M$) between 5000 and 10000. Compared to many other popular mass analyzers, there is only a shallow drop-off in resolution with increasing m/z , which is one reason why TOF analyzers are very suitable for native MS studies. Despite the numbers for the mass resolution quoted above, the effective resolution on a large protein complex' charge state reported on these instruments is usually no higher than 1500 and mostly only a few hundreds. Some have argued that effusive expansion of the ion beam for large analytes (during transfer from high- to low-pressure ion guides) causes similar dispersion in the pusher region and therefore in the obtained flight-time profiles (48). However, the main cause for relatively wider peaks observed in native MS studies seems to be incomplete desolvation of the analyte (49). Many different neutral and/or ionic adducts (sodium, potassium, ammonium, acetate, water, etc.), differing only slightly in mass, can all be attached on the same protein complex. Rather than measuring a bare protein ion, one measures the unresolved envelope of the protein with a variety of buffer adducts with closely spaced masses, leading to extensive peak-broadening. Buffer adduct formation becomes increasingly likely with larger protein complexes, resulting in broader peaks for higher masses. Relatively little can be gained with an enhancement of the instrumental resolution of the mass analyzer, given that the underlying isotope distributions of all the different adduct peaks will also overlap extensively (and to achieve 'isotopic' resolution on such large analytes at high m/z is currently simply unfeasible). For this reason, the resolving power (and with that also the accessible mass range) in native MS studies could be greatly enhanced if buffer adduct formation could be further reduced, moving towards complete desolvation.

Although the TOF has traditionally been the mass analyzer of choice, Fourier transform ion cyclotron resonance (FT-ICR) and Orbitrap mass analyzers have also been used to study large protein assemblies. Several native MS studies using FT-ICR mass analyzers have been reported, including the study of protein-protein and protein-oligosaccharide interactions, some of the first systematic studies of gas-phase dissociation kinetics of protein ions, and the use of alternative activation schemes for high mass ions, such as electron capture dissociation (50-54). It was recently shown that the Orbitrap mass analyzer could also be mod-

ified to become suitable for native MS (55). Rather than measuring the flight-time, the Orbitrap measures the axial frequency of oscillation of trapped ions along a central electrode (56, 57). A schematic of the Orbitrap platform used in those studies is presented in Figure 2c. Unlike the TOF instruments, the front end of the native MS Orbitrap is not operated at elevated pressures. The only differences compared to its use for analyzing small molecules is tuning of the DC potentials of the ion lenses and trapping of the ions in the high-energy collision (HCD) cell (which is operated at elevated pressure using a heavier collision gas), before injection into the Orbitrap. It was shown that this Orbitrap platform has high sensitivity and superior resolving power in native MS applications (55, 58), which, for the reasons outlined above, is most probably due to the difference in the front-end of the instrument (which apparently promotes loss of buffer adducts), whereas the higher nominal resolution of the analyzer has only a minor contribution to its higher performance.

In the past 20 years, native MS has established itself as a valuable tool in studying protein complex composition, structure, and assembly (34). Mass analysis with a precision that is routinely below 0.01% has been conducted on various protein complexes. Recent examples include membrane protein complexes (solubilized in detergent-micelles) (59-61), which are particularly challenging for many techniques in structural biology, as well as megadalton virus capsids (62, 63), which are thought to be the biggest molecules accessible to the technique with the current state of instrumentation. The resolving power of the technique is often sufficient to monitor post-translational modifications (such as glycosylation and phosphorylation) and small ligand binding on intact protein complexes up to the megadalton range (16, 55, 58). Data acquisition takes on the order of minutes between samples, allowing kinetic measurements of relatively slow reactions in the minutes-to-hours timescale (64). Studies on amyloid formation and chaperone and virus assembly illustrate how multiple coexisting states can be separated, detected, and characterized simultaneously with native MS (65-67). The quadrupole mass filter on the hybrid QTOF setup allows one to study the gas-phase stability and subunit arrangement of individual components from these complex mixtures by tandem MS using collision induced dissociation (CID).

CID is a so-called slow-heating activation process, and it typically results in gradual unfolding of single subunits that are ejected from the complex with a particularly high number of charges (protons are mobile on the protein and rearrange to occupy the newly available surface on the unfolded subunit) (68). Depending on precursor ion charge and quaternary structure of the protein

complex, alternative dissociation pathways are also observed that include the dissociation of lower-charged compact subunits, dissociation of higher order multimers from a complex, or fragmentation of the peptide backbone (13). As a consequence of the slow heating of ions in CID and the associated unfolding of subunits, the amount of structural information that can be gained from CID is somewhat limited. Assuming that peripheral subunits tend to dissociate first, the order of dissociation observed in CID can provide some clues as to the arrangement and connectivity of subunits. However, this is only useful when studying hetero-multimeric assemblies and the concept of a “peripheral subunit” is meaningless in the context of symmetric oligomers and ring-like structures. Alternative dissociation pathways where higher-order multimers dissociate are much more informative on subunit arrangement and connectivity. These alternative dissociation pathways can be achieved by using an alternative activation method called surface induced dissociation (SID). In SID, ions collide with an inert surface to achieve dissociation of protein complexes. Because activation in SID is a single-step high-energy event, subunit contacts are broken without the extensive unfolding that is observed in CID (69). Hence, it is more likely with SID to generate subcomplexes in the gas phase that are informative for resolving the quaternary structure of the assembly (70, 71).

Within the framework of the native MS workflow, additional structural information about the protein complex under study can be obtained by combining mass spectrometry with ion mobility spectrometry, an approach referred to as IMMS (72). Ion mobility is sometimes also used as a stand-alone technique to analyze the size of protein complexes, as is the case with differential mobility analysis (see section on GEMMA, below). In IMMS, a hybrid instrument performs ion mobility separation and MS analysis of the protein ions in sequence (73-75). From the ion mobility of an ion, its collisional cross section can be calculated, a property that is largely determined by the rotationally averaged projection area of the analyte. The collisional cross section of an ion thus relates to its shape, and IMMS measurements have been used to determine the subunit arrangement of protein complexes and to monitor conformational changes. IMMS has also been pivotal in assessing the gas-phase structure of protein ions and has gone a long way to show that global aspects of solution-phase structure are preserved after desolvation (12).

The most popular drift-time separation techniques for native IMMS are drift tube ion mobility (DTIMS) and travelling wave ion mobility (TWIMS). In DTIMS, ions are separated on the basis of mobility by applying a static field on a drift tube that is densely filled with a neutral buffer gas. In TWIMS, ions are

separated in a gas-filled stacked ring ion guide under the influence of a travelling wave potential. Whereas the collisional cross section can be directly deduced from DTIMS measurements, TWIMS requires careful calibration with known standards from DTIMS measurements (73, 76). The collisional cross section has been used in two important ways to assess protein complex structure in the gas phase. In one approach, it is determined how the collisional cross section scales with the molecular weight of the protein assembly. This trend allows one to make the distinction between globular, ring-like, or extended sheet-like structures and has been used in the study of amyloid and virus assembly (65-67). In a more systematic approach, experimental cross sections are compared to theoretically modeled projections of atomic or coarse-grained models of protein structures obtained by X-ray crystallography or electron microscopy (46, 77). By systematically scanning all possible arrangements of protein subunits in an unknown structure, the experimental cross section can be used as a constraint to exclude a large space of possible subunit arrangements.

Highlighted recent applications of native mass spectrometry.

F- and V-type ATPases.

A recent break-through in native MS has been the analysis of typically insoluble intact integral membrane protein complexes (59, 78). Membrane proteins are notoriously difficult to handle because of their poor solubility, which can be overcome by the use of detergent micelles. However, most ionic detergents are incompatible with ESI, but several non-ionic detergents have recently been described to facilitate the analysis of membrane protein complexes with native MS (61). Most notably, recent studies on a variety of ATPases demonstrate how stoichiometry, subunit arrangement, and even lipid and nucleotide binding can be uncovered with native MS (See Figure 3) (60, 79). Related to this, attempts to analyze membrane proteins have also been successful using MALDI or LILBID MS, rather than nanoESI (30, 31).

Bacterial immune response related protein complexes.

Recently, it was found that bacteria have defense systems against viral infections, generally known as clustered regularly interspaced short palindromic repeats (CRISPR) systems. In this defense mechanism, a protein-RNA complex plays a pivotal role, termed Cascade in *Escherichia coli*, which targets RNA molecules encoded in the genome of the bacteria towards viral DNA, where, upon binding, it hampers viral replication, thereby suppressing infection.

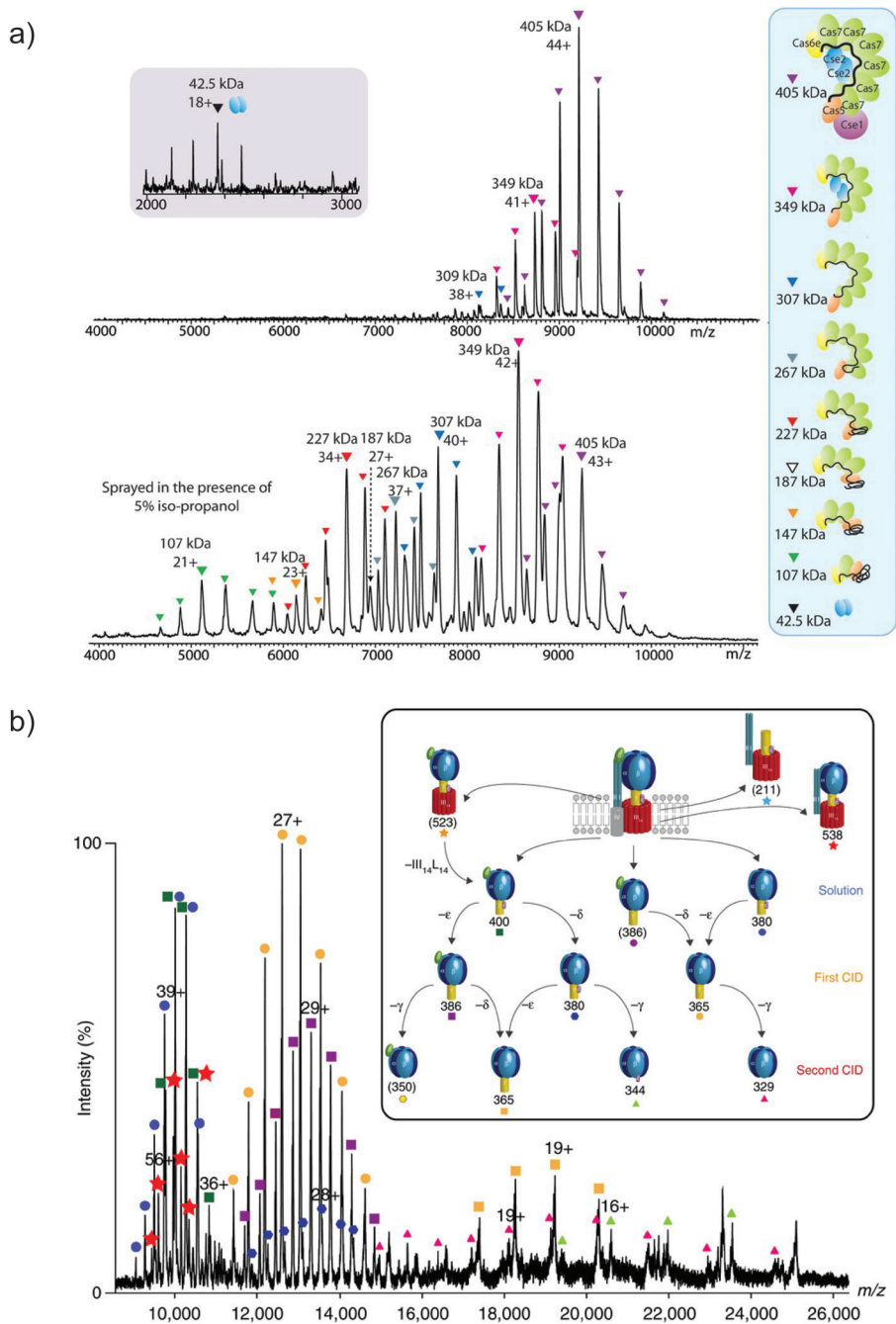


Figure 3. Highlighted application of Native Mass Spectrometry: subunit arrangement and connectivity from dissociation approaches. A) Native MS with solution phase dissociation to determine subunit connectivity of the CRISPR related complex Cascade. Taken from reference 80, with permission. B) Native MS of detergent solubilized F-ATPase membrane protein complexes. Subunit connectivity was deduced from CID pathways. Taken from reference 79, with permission.

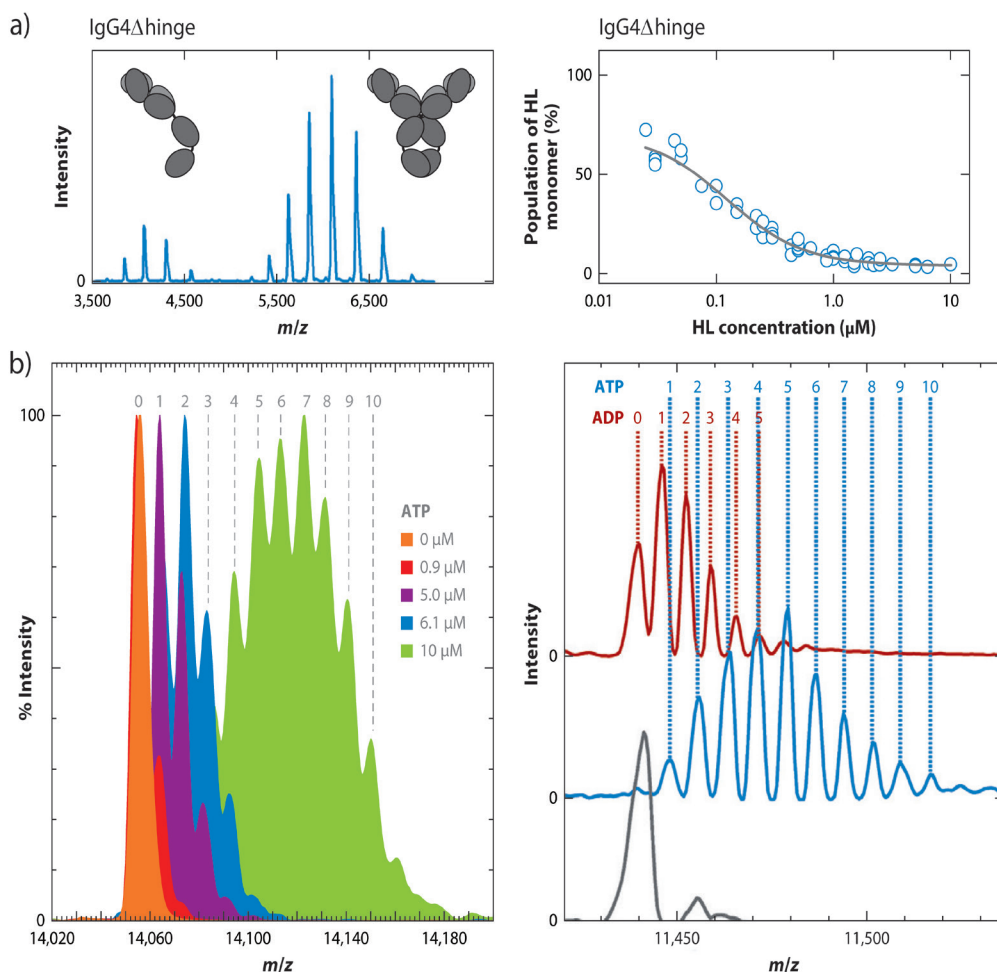


Figure 4. Highlighted applications of Native Mass Spectrometry: protein-protein and protein-ligand binding. A) Studying antibody dimerization with native MS. “HL” refers to the “hinge-less” antibody monomer. Taken from reference 88, with permission. B) (left) ATP binding to the chaperone GroEL in EDDA buffer, analyzed on a QTOF instrument. Taken from reference 16, with permission. (right) ADP/ATP binding on GroEL in ammonium acetate buffer, analyzed on an Orbitrap instrument. Taken from reference 55, with permission.

The stoichiometry and structures of a variety of such CRISPR-related complexes, that typically contain half a dozen different proteins and a crRNA sequence, have recently been elucidated with the aid of native MS by detailing the stoichiometry, demonstrating subunit connectivity with solution phase dissociation, and determining subunit arrangement and shape from IMMS (See Figure 3) (80-82). The resulting models of the CRISPR-related complexes were shown to be in excellent agreement with cryoEM reconstructions of the same complexes and a critical

first step in assigning those EM densities. These studies highlight how native MS serves as a valuable complement in the toolbox of structural biology. The use of solution-phase dissociation and IMMS to determine the topology of a protein complex has been reported in several studies and is becoming an integral part in the native MS workflow (46, 80, 83, 84).

Antibody dimerization.

Binding constants of protein-protein and protein-ligand interactions can be determined from titration experiments, using native MS as a read-out to quantify binding (85-87). As mentioned above, care needs to be taken to ensure that no ESI- or MS-related artifacts are introduced in such analyses. A recent example of the use of native MS to determine binding affinities comes from a study where the determinants of antibody half-body dimerization were uncovered (See Figure 4) (88). The binding affinities for dimerization were determined in engineered half-molecules of the IgG4 antibody by native MS. Titration experiments with native MS were used to determine the effect of a range of mutations on the dimerization constant of the half-antibodies to reveal which regions are particularly important. Moreover, time-resolved measurements of antibody mixtures of different mass allowed the kinetics of subunit exchange to be determined. The mass-tagged mixing of subunits to determine subunit exchange between complexes has been reported on several occasions (89-91).

Small molecule binding to large protein assemblies: GroEL with ATP and ADP.

Many supramolecular assemblies rely on small molecule effectors or substrates for their biological activity. Binding of nucleotides such as ATP is very common but not easily studied in large protein complexes. Information on the ensemble state of a multiprotein complex for ATP occupancy is particularly challenging for many techniques with relatively low resolution. It was recently shown that binding of individual ATP/ADP molecules could be resolved on the 800 kDa GroEL chaperone using the modified native MS Orbitrap platform, which would facilitate detailed studies on small molecule binding on supramolecular assemblies (See Figure 4) (55). Binding of ADP or ATP could be easily distinguished, even though this represents a mass difference of only 80 Da on a total mass of 800 kDa (i.e., 0.01%). With the use of an EDDA solution, rather than the more common ammonium acetate, it was recently shown that ATP binding could also be resolved on a more conventional QTOF instrument (16). Here, it was demonstrated that native MS can be used to characterize the ensemble state of ATP binding on GroEL, and this information was used to determine the allosteric mechanism of ATP binding. These data clearly demonstrate that native MS may

provide an ideal tool to monitor effector and/or drug binding to proteins and protein assemblies.

Alternative approaches in the determination of the size and molecular mass of protein assemblies.

As described above, TOF MS detection is still predominant in native MS applications, but alternative means to sense and record a mass of a particle are emerging and being explored. They present interesting alternatives for the analysis of protein complexes in the mass range between several kilodaltons to megadaltons.

Charge and image current detection for single-particle mass spectrometry.

Protein complexes acquire tens to hundreds of charges in ESI. This high number of charges makes it possible to detect single ions by measuring the image current they induce on a variety of detection plates, such as used in charge detection MS coupled to TOF, FT-ICR, and Orbitrap MS. In FT-ICR and Orbitrap mass analyzers, imaging currents are typically induced and detected when an ion packet, containing typically thousands of ions, is in coherent orbiting motion. After being amplified, these currents can be digitized, after which these time-domain signals are Fourier-transformed to yield a mass spectrum. To induce a recordable image current, ~100 charges are needed. Smith *et al.* pioneered the application of FT-ICR to measure the masses of single ions of macromolecular protein assemblies and DNA (92). They demonstrated the detection of single ions of BSA dimers carrying only 60 charges (93). Moreover, using FT-ICR and individual ion detection, large pieces of ~100 MDa DNA and circular double-stranded plasmids (~1.9 MDa) extracted from *Escherichia coli* could be analyzed (94). The plasmid ions trapped in the ICR were reacted with acetic acid molecules to induce shifts in the charge-state. Measurements of the multiple peaks arising from the charge-state shifting provided masses of the individual ions with an average accuracy of 99.8%. More recently, Makarov *et al.* have shown that single ion detection can also be achieved for highly charged proteins and protein assemblies using Orbitrap mass analyzer (55, 95).

Several groups have explored and constructed in-house charge detection devices for the analysis of large ions in TOF mass spectrometers. Pioneering work by Fuerstenau and Benner provided an instrument capable of analyzing highly charged macromolecular particles ionized by ESI that contained a very sensitive amplifier, which detects the charge on an ion as it passes through a tube detector (96). The tube (length of 36.1 mm and a bore of 6.35 mm) acts both as a TOF

mass analyzer and as a detector. It must be aligned with the ion beam axis, allowing ions to pass through the tube one at a time. This detected charge is then amplified and measured. This approach has the advantage of being simple and cheap to produce. A velocity measurement of the ion together with its known electrostatic energy provides the ion's mass-to-charge ratio. They demonstrated initially that, with this instrument, the molecular weight of single ions of DNA with masses ~ 1 MDa and charge numbers in excess of 400 could be measured. Later on, it was also used to analyze intact viruses (introduced by ESI), namely the rice yellow mottle virus that consists of a single-stranded RNA surrounded by a homogeneous protein capsid with a mass of 6.5 MDa, and the rod-shaped tobacco mosaic virus with a mass of ~ 40 MDa (97).

More recently, a few groups revisited the concept of charge detection MS for the analysis of macromolecular ions using a (linear) array of charge detectors enabling multiple image charge measurements per ion (98-100). These detectors consist of a multiple collinear tubes differentially isolated and connected to amplifiers to reduce the noise and improve the detection limit. For instance, the image charge detection mass spectrometer of Smith *et al.* encompassed an array of 22 charge detection tubes, arranged coaxially and divided into two sets of 11 detectors. They claimed that the correlation approach they used to analyze the output from the image charge detectors provides advantages over using a Fourier transform in terms of signal-to-noise ratio. Using correlation analysis, the achievable noise level with the 22 detectors was around $10e$ for a 500 m/s ion. They used the device to measure poly ethylene glycol (PEG) and found that the measured charge and molecular weight was in agreement with the expected polymer size of 300 kDa. Later on, it was shown that cytochrome C and ADH monomers can be detected and mass analyzed in their device.

The charge detection approaches outlined above seem to be applicable to protein complexes larger than those accessible to the wider reported native MS approach. These techniques cover a very wide mass range, but they still suffer from a comparatively low precision. The uncertainty in the obtained masses makes for a less powerful constraint in determining complex stoichiometry and composition. Whereas 'single-ion detection' does have a ring of sensitivity to it, the reality is that transfer of the analytes into the MS device is often the bottleneck, such that solutions in the high micromolar range are required for analysis. The generation of ions by ESI, combined with single ion detection, does make it possible to uncover multiple coexisting populations from heterogeneous mixtures, and therefore give these techniques a distinct advantage over conventional wet-lab techniques for the determination of protein complex size and mass. In addition,

whereas resolution of subsequent charge states is an absolute necessity for precise mass determination in native MS, single-particle charge-detection MS will yield a mass for every ion, no matter how heterogeneous the mixture of analytes is, which can often be problematic for 'conventional' native MS, especially with larger, poorly desolvated ions. The ability of charge-detection MS to deal with particularly heterogeneous mixtures, especially of large protein assemblies, is illustrated in recent work on virus capsids by Jarrold and co-workers (101, 102).

Nanomechanical mass sensors.

As another means to analyze macromolecules and nanoparticles, resonant micrometer scaled cantilevers can be used as mass sensors (103). Cantilever based mass sensors have been shown to ensure the sensitivity to measure the mass of not only single large protein assemblies, nanoparticles, and cells, but also of small molecules like naphthalene and even Xe atoms (104). In nano(electro)mechanical-mass spectrometry (NEMS), analytes such as DNA (105), protein assemblies, or nanoparticles are introduced into the device by ESI, followed by desolvation and guidance by ion optics towards the detector (104, 106, 107). The detector makes the major difference, as it is an ultrahigh-frequency nanoelectromechanical resonator. The vibrational frequency of a NEMS resonator is a delicate function of its total mass. Small variations in mass, as induced by adsorption of analytes onto the resonator, can measurably alter its resonant frequency. Even individual protein molecules induce an abrupt jump in the resonant frequency when they adsorb onto the sensor. These frequency shifts are proportional to the mass and position of adsorption on the resonator and can be recorded in real time (108). Individual events can be detected by monitoring successive frequency shifts when a protein assembly of interest is electrosprayed into the instrument. Presumably only a small fraction of the electrosprayed molecules make it to the small resonator surface, and thus the overall detection sensitivity is likely not yet that great. The NEMS detector seems to have a nearly unlimited mass range, capable of analyzing particles in the tens of MDa range. A current limitation seems to be the mass resolution and mass accuracy of such devices.

A nice recent application of NEMS has been in the analysis of human IgM antibodies (See Figure 5). IgM forms macromolecular complexes in serum, whereby biologically-active isoforms can be tetrameric, pentameric, hexameric, or dipentameric assemblies of ~190 kDa subunits. For the prevalent pentamer isoform, an additional small protein (the J chain) helps link the assembly and contributes ~15 kDa to the total ~960 kDa mass of the complex. The NEMS spectrum, accumulated from 74 single particle spectra, reveals particles of quite a few different masses but clearly shows the pentameric IgM complex as the most abundant

signal at 1.03 ± 0.05 MDa. Moreover, a dimerized pentameric complex at 2.09 ± 0.05 MDa is also present. These values are very close to the anticipated values 0.96 MDa and 1.92 MDa, which is remarkable as no calibration other than using the nominal mask dimensions of the mass sensor was applied. It is impressive that such data can already be acquired from analyzing only 74 particles, although it would also be interesting to see whether the convoluted spectrum still reveals the additional presence of multiple oligomers (trimer, tetramer, etc.) when the number of analyzed particles, and thus the statistics, could be increased.

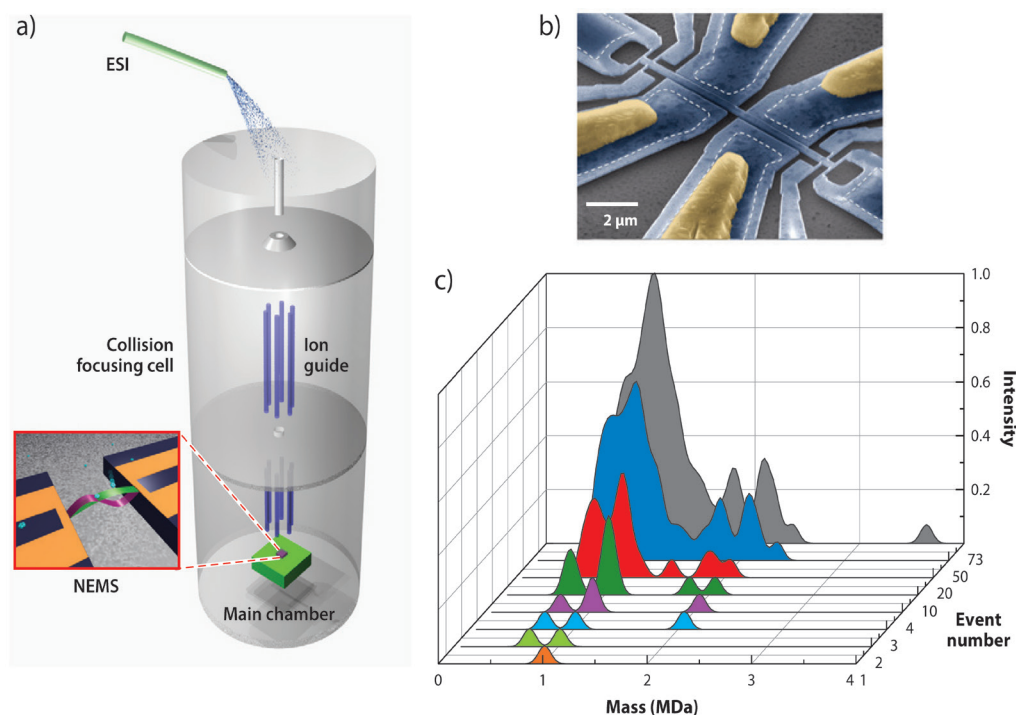


Figure 5. Single particle mass spectrometry with NEMS. A) Schematic overview of a NEMS mass spectrometer. B) Scanning Electron Micrograph of the NEMS device. The white dotted line shows the boundaries of the region beneath the suspended device that anchors it to the substrate. Yellow regions represent Al/Si gate contacts. Narrow gauges near the ends of the beam become strained with the motion of the beam, thereby transducing mechanical motion into electric resistance. Scale bar, 2 μm . C) NEMS mass spectrum of IgM oligomers, showing the cumulative spectrum for an increasing number of measured events. Taken from reference 106, with permission.

Gas-phase electrophoretic mobility molecular analyzers (GEMMA).

Representing another interesting alternative to TOF analyzers for the analysis of macromolecular assemblies, ESI has been combined with gas-phase electro-

phoretic mobility molecular analyzers (GEMMA) (109, 110). Therein, the high charge of particles as obtained via ESI is first reduced by means of bipolar ionized air (typically generated by an α -particle source) to yield predominantly singly charged ions. These charge-reduced species are subsequently separated and sized by their electrophoretic mobility in an ion mobility drift cell. In GEMMA, the particles are typically detected using a condensation particle counter. The resulting data can be converted into electrophoretic mobility diameters by applying the Millikan equation. These diameters can be converted to a mass spectrum as well, due to the generally good correlation between mobility size and molecular weight, although like in IMMS, exceptions to this rule may apply for non-globular structures, such as shell-like viral capsid intermediates (65). The GEMMA analyzer has an extended size range and has been successfully used for particles in size ranging from 3 nm to 100 nm, which covers a mass range of a few kDa (small proteins) to 100 MDa (whole viruses or even cell organelles).

Applications of the GEMMA analyzer have ranged from analyzing antibody aggregation, macromolecular protein complexes (111), synthetic polymers (112), intact viruses (113), and lipoparticles (114). GEMMA analysis was successfully applied for instance on the 4.6 MDa cowpea chlorotic mottle virus (CCMV) (113) and the tobacco mosaic virus (114). Although the mass resolution of the GEMMA instrument is still too low to enable accurate mass measurement, GEMMA does provide 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 as its measured diameter resembled that of the particle as measured by electron microscopy.

Loo *et al.* reported a direct comparison of native MS using a QTOF type mass analyzer with GEMMA for studying the macromolecular organization and structure of the 28-subunit 20S proteasome from *Methanosarcina thermophila* and the mammalian (rabbit) proteasome (111). ESI-MS measurements with a QTOF analyzer of the intact 690 kDa proteasome were consistent with the expected $\alpha 14\beta 14$ stoichiometry. Collisionally activated dissociation of the 20S gas-phase complex was applied on the QTOF resulting in the loss of only α -subunits, consistent with the known $\alpha 7\beta 7\beta 7\alpha 7$ topology. Moreover, the analysis of the binding of a reversible inhibitor to the 20S proteasome showed the expected stoichiometry of one inhibitor for each β -subunit. Using the GEMMA approach, an electrophoretic diameter of 15 nm could be measured for the $\alpha 7\beta 7\beta 7\alpha 7$ complex, in concordance with the diameter estimated from crystallographic measured electron densities. The authors concluded from their work that elements of the gas-phase structure of large protein complexes are preserved upon desol-

vation and that native MS and IMMS analysis can be used complementary and reveal structural details of the solution protein complex.

Although GEMMA does not provide the mass accuracy as potentially achievable by native MS using TOF analyzers, it is not that much limited by micro-heterogeneity of the samples. Illustrative for that is the study by Allmaier *et al.* (114), which presented differential analysis of intact very-low-density (approximately 35 nm), low-density (approximately 22 nm), and high-density lipoparticles (approximately 10 nm), which represent multifaceted hetero-complexes consisting of cholesterol, lipids, and proteins in different ratios. The measured EM diameter and narrowness of the peaks are indicative of the size and molecular complexity of the analyzed particles. HDL is the smallest/densest lipoparticle due to its high protein content (50% proteins, 25% phospholipids, and 15% cholesterinesters). The broad distribution and bigger size of the VLDL particles correlates to its task of transporting endogenous lipids, and it is estimated to contain 50% triacylglycerols, 20% phospholipids, 10% cholesterinesters, and 10% proteins. These particles could be easily differentiated by GEMMA analysis but would likely present difficulties in native MS as their micro-heterogeneity would hamper charge state resolution and thus accurate mass assignments.

Concluding remarks.

Several electrospray-based analytical methodologies used for the size and mass analysis of naturally occurring as well as synthetically prepared nanoparticles have been reviewed. As the importance of these particles is evident, in traditional biology, as well as in nanotechnology and synthetic biology, the methods described here are essential for better particle characterization and understanding. Improvements are still needed in the areas of sensitivity, selectivity, specificity, and speed. At present, it seems that extending the accessible mass range to over several tens of MDa comes at the expense of mass accuracy, and thus specificity. Additionally, although single particle analysis may provide certain advantages, current approaches still lack high sensitivity, thus requiring long analysis times. Hybrid technologies may emerge that provide valuable new approaches, combining either some of the analytical methods described above, or combining those methods directly with for instance electron microscopy or (gas-phase) X-ray diffraction (115, 116). Concluding, present-day ESI-based approaches have been very beneficial for the size and mass analysis of large protein assemblies and other nanoparticles, but there is still plenty of room and need for further future technology developments.

Acknowledgements.

This work was supported by the Netherlands Organization for Scientific Research (NWO) with ALW-ECHO (819.02.10) and the “Projectruimte” grant from Fundamenteel Onderzoek der Materie (FOM 12PR3033) to AJRH. We thank the Netherlands Proteomics Centre, embedded in the Netherlands Genomics Initiative, for financial support.

References.

1. Walzthoeni, T., A. Leitner, F. Stengel, and R. Aebersold. 2013. Mass spectrometry supported determination of protein complex structure. *Curr. Opin. Struct. Biol.* 23:252-260.
2. Gingras, A. -, M. Gstaiger, B. Raught, and R. Aebersold. 2007. Analysis of protein complexes using mass spectrometry. *Nature Reviews Molecular Cell Biology.* 8:645-654.
3. Gavin, A. -, P. Aloy, P. Grandi, R. Krause, M. Boesche, M. Marzioch, C. Rau, L. J. Jensen, S. Bastuck, B. Dümpelfeld, A. Edelmann, M. -. Heurtier, V. Hoffman, C. Hoefert, K. Klein, M. Hudak, A. -. Michon, M. Schelder, M. Schirle, M. Remor, T. Rudi, S. Hooper, A. Bauer, T. Bouwmeester, G. Casari, G. Drewes, G. Neubauer, J. M. Rick, B. Kuster, P. Bork, R. B. Russell, and G. Superti-Furga. 2006. Proteome survey reveals modularity of the yeast cell machinery. *Nature.* 440:631-636.
4. Gavin, A. -, M. Bösche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J. M. Rick, A. -. Michon, C. -. Cruciat, M. Remor, C. Höfert, M. Schelder, M. Brajenovic, H. Ruffner, A. Merino, K. Klein, M. Hudak, D. Dickson, T. Rudi, V. Gnau, A. Bauch, S. Bastuck, B. Huhse, C. Leutwein, M. -. Heurtier, R. R. Copley, A. Edelmann, E. Querfurth, V. Rybin, G. Drewes, M. Raida, T. Bouwmeester, P. Bork, B. Seraphin, B. Kuster, G. Neubauer, and G. Superti-Furga. 2002. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature.* 415:141-147.
5. Fenn, J. B., M. Mann, C. K. Meng, S. F. Wong, and C. M. Whitehouse. 1989. Electrospray ionization for mass spectrometry of large biomolecules. *Science.* 246:64-71.
6. Fenn, J. B. 2003. Electrospray wings for molecular elephants (Nobel lecture). *Angewandte Chemie - International Edition.* 42:3871-3894.
7. Dole, M., L. L. Mack, R. L. Hines, D. O. Chemistry, R. C. Mobley, L. D. Ferguson, and M. B. Alice. 1968. Molecular beams of macroions. *J. Chem. Phys.* 49:2240-2249.
8. Wilm, M. 2011. Principles of electrospray ionization. *Molecular and Cellular Proteomics.* 10:M1111.009407.
9. Wilm, M. S., and M. Mann. 1994. Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last? *International Journal of Mass Spectrometry and Ion Processes.* 136:167-180.
10. Fernandez De La Mora, J. 2000. Electrospray ionization of large multiply

charged species proceeds via Dole's charged residue mechanism. *Anal. Chim. Acta.* 406:93-104.

11. Breuker, K., and F. W. McLafferty. 2008. Stepwise evolution of protein native structure with electrospray into the gas phase, 10-12 to 102 s. *Proc. Natl. Acad. Sci. U. S. A.* 105:18145-18152.

12. Ruotolo, B. T., K. Giles, I. Campuzano, A. M. Sandercock, R. H. Bateman, and C. V. Robinson. 2005. Biochemistry: Evidence for macromolecular protein rings in the absence of bulk water. *Science.* 310:1658-1661.

13. Hall, Z., A. Politis, M. F. Bush, L. J. Smith, and C. V. Robinson. 2012. Charge-state dependent compaction and dissociation of protein complexes: Insights from ion mobility and molecular dynamics. *J. Am. Chem. Soc.* 134:3429-3438.

14. Wilm, M., and M. Mann. 1996. Analytical properties of the nanoelectrospray ion source. *Anal. Chem.* 68:1-8.

15. Hedges, J. B., S. Vahidi, X. Yue, and L. Konermann. 2013. Effects of Ammonium Bicarbonate on the Electrospray Mass Spectra of Proteins: Evidence for Bubble-Induced Unfolding. *Anal. Chem.* 85:6469-6476.

16. Dyachenko, A., R. Gruber, L. Shimon, A. Horovitz, and M. Sharon. 2013. Allosteric mechanisms can be distinguished using structural mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* 110:7235-7239.

17. Kaltashov, I. A., and A. Mohimen. 2005. Estimates of protein surface areas in solution by electrospray ionization mass spectrometry. *Anal. Chem.* 77:5370-5379.

18. Chowdhury, S. K., V. Katta, and B. T. Chait. 1990. Probing conformational changes in proteins by mass spectrometry. *J. Am. Chem. Soc.* 112:9012-9013.

19. Kaltashov, I. A., and R. R. Abzalimov. 2008. Do Ionic Charges in ESI MS Provide Useful Information on Macromolecular Structure? *J. Am. Soc. Mass Spectrom.* 19:1239-1246.

20. Catalina, M. I., R. H. H. Van Den Heuvel, E. Van Duijn, and A. J. R. Heck. 2005. Decharging of globular proteins and protein complexes in electrospray. *Chemistry - A European Journal.* 11:960-968.

21. Flick, T. G., and E. R. Williams. 2012. Supercharging with trivalent metal ions in native mass spectrometry. *J. Am. Soc. Mass Spectrom.* 23:1885-1895.

22. Lomeli, S. H., I. X. Peng, S. Yin, R. R. Ogorzalek Loo, and J. A. Loo. 2010. New Reagents for Increasing ESI Multiple Charging of Proteins and Protein Complexes. *J. Am. Soc. Mass Spectrom.* 21:127-131.

23. Lomeli, S. H., S. Yin, R. R. Ogorzalek Loo, and J. A. Loo. 2009. Increasing Charge While Preserving Noncovalent Protein Complexes for ESI-MS. *J. Am. Soc. Mass Spectrom.* 20:593-596.
24. Van Duijn, E., A. Barendregt, S. Synowsky, C. Versluis, and A. J. R. Heck. 2009. Chaperonin complexes monitored by ion mobility mass spectrometry. *J. Am. Chem. Soc.* 131:1452-1459.
25. Mann, M., C. K. Meng, and J. B. Fenn. 1989. Interpreting mass spectra of multiply charged ions. *Anal. Chem.* 61:1702-1708.
26. Morgner, N., and C. V. Robinson. 2012. Mass ign: An assignment strategy for maximizing information from the mass spectra of heterogeneous protein assemblies. *Anal. Chem.* 84:2939-2948.
27. Van Breukelen, B., A. Barendregt, A. J. R. Heck, and R. H. H. Van Den Heuvel. 2006. Resolving stoichiometries and oligomeric states of glutamate synthase protein complexes with curve fitting and simulation of electrospray mass spectra. *Rapid Communications in Mass Spectrometry.* 20:2490-2496.
28. Tseng, Y. -, C. Uetrecht, A. J. R. Heck, and W. -. Peng. 2011. Interpreting the charge state assignment in electrospray mass spectra of bioparticles. *Anal. Chem.* 83:1960-1968.
29. Sanglier, S., E. Leize, A. Dorsselaer, and F. Zal. 2003. Comparative ESI-MS study of ~2.2 MDa native hemocyanins from deep-sea and shore crabs: From protein oligomeric state to biotope. *J. Am. Soc. Mass Spectrom.* 14:419-429.
30. Chen, F., S. Gerber, K. Heuser, V. M. Korkhov, C. Lizak, S. Mireku, K. P. Locher, and R. Zenobi. 2013. High-mass matrix-assisted laser desorption ionization-mass spectrometry of integral membrane proteins and their complexes. *Anal. Chem.* 85:3483-3488.
31. Hoffmann, J., L. Aslimovska, C. Bamann, C. Glaubitz, E. Bamberg, and B. Brutschy. 2010. Studying the stoichiometries of membrane proteins by mass spectrometry: Microbial rhodopsins and a potassium ion channel. *Physical Chemistry Chemical Physics.* 12:3480-3485.
32. Morgner, N., J. Hoffmann, H. -. Barth, T. Meier, and B. Brutschy. 2008. LIL-BID-mass spectrometry applied to the mass analysis of RNA polymerase II and an F1Fo-ATP synthase. *International Journal of Mass Spectrometry.* 277:309-313.
33. Van Den Heuvel, R. H. H., and A. J. R. Heck. 2004. Native protein mass spectrometry: From intact oligomers to functional machineries. *Curr. Opin. Chem. Biol.* 8:519-526.

34. Sharon, M. 2013. Structural MS pulls its weight. *Science*. 340:1059-1060.
35. Sobott, F., H. Hernández, M. G. McCammon, M. A. Tito, and C. V. Robinson. 2002. A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Anal. Chem.* 74:1402-1407.
36. Kozlovski, V. I., L. J. Donald, V. M. Collado, V. Spicer, A. V. Loboda, I. V. Chernushevich, W. Ens, and K. G. Standing. 2011. A TOF mass spectrometer for the study of noncovalent complexes. *International Journal of Mass Spectrometry*. 308:118-125.
37. Van Den Heuvel, R. H. H., E. Van Duijn, H. Mazon, S. A. Synowsky, K. Lorenzen, C. Versluis, S. J. J. Brouns, D. Langridge, J. Van Der Oost, J. Hoyes, and A. J. R. Heck. 2006. Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. *Anal. Chem.* 78:7473-7483.
38. Tahallah, N., M. Pinkse, C. S. Maier, and A. J. R. Heck. 2001. The effect of the source pressure on the abundance of ions of noncovalent protein assemblies in an electrospray ionization orthogonal time-of-flight instrument. *Rapid Communications in Mass Spectrometry*. 15:596-601.
39. Schmidt, A., U. Bahr, and M. Karas. 2001. Influence of pressure in the first pumping stage on analyte desolvation and fragmentation in Nano-ESI MS. *Anal. Chem.* 73:6040-6046.
40. Chernushevich, I. V., and B. A. Thomson. 2004. Collisional Cooling of Large Ions in Electrospray Mass Spectrometry. *Anal. Chem.* 76:1754-1760.
41. Benesch, J. L. P. 2009. Collisional Activation of Protein Complexes: Picking Up the Pieces. *J. Am. Soc. Mass Spectrom.* 20:341-348.
42. Benesch, J. L. P., J. A. Aquilina, B. T. Ruotolo, F. Sobott, and C. V. Robinson. 2006. Tandem Mass Spectrometry Reveals the Quaternary Organization of Macromolecular Assemblies. *Chemistry and Biology*. 13:597-605.
43. Zhang, Z., Y. Zheng, H. Mazon, E. Milgrom, N. Kitagawa, E. Kish-Trier, A. J. R. Heck, P. M. Kane, and S. Wilkens. 2008. Structure of the yeast vacuolar AT-Pase. *J. Biol. Chem.* 283:35983-35995.
44. Zhou, M., A. M. Sandercock, C. S. Fraser, G. Ridlova, E. Stephens, M. R. Schenauer, T. Yokoi-Fong, D. Barsky, J. A. Leary, J. W. Hershey, J. A. Doudna, and C. V. Robinson. 2008. Mass spectrometry reveals modularity and a complete subunit interaction map of the eukaryotic translation factor eIF3. *Proc. Natl. Acad. Sci. U. S. A.* 105:18139-18144.
45. Lorenzen, K., A. Vannini, P. Cramer, and A. J. R. Heck. 2007. Structural Biology of RNA Polymerase III: Mass Spectrometry Elucidates Subcomplex Archi-

tecture. *Structure*. 15:1237-1245.

46. Hall, Z., A. Politis, and C. V. Robinson. 2012. Structural modeling of heteromeric protein complexes from disassembly pathways and ion mobility-mass spectrometry. *Structure*. 20:1596-1609.

47. Benesch, J. L. P., B. T. Ruotolo, F. Sobott, J. Wildgoose, A. Gilbert, R. Bateman, and C. V. Robinson. 2009. Quadrupole-time-of-flight mass spectrometer modified for higher-energy dissociation reduces protein assemblies to peptide fragments. *Anal. Chem.* 81:1270-1274.

48. Lee, J., and P. T. A. Reilly. 2011. Limitation of time-of-flight resolution in the ultra high mass range. *Anal. Chem.* 83:5831-5833.

49. Lössl, P., J. Snijder, and A. J. R. Heck. 2014. Boundaries of mass resolution in native mass spectrometry. *J. Am. Soc. Mass Spectrom.* 25:906-917.

50. Felitsyn, N., E. N. Kitova, and J. S. Klassen. 2001. Thermal decomposition of a gaseous multiprotein complex studied by blackbody infrared radiative dissociation. Investigating the origin of the asymmetric dissociation behavior. *Anal. Chem.* 73:4647-4661.

51. Kitova, E. N., D. R. Bundle, and J. S. Klassen. 2002. Thermal dissociation of protein - Oligosaccharide complexes in the gas phase: Mapping the intrinsic intermolecular interactions. *J. Am. Chem. Soc.* 124:5902-5913.

52. Zhang, H., W. Cui, J. Wen, R. E. Blankenship, and M. L. Gross. 2010. Native electrospray and electron-capture dissociation in fticr mass spectrometry provide top-down sequencing of a protein component in an intact protein assembly. *J. Am. Soc. Mass Spectrom.* 21:1966-1968.

53. Zhang, H., W. Cui, J. Wen, R. E. Blankenship, and M. L. Gross. 2011. Native electrospray and electron-capture dissociation FTICR mass spectrometry for top-down studies of protein assemblies. *Anal. Chem.* 83:5598-5606.

54. Yin, S., and J. A. Loo. 2011. Top-down mass spectrometry of supercharged native protein-ligand complexes. *International Journal of Mass Spectrometry*. 300:118-122.

55. Rose, R. J., E. Damoc, E. Denisov, A. Makarov, and A. J. R. Heck. 2012. High-sensitivity Orbitrap mass analysis of intact macromolecular assemblies. *Nature Methods*. 9:1084-1086.

56. Makarov, A. 2000. Electrostatic axially harmonic orbital trapping: A high-performance technique of mass analysis. *Anal. Chem.* 72:1156-1162.

57. Perry, R. H., R. G. Cooks, and R. J. Noll. 2008. Orbitrap mass spectrometry:

- Instrumentation, ion motion and applications. *Mass Spectrom. Rev.* 27:661-699.
58. Rosati, S., R. J. Rose, N. J. Thompson, E. Van Duijn, E. Damoc, E. Denisov, A. Makarov, and A. J. R. Heck. 2012. Exploring an orbitrap analyzer for the characterization of intact antibodies by native mass spectrometry. *Angewandte Chemie - International Edition.* 51:12992-12996.
59. Barrera, N. P., N. Di Bartolo, P. J. Booth, and C. V. Robinson. 2008. Micelles protect membrane complexes from solution to vacuum. *Science.* 321:243-246.
60. Zhou, M., N. Morgner, N. P. Barrera, A. Politis, S. C. Isaacson, D. Matk-Vinkovic, T. Murata, R. A. Bernal, D. Stock, and C. V. Robinson. 2011. Mass spectrometry of intact V-type atpases reveals bound lipids and the effects of nucleotide binding. *Science.* 334:380-385.
61. Laganowsky, A., E. Reading, J. T. S. Hopper, and C. V. Robinson. 2013. Mass spectrometry of intact membrane protein complexes. *Nature Protocols.* 8:639-651.
62. Uetrecht, C., C. Versluis, N. R. Watts, W. H. Roos, G. J. L. Wuite, P. T. Wingfield, A. C. Steven, and A. J. R. Heck. 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.* 105:9216-9220.
63. Shoemaker, G. K., E. Van Duijn, S. E. Crawford, C. Uetrecht, M. Baclayon, W. H. Roos, G. J. L. Wuite, M. K. Estes, B. V. V. Prasad, and A. J. R. Heck. 2010. Norwalk virus assembly and stability monitored by mass spectrometry. *Molecular and Cellular Proteomics.* 9:1742-1751.
64. Rose, R. J., D. Verger, T. Daviter, H. Remaut, E. Paci, G. Waksman, A. E. Ashcroft, and S. E. Radford. 2008. Unraveling the molecular basis of subunit specificity in P pilus assembly by mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* 105:12873-12878.
65. Uetrecht, C., I. M. Barbu, G. K. Shoemaker, E. Van Duijn, and A. J. R. Heck. 2011. Interrogating viral capsid assembly with ion mobility-mass spectrometry. *Nature Chemistry.* 3:126-132.
66. Smith, D. P., L. A. Woods, S. E. Radford, and A. E. Ashcroft. 2011. Structure and dynamics of oligomeric intermediates in β 2- microglobulin self-assembly. *Biophys. J.* 101:1238-1247.
67. Bernstein, S. L., N. F. Dupuis, N. D. Lazo, T. Wyttenbach, M. M. Condrón, G. Bitan, D. B. Teplow, J. -. Shea, B. T. Ruotolo, C. V. Robinson, and M. T. Bowers. 2009. Amyloid- β 2 protein oligomerization and the importance of tetramers and dodecamers in the aetiology of Alzheimer's disease. *Nature Chemistry.* 1:326-

331.

68. Jurchen, J. C., and E. R. Williams. 2003. Origin of asymmetric charge partitioning in the dissociation of gas-phase protein homodimers. *J. Am. Chem. Soc.* 125:2817-2826.

69. Zhou, M., S. Dagan, and V. H. Wysocki. 2012. Protein subunits released by surface collisions of noncovalent complexes: Nativelike compact structures revealed by ion mobility mass spectrometry. *Angewandte Chemie - International Edition.* 51:4336-4339.

70. Wysocki, V. H., C. M. Jones, A. S. Galhena, and A. E. Blackwell. 2008. Surface-Induced Dissociation Shows Potential to Be More Informative Than Collision-Induced Dissociation for Structural Studies of Large Systems. *J. Am. Soc. Mass Spectrom.* 19:903-913.

71. Blackwell, A. E., E. D. Dodds, V. Bandarian, and V. H. Wysocki. 2011. Revealing the quaternary structure of a heterogeneous noncovalent protein complex through surface-induced dissociation. *Anal. Chem.* 83:2862-2865.

72. Uetrecht, C., R. J. Rose, E. Van Duijn, K. Lorenzen, and A. J. R. Heck. 2010. Ion mobility mass spectrometry of proteins and protein assemblies. *Chem. Soc. Rev.* 39:1633-1655.

73. Bush, M. F., Z. Hall, K. Giles, J. Hoyes, C. V. Robinson, and B. T. Ruotolo. 2010. Collision cross sections of proteins and their complexes: A calibration framework and database for gas-phase structural biology. *Anal. Chem.* 82:9557-9565.

74. Clemmer, D. E., and M. F. Jarrold. 1997. Ion mobility measurements and their applications to clusters and biomolecules. *Journal of Mass Spectrometry.* 32:577-592.

75. Pringle, S. D., K. Giles, J. L. Wildgoose, J. P. Williams, S. E. Slade, K. Thalassinou, R. H. Bateman, M. T. Bowers, and J. H. Scrivens. 2007. An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *International Journal of Mass Spectrometry.* 261:1-12.

76. Ruotolo, B. T., J. L. P. Benesch, A. M. Sandercock, S. -. Hyung, and C. V. Robinson. 2008. Ion mobility-mass spectrometry analysis of large protein complexes. *Nature Protocols.* 3:1139-1152.

77. Politis, A., A. Y. Park, S. -. Hyung, D. Barsky, B. T. Ruotolo, and C. V. Robinson. 2010. Integrating ion mobility mass spectrometry with molecular modelling to determine the architecture of multiprotein complexes. *PLoS ONE.* 5:e12080.

78. Barrera, N. P., S. C. Isaacson, M. Zhou, V. N. Bavro, A. Welch, T. A. Schaedler, M. A. Seeger, R. N. Miguel, V. M. Korkhov, H. W. van Veen, H. Venter, A. R. Walmsley, C. G. Tate, and C. V. Robinson. 2009. Mass spectrometry of membrane transporters reveals subunit stoichiometry and interactions. *Nature Methods*. 6:585-587.
79. Schmidt, C., M. Zhou, H. Marriott, N. Morgner, A. Politis, and C. V. Robinson. 2013. Comparative cross-linking and mass spectrometry of an intact F-type ATPase suggest a role for phosphorylation. *Nature Communications*. 4:1985.
80. Van Duijn, E., I. M. Barbu, A. Barendregt, M. M. Jore, B. Wiedenheft, M. Lundgren, E. R. Westra, S. J. J. Brouns, J. A. Doudna, J. Van Der Oost, and A. J. R. Heck. 2012. Native tandem and ion mobility mass spectrometry highlight structural and modular similarities in clustered-regularly-interspaced shot-palindromic-repeats (CRISPR)-associated protein complexes from *Escherichia coli* and *Pseudomonas aeruginosa*. *Molecular and Cellular Proteomics*. 11:1430-1441.
81. Wiedenheft, B., E. Van Duijn, J. Bultema, S. Waghmare, K. Zhou, A. Barendregt, W. Westphal, A. Heck, E. Boekema, M. Dickman, and J. A. Doudna. 2011. RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc. Natl. Acad. Sci. U. S. A.* 108:10092-10097.
82. Jore, M. M., M. Lundgren, E. Van Duijn, J. B. Bultema, E. R. Westra, S. P. Waghmare, B. Wiedenheft, U. Pul, R. Wurm, R. Wagner, M. R. Beijer, A. Barendregt, K. Zhou, A. P. L. Snijders, M. J. Dickman, J. A. Doudna, E. J. Boekema, A. J. R. Heck, J. Van Der Oost, and S. J. J. Brouns. 2011. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nature Structural and Molecular Biology*. 18:529-536.
83. Hall, Z., H. Hernández, J. Marsh, S. Teichmann, and C. Robinson. 2013. The Role of Salt Bridges, Charge Density, and Subunit Flexibility in Determining Disassembly Routes of Protein Complexes. *Structure*. 21:1325-1337.
84. Marsh, J. A., H. Hernández, Z. Hall, S. E. Ahnert, T. Perica, C. V. Robinson, and S. A. Teichmann. 2013. Protein complexes are under evolutionary selection to assemble via ordered pathways. *Cell*. 153:461-470.
85. Kitova, E. N., A. El-Hawiet, P. D. Schmier, and J. S. Klassen. 2012. Reliable determinations of protein-ligand interactions by direct ESI-MS measurements. Are we there yet? *J. Am. Soc. Mass Spectrom.* 23:431-441.
86. Boeri Erba, E., K. Barylyuk, Y. Yang, and R. Zenobi. 2011. Quantifying protein-protein interactions within noncovalent complexes using electrospray ionization mass spectrometry. *Anal. Chem.* 83:9251-9259.

87. El-Hawiet, A., E. N. Kitova, D. Arutyunov, D. J. Simpson, C. M. Szymanski, and J. S. Klassen. 2012. Quantifying ligand binding to large protein complexes using electrospray ionization mass spectrometry. *Anal. Chem.* 84:3867-3870.
88. Rose, R. J., A. F. Labrijn, E. T. J. Van Den Bremer, S. Loverix, I. Lasters, P. H. C. Van Berkel, J. G. J. Van De Winkel, J. Schuurman, P. W. H. I. Parren, and A. J. R. Heck. 2011. Quantitative analysis of the interaction strength and dynamics of human IgG4 half molecules by native mass spectrometry. *Structure.* 19:1274-1282.
89. Deroo, S., S. - Hyung, J. Marcoux, Y. Gordiyenko, R. K. Koripella, S. Sanyal, and C. V. Robinson. 2012. Mechanism and rates of exchange of L7/L12 between ribosomes and the effects of binding EF-G. *ACS Chemical Biology.* 7:1120-1127.
90. Sobott, F., J. L. P. Benesch, E. Vierling, and C. V. Robinson. 2002. Subunit exchange of multimeric protein complexes: Real-time monitoring of subunit exchange between small heat shock proteins by using electrospray mass spectrometry. *J. Biol. Chem.* 277:38921-38929.
91. Uetrecht, C., N. R. Watts, S. J. Stahl, P. T. Wingfield, A. C. Steven, and A. J. R. Heck. 2010. Subunit exchange rates in Hepatitis B virus capsids are geometry- and temperature-dependent. *Physical Chemistry Chemical Physics.* 12:13368-13371.
92. Bruce, J. E., X. Cheng, R. Bakhtiar, Q. Wu, S. A. Hofstadler, G. A. Anderson, and R. D. Smith. 1994. Trapping, detection, and mass measurement of individual ions in a Fourier transform ion cyclotron resonance mass spectrometer. *J. Am. Chem. Soc.* 116:7839-7847.
93. Cheng, X., R. Bakhtiar, S. Van Orden, and R. D. Smith. 1994. Charge-state shifting of individual multiply-charged ions of bovine albumin dimer and molecular weight determination using an individual-ion approach. *Anal. Chem.* 66:2084-2087.
94. Cheng, X., D. G. Camp II, Q. Wu, R. Bakhtiar, D. L. Springer, B. J. Morris, J. E. Bruce, G. A. Anderson, C. G. Edmonds, and R. D. Smith. 1996. Molecular weight determination of plasmid DNA using electrospray ionization mass spectrometry. *Nucleic Acids Res.* 24:2183-2189.
95. Makarov, A., and E. Denisov. 2009. Dynamics of Ions of Intact Proteins in the Orbitrap Mass Analyzer. *J. Am. Soc. Mass Spectrom.* 20:1486-1495.
96. Fuerstenau, S. D., and W. H. Benner. 1995. Molecular weight determination of megadalton DNA electrospray ions using charge detection time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry.* 9:1528-1538.

97. Fuerstenau, S. D., W. H. Benner, J. J. Thomas, C. Brugidou, B. Bothner, and G. Siuzdak. 2001. Mass spectrometry of an intact virus. *Angewandte Chemie - International Edition*. 40:542-544.
98. Gamero-Castão, M. 2007. Induction charge detector with multiple sensing stages. *Rev. Sci. Instrum.* 78:043301.
99. Smith, J. W., E. E. Siegel, J. T. Maze, and M. F. Jarrold. 2011. Image charge detection mass spectrometry: Pushing the envelope with sensitivity and accuracy. *Anal. Chem.* 83:950-956.
100. Contino, N. C., E. E. Pierson, D. Z. Keifer, and M. F. Jarrold. 2013. Charge detection mass spectrometry with resolved charge states. *J. Am. Soc. Mass Spectrom.* 24:101-108.
101. Keifer, D. Z., E. E. Pierson, J. A. Hogan, G. J. Bedwell, P. E. Prevelige, and M. F. Jarrold. 2014. Charge detection mass spectrometry of bacteriophage P22 procapsid distributions above 20 MDa. *Rapid Communications in Mass Spectrometry*. 28:483-488.
102. Pierson, E. E., D. Z. Keifer, L. Selzer, L. S. Lee, N. C. Contino, J. C. -. Wang, A. Zlotnick, and M. F. Jarrold. 2014. Detection of Late Intermediates in Virus Capsid Assembly by Charge Detection Mass Spectrometry. *J. Am. Chem. Soc.* 136:3536-3541.
103. Chen, G. Y., T. Thundat, E. A. Wachter, and R. J. Warmack. 1995. Adsorption-induced surface stress and its effects on resonance frequency of microcantilevers. *J. Appl. Phys.* 77:3618-3622.
104. Chaste, J., A. Eichler, J. Moser, G. Ceballos, R. Rurali, and A. Bachtold. 2012. A nanomechanical mass sensor with yoctogram resolution. *Nature Nanotechnology*. 7:301-304.
105. Jiang, C., B. Chen, J. -. Li, and K. -. Zhu. 2011. Mass spectrometry based on a coupled cooperpair box and nanomechanical resonator system. *Nanoscale Research Letters*. 6:1-8.
106. Hanay, M. S., S. Kelber, A. K. Naik, D. Chi, S. Hentz, E. C. Bullard, E. Colinet, L. Duraffourg, and M. L. Roukes. 2012. Single-protein nanomechanical mass spectrometry in real time. *Nature Nanotechnology*. 7:602-608.
107. Naik, A. K., M. S. Hanay, W. K. Hiebert, X. L. Feng, and M. L. Roukes. 2009. Towards single-molecule nanomechanical mass spectrometry. *Nature Nanotechnology*. 4:445-450.
108. Gil-Santos, E., D. Ramos, J. Martínez, M. Fernández-Regúlez, R. García, Á. San Paulo, M. Calleja, and J. Tamayo. 2010. Nanomechanical mass sensing and

stiffness spectrometry based on two-dimensional vibrations of resonant nanowires. *Nature Nanotechnology*. 5:641-645.

109. Kaufman, S. L. 1998. Analysis of biomolecules using electrospray and nanoparticle methods: The gas-phase electrophoretic mobility molecular analyzer (GEMMA). *J. Aerosol Sci.* 29:537-552.

110. Bacher, G., W. W. Szymanski, S. L. Kaufman, P. Zllner, D. Blaas, and G. Allmaier. 2001. Charge-reduced nano electrospray ionization combined with differential mobility analysis of peptides, proteins, glycoproteins, noncovalent protein complexes and viruses. *Journal of Mass Spectrometry*. 36:1038-1052.

111. Loo, J. A., B. Berhane, C. S. Kaddis, K. M. Wooding, Y. Xie, S. L. Kaufman, and I. V. Chernushevich. 2005. Electrospray ionization mass spectrometry and ion mobility analysis of the 20S proteasome complex. *J. Am. Soc. Mass Spectrom.* 16:998-1008.

112. Kemptner, J., M. Marchetti Deschmann, J. Siekmann, P. L. Turecek, H. P. Schwarz, and G. Allmaier. 2010. GEMMA and MALDI-TOF MS of reactive PEGs for pharmaceutical applications. *J. Pharm. Biomed. Anal.* 52:432-437.

113. Kaddis, C. S., S. H. Lomeli, S. Yin, B. Berhane, M. I. Apostol, V. A. Kickhoefer, L. H. Rome, and J. A. Loo. 2007. Sizing Large Proteins and Protein Complexes by Electrospray Ionization Mass Spectrometry and Ion Mobility. *J. Am. Soc. Mass Spectrom.* 18:1206-1216.

114. Allmaier, G., C. Laschober, and W. W. Szymanski. 2008. Nano ES GEMMA and PDMA, New Tools for the Analysis of Nanobioparticles-Protein Complexes, Lipoparticles, and Viruses. *J. Am. Soc. Mass Spectrom.* 19:1062-1068.

115. Chapman, H. N., P. Fromme, A. Barty, T. A. White, R. A. Kirian, A. Aquila, M. S. Hunter, J. Schulz, D. P. Deponete, U. Weierstall, R. B. Doak, F. R. N. C. Maia, A. V. Martin, I. Schlichting, L. Lomb, N. Coppola, R. L. Shoeman, S. W. Epp, R. Hartmann, D. Rolles, A. Rudenko, L. Foucar, N. Kimmel, G. Weidenspointner, P. Holl, M. Liang, M. Barthelmess, C. Caleman, S. Boutet, M. J. Bogan, J. Krzywinski, C. Bostedt, S. Bajt, L. Gumprecht, B. Rudek, B. Erk, C. Schmidt, A. Hömke, C. Reich, D. Pietschner, L. Ströder, G. Hauser, H. Gorke, J. Ullrich, S. Herrmann, G. Schaller, F. Schopper, H. Soltau, K. -. Kühnel, M. Messerschmidt, J. D. Bozek, S. P. Hau-Riege, M. Frank, C. Y. Hampton, R. G. Sierra, D. Starodub, G. J. Williams, J. Hajdu, N. Timneanu, M. M. Seibert, J. Andreasson, A. Rocker, O. Jönsson, M. Svenda, S. Stern, K. Nass, R. Andritschke, C. -. Schröter, F. Krasniqi, M. Bott, K. E. Schmidt, X. Wang, I. Grotjohann, J. M. Holton, T. R. M. Barends, R. Neutze, S. Marchesini, R. Fromme, S. Schorb, D. Rupp, M. Adolph, T. Gorkhover, I. Andersson, H. Hirsemann, G. Potdevin, H. Graafsma, B. Nilsson, and

J. C. H. Spence. 2011. Femtosecond X-ray protein nanocrystallography. *Nature*. 470:73-78.

116. Benesch, J. L. P., B. T. Ruotolo, D. A. Simmons, N. P. Barrera, N. Morgner, L. Wang, H. R. Saibil, and C. V. Robinson. 2010. Separating and visualising protein assemblies by means of preparative mass spectrometry and microscopy. *J. Struct. Biol.* 172:161-168.

Chapter 2:

Challenging the current upper mass limit of native MS.

Joost Snijder^{1,2}, Rebecca J. Burnley^{1,2}, David Veesler³, Jack E. Johnson³, Albert J.R. Heck^{1,2}

¹ Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH, Utrecht, the Netherlands.

² Netherlands Proteomics Centre, Padualaan 8, 3584 CH, Utrecht, The Netherlands.

³ Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA 92037, USA.

Based on:

- Snijder, J., Rose, R. J., Veesler, D., Johnson, J. E. & Heck, A. J. R. Studying 18 MDa virus assemblies with native mass spectrometry. *Angewandte Chemie - International Edition* 52, 4020-4023 (2013).
- Veesler, D., *et al.* Architecture of a dsDNA viral capsid in complex with its maturation protease. *Structure* 22, 230-237 (2014).

Abstract.

Native mass spectrometry has been developed to study intact megadalton virus capsids in recent years. Here, we explore the mass limits of native MS by studying 420 subunit $T=7$ icosahedral virus capsids from the bacteriophage HK97. We show that these 18 megadalton assemblies can be analyzed on a modified QTOF instrument, commonly used in native MS studies. By analyzing the obtained peak shapes, and considering factors such as instrument resolution and sample heterogeneity, we reach the conclusion that these 18 megadalton virus particles are among the biggest assemblies that can be studied by native MS on the current generation of instruments. Having established native MS as a possible tool to characterize capsids above 10 megadalton, we analyze and compare capsids from three different stages of the maturation pathway of HK97. Our data provide new estimates of the number of viral protease molecules that are encapsulated in HK97 procapsids and allow us to assess the efficiency of the maturation cleavage by the viral protease on the level of the intact mature capsid.

Native MS has previously been applied to the study of intact viruses and virus capsids. Apart from their role in viral pathogenesis, capsids are interesting because of their potential application in nanotechnology and medicine: as nano-containers, -reactors and assembly scaffolds (1). Native MS enabled the study of intact virus capsids, quantifying (non-native) cargo encapsulation and monitoring capsid assembly (2-5). These previous studies focused on relatively small capsids with Caspar-Klug triangulation numbers of $T=1, 3$ and 4 , ranging up to ~ 10 MDa in mass. Although the 10 MDa norovirus capsids could be detected with MS, their precise mass could not be determined due to a lack of resolution of individual charge states (3). Charge-resolved spectra were limited to particles up to ~ 5 MDa, which was hypothesized to be the upper size limit achievable.

Most current native MS studies are carried out on quadrupole time-of-flight (QTOF) instruments, operating at higher pressures to facilitate transmission of large ions (6-10). The QTOF is typically the setup of choice because of the superior sensitivity of the TOF analyzer at high mass-to-charge ratio. Although there is no theoretical upper mass limit for a TOF analyzer, achieving good quality spectra of assemblies that are several megadaltons in mass has proven not to be trivial. Here we demonstrate the use of a modified QTOF for the analysis of intact 18 MDa capsids from the bacteriophage HK97. HK97 is an important model system for studying bacteriophage assembly and maturation (11). *In vitro*, the $T=7$ capsid assembles from a mixture of pentameric and hexameric capsomers (made of the capsid protein gp5) and the viral protease gp4 to form a first icosahedral intermediate termed prohead-1. Upon assembly, the encapsidated protease molecules are activated and prime the maturation process that is characterized by a swelling transition, yielding the mature capsid (Figure 1a) (12). Overexpression of the capsid protein in the absence of the viral protease results in the formation of a virtually identical icosahedral particle stalled at the prohead-1 stage (prohead-1^{gp5}) (13). We report here the analysis by native MS of prohead-1^{gp5} as well as of the free capsomers (resulting from high pH induced disassembly), providing for the first time mass spectra with charge-state resolution for a particle of such high mass. Having established native MS as a feasible approach to study such large protein assemblies, we analyzed prohead-1 particles loaded with gp4 as well as mature head-2 particles, to estimate the copy number of encapsulated proteases and the efficiency of cleavage of the scaffold domain of gp5, respectively. These analyses provide new insights into the maturation pathway of bacteriophage HK97 and to the factors that ultimately limit native MS analysis of megadalton protein complexes.

Results.

Figure 1 shows mass spectra of free capsomers and the prohead-1^{gp5} particle. The fraction of pentons in the capsomer mixture, judged by MS signal is 0.25 ± 0.04 (SD, $n=3$) (Figure 1b), which is slightly higher than expected as the capsid consists of 12 pentons and 60 hexons. Partially disassembled complexes and free capsid protein (gp5) monomers also appear in the spectrum, when the capsomer mixture is analyzed under alkaline conditions, allowing us to determine an accurate monomer mass of 42151 ± 1 Da (Supplementary Figure S1). From the expressed capsid protein sequence, a theoretical mass of 42243 Da was calculated, indicating that the N-terminal methionine is cleaved and that the resulting N-terminus is acetylated (yielding a theoretical mass of 42154 Da). The masses determined for the pentameric and hexameric capsomers are 210807 ± 9 Da and 252973 ± 9 Da, respectively, corresponding to 5.001 and 6.002 times the experimentally determined monomer mass. Prohead-1^{gp5} consists of 420 gp5 copies and its expected mass is thus close to 18 MDa. As illustrated in Figure 1c, the particle could be detected in our mass spectrum despite its massive size. Transmission of such large particles could be improved two-fold by tuning the ‘Steering’ parameter of the mass spectrometer, which controls the angle of the ion beam into the pusher region of the TOF analyzer (Supplementary Figure S2).

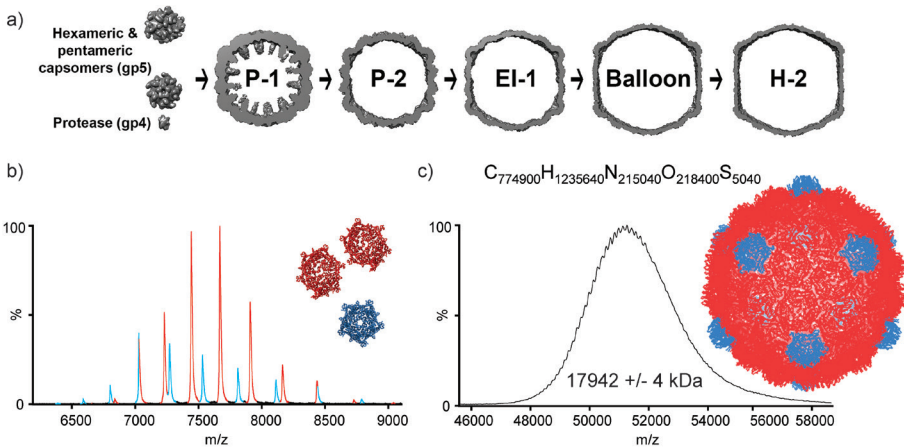


Figure 1. Monitoring assembly of HK97 capsids with native ESI-MS. a) Assembly and maturation pathway of HK97. b) Free capsomers with penton signal in blue and hexon signal in red. c) Intact prohead-1 particle. A well-resolved series of charge states is observed allowing the accurate mass calculation.

Poor desolvation of large protein complexes often results in poor spectral quality. To improve desolvation and resolve individual charge states, we applied the maximum collision voltage achievable on our instrument (400 V) (14). In ammonium acetate buffer of a typically used ionic strength (50-200 mM) this was still not sufficient to achieve resolution of individual charge states. Combining a decreased buffer ionic strength (12.5 mM) with application of the maximum collision voltage of 400 V using xenon as collision gas, a series of charge states clearly became visible (Figure 1c). Under these conditions, we achieved a base peak intensity of ~10 counts per 2 s scan, requiring ~50 minutes of acquisition time for one spectrum.

As originally demonstrated by Fenn and Mann, charge states can be assigned from the mass-to-charge ratios of adjacent peaks in order to calculate a mass (15). This is trivial for smaller protein complexes, but the charge state assignment is increasingly ambiguous with larger, poorly desolvated complexes. The strategy employed here for charge state assignment was to systematically test a wide range of charge state assignments and look for an assignment that results in a minimum variation in calculated mass within the same spectrum (see Supplementary Figure S3). This is essentially the strategy that is employed in the recently described LeastMass software (16). This procedure was repeated for three independent measurements and the charge state assignments were checked for consistency. This charge state assignment strategy yielded an average mass of 17942 ± 4 kDa across the triplicate experiments. This is a deviation of +1.3% compared to the prohead-1^{SP5} theoretical mass (17742 kDa) and we attribute this discrepancy to incomplete desolvation.

This result demonstrates that native MS can be used to study larger systems such as $T=7$ capsids in the case of HK97. This opens up new avenues for studying a wider variety of capsid structures in their role as pathogens and nanotechnological platforms. This result also confirms that native MS on a QTOF instrument has a high upper mass limit for the analysis of protein complexes. As the individual charge states are just resolved in the spectrum of the prohead-1^{SP5} particle, we now estimate that the upper mass limit is at ~20 MDa. To further illustrate that desolvation, and not the instrument hardware, is the limiting factor for protein complex analysis with native MS, we calculated the theoretical peak width of capsomers and prohead-1^{SP5}, taking into account the inherent isotope distribution of the assembly and the resolution of the mass analyzer.

The inherent width of a charge state in the mass-to-charge ratio dimension is proportional to the width of the isotope distribution and inversely proportional

to the number of charges. The isotope distribution is binomial and hence scales to the square root of mass. The number of charges is also expected to scale to the square root of mass, assuming the charged residue model of electrospray ionization (17, 18). However, here we show that it scales slightly stronger empirically ($\text{mass}^{0.55}$), from which it follows that peaks get narrower with increasing size of the complex (see Supplementary Figure S4). The instrumental resolution of our QTOF instrument was empirically determined from the width of singly charged, CsI-clusters that do not contain natural isotope variants (see Supplementary Figure S5). By nanoESI, CsI clusters were generated and the determined resolution is approximately constant across the m/z -range at a value of around 7000 at full-width at half-maximum (FWHM) on the QTOF employed here.

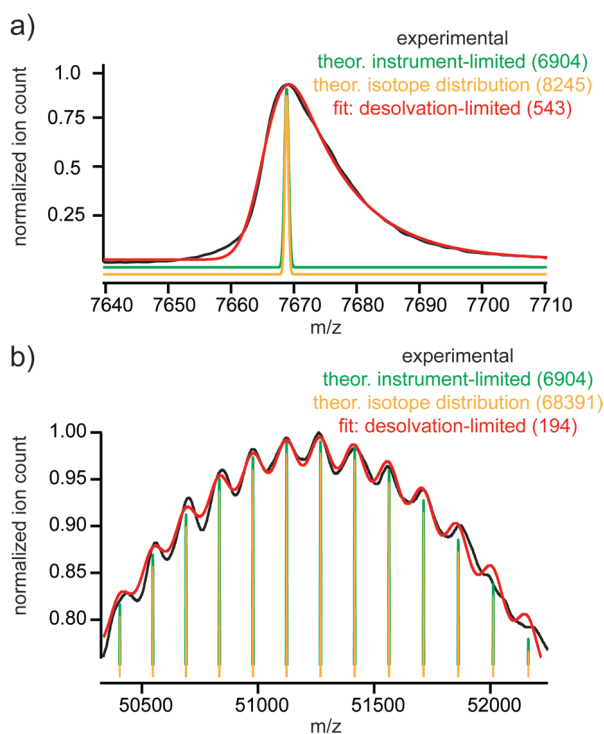


Figure 2. Incomplete desolvation is the limiting factor in attaining narrow peak width in native MS of protein complexes. Experimental peaks of hexons (a) and Prohead-1^{gp5} (b) are compared to theoretical peaks with widths defined by instrument resolution and the natural isotope distribution. The ‘incomplete desolvation limited’ curves are simulated to fit the experimental data (m/z positions of the peaks are fixed and defined by the experimentally determined mass, the width is fitted to the signal). The numbers in parentheses indicate the effective resolution (FWHM).

We compared the experimentally acquired spectra of free capsomers and prohead-1^{SP5} with the theoretical peak-widths defined by either the natural isotope distribution or the instrument resolution (Figure 2.). It is obvious from this comparison that both the capsomers and prohead-1^{SP5} particle have much broader peaks than predicted based on instrument resolution or natural isotope distribution. The peaks of the capsomers are also clearly asymmetric, a feature that likely originates from superposed binomial distributions of buffer and salt adducts. The experimental resolution, expressed as their mass-to-charge ratio divided by the peak width (FWHM), for capsomers and prohead-1^{SP5} are approximately 500 and 200, respectively. This difference between the two assemblies demonstrates how desolvation becomes less efficient with increasing size of the complex.

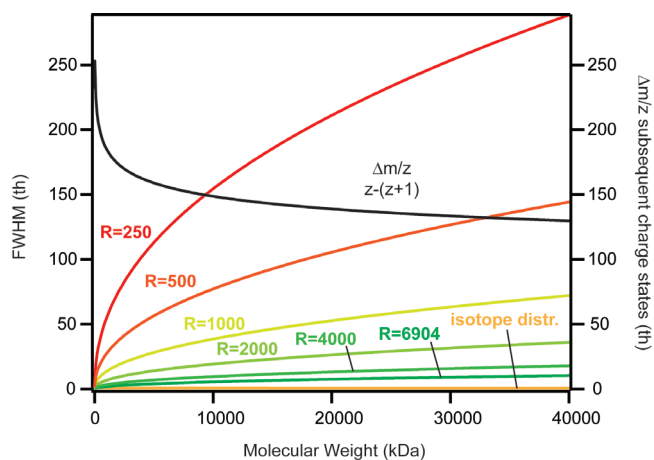


Figure 3. Theoretical upper mass limit of native MS studies of protein complexes as a function of effective resolution. Shown are the separation between subsequent charge states (black line, right-hand y-axis) and the calculated peak width as a function of molecular weight (colored lines, left-hand y-axis).

These investigations indicate that native MS would benefit from more efficient desolvation. It would greatly increase the effective resolving power of the instrument as well as eliminate any practical upper mass limit other than transmission and detection efficiency (we argue that the superior resolving power of the Orbitrap Exactive platform that was recently adapted for native MS is also primarily attributable to more efficient desolvation (19)). This is illustrated in Figure 3, where we calculated the expected separation between consecutive charge states (as determined from the empirical charging behavior of protein complexes) and compare this with peak width at increasing effective resolution (FWHM). The intersection between these curves is essentially the point at which consecutive

charge states are separated at half-maximum intensity. The peak width as determined from the isotope distribution is always smaller than the separation between consecutive charge states. From this it follows that nanoESI does not impose any inherent upper mass limit on protein assemblies that can be analyzed with charge-state resolution using MS. The instrument resolution is not a limiting factor either. Even at the moderate resolution of 7000 exhibited by the modified QTOF used in the experiments described above, the point where peak separation and peak width start to overlap is around ~ 40 gigadalton. Considering the decrease in effective resolution caused by incomplete desolvation of the specimen, the limit drops from gigadaltons to megadaltons. Our results demonstrate that native MS can be used to study protein complexes up to ~ 20 MDa with charge state resolution on a QTOF instrument.

Having established native MS as a suitable tool to study such large protein complexes as the HK97 procapsid, we proceeded to analyze particles representing other stages from the HK97 maturation pathway. The stoichiometry defining the number of protease molecules incorporated into the procapsid of dsDNA viruses has been elusive. Densitometry analysis of SDS protein gels of HK97 prohead-1 with a non-functional protease indicated ~ 50 protease molecules per particle (20). We tried to refine this estimate here by native MS by analyzing the Prohead-1 capsid, loaded with an inactive mutant form of gp4. As described above, the spectrum for the protease-free procapsid allowed resolution of individual charge states and yielded a mass estimate of 17.942 ± 0.004 MDa (Figure 4a, top). Analysis of the protease-containing particle resulted in an unresolved spectrum at a higher m/z than the empty prohead-1 (Figure 4a, middle). In this case a mass estimate of the particle requires assumptions about the charging behavior of the procapsid particles in electrospray ionization. A plot of mass vs. charge that we obtained in the lab shows a power-law function with strong correlation ($R^2=0.9965$, see Figure 4b). The average residual from the obtained fit is 2 ± 3 charges, with no clear mass dependence, indicating that we can reasonably estimate z within approximately 5 units. The m/z position of the particle is determined at 53190 ± 180 (average \pm standard deviation, $n=5$). We can estimate the mass as: $m=[m/z]*z$, with an error that is estimated as the sum of the relative errors on $[m/z]$ and z . We thereby estimate the mass of the protease-filled prohead-1 particle at approximately 21.4 ± 0.34 MDa. This corresponds to 144 ± 14 copies of the protease. It should be noted that this estimate only holds under the assumption of normal charging. For instance, as charge is most likely distributed on the exterior of a droplet in the final stages of electrospray ionization, the internal surface of the capsid might not be equally accessible to charge.

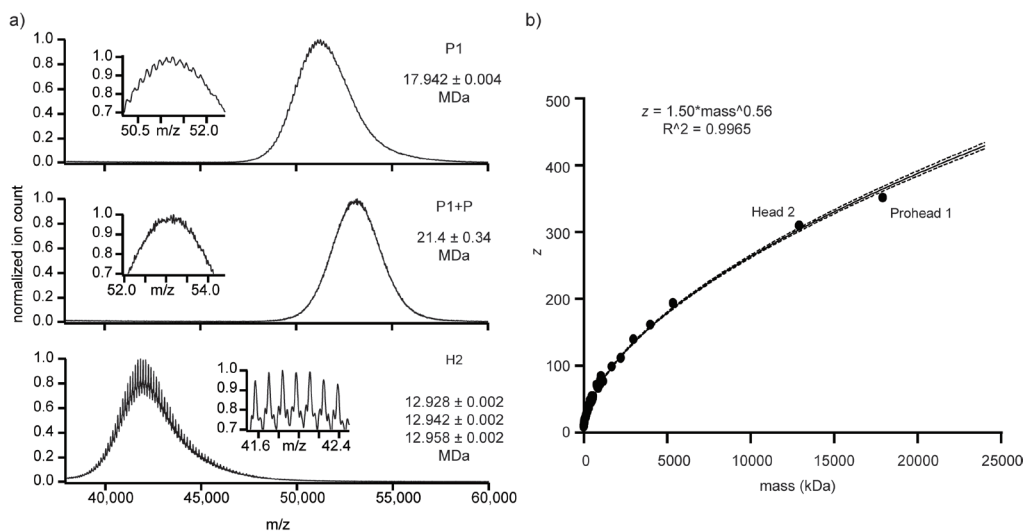


Figure 4. Native mass-spectrometry quantification of protease packaging and cleavage efficiency. a) Native MS spectra of HK97 capsids. (top) The mass spectrum obtained for the protease-free procapsid indicates a mass of 17.9 ± 0.004 MDa while the theoretical one is 17.7 MDa (mass accuracy of 1%). (middle) The mass spectrum obtained for the protease-containing procapsid is noisier and yields a mass estimate of 21.4 ± 0.34 MDa. (bottom) Mass spectrum of Head-2 reveals three distributions with a mass of approximately 12.9 MDa. The masses are consistent with a highly efficient maturation cleavage but clearly reveal some residual material in the Head-2 particles. (insets) Zoom-in of the indicated m/z region to reveal the fine structures on the signal. b) Charging behavior of large protein complexes in nanoESI native MS experiments. Each dot represents one unique protein complex measures under similar conditions. The black line represents the fitted power function indicated on the graph, the dashed black lines indicate the 95% confidence interval.

To confirm this estimate of the number of packaged protease molecules, we also performed a label-free quantitative proteomics experiment, in which the protease-containing particle was digested and analyzed by LC-MS/MS. We used the number of peptide-spectrum-matches (PSM) as a quantitative measure for protein abundance. Some issues may arise in PSM based quantitation, especially of low abundant proteins that are identified with a low number of peptides in complex background matrix, which results in high stochastic fluctuations in the obtained PSM's. However, the prohead-1 with protease sample is a very simple mixture, generating a total of ~ 70 identified peptides. Therefore, no issues are expected based on sample complexity. In addition, the total number of PSM's is based on 50 and 20 unique peptides for gp5 and gp4, respectively. Both proteins generate a comparable set of tryptic peptides, given the distribution of cleavable sites in the sequence, such that little bias in observable peptides is expected. Based on replicate injections from the same digest we arrive at an estimate of 104

± 18 copies (average \pm standard deviation), indicating that instrumental reproducibility is $\sim 20\%$. From six independent digest from two independent purifications, we arrive at an estimate of 116 ± 23 copies (average \pm standard deviation), which agrees reasonably well with the native MS estimate.

It should be noted that the lack of resolution in the native MS experiment is a clear indication of heterogeneity in the number of packaged protease molecules. The varying amount of protease in the particles could also lead to a heterogeneous population of mature particles following proteolysis. We analyzed the mature head-2 particle by native MS and observed a large shift in mass to 12.9 MDa (Figure 4a, bottom). Closer inspection of the signal revealed that at least three overlapping distributions could be distinguished. Their mass deviates by approximately 10, 24 and 39 kDa from a theoretically fully cleaved head-2 particle, indicating that at least 99.5% of all delta-domains were cleaved. Native MS thus revealed how maturation proceeds with extraordinarily high efficiency in HK97, but also revealed a small amount of residual material that is probably not detectable by conventional methods. The residual masses may correspond to one copy of the protease molecule that could not exit the particle (24 kDa), or perhaps to partly cleaved fragments of the delta-domain (10 kDa).

Discussion.

We have demonstrated here for the first time that protein complexes as big as 18 megadalton can be analyzed by native MS on a modified QTOF platform with resolved charge states. The spectra of the empty prohead-1 and head-2 particles are to our knowledge the first virus capsids with such high triangulation numbers ($T=7$) analyzed by native MS. These results thereby demonstrate that the native MS approach to uncovering the quaternary structure of protein complexes applies to a bigger range of structures than ever before. Detailed analysis of the acquired spectra of HK97 capsids has revealed which are the likely limiting factors for the accessible mass range of native MS. In particular, analysis of the experimentally achieved peak widths indicated that poor desolvation, or equivalently, the formation of solvent adducts with the protein complex ultimately limits whether sufficient resolution can be achieved to analyze protein complexes of increasing mass.

The current generation of native MS instruments may be improved for analysis of these very large complexes in any number of ways. As our experiments revealed

poor desolvation to be a main limiting factor, trying to improve desolvation is our main concern. The spectra of HK97 shown here have all been acquired at maximum collision voltages. No significant degree of collision induced dissociation is observed at this point, suggesting that further extending the applicable range of collision voltages might promote further desolvation, thereby increasing effective resolving power and perhaps further extending the accessible mass range. Moreover, extending the capability of ion activation would ultimately also allow the study of the gas-phase dissociation behavior of these ions. On the same note, the current generation of quadrupole mass filters that the instruments are equipped with are also not suited for selection of ions with m/z above 32000. Further extending the accessible mass range of the quadrupole would therefore also open up the possibility of doing tandem MS on these large protein assemblies.

The need for improved resolving power is illustrated by the obtained results for the protease-filled particle. According to our estimates, the mass of the particle increases by less than 10% compared to the empty prohead-1 capsid. Yet, no series of charge states can be distinguished on the signal, indicating that the spectrum is overly crowded with overlapping charge state distributions corresponding to the particle with a variable number of gp4. This does not only mean that we are currently unable to resolve such heterogeneous ensembles of closely spaced masses (note that one additional copy of gp4 represent a mere 0.01% increase in mass), but also that the average mass of the ensemble can only be guessed from the m/z position alone (assuming a typical charging behavior). A more confident and absolute estimate of the average number of encapsulated gp4, and an estimate of the spread in the ensemble, would thus require higher resolving power – better desolvation of the ions. The need for higher resolution might be circumvented in single particle charge detection MS, as was recently demonstrated for a very similar problem: estimating the copy number of encapsulated scaffold protein in the capsid of bacteriophage P22 (21). Though this method still suffers from a relatively low precision in the overall mass determination, the independent measure of charge that it provides gets rid of the need for resolved charge states in the m/z measurement. Note, however, that in our native MS spectrum of the slightly smaller head-2 capsid, very closely spaced masses are clearly resolved. This indicates that we are operating in a transition region with regards to the need for efficient desolvation and that sufficient improvements to the instrument's performance in this regard should be feasible. From the results obtained in the current study, which represent analyses of some of the largest known structures by native MS, it therefore seems that further improving resolving power and extending the mass range are well within reach.

Methods.

Native MS.

HK97 prohead-1^{gp5} particles were purified as described previously with minor modifications (13, 22). Free capsomers were obtained by incubating prohead-1^{gp5} particles overnight in 100 mM CHES (pH 9.5) 2 M KCl. Samples were prepared for mass spectrometry by five rounds of dilution and concentration to exchange the buffer to ammonium acetate, using Vivaspin 500 10K MWCO centrifugal filter units. Free capsomers were analyzed in 50 mM pH 7.4 ammonium acetate buffer, partially denatured capsomers in 50 mM pH 9.5 and intact prohead-1^{gp5} in 12.5 mM pH 7.4. Aliquots of 1-2 μ L were loaded into gold-coated borosilicate capillaries, prepared in-house, for nano-electrospray ionization.

Samples were analyzed on a modified QTOF II instrument optimized for high-mass protein analysis (7-9). The instrument operates at elevated pressure in the source region and hexapole ion guide, which results in collisional cooling of the ions and improves transmission. The measurements were performed with xenon as collision gas and the instrument has been modified to allow an extra 200 V of collision energy to be applied (up to 400 V). The instrument is also equipped with a 32,000 Th quadrupole. The source conditions for both capsomers and prohead-1^{gp5} were as follows: capillary 1300-1500 V, cone 160 V, extraction cone 0 V and 10 mbar backing pressure. The pressure in the collision cell was 2×10^{-2} mbar, 60 V collision energy was applied to free capsomers, and 400 V to prohead-1^{gp5}. The 'steering' parameter was adjusted for prohead-1^{gp5}, which resulted in better transmission of these large ions. Spectra of free capsomers were calibrated externally based on a spectrum of a 25 mg/mL cesium iodide solution. No CsI clusters were observed in the m/z range of the prohead-1^{gp5} spectral signal, thus it was not possible to mass calibrate these spectra.

Data Analysis.

Masses were determined by calculating the number of charges on subsequent charge states from mass-to-charge peak tops, extracted manually from Masslynx software. Assuming that the number of charges, z , is negligible compared to the mass of the complex, this number can be estimated with (15):

$$z = [m/z]_z / ([m/z]_z - [m/z]_{z+1}).$$

This gives accurate near-integer solutions for smaller protein complexes with well resolved peaks, but this is not the case for large complexes with poorly resolved peaks. Therefore, the solution to the above equation was only used as a

first estimate for prohead-1^{gp5} spectra. Then, we employed a similar strategy to the one used in the recently described LeastMass software (16), as described in the main text. The relative signal of pentons in the capsomer mixture was determined from triplicate analysis. The spectra were deconvoluted using the MaxEnt1 function of Masslynx. The peak areas for hexons and pentons were extracted from this deconvoluted spectrum, and the fraction of pentons was averaged over the three analyses.

The charging behavior of protein complexes in nanoESI was determined from experiment-based charge state assignments of 73 native protein complexes, including HK97 pentons, hexons and prohead-1^{gp5}. The most intense charge state, z , was plotted as a function of molecular weight, M , which follows a power function of the form:

$$z = a M^b.$$

Whereas the charged residue model of electrospray ionization predicts that $b = 0.5$, the charging scales slightly stronger with molecular weight empirically, such that $b = 0.55$ (17, 23). The width of the isotope distribution of HK97 complexes was determined by simulating the isotope distribution for $z = 1$ with Masslynx. The *FWHM* of the distribution was extracted manually from the simulated distributions of increasing gp5 stoichiometry, up to $n = 420$ (i.e. the prohead-1^{gp5} particle). The *FWHM* of the isotope distribution scales with the square root of M :

$$FWHM = c M^{0.5}.$$

The square root scaling of *FWHM* to M arises from the fact that the isotope distribution is binomial, the factor c is proportional to the likelihood of stable heavy isotopes being present. The inherent peak width of a charge state in the m/z dimension is thus proportional to the width of the isotope distribution and inversely proportional to the number of charges:

$$Width = FWHM / z = (cM^{0.5}) / (aM^b) = (c/a)M^{(0.5 - b)}.$$

As we determined that the exponent $b > 0.5$, it follows that the inherent width of peaks generated from protein complexes with nanoESI get increasingly narrow with increasing mass of the complex.

The experimentally determined trend for charging of protein complexes in nanoESI was used to determine the expected peak widths for a given resolution, R , by estimating the m/z position:

$$\text{Width} = [m/z]z / R = [M / (aM \wedge b)] / R.$$

Similarly, the expected separation between consecutive charge states as a function of mass was calculated as:

$$[m/z]z - [m/z]z+1 = [M / (aM \wedge b)] - [M / (aM \wedge b+1)].$$

With the width of the peak calculated at half-maximum, peaks are separated at half height when the width is equal to the difference in peak position.

Quantitative Proteomic analysis.

Prohead-1 with protease was digested with LysC and Trypsin following denaturation in sodium deoxycholate (SDC) and reduction/alkylation with DTT and iodo-acetamide (IAA). Briefly, Prohead-1 with protease in 2.5% SDC in 50 mM AmBic was heated for 5 minutes at 95°C. Reduction was performed in 10 mM DTT at 56°C, followed by alkylation in 20 mM IAA at room temperature in the dark. Samples were digested using 1:75 LysC for 3.5 hrs at 37°C, before dilution to 1% SDC and overnight digestion with 1:50 Trypsin at 37°C. Samples were recovered from SDC using liquid/liquid extraction with ethylacetate. Samples were dried in speedvac and redissolved in aqueous solution containing 10% formic acid and 5% DMSO. Peptides were separated by reversed phase LC coupled on-line to an LTQ or LTQ-Orbitrap XL for MS/MS analysis. The nano-LC consists of an Agilent 1200 series LC system equipped with a 20 mm ReproSil- Pur C18-AQ (Dr. Maisch GmbH) trapping column (packed in-house, i.d., 100 μm ; resin, 5 μm) and a 400 mm ReproSil-Pur C18-AQ (Dr. Maisch GmbH) analytical column (packed in-house, i.d., 50 μm ; resin, 3 μm) arranged in a vented-column configuration. The flow was passively split to 100 nl/min. We used a standard 45 minute gradient from 7-30% acetonitrile in aqueous 0.1% formic acid. Protein abundance was quantified by spectral counting. The reported numbers are scaled by molecular weight and normalized with the known copy number of 420 for gp5.

Acknowledgements.

This project was supported by grants from the NIH (R01 AI040101) as well as

a FP7 Marie-Curie IOF fellowship (273427) attributed to D.V.; J.S. and A.J.R.H. are supported by the Netherlands Proteomics Centre.

Supplementary information.

Supplementary information can be found in the appendix.

References.

1. Douglas, T., and M. Young. 2006. Viruses: Making friends with old foes. *Science*. 312:873-875.
2. Uetrecht, C., C. Versluis, N. R. Watts, W. H. Roos, G. J. L. Wuite, P. T. Wingfield, A. C. Steven, and A. J. R. Heck. 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.* 105:9216-9220.
3. Shoemaker, G. K., E. Van Duijn, S. E. Crawford, C. Uetrecht, M. Baclayon, W. H. Roos, G. J. L. Wuite, M. K. Estes, B. V. V. Prasad, and A. J. R. Heck. 2010. Norwalk virus assembly and stability monitored by mass spectrometry. *Molecular and Cellular Proteomics*. 9:1742-1751.
4. Uetrecht, C., I. M. Barbu, G. K. Shoemaker, E. Van Duijn, and A. J. R. Heck. 2011. Interrogating viral capsid assembly with ion mobility-mass spectrometry. *Nature Chemistry*. 3:126-132.
5. Brasch, M., A. De La Escosura, Y. Ma, C. Uetrecht, A. J. R. Heck, T. Torres, and J. J. L. M. Cornelissen. 2011. Encapsulation of phthalocyanine supramolecular stacks into virus-like particles. *J. Am. Chem. Soc.* 133:6878-6881.
6. Chernushevich, I. V., and B. A. Thomson. 2004. Collisional Cooling of Large Ions in Electrospray Mass Spectrometry. *Anal. Chem.* 76:1754-1760.
7. Tahallah, N., M. Pinkse, C. S. Maier, and A. J. R. Heck. 2001. The effect of the source pressure on the abundance of ions of noncovalent protein assemblies in an electrospray ionization orthogonal time-of-flight instrument. *Rapid Communications in Mass Spectrometry*. 15:596-601.
8. Lorenzen, K., C. Versluis, E. van Duijn, R. H. H. van den Heuvel, and A. J. R. Heck. 2007. Optimizing macromolecular tandem mass spectrometry of large non-covalent complexes using heavy collision gases. *International Journal of Mass Spectrometry*. 268:198-206.
9. Van Den Heuvel, R. H. H., E. Van Duijn, H. Mazon, S. A. Synowsky, K. Lorenzen, C. Versluis, S. J. J. Brouns, D. Langridge, J. Van Der Oost, J. Hoyes, and A. J. R. Heck. 2006. Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. *Anal. Chem.* 78:7473-7483.
10. Sobott, F., H. Hernández, M. G. McCammon, M. A. Tito, and C. V. Robinson. 2002. A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Anal. Chem.* 74:1402-1407.
11. Johnson, J. E. 2010. Virus particle maturation: Insights into elegantly pro-

grammed nanomachines. *Curr. Opin. Struct. Biol.* 20:210-216.

12. Hendrix, R. W., and J. E. Johnson. 2012. Bacteriophage HK97 capsid assembly and maturation. *Advances in Experimental Medicine and Biology.* 726:351-363.

13. Huang, R. K., R. Khayat, K. K. Lee, I. Gertsman, R. L. Duda, R. W. Hendrix, and J. E. Johnson. 2011. The prohead-I structure of bacteriophage HK97: Implications for scaffold-mediated control of particle assembly and maturation. *J. Mol. Biol.* 408:541-554.

14. Benesch, J. L. P., B. T. Ruotolo, F. Sobott, J. Wildgoose, A. Gilbert, R. Bateman, and C. V. Robinson. 2009. Quadrupole-time-of-flight mass spectrometer modified for higher-energy dissociation reduces protein assemblies to peptide fragments. *Anal. Chem.* 81:1270-1274.

15. Mann, M., C. K. Meng, and J. B. Fenn. 1989. Interpreting mass spectra of multiply charged ions. *Anal. Chem.* 61:1702-1708.

16. Tseng, Y. -, C. Uetrecht, A. J. R. Heck, and W. -. Peng. 2011. Interpreting the charge state assignment in electrospray mass spectra of bioparticles. *Anal. Chem.* 83:1960-1968.

17. Fernandez De La Mora, J. 2000. Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism. *Anal. Chim. Acta.* 406:93-104.

18. Dole, M., L. L. Mack, R. L. Hines, D. O. Chemistry, R. C. Mobley, L. D. Ferguson, and M. B. Alice. 1968. Molecular beams of macroions. *J. Chem. Phys.* 49:2240-2249.

19. Rose, R. J., E. Damoc, E. Denisov, A. Makarov, and A. J. R. Heck. 2012. High-sensitivity Orbitrap mass analysis of intact macromolecular assemblies. *Nature Methods.* 9:1084-1086.

20. Hendrix, R. W., and R. L. Duda. 1998. Bacteriophage HK97 head assembly: a protein ballet. *Adv. Virus Res.* 50:235-288.

21. Keifer, D. Z., E. E. Pierson, J. A. Hogan, G. J. Bedwell, P. E. Prevelige, and M. F. Jarrold. 2014. Charge detection mass spectrometry of bacteriophage P22 procapsid distributions above 20 MDa. *Rapid Communications in Mass Spectrometry.* 28:483-488.

22. Veessler, D., J. Quispe, N. Grigorieff, C. S. Potter, B. Carragher, and J. E. Johnson. 2012. Maturation in action: CryoEM study of a viral capsid caught during expansion. *Structure.* 20:1384-1390.

23. Dole, M., L. L. Mack, R. L. Hines, D. O. Chemistry, R. C. Mobley, L. D. Ferguson, and M. B. Alice. 1968. Molecular beams of macroions. *J. Chem. Phys.* 49:2240-2249.

Chapter 3:

A new platform with improved resolution for native MS of virus particles.

Joost Snijder^{1,2}, Michiel van de Waterbeemd^{1,2}, Eugen Damoc³, Eduard Denisov³, Dmitry Grinfeld³, Antonette Bennett⁴, Mavis Agbandje-McKenna⁴, Alexander Makarov^{1,2,3}, Albert J.R. Heck^{1,2}

¹ Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.

² Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands.

³ Thermo Fisher Scientific (Bremen), Bremen, Germany.

⁴ Department of Biochemistry and Molecular Biology, Center for Structural Biology, McKnight Brain Institute, College of Medicine, University of Florida, Gainesville, Florida, USA.

Based on:

Snijder, J., *et al.* Defining the stoichiometry and cargo load of viral and bacterial nanoparticles by orbitrap mass spectrometry. *J.Am.Chem.Soc.* 136, 7295-7299 (2014).

Abstract.

Accurate mass analysis can provide useful information on the stoichiometry and composition of protein-based particles, such as virus-like assemblies. For applications in nanotechnology and –medicine, such nanoparticles are loaded with foreign cargos, making accurate mass information essential to define the cargo load. Here we describe modifications to an Orbitrap mass spectrometer that enable high mass analysis of several virus-like nanoparticles up to 4.5 megadalton in mass. This allows the accurate determination of the composition of virus-like particles. The modified instrument is utilized to determine the cargo load of bacterial encapsulin nanoparticles that were engineered to encapsulate foreign cargo proteins. We find that encapsulin packages from 8 up to 12 cargo proteins, thereby quantifying cargo load but also showing the ensemble spread. In addition, we determined the previously unknown stoichiometry of the three different splice variants of the capsid protein in adeno-associated virus (AAV) capsids, showing that symmetry is broken, and assembly is heterogeneous and stochastic. These results demonstrate the potential of high-resolution mass analysis of protein-based nanoparticles, with widespread applications in chemical biology and nanotechnology.

Native MS is traditionally and still predominantly performed on modified (quadrupole-) time-of-flight (TOF) instruments, where the operating pressures of several pumping stages are increased to improve transmission of high m/z ions via collisional cooling (1-6). More recently, an Orbitrap™-based platform was developed for application in native MS, demonstrating the use of the mass analyzer in detection of non-covalent protein complex ions up to approximately 20000 m/z with improved resolving power compared to traditional TOF instruments modified for native MS (7-10). It was shown that Orbitrap-based analysis offers great potential for characterizing micro-heterogeneity in large protein complexes, such as small ligand binding on the 800 kDa GroEL chaperonin and glycosylation profiling on intact proteins under native conditions (7, 11-14). Here we seek to further extend and explore the upper mass limits of the Orbitrap platform with the purpose of analyzing virus like assemblies.

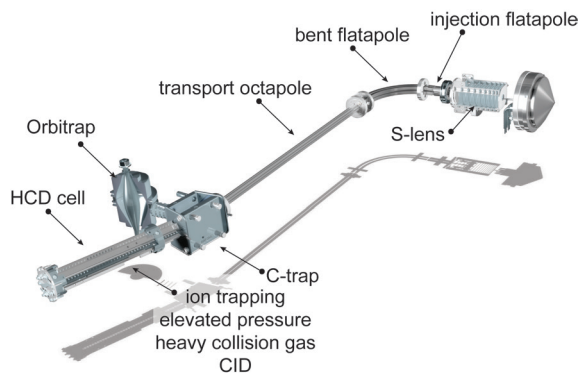


Figure 1. Schematic of the Exactive Plus instrument.

Results.

Based on theoretical considerations regarding the ion optics of the instrument (see Supplementary text and Supplementary Figures S1 and S2), we hypothesized that more efficient focusing of ions, especially in the C-trap, would be beneficial for transmission of high m/z ions (>20000). A schematic of the instrument is presented in Figure 1. We sought to more effectively focus high mass ions in the front-end ion guides and C-trap by lowering their RF frequency. The effects of reducing the RF frequencies of the front-end ion guides on high m/z ion transmission were tested on large Cesium Iodide clusters (up to 20,000 m/z) and GroEL CID product ions (ranging from 15,000 to 40,000 m/z). A 5-10 fold

increase in ion transmission could be demonstrated for CsI clusters at 20,000 m/z and GroEL product ions at 30,000 m/z . Moreover, the instrument modifications allowed detection of GroEL product ions up to 40,000 m/z , compared to 30,000 m/z in the standard configuration. At the transient time of 64 ms that was used to analyze GroEL, intact 14-mer ions are detected with an effective resolution of ~ 1500 ($M/\Delta M$, FWHM), whereas there is a shallow decline in effective resolution for the higher m/z 13-mer and 12-mer ions, down to ~ 1000 at m/z 40000 (see Supplementary Table S1 and Figures S3-S5).

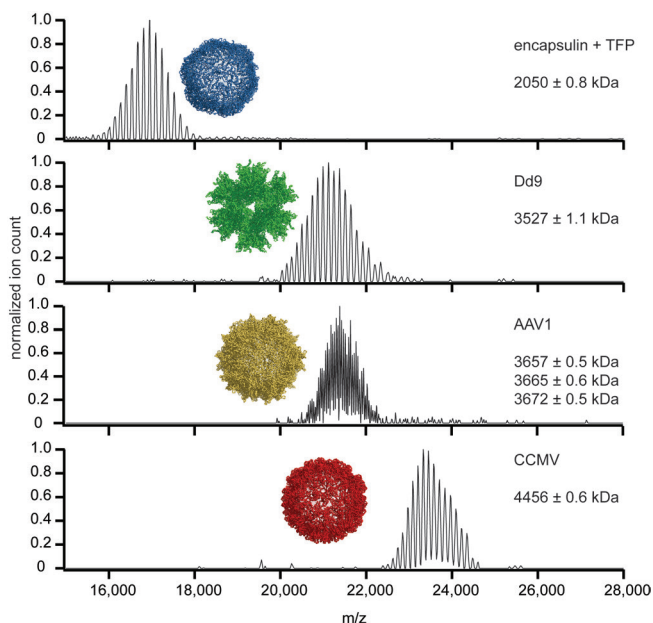


Figure 2. Native MS of virus-like particles on the modified Exactive Plus. Shown are encapsulin (top), adenovirus dodecahedron (top-middle), adeno-associated virus serotype 1 (bottom-middle) and cowpea chlorotic mottle virus (bottom). Corresponding structures are also shown on the graphs. The quoted masses represent the average \pm standard deviation over all charge states in the spectrum. For detailed peak assignments see Supplementary Table S2.

We further explored the performance of the modified instrument for the analysis of megadalton (MDa) virus assemblies (Figure 2). Therefore, we analyzed an array of particles, namely the bacterial nanocontainer encapsulin (2.1 MDa, $d \sim 24$ nm), the adenovirus dodecahedron (Dd, 3.5 MDa, $d \sim 22$ nm), adeno-associated virus serotype 1 (AAV1, 3.7 MDa, $d \sim 27$ nm) and cowpea chlorotic mottle virus (CCMV, 4.5 MDa, $d \sim 28$ nm). Encapsulin, Dd, AAV1 and CCMV could all be mass analyzed on the modified instrument (see Figure 2). All the detected ion signals originate from well-defined charge state distributions that are nearly

baseline resolved for all species. The standard deviations on the obtained masses are all on the order of 0.01%. Details of the peak assignments are provided in Supplementary Table S2. These MS analyses are considerably more precise and accurate than conventional techniques that are used to determine the size and mass of protein nanoparticles, such as size exclusion chromatography, light-scattering based techniques or gel-based assays. Although attempted, we were so far not able to analyze particles larger than CCMV, such as the 13 megadalton HK97 head II particle discussed in Chapter 2, on the Orbitrap-based platform. It is estimated that the upper limit of the instrument for intact protein complexes is currently at 25,000–30,000 m/z , corresponding to a mass of roughly 5 megadalton in positive-mode nano electrospray. Because all ions generated in the source must pass the C-trap twice, which we identified as a bottleneck for high-mass transmission, this limit is lower compared to the highest m/z that can be observed for HCD product ions, which only pass the C-trap once after being generated in the HCD cell.

Encapsulin is a recently discovered bacterial nanoparticle, consisting of 60 copies of a capsid-like protein that form a $T=1$ icosohedral capsid-like particle that encapsulates a functional enzyme *in vivo* in bacteria (15) (See also Chapter 6). More recently it was shown that encapsulin can also be utilized as a nanocontainer to package non-native cargo proteins (16). To demonstrate the utility of the modified Orbitrap platform in chemical biology and nanotechnology, we analyzed encapsulin loaded with foreign fluorescent cargo proteins (Figure 3). Whereas the spectrum in Figure 2 seems to represent a single charge state distribution, we observed that it is actually the sum of many very closely overlapping distributions, originating from the encapsulin nanoparticle with a variable number of the cargo proteins encapsulated. We discovered this only when we applied high energy to the ions in the HCD cell, promoting dissociation, whereby subsequently up to two encapsulin monomeric subunits became expelled from the intact precursor ions. In the resulting high m/z fragment ions, the underlying distributions of the different species became resolved (Figure 3).

The resulting masses correspond to an integer number of loaded cargo proteins, thereby verifying the more accurate mass assignments on the HCD product ions compared to the intact parent ions (see Supplementary Table S3). We observe charge state distributions corresponding to nanoparticles of the 60-mer encapsulin, encapsulating between exactly 8 and 12 copies of foreign fluorescent cargo molecule. The intensity weighted average number of encapsulated cargo molecules is determined at 10.6 TFP per encapsulin nanoparticle. The results show that the modified Orbitrap platform can be used to quantify the encapsulated

cargo in this ~ 2 MDa bacterial nanocontainer with sufficient resolution to also characterize the ensemble spread. To validate the above quantitation procedure, we used the extracted intensity information of the individual encapsulin-cargo stoichiometries to reconstruct a theoretical spectrum for the intact particle (Supplementary Figure S6). We used an in-house developed software package SOMMS (SOLving complex Macromolecular Mass Spectra) to generate spectra of encapsulin 60-mer with 8-12 copies of the cargo molecule (17). These spectra illustrate how the very close overlap between the different charge state distributions results in one unresolved ion series (the required resolution to separate these peaks at half-height is 2500-3000, currently beyond the capabilities of the instrument). Summing the theoretical spectra together reasonably reproduces the experimental spectrum of the undissociated encapsulin with the determined mixed number of fluorescent cargo proteins encapsulated.

The Dd and AAV1 particles have similar masses of around 3.5 MDa. Still, the corresponding mass spectra shown in Figure 2 look strikingly different. Dd has been proposed as a non-infective vector for gene transfer and is structurally a dodecahedron made of adenovirus pentons, i.e. a 60-mer homo-oligomeric assembly, having a single defined mass. AAV1 has a $T=1$ icosahedral capsid and is extensively used as a vector for gene therapy and vaccine delivery (18, 19). It is the first approved gene therapy vector in the Western world (20). In AAVs alternative splicing and differential codon usage of a single capsid gene produces three variants of the capsid viral protein, VP1/VP2/VP3. The VP3 sequence is common between all three splice variants, VP2 and VP1 have N-terminal longer sequences, with VP1 containing a phospholipase domain in its unique region. The exact amounts of VP1/VP2/VP3 in the capsid are unknown, but estimated to be 1/1/10, based on densitometry analyses of the capsid proteins resolved on SDS-PAGE (21-23).

Despite many reported crystal structures and cryo-electron microscopy reconstructions of several AAV serotypes, it is currently not known whether there is any defined VP1/VP2/VP3 stoichiometry in the capsid, as only the common part between the three splice variants (VP3) is clearly resolved in those structures. In our mass spectrum of AAV1 capsids, we were able to resolve three series of peaks (see Figure 4). From the calculated masses, we can determine the copy numbers of VP1/VP2/VP3, also applying the constraint that the total copy number in the capsid is fixed at $n=60$. Based on the sequences of VP1/VP2/VP3, their monomer masses are calculated to be 81375/66225/59606 Da, respectively. Based on these theoretical masses, the experimental masses of the intact capsids were used to estimate the VP1/VP2/VP3 copy numbers, where we considered

every theoretical mass within 2 standard deviations of the experimental mass to be a possible match.

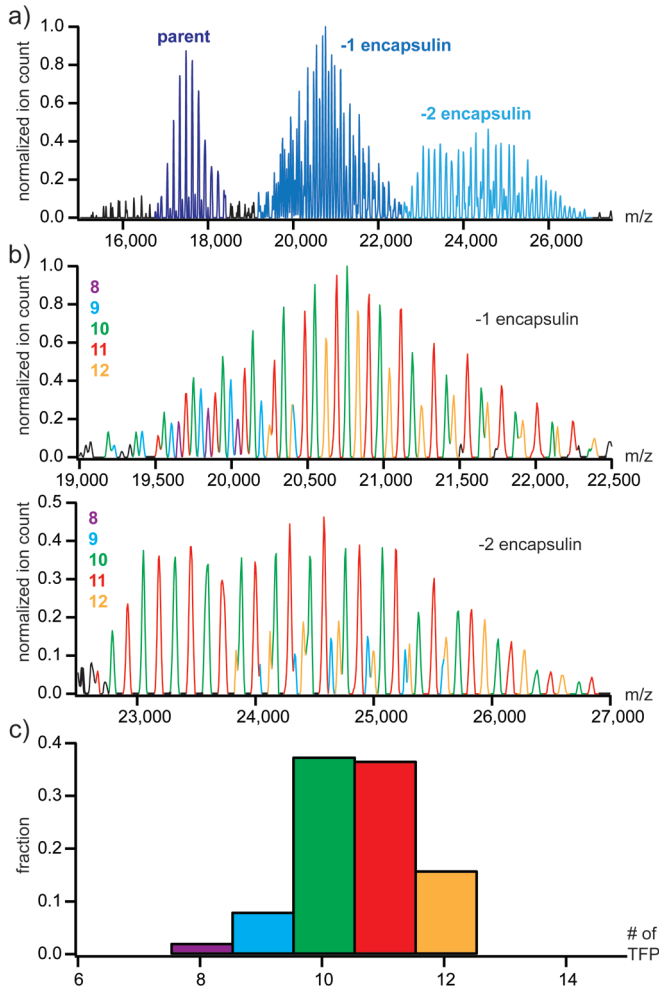


Figure 3. Identifying and quantifying foreign cargo encapsulation (Teal Fluorescent Protein, TFP) in bacterial encapsulin. a) Native MS spectrum of encapsulin at high collision energy, showing dissociation of up to two encapsulin monomeric subunits. b) Zoom-in of the peaks corresponding to the first (top) and second (bottom) dissociation products, colors according to number of encapsulated cargo molecules. c) Total intensity of all identified encapsulin-cargo stoichiometries. For detailed peak assignments see Supplementary Table S3.

The masses listed in Figure 4 correspond to capsids with a single copy of VP1 and between 9 and 11 copies of VP2 (with all remaining subunits VP3). For all three series of peaks, the charge state assignment is relatively ambiguous. We

get a similar standard deviation of 0.5 kDa on the calculated masses, regardless of whether the masses are calculated using the charge-state range of 165-178 or 164-177 (a bigger shift in charge state assignment does result in bigger errors). By simulating theoretical spectra of AAV1 capsids with different VP1/VP2/VP3 stoichiometries using SOMMS, it became apparent that there is substantial peak overlap between capsids with different VP1 and VP2 copy numbers and that it is therefore fundamentally impossible to resolve more than the three series of peaks (Figure 4b). Considering the two alternative charge state assignments, we calculated a total of six masses, matching a total of 8 stoichiometries (see Supplementary Table S4). From these calculations we extract that there are 0-2 copies of VP1, 8-11 copies of VP2 and 48-51 copies of VP3 in individual AAV1 particles. These results demonstrate for the first time that there is no defined VP1/VP2/VP3 stoichiometry for AAV1, and therefore suggest that assembly is stochastic such that the relative amount of VP1/VP2/VP3 that is incorporated in the capsid might depend mainly on their relative expression levels. The variable copy numbers of VP1/VP2/VP3 offer an explanation for the lack of VP1/VP2 density in the structural analyses of AAV, as the lack of symmetry would preclude VP1/VP2 to clearly appear in reconstructions of the particles.

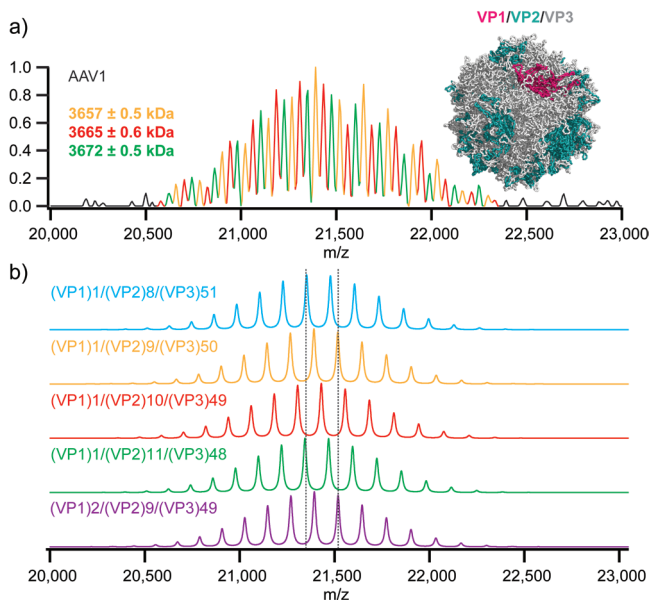


Figure 4. Determining the VP1/VP2/VP3 stoichiometry in AAV1. a) Experimental spectrum of AAV1 capsids recorded at 32 ms transient time. The AAV1 capsid structure is shown with scattered copies of VP1/VP2. b) Simulated spectra of AAV1 capsids with different VP1/VP2/VP3 stoichiometries. The simulated spectra illustrate how there is partial overlap between peaks of capsids with different VP1 and VP2 copy numbers.

To illustrate the benefit of the MS-based approach described here, examples of SDS-PAGE and negative stain electron microscopy analysis of AAV1 are provided in Supplementary Figure S7. These represent widely used techniques for the analysis of AAV and other protein-based nanoparticles in general. Whereas SDS-PAGE does provide good separation of VP1/VP2/VP3, such an experiment only provides an ensemble average regarding the relative amounts of the three different capsid proteins, because the particles are analyzed under denaturing conditions. Techniques such as negative stain electron microscopy allow the still intact particles to be analyzed, but despite the relatively high spatial resolution offers no information about the relative amounts of VP1/2/3 in the absence of any distinct morphological features. SEC or gel shift assays similarly lack the resolution to resolve mass differences on the order of 0.1%. As discussed above, even when the particles can be analyzed with atomic-level resolution using X-ray crystallography, no density corresponding to the unique regions of VP1 and VP2 can be seen, because of the inherent variability between particles. For these reasons, the ability to determine with high precision and resolution the absolute masses of nanoparticles makes the MS-based approach described here a uniquely powerful analytical tool.

Discussion.

Most native MS platforms described so far employ a markedly different strategy for improving transmission of high mass ions. In most TOF-based platforms, the source regions and ion guides are operated at elevated pressure to improve transmission of high mass ions via collisional cooling (1-6). Whereas enhanced collisional cooling is in principle also possible on the Orbitrap-based instrument described here, there is a possible benefit of the strategy for improved high mass transmission that we employ. By increasing the quasi-potential well depth of the ion focusing devices, we prevent the introduction of excessive gas in regions of the instruments where this is unwanted. Most notably, it was recently demonstrated that the main reason for a decay of signal for larger ions in the analyzer is due to collisions with background gas (24). We identified the C-trap as a bottleneck for efficient high mass transmission, but collisional cooling in this region would lead to particularly high pressures in the Orbitrap as well, which we anticipate to lead to rapid decay of coherent and stable oscillation of the ions inside the Orbitrap analyzer.

The RF frequencies on the ion guides have been lowered by only 25% in the cur-

rent study, a drop that is ultimately limited by the design of the instruments electronics boards. Exploring the potential improvement in ion transmission with a further drop in RF frequency would require redesign of the instrument electronics. Other potential improvements could involve the design of the instrument's HCD cell. In its current design, the HCD cell appears to leak a substantial amount of collision gas into the Orbitrap analyzer, thereby limiting the stability of the transient, and therefore the obtained S/N and obtainable resolution. Designing an HCD cell with a dual pumping stage could offer a solution to this problem. More efficient pumping of the Orbitrap analyzer would be required to improve on this aspect, otherwise. Improving the instruments capacity to maintain the vacuum in the Orbitrap analyzer would also open up the possibility of exploring collisional cooling as a mechanism for improved ion transmission throughout the rest of the instrument. The ion path currently also involves a full turn in the HCD cell, which appears to be problematic, especially at high collision voltages. Apparently, some ions are too energetic and overshoot the exponential retarding field at the back of the HCD cell. Redesigning the ion optics of the cell in this respect might provide further improvements to the transmission of high m/z ions. Alternatively, one could explore the option of an HCD cell that precedes the C-Trap, such that no field at the back of the HCD cell would be required at all. Finally, the observed increase in high m/z ion transmission from changing the Source DC offset parameter suggest that ions are overshooting the bent flatapole, rather than making that turn. Further improving the focusing of ions here, or changing the angle of the turn altogether, might provide an additional means to boosting high m/z ion transmission and further extending the accessible mass range of the instrument.

In conclusion, we have identified limitations for high mass analysis in the Orbitrap-based platform for native MS and developed a successful strategy to improve the instruments sensitivity at high m/z . This allows the characterization of virus-like assemblies with superior mass resolving power, thereby facilitating analyses such as the quantification of cargo encapsulation in engineered virus like particles for nanotechnology, and the precise definition of protein stoichiometries in heterogeneous protein assemblies.

Methods.

Sample preparation.

Samples were prepared for mass spectrometry by buffer exchange to ammonium acetate, using Vivaspin 500 10 kDa MWCO centrifugal filter units. GroEL

samples were analyzed from 50 mM ammonium acetate pH 6.8. Encapsulin and Dodecahedron samples were analyzed from 150 mM ammonium acetate pH 6.8. The AAV1 samples were prepared as previously described (25, 26) and analyzed in 100 mM ammonium acetate pH 6.8. CCMV was analyzed from 50 mM ammonium acetate pH 5.0. Aliquots of 1-2 μL , at a final concentration of $\sim 2 \mu\text{M}$, were loaded into gold-coated borosilicate capillaries (prepared in-house) for nano-electrospray ionization.

Instrument modifications.

The Exactive Plus™ mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) was modified to include an adjustable gas supply for the HCD cell and analog filters were removed from the image current pre-amplifier to allow detection over the entire frequency range. The control software of the instrument was modified to allow the standard mass range of this instrument to be increased from m/z 50–6000 to m/z 400–40000. In addition, maximum RF voltages were applied to all RF multipoles including the C-trap. Instead of trapping in the C-trap, ions were allowed to enter the HCD cell and were stored there prior to their return back into the C-trap. Manual tuning of the voltage offset on the transport octapole was used for mass filtering of the incoming protein ions, as previously described (7). Frequency reduction on RF multipoles was implemented by adding high-voltage capacitors to corresponding RF coils and electronic boards automatically adjusted resonance frequency while keeping the amplitude constant. The RF frequency of the injection flatapole, bent flatapole, transport octapole, C-trap and HCD cell were lowered by 20-25%. In particular, the RF frequency of the C-trap was lowered from 3.2 MHz to 2.8 MHz, which, according to Supplementary equation 5, should lead to an extension of the upper mass limit from 29000–36000 to 37000–47000 m/z , and generally improve transmission of ions at high m/z .

Tuning of some crucial instrument parameters was necessary for the analysis of megadalton assemblies. Mainly, whereas the instrument is normally operated at a source DC offset of 25 V, this did not allow transmission of the tested virus particles. Instead, no offset (0 V) resulted in base peak intensities of $\sim 10^5$ – 10^6 at 100 ms injection time. In addition, the required high xenon pressure in the HCD cell for efficient transmission of the high mass ions resulted in elevated pressures in the Orbitrap compartment as well (10^{-9} mbar, compared to 10^{-11} in normal operating mode). This resulted in rather unstable and dampened transients, which gave rise to noisy spectra at longer transient times. Therefore, transient times were adjusted according to S/N to either 32 or 16 ms (compared to 64–256 ms in normal Orbitrap operating mode). For the same reason, transients of the same

acquisition were averaged before FT as microscans, rather than combining spectra after FT. It should be noted that ions at high m/z are close in frequency (~50 kHz) to the high-pass filter of the pre-amplifier of the image current detection. To prevent erroneous eFT calibration at high m/z , the pre-amplifier was replaced with a version equipped with a 22 kHz high-pass filter.

Notes.

Eu.D., Ed.D., D.G. and A.M. are employees of Thermo Fisher Scientific Inc., the company that produces and sells the Exactive Plus EMR instrument.

Acknowledgements.

We kindly acknowledge our collaborators for providing the virus and encapsulin samples. Encapsulin and CCMV samples were provided by W.F. Rurup, M. Brasch, M.S.T Koay and J.J.L.M. Cornelissen (MESA+ Institute for Nanotechnology, University of Twente, The Netherlands). Adenovirus Dodecahedron samples were provided by E. Szolajska and J. Chroboczek (IBB, Polish Academy of Sciences, Poland). J.S., M.vd.W. and A.J.R.H are supported by the Netherlands Proteomics Centre. M.vd.W. and A.J.R.H. are additionally supported by a Projectruimte grant (12PR3303-2) from Fundamenteel Onderzoek der Materie (FOM). A.B. and M.A.M. are supported by NIH R01 GM109524 and funds from the UF College of Medicine.

Supplementary information.

Supplementary information can be found in the appendix.

References.

1. Benesch, J. L. P., B. T. Ruotolo, F. Sobott, J. Wildgoose, A. Gilbert, R. Bateman, and C. V. Robinson. 2009. Quadrupole-time-of-flight mass spectrometer modified for higher-energy dissociation reduces protein assemblies to peptide fragments. *Anal. Chem.* 81:1270-1274.
2. Chernushevich, I. V., and B. A. Thomson. 2004. Collisional Cooling of Large Ions in Electrospray Mass Spectrometry. *Anal. Chem.* 76:1754-1760.
3. Kozlovski, V. I., L. J. Donald, V. M. Collado, V. Spicer, A. V. Loboda, I. V. Chernushevich, W. Ens, and K. G. Standing. 2011. A TOF mass spectrometer for the study of noncovalent complexes. *International Journal of Mass Spectrometry.* 308:118-125.
4. Krutchinsky, A. N., I. V. Chernushevich, V. L. Spicer, W. Ens, and K. G. Standing. 1998. Collisional damping interface for an electrospray ionization time-of-flight mass spectrometer. *J. Am. Soc. Mass Spectrom.* 9:569-579.
5. Van Den Heuvel, R. H. H., E. Van Duijn, H. Mazon, S. A. Synowsky, K. Lorenzen, C. Versluis, S. J. J. Brouns, D. Langridge, J. Van Der Oost, J. Hoyes, and A. J. R. Heck. 2006. Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. *Anal. Chem.* 78:7473-7483.
6. Sobott, F., H. Hernández, M. G. McCammon, M. A. Tito, and C. V. Robinson. 2002. A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Anal. Chem.* 74:1402-1407.
7. Rose, R. J., E. Damoc, E. Denisov, A. Makarov, and A. J. R. Heck. 2012. High-sensitivity Orbitrap mass analysis of intact macromolecular assemblies. *Nature Methods.* 9:1084-1086.
8. Makarov, A. 2000. Electrostatic axially harmonic orbital trapping: A high-performance technique of mass analysis. *Anal. Chem.* 72:1156-1162.
9. Zubarev, R. A., and A. Makarov. 2013. Orbitrap mass spectrometry. *Anal. Chem.* 85:5288-5296.
10. Belov, M. E., E. Damoc, E. Denisov, P. D. Compton, S. Horning, A. A. Makarov, and N. L. Kelleher. 2013. From protein complexes to subunit backbone fragments: A multi-stage approach to native mass spectrometry. *Anal. Chem.* 85:11163-11173.
11. Rosati, S., E. T. Van Den Bremer, J. Schuurman, P. W. Parren, J. P. Kamerling, and A. J. Heck. 2013. In-depth qualitative and quantitative analysis of composite glycosylation profiles and other micro-heterogeneity on intact monoclonal anti-

bodies by high-resolution native mass spectrometry using a modified Orbitrap. *MAbs*. 5:917-924.

12. Rosati, S., R. J. Rose, N. J. Thompson, E. Van Duijn, E. Damoc, E. Denisov, A. Makarov, and A. J. R. Heck. 2012. Exploring an orbitrap analyzer for the characterization of intact antibodies by native mass spectrometry. *Angewandte Chemie - International Edition*. 51:12992-12996.

13. Yang, Y., A. Barendregt, J. P. Kamerling, and A. J. R. Heck. 2013. Analyzing protein micro-heterogeneity in chicken ovalbumin by high-resolution native mass spectrometry exposes qualitatively and semi-quantitatively 59 proteoforms. *Anal. Chem*. 85:12037-12045.

14. Thompson, N. J., L. J. A. Hendriks, J. De Kruif, M. Throsby, and A. J. R. Heck. 2014. Complex mixtures of antibodies generated from a single production qualitatively and quantitatively evaluated by native Orbitrap mass spectrometry. *MAbs*. 6:197-203.

15. Sutter, M., D. Boehringer, S. Gutmann, S. Günther, D. Prangishvili, M. J. Loessner, K. O. Stetter, E. Weber-Ban, and N. Ban. 2008. Structural basis of enzyme encapsulation into a bacterial nanocompartment. *Nature Structural and Molecular Biology*. 15:939-947.

16. Rurup, W. F., J. Snijder, M. S. T. Koay, A. J. R. Heck, and J. J. L. M. Cornelissen. 2014. Self-sorting of foreign proteins in a bacterial nanocompartment. *J. Am. Chem. Soc*. 136:3828-3832.

17. Van Breukelen, B., A. Barendregt, A. J. R. Heck, and R. H. H. Van Den Heuvel. 2006. Resolving stoichiometries and oligomeric states of glutamate synthase protein complexes with curve fitting and simulation of electrospray mass spectra. *Rapid Communications in Mass Spectrometry*. 20:2490-2496.

18. Agbandje-McKenna, M., and J. Kleinschmidt. 2011. AAV capsid structure and cell interactions. *Methods in Molecular Biology*. 807:47-92.

19. Mingozzi, F., and K. A. High. 2011. Therapeutic in vivo gene transfer for genetic disease using AAV: Progress and challenges. *Nature Reviews Genetics*. 12:341-355.

20. Melchiorri, D., L. Pani, P. Gasparini, G. Cossu, J. Ancans, J. J. Borg, C. Draï, P. Fiedor, E. Flory, I. Hudson, H. G. Leufkens, J. Müller-Berghaus, G. Narayanan, B. Neugebauer, J. Pokrotnieks, J. -. Robert, T. Salmonson, and C. K. Schneider. 2013. Regulatory evaluation of Glybera in Europe-two committees, one mission. *Nature Reviews Drug Discovery*. 12:719.

21. Rose, J. A., J. V. Maizel Jr., J. K. Inman, and A. J. Shatkin. 1971. Structural

proteins of adenovirus-associated viruses. *J. Virol.* 8:766-770.

22. Buller, R. M., and J. A. Rose. 1978. Characterization of adenovirus-associated virus-induced polypeptides in KB cells. *J. Virol.* 25:331-338.

23. Johnson, F. B., H. L. Ozer, and M. D. Hoggan. 1971. Structural proteins of adenovirus-associated virus type 3. *J. Virol.* 8:860-863.

24. Makarov, A., and E. Denisov. 2009. Dynamics of Ions of Intact Proteins in the Orbitrap Mass Analyzer. *J. Am. Soc. Mass Spectrom.* 20:1486-1495.

25. Miller, E. B., B. Gurda-Whitaker, L. Govindasamy, R. McKenna, S. Zolotukhin, N. Muzyczka, and M. Agbandje-McKenna. 2006. Production, purification and preliminary X-ray crystallographic studies of adeno-associated virus serotype 1. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications.* 62:1271-1274.

26. Zolotukhin, S., M. Potter, I. Zolotukhin, Y. Sakai, S. Loiler, T. J. Fraites Jr., V. A. Chiodo, T. Phillipsberg, N. Muzyczka, W. W. Hauswirth, T. R. Flotte, B. J. Byrne, and R. O. Snyder. 2002. Production and purification of serotype 1, 2, and 5 recombinant adeno-associated viral vectors. *Methods.* 28:158-167.

Chapter 4:

Mass spectrometry uncovers a novel component of mature adenovirus capsids.

Joost Snijder^{1,2}, Marco Benevento^{1,2}, Crystal L. Moyer³, Vijay Reddy⁴, Glen R. Nemerow³, Albert J.R. Heck^{1,2}

¹ Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.

² Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands.

³ The Department of Immunology and Microbial Science, The Scripps Research Institute, 10550 N. Torrey Pines Rd, La Jolla, CA 92037, USA

⁴ The Department of Molecular Biology, The Scripps Research Institute, 10550 N. Torrey Pines Rd, La Jolla, CA 92037, USA

Based on:

Snijder, J., *et al.* The cleaved N-Terminus of pVI binds peripentonal hexons in mature adenovirus. *J.Mol.Biol.* 426, 1971-1979 (2014).

Abstract.

Mature human adenovirus particles contain four minor capsid proteins, in addition to the three major capsid proteins (penton base, hexon and fiber) and several proteins associated with the genomic core of the virion. Of the minor capsid proteins, VI plays several crucial roles in the infection cycle of the virus, including hexon nuclear targeting during assembly, activation of the adenovirus proteinase (AVP) during maturation and endosome escape following cell entry. VI is translated as a precursor (pVI) that is cleaved at both the N- and C-termini by AVP. Whereas the role of the C-terminal fragment of pVI, pVIc, is well established as an important co-factor of AVP, the role of the N-terminal fragment, pVI_n, is currently elusive. In fact, the fate of pVI_n following proteolytic cleavage is completely unknown. Here, we use a combination of proteomics-based peptide identification, native mass spectrometry and hydrogen-deuterium exchange mass spectrometry to show that pVI_n is associated with mature human adenovirus, where it binds at the base of peripentonal hexons in a pH-dependent manner. Our findings suggest a possible role for pVI_n in targeting pVI to hexons for proper assembly of the virion and timely release of the membrane lytic mature VI molecule.

There are more than 60 known types of human adenovirus (HAdV), approximately one-third of which cause acute infection in humans. Infection is generally self-limiting, except in immunocompromised patients. Replication-defective HAdV is also used in numerous gene therapy and vaccine delivery applications (1, 2). HAdV is one of the largest known non-enveloped dsDNA viruses. It consists of a ~36 kb genome that is encapsidated in a pseudo $T=25$ icosahedral capsid (3-6). The HAdV virion is approximately 90 nm in diameter, with a total mass estimated at 150 MDa. The capsid of HAdV is composed of 240 trimeric hexons, 12 pentameric penton base molecules that occupy the vertices, and 12 trimeric fiber proteins that extend outward from the penton base and interact with host receptors to facilitate cell attachment. In addition, the mature capsid contains four cement proteins (IIIa, VI, VIII and IX), five proteins associated with the genomic core of the virion (V, VII, μ , IVa2 and terminal protein), as well as a few copies of the adenovirus proteinase (AVP).

As is common in dsDNA viruses, HAdV assembles with the aid of scaffolding proteins and subsequently matures into the infectious virion. AVP plays an essential role in HAdV maturation by processing many of the accessory proteins through cleavage at two consensus motifs, (I,L,M)XGG/X and (I,L,M)XGX/G (7). Upon assembly in the immature virus particle, AVP is inactive and bound to DNA (8). Complete activation of AVP requires both DNA-binding and pVI (8, 9). In immature virus particles, pVI slides along the viral genome through interactions of the c-terminus of pVI with the DNA (10, 11). As pVI encounters AVP, AVP cleaves the 11 C-terminal residues from pVI (pVIc), which then binds and covalently links to AVP via a disulfide bridge yielding maximum AVP activity (8, 10-16). AVP then processes the precursors of IIIa, VI, VII, VIII, μ and terminal protein into their mature forms (7). In addition to cleaving pVIc, AVP also cleaves the 33 amino-terminal residues of pVI to generate mature VI and pVI_n (i.e. residues 1-33 of pVI, see Figure 1) (17). In mature virions, VI is likely bound to peripentonal hexons on the capsid interior, though clear density for the molecule has not been definitively identified in current structural models (3, 4). The precursor pVI was shown to bind to hexon trimers with nanomolar affinity *in vitro* with a stoichiometry of 3:3 (15). Whether this stoichiometry also reflects the mode of binding in the virion is currently not known; however the disparity in copy number for hexon (720) and VI (360) monomers suggests this is not the case, at least in the mature virion. In addition to activation of AVP, pVI also plays an important role in nuclear import of the hexon (18). There are two nuclear localization (NLS) and two nuclear export signals (NES) on the full sequence of pVI. One particular NLS is located entirely on the sequence corresponding to pVIc, and this NLS is crucial for nuclear import of the hexon, in an

importin α/β dependent pathway.

Whereas pVI (pVIc in particular) plays an essential role in virus maturation, thereby priming the virion for efficient uncoating (19), VI is essential at a later stage during infection. HAdV enters host cells through receptor-mediated endocytosis. Following fiber-mediated cell attachment, subsequent interactions of the penton base with cell-surface integrins trigger endocytosis of the virion (20-24). Partial disassembly of the virion is initiated at the cell surface with release of the fiber (25, 26). Integrin binding weakens the vertex region of HAdV, which is thought to facilitate subsequent release of the penton base in early endosomes, triggered by the mildly acidic conditions encountered in that environment (21, 27). Release of the penton base opens the virion at its vertices, allowing release of VI into the endosome. VI was shown to have membrane lytic activity and is essential for escape from the endosome into the cytosol (28, 29). In addition to promoting endosome escape, VI was also reported to promote adenovirus gene expression by counteracting the Daxx-mediated host defense that suppresses viral gene expression (30).

Whereas the roles of pVIc and VI are emerging, it is currently unclear what role the 33-residue pVIn peptide has in the replication cycle of HAdV or if it remains part of the mature virion. Here we identify pVIn in mature virions by mass spectrometry (MS) based proteomics. We show with native MS that pVIn is associated with peripentonal hexons, which are released in complex from heat-disrupted virions. It is demonstrated that the interaction between pVIn and hexon is strongly pH-dependent. Finally, using hydrogen-deuterium exchange (HDX) MS, we show that pVIn binds to hexons in a region spanning residues 32-65 near the base of the hexon trimer in the capsid interior. Our findings pinpoint a previously unknown anchoring point for pVI during assembly of the virion within the peripentonal hexons.

Results.

pVIn is part of mature HAdV.

The virions used for mass spectrometry analyses were derived from double-banded cesium chloride ultracentrifugation that separates immature from mature virions (see “Materials and Methods”). We confirmed the purity of mature virions by performing SDS-PAGE to ensure the absence of uncleaved precursor polypeptides that are present in immature particles. In addition we determine the 260/280 ratio of the purified particles that is ~ 1.3 for fully mature particles

containing a complete viral genome. To identify the parts of pVI that are in the mature virion, we analyzed mature HAdV particles by MS-based proteomics. HAdV particles were proteolytically digested and subsequently analyzed by LC-MS/MS. All 13 proteins of the HAdV virion could be identified from this analysis (data not shown), including pVI. Of the full pVI sequence, ~70% was covered from peptide identifications (see Figure 1). This included most of the mature VI but also the full first 33 residues of the pVI sequence, i.e. pVIn: MEDIN-FASLAPRHGSRPFMGNWQDIGTSNMSGG (pVIn peptide identifications are listed in Supplementary Table S1). The protease (trypsin) used for digestion of the sample cleaves after lysine and arginine, but the C-terminal pVIn peptide terminates with a glycine residue. The observation that both the C-terminal half of pVIn and the N-terminus of VI were identified from semi-tryptic peptides, containing the AVP consensus terminal motif, suggests that the peptides that cover pVIn originate from the endogenously processed product. No peptides corresponding to pVIc were observed. This might simply mean that pVIc yields unobservable peptides in our analytical procedure and the lack of identification does not exclude that pVIc is still contained in the mature virion.

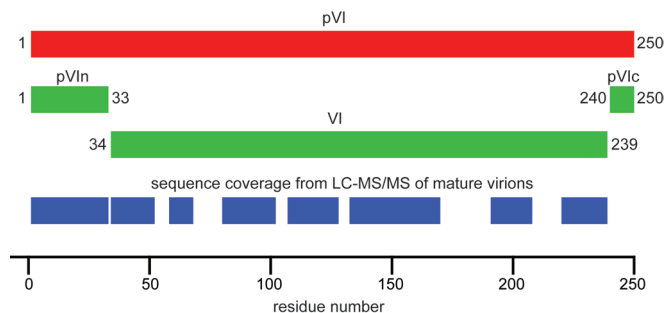


Figure 1. Schematic of pVI maturation cleavage and the obtained sequence coverage from LC-MS/MS analysis of mature virions. The 250 aa precursor pVI (red) is cleaved by AVP between residues 33/34 and 239/240 to form pVIn, VI and pVIc (green). Approximately 70% of the pVI sequence is covered by peptide identifications from mature virions, including the full pVIn sequence (blue).

Endogenous pVIn is released from HAdV during *in vitro* partial capsid disassembly.

In the mature virion, VI is likely located near the capsid vertices, interacting with the interior side of the peripentonal hexons (3, 4). The protein is released in early endosomes following partial capsid disassembly, triggered by the acidic pH of the compartment (28, 29) and perhaps also by association with α v integrins (21, 27). Moderate heating of mature virions results in a similar disassembly of

the particle, with nearly complete release of fiber and penton base, and partial release of IIIa, VI, VIII and peripentonal hexons (31-33). We separated the released components from the remaining viral cores using density centrifugation and analyzed the relative abundance of the pVIn peptides generated from heated virions using LC-MS/MS. Judging from peptide spectrum matches with pVII as an internal standard, approximately 80% of pVIn was found to be released from the heated virions (see Supplementary Table S2).

Endogeneous pVIn is bound to released hexon trimers.

The fraction of components that is released upon heating of mature HAdV was prepared for analysis by native MS in order to identify the subcomplexes that disassemble from the particle. Native MS is an emerging technique for the detailed mass analysis of non-covalent protein complexes (34, 35). Protein assemblies are transferred to an MS compatible buffer and analyzed by nano-electrospray ionization to yield highly precise and accurate mass information with unmatched resolving power.

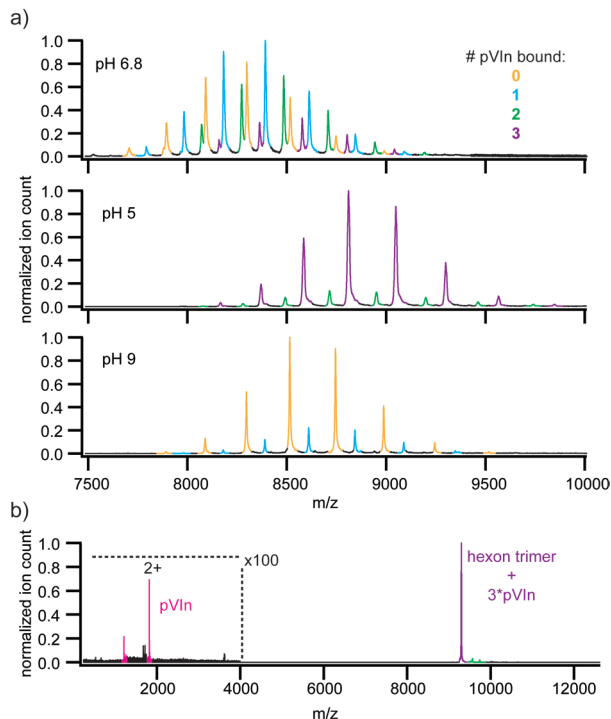


Figure 2. Native MS analysis of hexons from heat-disrupted virions show that pVIn is bound to peripentonal hexons in a pH-dependent manner. a) Hexon-pVIn complexes recovered from heat-disrupted virions were analyzed at the indicated pH. b) Tandem MS of 3:3 hexon:pVIn complex. The measured masses of the different detected hexon-pVIn assemblies are provided in supplementary table S2.

Hexon trimers were the primary molecules detected in these preparations when analyzed by native MS. We did not observe signals for the concomitantly released penton base and fibers, which may be due to losses occurring during the sample preparation for native MS and/or their relative lower abundance in the fraction of released proteins, as all released components (penton base, fiber, hexon, VI, pVIn, IIIa and VIII) could be identified from LC-MS/MS analysis of the same fraction (data not shown). The accurate mass measured for the hexon trimers (323.62 kDa, within 0.025% of the expected mass, see Supplementary table S3) allowed an unambiguous assignment. In addition to free hexon trimers, we observed up to three different hexon trimer complexes with an excess mass that matched within 0.025%, to the binding of 1, 2 or 3 copies of pVIn to the hexon trimer (see Figure 2a and Supplementary Table S3). Tandem MS of these putative hexon-pVIn complexes confirmed that the mass of the bound peptide corresponds closely to that of pVIn (see Figure 2b). At a collision voltage in excess of 100 V, we detected two charge states at low m/z that correspond to a mass of 3624.1 Da, compared to a theoretical pVIn mass of 3624.0 Da.

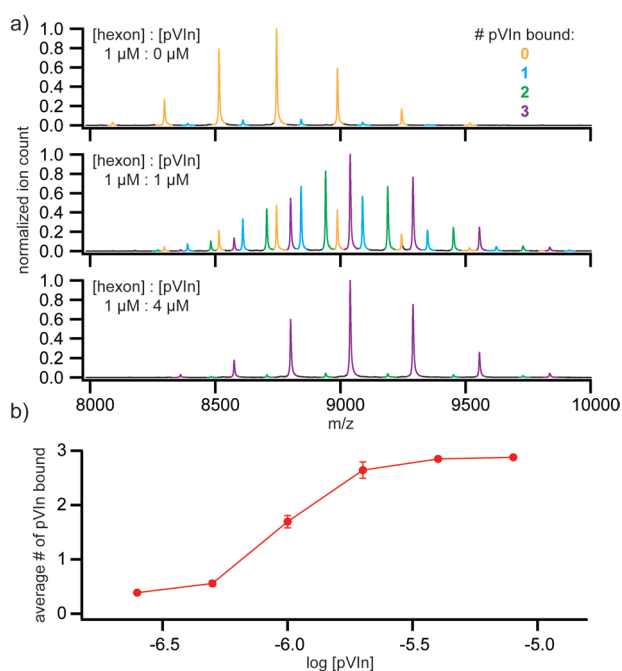


Figure 3. Reconstitution of hexon-pVIn using purified hexon and synthetic pVIn. a) Native MS spectra of hexon-pVIn mixtures. Note that some residual endogenous pVIn is already bound to the purified hexon. b) Average number of pVIn bound to hexon trimer, as a function of pVIn concentration. Points represent average \pm standard deviation from duplicate experiments.

When the released fraction was transferred to a pH 5 solution, nearly complete binding was observed, i.e. 3 pVIn molecules/hexon trimer. At pH 9, however, only free hexon trimers were recovered and detected by native MS. The sensitivity of the interaction to pH was utilized to isolate the bound pVIn peptide from recovered hexon trimers. LC-MS/MS analysis of the isolated peptide confirmed that it is indeed pVIn, as the full sequence could be uniquely defined from the observed fragment ions (see Supplementary Figures S1 and S2). Release of pVIn during partial capsid disassembly might suggest a possible role in endosome escape. However, no direct membrane lytic activity of pVIn was detected on SulfoB- loaded liposomes (see Supplementary Figure S3). We cannot exclude the possibility that pVIn might have an indirect role in endosome escape or perhaps in subsequent events in HAdV infection.

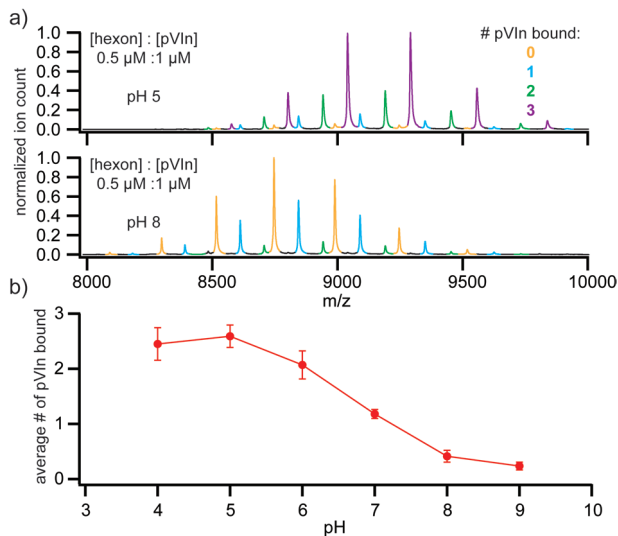


Figure 4. Strength of the hexon-pVIn interaction is pH-dependent. A) Native MS of 1:2 hexon:pVIn mixtures at the indicated pH. B) Average number of pVIn (1 μ M) bound to hexon trimer (0.5 μ M) as a function of pH. Points represent average \pm standard deviation from triplicate experiments.

The hexon-pVIn interaction is strongest at endosomal pH.

The observation that more hexon-pVIn complex could be recovered at acidic pH than at neutral or high pH strongly suggests that the affinity of the interaction is pH dependent. To test this more systematically, we reconstituted the hexon-pVIn complex from purified hexons and synthetic pVIn (see Figure 3). In a titration experiment, where the relative intensities of the differently bound forms are used as a measure of binding, we observed increased binding of pVIn with increasing concentration. Up to, but no more than, three copies of pVIn were bound to hexon trimers, even when pVIn was added in high excess, indi-

cating that the interaction is highly specific. We next tested the fraction of bound pVIn in a 1:2 mixture of hexon:pVIn in the pH-range of 4-9 (see Figure 4). We observed half-maximum binding at pH 6.5-7 and near-complete dissociation towards pH 8. These findings indicate that the interaction is strongest at endosomal pH but relatively weak at cytosolic pH. There are several basic residues in the sequence of pVIn, including one histidine at position 13, which may contribute to the observed effect of pH on the hexon-pVIn interaction.

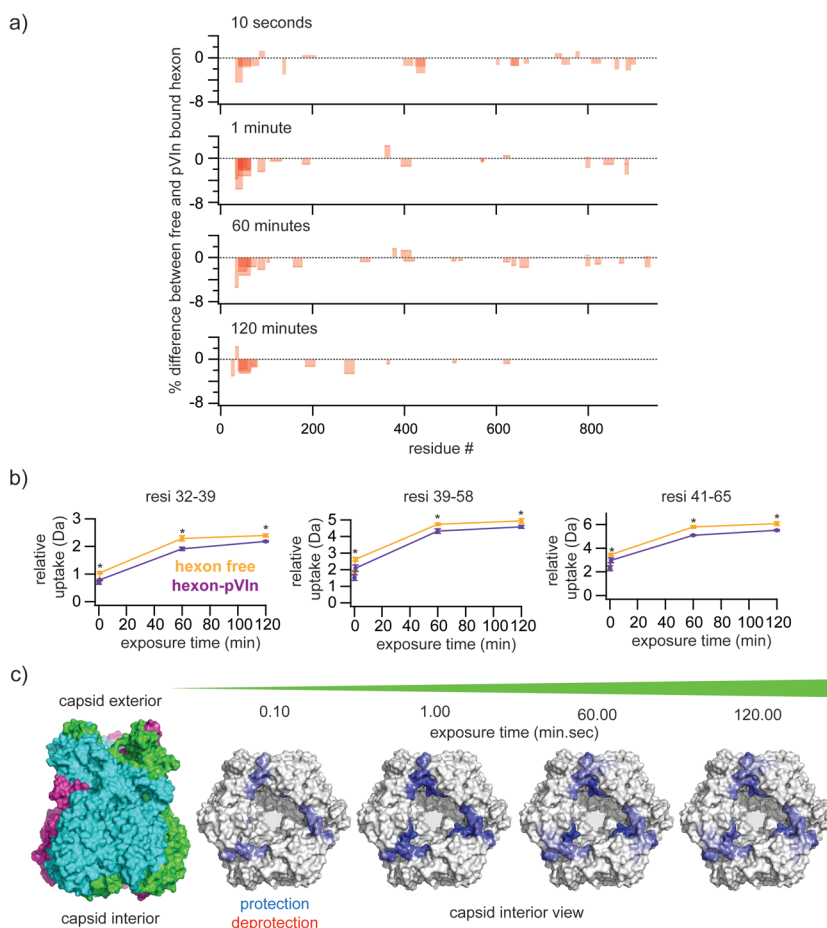


Figure 5. The pVIn binding site of hexon is located within the region spanning residues 32-65 at the base of the hexon cavity. A) Overview of peptides that exhibit significantly changed deuterium uptake between free and pVIn-bound hexon. Bars are transparent, such that darker regions represent overlapping peptides. B) Deuterium uptake as a function of time of peptides displaying a consistent and statistically relevant differential uptake between free hexon and hexon-pVIn assemblies. Points represent average \pm standard deviation from triplicate experiments. Asterisks indicate that $p < 0.05$ in unpaired two-tailed Student's t-test between two states at that time point. C) Changes in deuterium uptake mapped to the crystal structure of the hexon (PDB code: 1p30).

pVIn binds in the cavity of the hexon trimer.

In an attempt to localize the binding site of pVIn on the hexon trimer, we performed an HDX-MS experiment to monitor the changes in deuterium uptake in the hexon trimer upon binding of pVIn. After incubation of free hexon or hexon-pVIn in D_2O , the reaction was quenched, followed by online pepsin digestion and LC-MS/MS analysis. Peptides were identified from hexon in MilliQ, yielding 236 peptides of sufficient quality for quantifying deuterium uptake, covering 95% of the full hexon sequence (see Supplementary Figure S4). Exposure was carried out for 10 seconds and 1, 60 and 120 minutes in triplicate and changes in deuterium uptake were filtered to $p < 0.05$ using unpaired two-tailed Student's t-tests. Out of the over 200 peptides analyzed, we observed three peptides that showed a consistent, statistically significant change between the free hexon and hexon-pVIn (See Figure 5). For comparison, the deuterium uptake of the peptides that are affected by pVIn binding is shown alongside representative unaffected peptides in Supplementary Figure S5. The three peptides that exhibit protection upon pVIn binding localize to the same region of the hexon sequence and collectively span amino acid residues 32-65 of the hexon. These residues are in the extended N-terminal domain that interacts with the neighboring hexon monomer of the trimer, and line the rim of the hexon cavity on the capsid interior (Figure 5). These results provide evidence that residues 32-65 span the pVIn-binding region of the hexon.

Discussion.

Using a variety of primarily MS-based methods, we show here that the N-terminus of pVI, pVIn, is part of mature HAdV particles following cleavage by AVP. We demonstrate that pVIn interacts with residues 32-65 of the hexon, located near the rim of the cavity of the hexon trimer on the capsid interior. Full-length pVI binds with residues 48-74 and 233-239 (in addition to residues from pVIn) (17) to hexon trimers with high affinity (15). Our experiments indicate that the affinity of the hexon pVIn interaction is highest at endosomal pH but decreases significantly at higher pH. Given the pH-dependence of the interaction, it is perhaps surprising that pVIn stays bound to hexon, despite the relatively high pH of the extracellular environment, cytosol and nucleus. Considering the fact that pVIn is enclosed in the small inner volume of the capsid, local pH and concentration is hard to estimate. However, we expect that the local concentrations of both pVIn and hexon are effectively high, especially considering the volume that is excluded by the densely packed genome and additional core components of the virion. Partial disassembly of HAdV in early endosomes would increase the

volume in which pVIn is free to diffuse, thereby effectively decreasing the local concentrations and possibly allowing release of free pVIn into the endosome. However, we found that at pH 5, endogenous pVIn appears to still be fully retained on hexon trimers. If pVIn is retained on hexon trimers in the endosome, the transition to higher pH (~ 7.4) following endosome escape likely results in dissociation of the hexon-pVIn complex. It is currently unclear whether pVIn has any effect on other steps in the replication cycle of HAdV.

The exact copy number of pVI and for that matter pVIn is currently not well defined. The near-complete occupancy of endogenous pVIn-binding sites on hexons released from heated virions would suggest that the copy number is on the order of 180, assuming that the hexons that are recovered in our experiment are all peripentonal, and that no additional pVIn is present in heated particles (approximately 80% is released according to our LC-MS/MS analysis). Our native MS data do show that hexon trimers with sub-stoichiometric pVIn are also stable configurations, and it is thus possible that the actual copy number is lower. This estimate of ~180 copies of pVIn in mature virions also suggests that a subset of pVIn molecules are lost prior to full maturation, as the VI copy number is estimated to be 340-360. On the other hand, it is possible that additional copies of pVIn dissociate from heated particles without being bound to hexon, and would therefore not be detected in our native MS experiment, in which case the actual copy number of pVIn might be higher.

It is interesting to note that the purified hexons in our experiments had a very small residual amount of endogenous pVIn bound (see Figure 2a, approximately 3% of maximum occupancy). These hexons, isolated from virus-infected cells, are produced in excess of what is actually assembled into complete virions, hence suggesting that (partial) processing of pVI can possibly take place outside of the assembled virion, or that these represent “debris” from broken/disassembled virions.

The binding site of pVIn that we observed in our HDX-MS experiments had not been previously identified as a pVI binding site on hexons. Two cryo-electron microscopy reconstructions of the maturation defective *ts1* mutant of HAdV did show additional density in the hexon cavity, which was tentatively assigned to pVI, but no additional density in the vicinity of residues 32-65 of the hexon monomer has been reported (36, 37). This small 33 residue portion of pVI might simply not have been visible at the moderate resolution of these cryo-EM reconstructions, but rearrangement of pVIn upon cleavage is also a possible scenario.

In conclusion, we have shown that pVI_n is a component of mature HAdV virions. The fragment resides in the cavity of hexon trimers and is released in an *in vitro* model of partial capsid disassembly. The interaction of pVI_n with hexon points to a role for this peptide in AdV assembly. This is supported by the high sequence conservation of both the pVI_n fragment and hexon (residues 32-65) from both human and non-human AdVs. We speculate that cleavage by AVP between residues 33/34 of pVI is necessary for efficient release of the membrane lytic VI molecule in the endosome during entry. This might contribute to the uncoating defect of the maturation defective ts1 mutant. In this regard, mutagenesis of the AVP cleavage site for pVI_n in a WT HAdV background should clarify the role of AVP cleavage of pVI in efficient endosome escape. In the context of mature HAdV particles, pVI_n might simply be a remnant of hexon-pVI assembly, that is cleaved from pVI for efficient release of the mature product VI.

Methods.

Purification of Ad5F35 and hexon.

Replication-defective Ad5F35 and Ad5 hexon, were grown and isolated as previously described with some modification (38, 39). Briefly, a Nunc cell factory of 293β5 cells was infected with 300-500 Ad5F35 particles/cell (28). Between 48-60 hours post-infection, cells were harvested and resuspended in 8 ml of 10 mM Bis-Tris pH 6.5 and flash-frozen in liquid nitrogen. Cells were thawed at 37°C and 1 ml of Freon-113 (Sigma) was added. Complete lysis was achieved with two additional rounds of freeze/thaw. Clarified lysates were separated on 15-40% cesium chloride density gradients and the hexon and virus fractions were collected. Virus was dialyzed into DX10/10 buffer (40 mM Tris pH 8.1, 500 mM NaCl, 10% glycerol, 10% ethylene glycol, 2% sucrose and 1% mannitol). Hexon was dialyzed into 10 mM Bis-Tris pH 7.0 and further purified by ion exchange on a ResourceQ (GE Healthcare) column in 10 mM Bis-Tris pH 7.0 with a multi-step NaCl gradient to 1 M. Hexon fractions were pooled and concentrated in an Amicon centrifugal filter unit (Millipore) to 7-12 mg/ml. Aliquots of purified virus and hexon were frozen in liquid nitrogen and stored at -80°C until use.

In vitro partial capsid disassembly.

Large-scale partial disassembly of HAdV capsids suitable for native mass spectrometry was adapted from previously published protocols (31, 40). Ad5F35 (5 mg) was diluted in 7.5 mM HEPES pH 7.4, 50 mM NaCl to a final volume of 3 ml. To ensure complete disassembly, the virus was divided into six 500 μl aliquots and heated to 55°C for 15 min. The aliquots were pooled together and loaded

onto four 30-80% discontinuous Histodenz (Sigma) gradients dissolved in 20 mM Tris pH 7.4 and centrifuged at 198,000xg (ave) for 3 hrs. The supernatant fraction, containing the subset of Ad5F35 proteins released from the virions, was collected with a pipette from the top of the gradient in two 1 ml fraction, which were subsequently pooled together. Samples were stored frozen until analysis.

Membrane lytic activity of pVIn.

Synthetic pVIn peptide was dissolved at high concentration (>20 mg/ml) in MilliQ water. Recombinant protein VI114 (residues 34-114) was purified as described (41) and used as a positive control for membrane lytic activity. The L40Q mutant of VI114 has decreased membrane lytic activity compared to the wild-type protein and is used to show an intermediate phenotype (40). Bovine serum albumin (BSA, Sigma) was used as a negative control. Serial dilutions of protein were done in 50 mM Tris pH 8.0, 150 mM NaCl. The membrane lytic activity of pVIn and VI114 was measured in a liposome lysis assay as previously described (41).

Proteomics of mature Ad5F35.

Both the intact and heated HAdV particles were prepared for LC-MS/MS analysis by resuspending them in 50 mM ammonium bicarbonate, 5% (w/v) sodium deoxycholate (SDC, Sigma Aldrich) and heating at 90°C for 5 minutes. Proteins (~200 µg) for each enzymatic digestion, were first reduced for 30 minutes at 56°C and then alkylated for 1h in the dark. Proteins were first digested with Lys-C (Roche) for 4h at 37°C. After diluting the samples to a final SDC concentration of 0.5% (w/v), enzymatic digestion was performed overnight at 37°C adding Trypsin (Promega) in a substrate/enzyme ratio of 50:1 (w/w). Digestion was quenched by acidification with Formic acid (FA) to a final concentration of 10% and peptides were desalted by solid phase extraction (Sep-pack Vac C18 cartridges, Waters). Peptides were eluted in 80% acetonitrile, dried in a speedvac and then resuspended in 10% formic acid solution. Each adenovirus sample was analyzed by LC-MS/MS in triplicate. An EASY-nLC 1000 (Thermo Fischer Scientific) was equipped with a 20 mm Aqua C18 (Phenomenex) trapping column (packed in-house, 100 µm i.d., 5 µm particle size) and a 400 mm Zorbax SB-C18 (Agilent) analytical column (packed in-house, either 50 µm i.d. or 75 µm i.d., 1.8 µm particle size). Trapping and washing was performed at 10 µl/min for 4 min with solvent A (0.1 M acetic acid in water). Subsequently, peptides were transferred to the analytical column at about 150 nl/min and eluted with a gradient of 3-40% (v/v) solvent B (0.1 M acetic acid in 80% ACN) in 45 min. The eluent was sprayed via distal coated emitter tips butt-connected to the analytical column and the ion spray voltage was set to 1.7 kV. The mass spectrometers were operat-

ed in data-dependent mode, automatically switching between MS and MS/MS. The full-scan MS spectra (from m/z 350 to 1500) were acquired in the Orbitrap analyzer with a resolution of 60,000 FHMW at 400 m/z after accumulation to target value of $1e6$ in the linear ion trap (maximum injection time was 250 ms). After the survey scans, the ten most intense precursor ions at a threshold above 5000 were selected for MS/MS with an isolation width of 1.5 Da. Peptide fragmentation was carried out by using a decision tree performing higher collision dissociation (HCD) or electron transfer dissociation (ETD) depending on their charge state and m/z . HCD fragment ions readout was performed in the orbitrap analyzer with a resolution of 15,000 FHMW, activation time was of 0.1 ms and normalized collision energy (CE) of 32. ETD fragment ions readout was performed in the linear ion trap analyzer with an activation time of 50 ms.

MS raw data from the shotgun LC-MS/MS analyses were processed by Proteome Discoverer (version 1.3, Thermo Electron). Peptide identification was performed with Mascot 2.3 (Matrix Science) against a concatenated forward-decoy UniPROT database including the HAdV protein sequences and supplemented with all the frequently observed contaminants in MS. The following parameters were used: 6 ppm precursor mass tolerance, 0.6 Da (for ETD) and 0.05 Da (for HCD) fragment ion tolerance. In order to evaluate the viral protease activity we allowed the identification of peptides containing one non-specific cleavage and a maximum of 2 missed cleavages. Carbamidomethyl cysteine was allowed as fixed modification, while oxidized methionine and protein N-terminal acetylation were set as variable modifications. Finally, results were filtered using the following criteria: (i) mass deviations of ± 5 p.p.m., (ii) Mascot Ion Score of at least 20, (iii) a minimum of 6 amino-acid residues per peptide and (iv) position rank 1 in Mascot search. As a result, we obtained peptide FDRs below 1% for all the three peptide mixtures analyzed. Relative quantification was performed by spectral counts, which relies on the number of peptide spectrum matches (PSMs) specific for a certain protein (or protein domain). As an internal control for data normalization, we used the two DNA interacting proteins pVII and pTP as housekeeping proteins.

Native MS of hexon-pVIn.

The supernatant fraction from heat-disrupted virion density centrifugation or purified hexon trimer were buffer exchanged to 150 mM ammonium acetate at the indicated pH using Vivaspinn 500 10 kDa MWCO centrifugal filter units. Reconstituted hexon-pVIn from synthetic pVIn was prepared by mixing hexon in ammonium acetate with pVIn dissolved in MilliQ water. Samples were loaded into gold-coated boro-silicate capillaries, prepared in-house, for nano-electros-

pray ionization. Mass spectrometry was performed on a QTOF II instrument (Waters), modified for optimal transmission of high-mass ions (MS Vision) (42, 43). The instrument was operated at 10 mbar in the source region, and 1.5×10^{-2} mbar in the collision cell using xenon as collision gas. Capillary voltage was set at 1400-1500 V, sample cone voltage at 160 V. Hexon-pVIn complexes were analyzed at a collision voltage of 100 V, and tandem MS of hexon-pVIn was performed at a collision voltage of 120 V. Spectra were calibrated with cesium iodide.

Isolation and identification of endogenous pVIn from hexon-pVIn complexes.

We utilized the pH-dependence of the hexon-pVIn interaction to isolate pVIn from the complex. Hexon-pVIn was trapped on C18 ZipTip resin. The trapped complex was dissociated with a washing step, using 150 mM ammonium acetate pH 9, thereby trapping free pVIn on the C18 resin. After one additional washing step with 0.1% formic acid in water, pVIn was eluted from the resin with 80% acetonitrile/0.1% formic acid in water, while hexon was retained on the ZipTip. The sample was dried in a speedvac and redissolved in 10% formic acid/5% DMSO in water. The sample was analyzed by reversed phase LC coupled on-line to an LTQ or LTQ-Orbitrap Velos II for MS/MS analysis. The nano-LC consists of an Agilent 1200 series LC system equipped with a 20 mm ReproSil-Pur C18-AQ (Dr. Maisch GmbH) trapping column (packed in-house, i.d., 100 μ m; resin, 5 μ m) and a 400 mm ReproSil-Pur C18-AQ (Dr. Maisch GmbH) analytical column (packed in-house, i.d., 50 μ m; resin, 3 μ m). The flow was passively split to 100 nl/min. We used a standard 45 minute gradient from 7-30% acetonitrile in aqueous 0.1% formic acid. For MS/MS analysis on the LTQ-Orbitrap XL, all precursors were sequenced with both CID and ETD fragmentation and MS analysis in the ion trap. The data were searched with Mascot against a custom database containing all HAdV5 protein sequences, except for the fiber, which was replaced with the HAdV35 fiber sequence.

HDX-MS of hexon-pVIn complexes.

Free hexon or hexon-pVIn (at a 1:4 molar ratio) was prepared at a concentration of 10 μ M in 150 mM ammonium acetate pH 5. At the start of the exchange reaction, the sample was diluted six-fold into D₂O to a final volume of 30 μ L. Exposure was carried out for 10 seconds, 1, 60 and 120 minutes in triplicate at 25°C. The reaction was quenched by 1:1 mixing with ice-cold 4 M urea, 200 mM TCEP and adjusted with HCl to give a final pH upon mixing of 2.5. Immediately after quenching, the sample was injected into a Waters HDX/ nanoAcquity system for digestion on an online pepsin column (25 °C, at a flow-rate of 50 μ L.min⁻¹) followed by separation on a 10 minute RP-UPLC gradient at 0 °C and MS on a

Waters Xevo QTOF G2. For peptide identification, hexon was prepared under identical conditions in H₂O and analyzed using MSe data acquisition. Data for peptide identification was processed with ProteinLynx Global Server 2.5 software. Deuterium uptake was calculated compared to the control samples in H₂O using Waters DynamX 1.0.0 software. Back-exchange was estimated at approximately 30% in our workflow. No corrections for back-exchange were applied since the analysis focuses on relative changes in deuterium uptake, rather than on absolute levels. Observed changes in deuterium uptake were filtered to $p < 0.05$ in an unpaired two-tailed Student's t-test between free and bound hexon.

Acknowledgements.

We thank Tina-Marie Mullen for technical assistance. These studies were supported in part by NIH grants HL054352 to G.R.N. and 5T32 AI007354 for C.L.M. J.S., M.B, and A.J.R.H. are supported by the Netherlands Proteomics Centre, and by the European Community's Seventh Framework Programme (FP7/2007-2013) by the PRIME-XS project grant agreement number 262067.

Supplementary information.

Supplementary information can be found in the appendix.

References.

1. Kay, M. A. 2011. State-of-the-art gene-based therapies: The road ahead. *Nature Reviews Genetics*. 12:316-328.
2. Small, J. C., and H. C. J. Ertl. 2011. Viruses - From pathogens to vaccine carriers. *Current Opinion in Virology*. 1:241-245.
3. Reddy, V. S., S. K. Natchiar, P. L. Stewart, and G. R. Nemerow. 2010. Crystal structure of human adenovirus at 3.5 Å resolution. *Science*. 329:1071-1075.
4. Liu, H., L. Jin, S. B. S. Koh, I. Atanasov, S. Schein, L. Wu, and Z. H. Zhou. 2010. Atomic structure of human adenovirus by Cryo-EM reveals interactions among protein networks. *Science*. 329:1038-1043.
5. Nemerow, G. R., P. L. Stewart, and V. S. Reddy. 2012. Structure of human adenovirus. *Current Opinion in Virology*. 2:115-121.
6. Martín, C. S. 2012. Latest insights on adenovirus structure and assembly. *Viruses*. 4:847-877.
7. Diouri, M., H. Keyvani-Amineh, K. F. Geoghegan, and J. M. Weber. 1996. Cleavage efficiency by adenovirus protease is site-dependent. *J. Biol. Chem*. 271:32511-32514.
8. Mangel, W. F., W. J. McGrath, D. L. Toledo, and C. W. Anderson. 1993. Viral DNA and a viral peptide can act as cofactors of adenovirus virion proteinase activity. *Nature*. 361:274-275.
9. Webster, A., R. T. Hay, and G. Kemp. 1993. The adenovirus protease is activated by a virus-coded disulphide-linked peptide. *Cell*. 72:97-104.
10. Blainey, P. C., V. Graziano, A. J. Pérez-Berná, W. J. McGrath, S. J. Flint, C. S. Martín, X. S. Xie, and W. F. Mangel. 2013. Regulation of a viral proteinase by a peptide and DNA in one-dimensional space IV: Viral proteinase slides along dna to locate and process its substrates. *J. Biol. Chem*. 288:2092-2102.
11. Graziano, V., G. Luo, P. C. Blainey, A. J. Pérez-Berná, W. J. McGrath, S. J. Flint, C. S. Martín, X. S. Xie, and W. F. Mangel. 2013. Regulation of a viral proteinase by a peptide and DNA in one-dimensional space: II. adenovirus proteinase is activated in an unusual one-dimensional biochemical reaction. *J. Biol. Chem*. 288:2068-2080.
12. Baniecki, M. L., W. J. McGrath, S. M. McWhirter, C. Li, D. L. Toledo, P. Pellenca, D. L. Barnard, K. S. Thorn, and W. F. Mangel. 2001. Interaction of the human adenovirus proteinase with its 11-amino acid cofactor pVIc. *Biochemistry (N. Y.)*. 40:12349-12356.

13. Mangel, W. F., D. L. Toledo, M. T. Brown, J. H. Martin, and W. J. McGrath. 1996. Characterization of three components of human adenovirus proteinase activity in vitro. *J. Biol. Chem.* 271:536-543.
14. McGrath, J. W., M. L. Baniecki, C. Li, S. M. McWhirter, M. T. Brown, D. L. Toledo, and W. F. Mangel. 2001. Human adenovirus proteinase: DNA binding and stimulation of proteinase activity by DNA. *Biochemistry (N. Y.)*. 40:13237-13245.
15. Graziano, V., W. J. McGrath, M. Suomalainen, U. F. Greber, P. Freimuth, P. C. Blainey, G. Luo, X. S. Xie, and W. F. Mangel. 2013. Regulation of a viral proteinase by a peptide and DNA in one-dimensional space: I. binding to dna and to hexon of the precursor to protein VI, pVI, of human adenovirus. *J. Biol. Chem.* 288:2059-2067.
16. Baniecki, M. L., W. J. McGrath, and W. F. Mangel. 2013. Regulation of a viral proteinase by a peptide and DNA in one-dimensional space: III. atomic resolution structure of the nascent form of the adenovirus proteinase. *J. Biol. Chem.* 288:2081-2091.
17. Matthews, D. A., and W. C. Russell. 1995. Adenovirus protein-protein interactions: Molecular parameters governing the binding of protein VI to hexon and the activation of the adenovirus 23K protease. *J. Gen. Virol.* 76:1959-1969.
18. Wodrich, H., T. Guan, G. Cingolani, D. Von Seggern, G. Nemerow, and L. Gerace. 2003. Switch from capsid protein import to adenovirus assembly by cleavage of nuclear transport signals. *EMBO J.* 22:6245-6255.
19. Perez-Berna, A. J., A. Ortega-Esteban, R. Menendez-Conejero, D. C. Winkler, M. Menendez, A. C. Steven, S. J. Flint, P. J. de Pablo, and C. San Martin. 2012. The Role of Capsid Maturation on Adenovirus Priming for Sequential Uncoating. *J. Biol. Chem.* 287:31582-31595.
20. Mathias, P., M. Galleno, and G. R. Nemerow. 1998. Interactions of soluble recombinant integrin $\alpha\beta 5$ with human adenoviruses. *J. Virol.* 72:8669-8675.
21. Lindert, S., M. Silvestry, T. -. Mullen, G. R. Nemerow, and P. L. Stewart. 2009. Cryo-electron microscopy structure of an adenovirus-integrin complex indicates conformational changes in both penton base and integrin. *J. Virol.* 83:11491-11501.
22. Stewart, P. L., and G. R. Nemerow. 2007. Cell integrins: commonly used receptors for diverse viral pathogens. *Trends Microbiol.* 15:500-507.
23. Wickham, T. J., P. Mathias, D. A. Cheresch, and G. R. Nemerow. 1993. Integrins $\alpha(v)\beta 3$ and $\alpha(v)\beta 5$ promote adenovirus internalization but not virus attach-

ment. *Cell*. 73:309-319.

24. Chiu, C. Y., P. Mathias, G. R. Nemerow, and P. L. Stewart. 1999. Structure of adenovirus complexed with its internalization receptor, $\alpha(v)\beta 5$ integrin. *J. Virol.* 73:6759-6768.

25. Nakano, M. Y., K. Boucke, M. Suomalainen, R. P. Stidwill, and U. F. Greber. 2000. The first step of adenovirus type 2 disassembly occurs at the cell surface, independently of endocytosis and escape to the cytosol. *J. Virol.* 74:7085-7095.

26. Burckhardt, C. J., M. Suomalainen, P. Schoenenberger, K. Boucke, S. Hemmi, and U. F. Greber. 2011. Drifting motions of the adenovirus receptor CAR and immobile integrins initiate virus uncoating and membrane lytic protein exposure. *Cell Host and Microbe*. 10:105-117.

27. Snijder, J., V. S. Reddy, E. R. May, W. H. Roos, G. R. Nemerow, and G. J. L. Wuite. 2013. Integrin and defensin modulate the mechanical properties of adenovirus. *J. Virol.* 87:2756-2766.

28. Smith, J. G., M. Silvestry, S. Lindert, W. Lu, G. R. Nemerow, and P. L. Stewart. 2010. Insight into the mechanisms of adenovirus capsid disassembly from studies of defensin neutralization. *PLoS Pathogens*. 6:e1000959.

29. Wiethoff, C. M., H. Wodrich, L. Gerace, and G. R. Nemerow. 2005. Adenovirus protein VI mediates membrane disruption following capsid disassembly. *J. Virol.* 79:1992-2000.

30. Schreiner, S., R. Martinez, P. Groitl, F. Rayne, R. Vaillant, P. Wimmer, G. Bossis, T. Sternsdorf, L. Marcinowski, Z. Ruzsics, T. Dobner, and H. Wodrich. 2012. Transcriptional activation of the adenoviral genome is mediated by capsid protein VI. *PLoS Pathogens*. 8:e1002549.

31. Smith, J. G., and G. R. Nemerow. 2008. Mechanism of Adenovirus Neutralization by Human α -Defensins. *Cell Host and Microbe*. 3:11-19.

32. Rexroad, J., C. M. Wiethoff, A. P. Green, T. D. Kierstead, M. O. Scott, and C. R. Middaugh. 2003. Structural stability of adenovirus type 5. *J. Pharm. Sci.* 92:665-678.

33. Russell, W. C., R. C. Valentine, and H. G. Pereira. 1967. The effect of heat on the anatomy of the adenovirus. *J. Gen. Virol.* 1:509-522.

34. Van Den Heuvel, R. H. H., and A. J. R. Heck. 2004. Native protein mass spectrometry: From intact oligomers to functional machineries. *Curr. Opin. Chem. Biol.* 8:519-526.

35. Sharon, M. 2013. Structural MS pulls its weight. *Science*. 340:1059-1060.

36. Pérez-Berná, A. J., R. Marabini, S. H. W. Scheres, R. Menéndez-Conejero, I. P. Dmitriev, D. T. Curiel, W. F. Mangel, S. J. Flint, and C. S. Martín. 2009. Structure and Uncoating of Immature Adenovirus. *J. Mol. Biol.* 392:547-557.
37. Silvestry, M., S. Lindert, J. G. Smith, O. Maier, C. M. Wiethoff, G. R. Nemerow, and P. L. Stewart. 2009. Cryo-electron microscopy structure of adenovirus type 2 temperature-sensitive mutant 1 reveals insight into the cell entry defect. *J. Virol.* 83:7375-7383.
38. Reddy, V. S., S. K. Natchiar, L. Gritton, T. -. Mullen, P. L. Stewart, and G. R. Nemerow. 2010. Crystallization and preliminary X-ray diffraction analysis of human adenovirus. *Virology.* 402:209-214.
39. Rux, J. J., and R. M. Burnett. 2007. Large-scale purification and crystallization of adenovirus hexon. *Methods Mol. Med.* 131:231-250.
40. Moyer, C. L., C. M. Wiethoff, O. Maier, J. G. Smith, and G. R. Nemerow. 2011. Functional genetic and biophysical analyses of membrane disruption by human adenovirus. *J. Virol.* 85:2631-2641.
41. Moyer, C. L., and G. R. Nemerow. 2012. Disulfide-bond formation by a single cysteine mutation in adenovirus protein VI impairs capsid release and membrane lysis. *Virology.* 428:41-47.
42. Van Den Heuvel, R. H. H., E. Van Duijn, H. Mazon, S. A. Synowsky, K. Lorenzen, C. Versluis, S. J. J. Brouns, D. Langridge, J. Van Der Oost, J. Hoyes, and A. J. R. Heck. 2006. Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. *Anal. Chem.* 78:7473-7483.
43. Lorenzen, K., C. Versluis, E. van Duijn, R. H. H. van den Heuvel, and A. J. R. Heck. 2007. Optimizing macromolecular tandem mass spectrometry of large non-covalent complexes using heavy collision gases. *International Journal of Mass Spectrometry.* 268:198-206.

Chapter 5:

Probing the biophysical interplay between a viral genome and its capsid.

J. Snijder^{1,2,3}, C. Uetrecht^{1,2†}, R.J. Burnley^{1,2}, R. Sanchez-Eugenía⁴, G.A. Martí⁵, J. Agirre^{4,6}, D.M.A. Guérin^{4,6}, G.J.L. Wuite³, A.J.R. Heck^{1,2}, W.H. Roos³

¹ Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH, Utrecht, the Netherlands.

² Netherlands Proteomics Centre, Padualaan 8, 3584 CH, Utrecht, The Netherlands.

³ Natuur- en Sterrenkunde and LaserLab, Vrije Universiteit, De Boelelaan 1081, Amsterdam, The Netherlands

⁴ Unidad de Biofísica (CSIC, UPV-EHU) B° Sarriena S/N, 48940 Leioa, Bizkaia, Spain.

⁵ Centro de Estudios Parasitológicos y de Vectores (CEPAVE-CCT-La Plata-CONICET-UNLP). 2#584 (1900) La Plata, Argentina.

⁶ Fundación Biofísica Bizkaia. B° Sarriena S/N, 48940 Leioa, Bizkaia, Spain.

Based on:

Snijder, J., *et al.* Probing the biophysical interplay between a viral genome and its capsid. *Nature Chemistry* 5, 502-509 (2013).

Abstract.

The interaction between a viral capsid and its genome governs crucial steps in the life cycle of a virus such as assembly and genome uncoating. Tuning cargo-capsid interactions is also essential for successful design and cargo delivery in engineered viral systems. Here we investigate the interplay between cargo and capsid for the picorna-like *Triatoma* Virus using a combined native mass spectrometry and atomic force microscopy approach. We propose a topology and assembly model in which hetero-trimeric pentons consisting of 5 copies of structural proteins VP1, VP2 and VP3 are the free principal units of assembly. The inter-penton contacts are primarily established by VP2. The dual role of the genome is first to stabilize the densely packed virion and, upon an increase in pH, to trigger uncoating by relaxing the stabilizing interactions with the capsid. Uncoating occurs through a labile intermediate state of the virion which reversibly disassembles into pentons with concomitant release of protein VP4.

In addition to coding for viral genes, the genome of a virus plays an essential role through its interactions with the viral capsid. The genome often facilitates capsid assembly and once packaged, the genome stabilizes and reinforces the virus (1). This interplay between the capsid and its cargo is thereby a crucial factor in designing engineered viral systems (2). Hence, increased knowledge of genome-capsid interactions can enhance our understanding of viruses in their role as pathogens, drug-delivery vehicles and nanotechnological platforms. However, details of genome release (i.e. uncoating) and virion assembly are not easily accessible experimentally, due to the heterogeneity and transient nature of several co-occurring intermediate states. Most analytical and biochemical assays are unable to resolve these small sub-populations. The study of virus assembly and uncoating thus requires techniques that are capable of simultaneous detailed structural and physical characterization of components in complex mixtures.

In recent years, both native Mass Spectrometry (MS) and Atomic Force Microscopy (AFM) have emerged as alternative, powerful platforms for the study of viruses (3-6). Where more established structural biology techniques suffer, they effectively capture assembly intermediates, exposing details of viral assembly and maturation pathways. Key to these techniques is the high mass selectivity of MS and the single particle approach of AFM. We previously implemented Ion Mobility Spectrometry (IMS) to characterize the geometry of *in vitro* reconstituted capsid assembly intermediates (7). We demonstrate here that the combination of this IMS methodology with tandem MS analysis yields rich structural information on authentic virions, their genome-free capsid counterparts and the assembly intermediates. Essential aspects of the structural model are confirmed and further examined by high-resolution AFM imaging. Additionally, we apply AFM in its force spectroscopy mode to characterize in detail the process of genome uncoating.

Employing this powerful combination of native MS and AFM, we study the picorna-like Triatoma Virus (TrV). The ssRNA insect virus TrV is a member of the *Dicistroviridae* family and is closely related to common human pathogens such as Polio Virus (8). Infection is lethal to the insect host of TrV, *Triatoma infestans* (9, 10). As these triatomine insects are the most important vectors for Chagas disease (11), which was recently recognized by the WHO as one of several neglected tropical diseases with a considerable strain on public health in affected regions, TrV has potential applications as a biopesticide (10). TrV consists of a 9010 nt positive sense ssRNA genome that is encapsidated by 60 copies of the structural proteins VP2, VP4, VP3 and VP1 (from N-terminus to C-terminus on the precursor P1) (8, 12, 13). VP1/2/3 are the main structural proteins

of the capsid (PDB code: 3NAP), whereas the role of VP4 is currently unclear (14). The capsid proteins VP1/2/3 have a jelly-roll fold, typical of picorna(-like) viruses and a high degree of structural homology to the related Cricket Paralysis Virus (15). Five copies of the VP1/2/3 heterotrimer are arranged in five-fold symmetric sub-structures (capsomeres, called ‘pentons’), twelve of which form the fully closed capsid. VP1 is centred on the five-fold icosahedral symmetry axis and protrudes out from the capsid surface. VP2 and VP3 are situated on quasi-three-fold and two-fold related symmetry axes, whereby VP2 establishes inter-penton contacts via a domain-swapping interaction between capsomeres.

The insect host of TrV feeds on blood, and is adapted to cope with the toxic effects of heme by maintaining acidic pH in the midgut (16-19). In contrast to the acidic conditions in the midgut of *T. infestans*, the hindgut is maintained at alkaline pH (17, 20). Since TrV infection takes place via an oral-fecal route, large pH changes may play an important role in infectivity (21). Here we assess assembly, stability and genome release in TrV as a function of pH. We determine the mass of the intact TrV virion and capsid with unprecedented precision and, using naturally occurring infectious particles, we uncover an *in vitro* pathway of alkaline-triggered uncoating. We show how this pathway is regulated by genome-capsid interactions and we present a model of the dual role of the genome in capsid stabilization and genome uncoating.

Results.

A pathway of alkaline-triggered genome release.

We tested the pH-stability of TrV using native MS. The theoretical (as predicted from the sequence) and experimentally measured masses of the individual VP’s and the assemblies detected here by MS can be found in Supplementary Table S1. Figure 1a shows native mass spectra of purified TrV virions at pH 5, 7, 8 and 9. The extracted samples contain, as well as the full genome-packed virions, a small amount of naturally occurring empty capsids (n-empty) that are thought to arise from misprocessing of P1. This misprocessing, which is predominantly inefficient cleavage of VP0 into VP3/4 (14), and the absence of RNA in n-empty capsids are the main differences between this capsid particle and the virion. Using MS the n-empty capsids (~5.4 MDa) can be well separated from the full virions (~8.3 MDa). The MS data reveal that whereas TrV is acid-stable, alkaline conditions do trigger release of the genome, as can be seen from the loss of virion signal. The transition from virion to empty capsid occurs predominantly between pH 8 and pH 9 (Figure 1b).

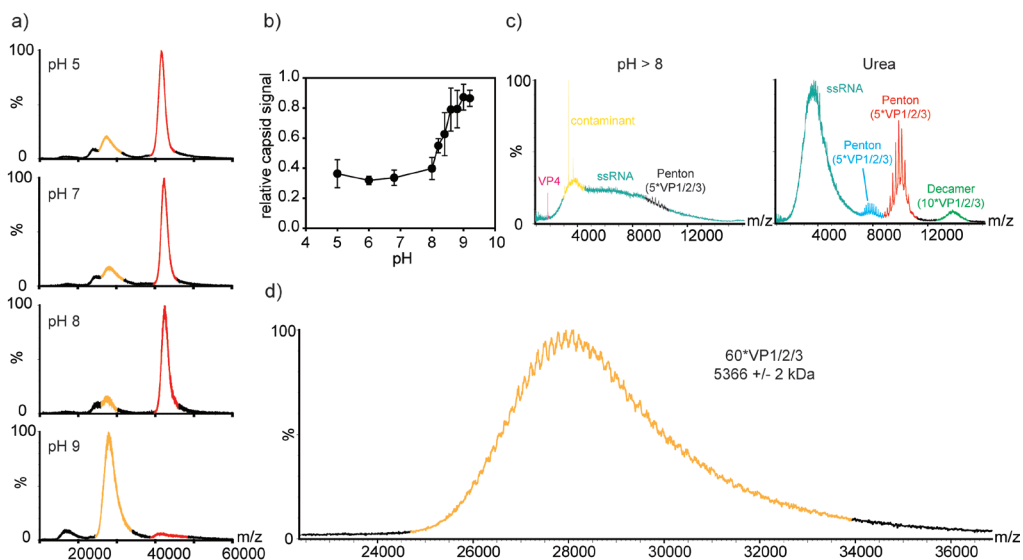


Figure 1. Alkaline-triggered uncoating of TrV monitored by native mass spectrometry. a) Spectra of TrV virions at different pH. Signal corresponding to virion is highlighted in red ($m/z \sim 40000$; $M_w \sim 8.3$ MDa), that to empty capsids in yellow ($m/z \sim 28000$; $M_w \sim 5.4$ MDa). Note that the shoulder on the capsid peak at pH 5-7, which shifts towards lower m/z at high pH is not an “authentic” ion signal, but a time-of-flight related artifact that arises for high mass complexes. It relates to the real signals of virion and capsid, respectively. b) Relative amount of capsid versus virion signal as a function of pH. Data points are the average of triplicate experiments, error bars represent SD. c) Species detected at lower m/z , associated with alkaline-triggered uncoating or denaturation. As color coded, ions originating from RNA, VP4 and pentons are detected. Pentons and penton-dimers are also the main disassembly products of urea-induced denatured virions. The additional charge state distribution of pentons in urea is indicative of unfolding. d) Mass spectrum of the r-empty capsid, with elevated collision voltage (180-200 V) enhancing desolvation.

In the low m/z range of a virion sample at pH >8, several disassembly products are detected (Figure 1c). The unresolved high baseline signals in these spectra originates from degradation products of the ssRNA genome, as validated by tandem MS (Supplementary Figure S1). There are additional peaks observed that originate from the VP4 protein. The samples also contain a 36 kDa contaminant that is already present under neutral conditions and therefore not related to genome release (Supplementary Figure S2). Finally, several peaks belonging to free pentons ($5^*VP1/2/3$) are detected. We hypothesized that the pentons are assembly intermediates that originate from disassembled virions but subsequently reassemble into empty capsids (denoted as “r-empty capsid” to distinguish them from the n-empty capsids). It was recently demonstrated for the related Bovine Enterovirus that pentons of VP proteins can assemble into empty capsids under

high ionic strength conditions (22). We confirm that pentons of TrV are indeed assembly intermediates of empty capsid by demonstrating that the alkaline-triggered uncoating reaction yields more free penton under low ionic strength conditions and that these pentons assemble into empty capsids when the ionic strength is subsequently increased (see Supplementary Figure S3). These results indicate that alkaline-triggered genome release occurs through reversible disassembly of the virion into pentons and is accompanied by release of VP4. Pentons are also formed, as well as some dimers of pentons, following denaturation of the virion in 8M urea. This confirms that the penton is a fundamental and very stable structural unit of the capsid. In contrast to the alkaline-induced pentons, urea-denatured penton sub-complexes are no longer assembly competent, as no intact capsid is detected after removal of urea (Supplementary Figure S4). The presence of disassembly products in high pH and after denaturation in urea could be further confirmed using native PAGE (Supplementary Figure S5).

Raising the MS collision voltage (180-200 V using xenon as collision gas) enhances desolvation (23) and here resulted in a clearly resolved series of charge states for the r-empty capsid (Figure 1d), allowing a precise mass determination of 5366 ± 2 kDa. This is in good agreement with a calculated mass of $60 \times \text{VP1}/2/3$ (the experimental mass deviates from the theoretical by +0.1%, likely due to incomplete desolvation and residual binding of small molecules and counter ions). The analysis is also highly reproducible as five replicate analyses result in an average mass of 5362 ± 2 kDa. This provides direct evidence that reassembly is complete, albeit without the incorporation of RNA or VP4. The mass assignment and exact stoichiometry of the r-empty capsids could be further confirmed with tandem MS at higher collision voltage (260 V) (Supplementary Figure S6).

We observed no differences in peak position of the n-empty capsid and remaining virion with increasing pH (virion peak $41,700 \pm 140$ m/z at pH 6.8 and $41,600 \pm 130$ m/z at pH 9.0; n-empty capsid 29420 ± 260 m/z at pH 6.8 and 29710 ± 200 m/z at pH 9.0) and no detectable disassembly products of the n-empty capsids at pH 9 (Figure 1a and Supplementary Figure S7). There is also no significant change in the auto-fluorescence of the capsid proteins at elevated pH (Supplementary Figure S8). Therefore, no conformational changes to the effect of a different chargeable surface area or changes in the vicinity of aromatic residues accompany genome release. Alkaline-triggered uncoating in TrV is thus uniquely triggered by the presence of the ssRNA. We also calculated the binding energies between VP1/2/3 protomers within and between pentons at pH 7 and pH 9 as modelled in the crystal structure of TrV (Supplementary Figure S9). The intramolecular interactions are stronger within a penton than between pentons

and there are no substantial changes with increasing pH. This both explains the occurrence of the penton as a fundamental structural unit of the capsid and confirms a unique role for the ssRNA in alkaline-triggered uncoating.

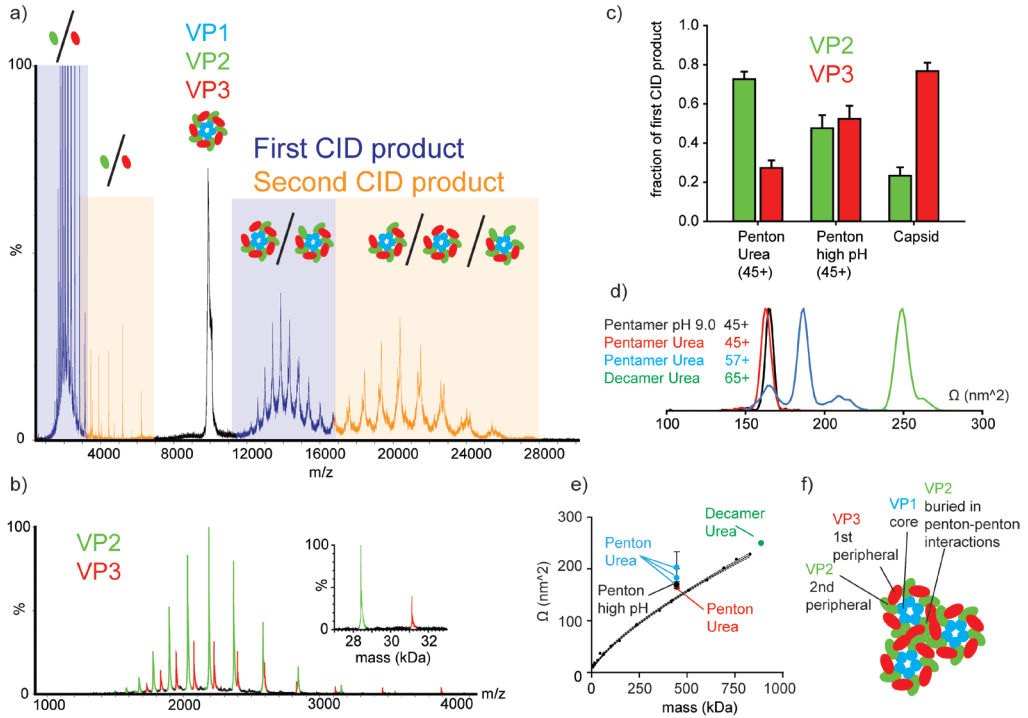


Figure 2. Native mass spectrometry based topology model of TrV. a) Tandem MS spectrum of a denatured penton at 160 V collision voltage, showing the sequential loss of two subunits. b) Zoom-in of the ejected subunits from the denatured pentons. The first CID products consist solely of VP2 and VP3, showing that VP1 is a core subunit while VP2/VP3 are peripheral. The relative amounts of each subunit are estimated from integrated peak areas, as shown in the inset. c) Relative amounts of dissociated VP's from indicated complexes. The selected charge state is indicated in parentheses. Average from triplicate experiments, error bars represent SD. Urea induced pentons preferentially eliminate VP2 under CID, the r-empty capsids more easily lose VP3 d) Collision Cross Section (Ω) profiles extracted from the ion mobility MS data of pentons and penton-dimers. e) Plot of Ω vs. molecular weight for a range of protein complexes analyzed under identical conditions by IMS. Solid black circles represent globular proteins and protein complexes. The black line is a power function fitted to these points ($\Omega = 2.41 \cdot \text{mass}^{0.67}$, $R^2=0.99$). TrV pentons/penton-dimers are plotted as averages from triplicate experiments, error bars represent $2 \cdot \text{SD}$. The pentons fall significantly outside the confidence interval for globular complexes suggesting that they have retained a sheet-like structure. f) MS-based topology model of TrV. Penton assembly intermediates are composed of 5 copies of VP1/2/3. VP1 is the core subunit oriented along the 5-fold icosahedral symmetry axis. VP2/3 are peripheral subunits. VP2 is the main anchor for assembly and buried in penton-penton contacts.

MS-based topology models of TrV.

With Collision Induced Dissociation (CID) and Ion Mobility Spectrometry (IMS) of protein complexes, topology models can be reconstructed, even of low-abundant components in complex mixtures(24, 25). In CID, peripheral subunits tend to dissociate more readily than core-subunits. Thus, by carefully examining the dissociation behaviour of selected protein complexes during tandem MS, subunit locations can be determined. A tandem mass spectrum of individual TrV pentons is shown in Figure 2a, revealing the sequential loss of two monomeric subunits after activation with CID. Figure 2b shows a zoom-in of the ejected subunits during the first CID event. There is no detectible elimination of VP1, which suggests VP1 is a core subunit, whereas the facile elimination of VP2 and VP3 indicates that they are situated more towards the periphery of the penton.

The relative amount of dissociated VP2 and VP3 can be determined by integrating the ion intensities over all their respective charge states (Figure 2c). We performed CID on the pentons either formed by alkaline treatment or urea treatment, and on the intact r-empty capsids. The assembly-competent pentons, generated by alkaline-triggered uncoating, revealed no preferential dissociation of either VP2 or VP3. In the r-empty capsids, we observed substantially less elimination of VP2 compared to VP3, indicating that VP2 is more buried and stabilized by inter-penton contacts. The assembly-incompetent, denatured pentons show the opposite trend. VP2 is preferentially eliminated, indicating diminished contacts or partial unfolding of the subunit. Therefore our CID data reveal the central role of VP2 for assembly of the TrV capsid.

Using IMS, a coarse geometry of protein assemblies can be determined and conformational sub-states can be distinguished by determining a rotationally averaged collisional cross section (Ω) (25). It was previously shown that IMS can be used to distinguish between sheet-like and globular structures of viral assembly intermediates by observing how cross sections scale with molecular weight (7, 26). We performed Travelling Wave IMS on sub-complexes of TrV to determine the geometry of the assemblies (Figure 2d and Supplementary Table S2). Lower charged pentons and penton-dimers show narrow peaks indicating folded and homogeneous conformations. Higher charged denatured pentons display multiple larger conformations, consistent with their more unfolded nature. Comparing the obtained cross sections of the pentons to those from a large set of globular proteins shows that the pentons are significantly larger than globular proteins of equal molecular weight (Figure 2e). This is indicative of sheet-like (i.e. capsomer) structures. Whereas pentons appear as capsomer structures, the

denatured penton-dimers follow more closely the same trend as globular proteins. This suggests that penton-dimers hinge on the penton-penton interface and collapse into a less extended structure (Supplementary Figure S10). The combined CID and IMS analysis of TrV results in the topology model presented in Figure 2f. Assembly intermediates of TrV are five-fold symmetric structures with VP1 located at the core around the symmetry axis. VP2 and VP3 are located at the periphery of the complex. VP2 is the main anchor for establishing inter-penton contacts and is crucial for assembly of complete capsids.

Mechanical properties of TrV reveal the mechanism of pH-triggered uncoating.

To probe the capsid-genome interplay further, the elastic strength and mechanical resilience of TrV capsid and virion was studied using AFM nanoindentation. Using this approach, it has previously been demonstrated that genome packaging has a stabilizing effect on the capsids of a diverse set of viruses (27-29). Our results described above indicate that the ssRNA triggers disassembly of the TrV capsid under alkaline conditions, resulting in release of the genome. Therefore, TrV was probed with AFM nanoindentation in the pH range 7-9 to see whether there was a mechanical basis for alkaline-triggered uncoating. A typical Force Distance Curve (FDC) of TrV can be divided into three stages (Figure 3a). After imaging, the tip of the cantilever is positioned over the center of the particle and pushed into the capsid. There is zero force with z-displacement until the tip and sample make contact. As the tip pushes down on the particle, there is a linear increase in force from which the spring constant k of the particle is determined. A higher load results in mechanical failure. This can be seen from a sharp transition in the FDC that is referred to as the breaking force, F_{break} .

TrV virions and n-empty capsids appear as spherical particles of ~33 nm in AFM imaging (Figure 3b and Supplementary Table S3). The five-fold protrusions of VP1 can be distinguished on the capsid surface (30). Similar to force-induced disassembly of minute virus of mice (31), imaging after nanoindentation reveals that mechanical failure of TrV results in capsid disassembly. Instead of a visible capsid, there are now up to twelve smaller particles. Upon closer inspection, each of these smaller breakdown products consists of five smaller subunits arranged in a five-fold rotation. The dimensions of these particles are strikingly similar to pentons of $5 \times VP1/2/3$ (Supplementary Table S4). This indicates that pentons are the mechanical building blocks of TrV. Similar particles can be observed in AFM images of the virion at pH 9 before nanoindentation, indicating both that pentons are assembly intermediates for alkaline-triggered uncoating and that they

have retained the capsid-like capsomer structure in isolation, confirming our ion mobility findings (Supplementary Figure S11 and Supplementary Table S4).

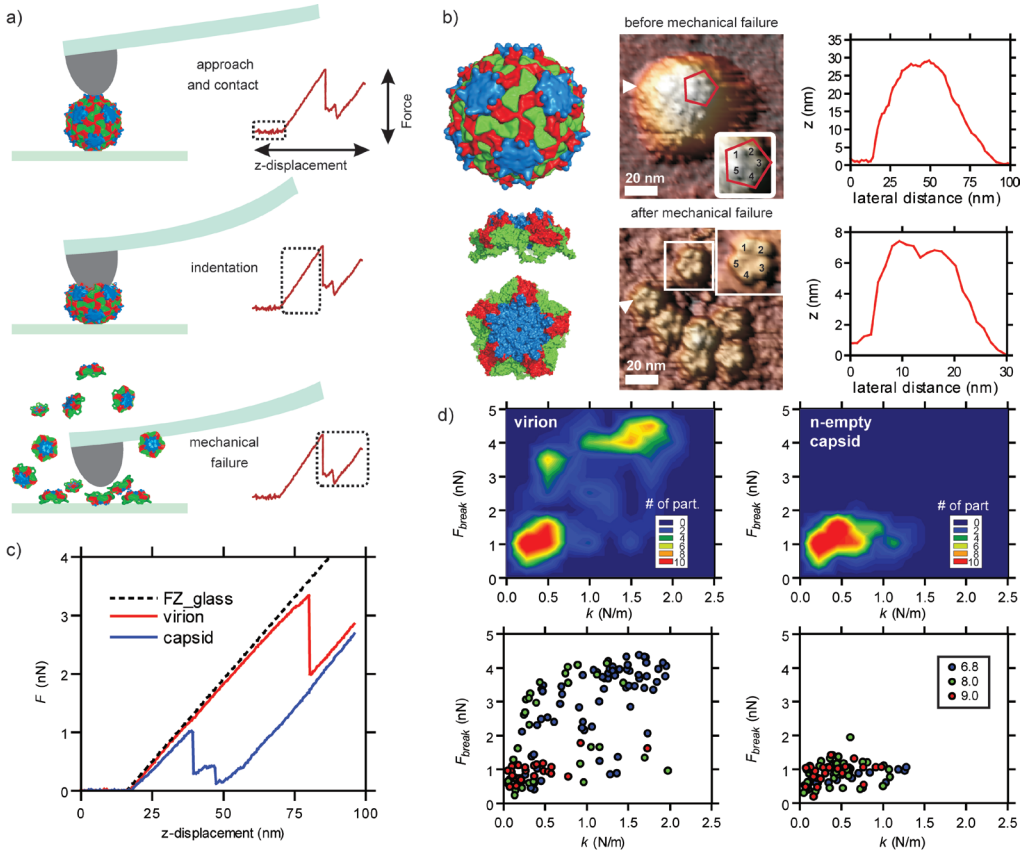


Figure 3. A mechanical basis for alkaline-triggered uncoating of TrV. a) Schematic of AFM nanoindentation. b) Before nanoindentation, TrV appears as round particles of ~ 33 nm. Surface protrusions of VP1 are also visible. Mechanical failure results in complete disassembly into pentons. The corresponding height-profiles are taken along the line of the white arrowheads. Surface renderings of intact capsids and penton are shown for reference with VP1 in blue, VP2 in green and VP3 in red (PDB code: 3NAP). c) Force Distance Curves of virion and n-empty capsids at neutral pH. d) Top panels show the 2D density distribution (k vs. F_{break}) of single virions (left) vs. n-empty capsids (right), at pH 6.8, pH 8 and pH 9. Bottom panels show the individual points at different pH as indicated, from which the density distributions were calculated. The virion under neutral pH has a characteristically higher spring constant and breaking force than the empty capsid. With increasing pH, the spring constant decreases but the high breaking force is maintained, resulting in a discrete intermediate state. Consistent with genome release to yield empty capsids, all particles are mechanically indistinguishable from n-empty capsids at pH 9.

At neutral pH, virions are both stiffer (higher k) and more resilient to high forces than empty capsids (Figure 3c.). Moving towards increasingly alkaline conditions the mechanical response of the virion shifts towards that of the empty capsid, via a discrete intermediate state (Figure 3d). The intermediate state is most populated at pH 8. From our MS analyses described above, we know that this intermediate still encloses the genome (Figure 1a). The intermediate is mechanically characterized by a spring constant equal to the empty capsid, and a breaking force similar to the intact virion. The r-empty capsids that arise from alkaline-triggered uncoating are mechanically indistinguishable from n-empty capsids at pH 9. The mechanical properties of virions, the uncoating intermediate and empty capsids are stated in Supplementary Table S5. The transition from virion to intermediate can be reversed by lowering the pH back to neutral, the transition from virion/intermediate to r-empty capsid is irreversible (Supplementary Figure S12). As with the n-empty capsid, the mechanical response of the r-empty capsid is equal at pH 7 and pH 9. As an additional check that reassembly takes place, we assessed the particle density of TrV on the AFM substrate, showing that it is equal at all pH (Supplementary Figure S13).

The material properties of icosahedral capsids can be approximated within the framework of continuum thin shell theory (5). Here, the spring constant, k , of a shell with radius R , and thickness h is uniquely related to the Young's modulus E of the capsid material, following $k = \alpha Eh^2/R$, where α is a proportionality factor that is typically close to 1. The spring constant of r-empty capsids is on average 0.43 N/m and the average shell radius and thickness as described in the crystal structure of TrV (PDB code: 3NAP) are 13.4 and 3.4 nm, respectively. These values yield a Young's modulus of $E=0.54$ GPa for the empty capsids. We know that the elasticity of the capsid is unaltered by alkaline pH and the measured particle height is also unaltered (this was additionally confirmed by IMS of the intact capsids and virion, Supplementary Table S6). Hence, continuum thin shell theory suggests that the increased spring constant of the virion is due to an effective increase in shell thickness. With a spring constant of 1.46 N/m, the shell thickness of the virion is estimated at 6.0 nm, i.e. a genome-induced effective increase of 2.6 nm compared to r-empty capsids.

Dicistroviruses package an enormous amount of genomic material compared to other ssRNA viruses such as the bromoviruses (1). TrV packages over 9 kb of ssRNA compared to just 3 kb in CCMV, while they both have a comparable diameter of ~ 30 nm. The packing density in TrV is even higher than in bacteriophage λ , which is thought to possess a high internal pressure of several tens of atmospheres due to the high density of dsDNA it encapsidates (based on the internal

volumes and number of nucleotides/basepairs, the charge-densities compare as -1.03 vs. -1.34 e/nm^3 in λ and TrV, respectively, see Supplementary Methods). Confining such a large molecule to the relatively small inner volume of the capsid is bound to come at a high energetic cost, requiring for instance significant dehydration of the ssRNA (32). Furthermore, the observation of polyamines in picornaviruses suggests that condensation of the genome onto counter ions is required to facilitate genome packaging (33). The effective strengthening of the virion at neutral pH is probably the result of a combination of non-specific interactions of the RNA with positive patches on the inner capsid wall and the condensation of the RNA by counter ions, confining the densely packed RNA to the inner capsid volume. We confirmed that there is an electrostatic component that determines the higher spring constant of the virion compared to empty capsids by testing the response to nanoindentation in high concentrations of $MgCl_2$ (Supplementary Figure S14). We found that in the presence of magnesium, there was a marked decrease in the spring constant of the virion, as previously observed for phage λ (27). This result shows that either magnesium ions condense the genome, thereby reducing the forces that drive the genome in close contact with the inner capsid wall, or that magnesium ions screen the interaction between capsid and genome.

The presence of densely packed RNA explains how both the virion and alkaline-triggered uncoating intermediate have a higher breaking force than empty capsids. When applying forces to the shells, the enclosed RNA will resist actual indentation and instead the capsid deforms into an oblate sphere (maximizing inner volume to accommodate the RNA). It was shown in molecular dynamics simulations of AFM nanoindentation experiments of similarly sized particles that irreversible distortion is dominated by locally imposed curvature (34). The maximum imposed curvature will be smaller for deformation into oblate spheres than in true indentation, which explains the high breaking forces of the virion and uncoating intermediate compared to empty capsids.

Discussion.

The data from our AFM and MS experiments suggest the following model for alkaline-triggered uncoating in TrV. Under neutral conditions, the large genome stabilizes the capsid. As the pH rises, the genome-capsid interactions are lost and the condensing strength of counter ions decreases, however the genome is still contained. This leads to a labile shell at the very limit of its packaging capacity. Upon further increase in pH, this labile intermediate is unable to resist the elec-

trostatic and entropic forces that drive the genome outwards and it eventually bursts and disassembles into pentons of $5 \times VP1/2/3$. The genome and VP4 are released into solution and pentons reassemble into empty capsids. A schematic overview of the model is presented in figure 4. Alternatively, the genome and VP4 may be expelled from an otherwise intact capsid, as observed for polio virus (35). This does not, however, account for the presence of pentons during the uncoating reaction. Also, the absence of pentons from samples of n-empty capsids deems it unlikely that pentons arise from a penton-capsid equilibrium that is initiated upon genome release.

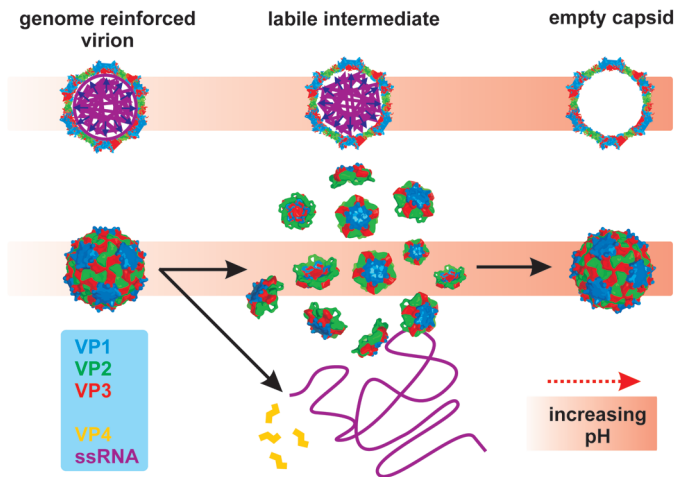


Figure 4. Schematic of the alkaline-triggered uncoating of TrV. Under neutral pH, TrV confines a very large genome. This comes at a high energetic cost, but the ssRNA stabilizes the capsid, thereby preventing premature uncoating. At higher pH, this stabilizing interaction is lost and electrostatic self-repulsion of the ssRNA increases due to loss of charge on counterions; the capsid bursts and falls apart into pentons. The genome and VP4 are released into solution and pentons reassemble into empty capsids.

Whether the observed pathway of alkaline-triggered uncoating is representative of uncoating *in vivo* remains unresolved as most aspects of genome-release are still unknown for picorna(-like) viruses. The acid stability of TrV excludes low endosomal pH as a sufficient trigger for uncoating. Other factors are thus required. Also, it was recently shown that empty TrV capsids produced by heating a virion sample do not display any morphological change similar to those described for either Poliovirus or Human Rhinovirus type 2 (36). The mechanism of uncoating for TrV is thus likely distinct from picornaviruses. Given the infection route of TrV, it seems only logical that the virus should be acid-stable. TrV infects its host through an oral-fecal route (21). It passes the acidic crop

(stomach) of the insect before reaching the mid- and hindgut, where it enters and infects cells (37). The gut of the insect reaches a higher pH upon feeding, suggesting that timing virus release from infected cells is required (17, 20). The acid stability of the virion thus prevents premature uncoating. Non-enveloped ssRNA viruses do not typically encounter alkaline conditions during cell-entry. However, the alkaline conditions (up to pH 8.9) reached in the intestinal tract of the insect may induce a softened, labile state of TrV that is required for effective infection. The requirement of a softened state for successful infection was previously shown for retroviral particles (38).

The genome size of picorna- and picorna-like viruses is typically in the range of 7-10 kb. The observation that the capsid of TrV has to accommodate an enormous amount of ssRNA has important implications for the assembly and stability of the virion. It was recently shown that mutations towards a bipartite genome in Foot-and-Mouth Disease Virus confer an enhanced stability to the virion and increase fitness (32). Genome size is thus limiting to viral replication due to the unfavorable state of confining a large negatively charged polymer in a small volume. This barrier will also have to be overcome during genome packaging. The pathway of virion assembly is largely unknown for picorna(-like) viruses. Simple co-assembly of ssRNA and pentons will not be energetically feasible and requires condensation of the genome on counter ions. Our results show that the packaging capacity of viral capsids is directly related to the stability of the loaded particle, which suggests that the use of viruses as nanocontainers is limited not only by the packaging volume but also by capsid-cargo interactions and destabilizing effects of heavy loads.

In summary, the combination of native MS and AFM allowed us to comprehensively study uncoating in terms of the biophysical properties of TrV. Our current model for the mechanism of uncoating is uniquely defined by an intermediate state uncovered with AFM by virtue of the single-molecule approach. However, this state could only be defined as intermediate to empty capsids because native MS showed that it still contained the ssRNA. Conversely, native MS uncovered the pathway of uncoating via reversible disassembly of the virion, and a topology model of the assembly intermediate was confirmed by AFM. This report shows how native MS and AFM are highly complementary tools for the structural and biophysical characterization of viruses, enabling us to discover a novel *in vitro* uncoating pathway for dicistroviruses.

Methods.

TrV purification.

TrV was purified from *Triatoma infestans* feces using previously described protocols (14).

Mass Spectrometry.

Samples were buffer exchanged to 150-200 mM ammonium acetate at the indicated pH and a final concentration of 5-10 μM (based on the heterotrimer). A 1-2 μL aliquot was loaded into gold-coated capillaries for nano-electrospray ionization (nanoESI). Capillaries for nanoESI were home-made from borosilicate glass tubes of 1.2 mm OD and 0.68 mm ID (World Precision Instruments, USA) using a P-97 micropipette puller (Sutter Instruments, USA) and gold-coated using an Edwards Scancoat (Edwards Laboratories, USA) six Pirani 501 sputter coater. MS was performed on modified QTOF I and QTOF II instruments (Waters, UK and MS Vision, Netherlands) (39). Spectra were mass calibrated with cesium iodide. Capillary voltage was in the range of 1300-1500 V, the sample cone, 120-200 V. To optimize transmission of larger ions, the backing pressure was elevated to ~ 10 mbar (40). xenon was used as collision gas at a pressure of $1-2 \times 10^{-2}$ mbar (41).

Ion Mobility MS was performed on a Synapt G1 instrument (Waters, UK) (42). Source conditions were as described above. Xenon was used as collision gas, and nitrogen as buffer gas in the TWIMS cell. For analysis of the penton and decamer complexes, as well as the corresponding globular protein standards, IMS was performed using a ramped waveheight of 2-12 V, ramped over 60% of the cycle. The nitrogen gas-pressure in the TWIMS cell was 5.55×10^{-1} mbar. Cross sections were calibrated based on known cross sections of denatured ubiquitin, cytochrome c and myoglobin (5 μM in 50/45/5 water/acetonitrile/formic acid) and native GroEL (5 μM in 75 mM ammonium acetate pH 6.8) (26, 43). *In silico* cross section calculations (PA – Projection Approximation) were performed with MobCal, which gave similar results to the Driftscope PA algorithm (44, 45).

Due to their size, Ω 's for TrV virions and capsids had to be calibrated using native GroEL (as above) and HBV capsids as standards (46). The integrity of the calibrants under the conditions used was verified by calibration of HBV data with GroEL and comparison to the previously published results. Measurements were performed at 1.35 kV capillary and 150 V cone voltage with 6 mbar backing pressure. Additional settings were as follows: 50 V in the trap, 75 V in the transfer, both at 2.3×10^{-2} mbar xenon with 500 μs trap release time; 70 V over the IMS

cell running at 250 m/s wave velocity and 15-30 V full cycle wave height ramping at 4.6×10^{-1} mbar nitrogen.

Atomic Force Microscopy.

Virus samples were analyzed in 200 mM ammonium acetate at the indicated pH. Silanized glass slides were used as substrate for AFM. Glass slides were prepared as described (47). Briefly, after thorough rinsing with MilliQ water, the slides were incubated overnight in an ethanol/water (90/10%) bath, saturated with potassium hydroxide. After another round of rinsing in MilliQ water, the glass slides were dried and incubated overnight in a hexamethyldisilazane vapour. A 100 μ L droplet of virus solution was incubated on the silanized substrate for exactly 30min before adding another 100 μ L of ammonium acetate buffer, wetting the AFM tip and mounting the head on the sample. We used Olympus OMCL-RC800PSA silicon-nitride cantilevers with a nominal spring constant of 0.05 N/m and nominal tip radius of 15 nm. Cantilevers were calibrated using the method of Sader *et al.* (48), giving an average value of 0.056 ± 0.006 N/m. Imaging and nanoindentation were performed on a Nanotec Electronica AFM operated in jumping mode. The average maximum imaging force was ~ 50 pN. Nanoindentation was performed at a probe velocity of ~ 60 nm/s and data were processed using a home-built Labview application (49). Image processing was performed using WSxM software (50).

Biochemical analysis and binding energy calculations.

Experimental procedures for native PAGE and autofluorescence analysis of TrV, as well as the procedure used to calculate the binding energies, can be found in the Supplementary Information.

Acknowledgements.

Correspondence and requests for materials should be addressed to WHR or AJRH. GJLW is supported by a NanoSci-E+ grant, by FOM under the “Physics of the Genome” program and by a Vici grant from NWO. JS, CU, RJR and AJR are supported by the Netherlands Proteomics Centre. AJRH is further supported by the Netherlands Organization for Scientific Research (NWO) under ALW-ECHO (819.02.10). RSE is recipient of a predoctoral fellowship from the Basque Government (BG). GAM is partially supported by PICT 2008-0035 and V188 (UNLP). DMAG is supported by CYTED (209RT0364), by BG (MV-2012-2-41; AE-2012-1-44, SPE1 1FB001), and by MICINN (BFU2007-62062). We thank the Scientific Calculation Service of the UPV/EHU (SGIker) for computation

resources.

Supplementary information.

Supplementary information can be found in the appendix.

References.

1. Speir, J. A., and J. E. Johnson. 2012. Nucleic acid packaging in viruses. *Curr. Opin. Struct. Biol.* 22:65-71.
2. Douglas, T., and M. Young. 2006. Viruses: Making friends with old foes. *Science.* 312:873-875.
3. Mateu, M. G. 2012. Mechanical properties of viruses analyzed by atomic force microscopy: A virological perspective. *Virus Res.* 168:1-22.
4. Morton, V. L., P. G. Stockley, N. J. Stonehouse, and A. E. Ashcroft. 2008. Insights into virus capsid assembly from non-covalent mass spectrometry. *Mass Spectrom. Rev.* 27:575-595.
5. Roos, W. H., R. Bruinsma, and G. J. L. Wuite. 2010. Physical virology. *Nature Physics.* 6:733-743.
6. Uetrecht, C., and A. J. R. Heck. 2011. Modern Biomolecular Mass Spectrometry and its Role in Studying Virus Structure, Dynamics, and Assembly. *Angewandte Chemie International Edition.* 50:8248-8262. doi: 10.1002/anie.201008120.
7. Uetrecht, C., I. M. Barbu, G. K. Shoemaker, E. Van Duijn, and A. J. R. Heck. 2011. Interrogating viral capsid assembly with ion mobility-mass spectrometry. *Nature Chemistry.* 3:126-132.
8. Bonning, B. C., and W. A. Miller. 2010. Dicrostoviruses. *Annual Review of Entomology.* 55:129-150.
9. Marti, G. A., M. G. Echeverria, M. L. Susevich, J. J. Becnel, S. A. Pelizza, and J. J. García. 2009. Prevalence and distribution of parasites and pathogens of Triatominae from Argentina, with emphasis on *Triatoma infestans* and *Triatoma* virus TrV. *J. Invertebr. Pathol.* 102:233-237.
10. Muscio, O. A., J. L. La Torre, M. A. Bonder, and E. A. Scodeller. 1997. *Triatoma* Virus Pathogenicity in Laboratory Colonies of *Triatoma infestans* (Hemiptera: Reduviidae). *J. Med. Entomol.* 34:253-256.
11. Rassi Jr., A., A. Rassi, and J. A. Marin-Neto. 2010. Chagas disease. *The Lancet.* 375:1388-1402.
12. Muscio, O. A., J. L. La Torre, and E. A. Scodeller. 1988. Characterization of *Triatoma* virus, a picorna-like virus isolated from the triatomine bug *Triatoma infestans*. *J. Gen. Virol.* 69:2929-2934.
13. Czibener, C., J. L. La Torre, O. A. Muscio, R. A. Ugalde, and E. A. Scodeller. 2000. Nucleotide sequence analysis of *Triatoma* virus shows that it is a member

of a novel group of insect RNA viruses. *J. Gen. Virol.* 81:1149-1154.

14. Agirre, J., K. Aloria, J. M. Arizmendi, I. Iloro, F. Elortza, R. Sánchez-Eugenia, G. A. Marti, E. Neumann, F. A. Rey, and D. M. A. Guérin. 2011. Capsid protein identification and analysis of mature *Triatoma* virus (TrV) virions and naturally occurring empty particles. *Virology*. 409:91-101.

15. Tate, J., L. Liljas, P. Scotti, P. Christian, Lin Tianwei, and J. E. Johnson. 1999. The crystal structure of cricket paralysis virus: The first view of a new virus family. *Nat. Struct. Biol.* 6:765-774.

16. Kumar, S., and U. Bandyopadhyay. 2005. Free heme toxicity and its detoxification systems in human. *Toxicol. Lett.* 157:175-188.

17. Schaub, G. A. 2009. Chapter 4 Interactions of Trypanosomatids and Triatomines, p. 177-242. In Stephen J. Simpson and Jeacuterocircme Casas (ed.), *Advances in Insect Physiology*. Volume 37. Academic Press.

18. Oliveira, M. F., A. C. P. Gandara, C. M. S. Braga, J. R. Silva, F. B. Mury, M. Dansa-Petretski, D. Menezes, M. A. Vannier-Santos, and P. L. Oliveira. 2007. Heme crystallization in the midgut of triatomine insects. *Comparative Biochemistry and Physiology - C Toxicology and Pharmacology*. 146:168-174.

19. Stiebler, R., B. L. Timm, P. L. Oliveira, G. R. Hearne, T. J. Egan, and M. F. Oliveira. 2010. On the physico-chemical and physiological requirements of hemozoin formation promoted by perimicrovillar membranes in *Rhodnius prolixus* midgut. *Insect Biochem. Mol. Biol.* 40:284-292.

20. Kollien, A. H., T. Grospletsch, T. Kleffmann, I. Zerbst-Boroffka, and G. A. Schaub. 2001. Ionic composition of the rectal contents and excreta of the reduviid bug *Triatoma infestans*. *J. Insect Physiol.* 47:739-747.

21. Muscio, O., M. A. Bonder, J. L. La Torre, and E. A. Scodeller. 2000. Horizontal transmission of *Triatoma* virus through the fecal-oral route in *Triatoma infestans* (Hemiptera: Triatomidae). *J. Med. Entomol.* 37:271-275.

22. Li, C., J. C. -. Wang, M. W. Taylor, and A. Zlotnick. 2012. In vitro assembly of an empty picornavirus capsid follows a dodecahedral path. *J. Virol.* 86:13062-13069.

23. Benesch, J. L. P. 2009. Collisional Activation of Protein Complexes: Picking Up the Pieces. *J. Am. Soc. Mass Spectrom.* 20:341-348.

24. Benesch, J. L. P., B. T. Ruotolo, D. A. Simmons, and C. V. Robinsons. 2007. Protein complexes in the gas phase: Technology for structural genomics and proteomics. *Chem. Rev.* 107:3544-3567.

25. Uetrecht, C., R. J. Rose, E. Van Duijn, K. Lorenzen, and A. J. R. Heck. 2010. Ion mobility mass spectrometry of proteins and protein assemblies. *Chem. Soc. Rev.* 39:1633-1655.
26. Ruotolo, B. T., J. L. P. Benesch, A. M. Sandercock, S. - Hyung, and C. V. Robinson. 2008. Ion mobility-mass spectrometry analysis of large protein complexes. *Nature Protocols.* 3:1139-1152.
27. Evilevitch, A., W. H. Roos, I. L. Ivanovska, M. Jeembaeva, B. Jönsson, and G. J. L. Wuite. 2011. Effects of salts on internal DNA pressure and mechanical properties of phage capsids. *J. Mol. Biol.* 405:18-23.
28. Carrasco, C., A. Carreira, I. A. T. Schaap, P. A. Serena, J. Gómez-Herrero, M. G. Mateu, and P. J. De Pablo. 2006. DNA-mediated anisotropic mechanical reinforcement of a virus. *Proc. Natl. Acad. Sci. U. S. A.* 103:13706-13711.
29. Michel, J. P., I. L. Ivanovska, M. M. Gibbons, W. S. Klug, C. M. Knobler, G. J. L. Wuite, and C. F. Schmidt. 2006. Nanoindentation studies of full and empty viral capsids and the effects of capsid protein mutations on elasticity and strength. *Proc. Natl. Acad. Sci. U. S. A.* 103:6184-6189.
30. Estrozi, L. F., E. Neumann, G. Squires, G. Rozas-Dennis, M. Costabel, F. A. Rey, D. M. A. Guérin, and J. Navaza. 2008. Phasing of the Triatoma virus diffraction data using a cryo-electron microscopy reconstruction. *Virology.* 375:85-93.
31. Castellanos, M., R. Pérez, P. J. P. Carrillo, P. J. De Pablo, and M. G. Mateu. 2012. Mechanical disassembly of single virus particles reveals kinetic intermediates predicted by theory. *Biophys. J.* 102:2615-2624.
32. Ojosnegros, S., J. García-Arriaza, C. Escarmís, S. C. Manrubia, C. Perales, A. Arias, M. G. Mateu, and E. Domingo. 2011. Viral genome segmentation can result from a trade-off between genetic content and particle stability. *PLoS Genetics.* 7:e1001344.
33. Fout, G. S., K. C. Medappa, J. E. Mapoles, and R. R. Rueckert. 1984. Radiochemical determination of polyamines in poliovirus and human rhinovirus 14. *J. Biol. Chem.* 259:3639-3643.
34. Roos, W. H., M. M. Gibbons, A. Arkhipov, C. Uetrecht, N. R. Watts, P. T. Wingfield, A. C. Steven, A. J. R. Heck, K. Schulten, W. S. Klug, and G. J. L. Wuite. 2010. Squeezing protein shells: How continuum elastic models, molecular dynamics simulations, and experiments coalesce at the nanoscale. *Biophys. J.* 99:1175-1181.
35. Bostina, M., H. Levy, D. J. Filman, and J. M. Hogle. 2011. Poliovirus RNA is released from the capsid near a twofold symmetry axis. *J. Virol.* 85:776-783.

36. Agirre, J., G. Goret, M. Legoff, R. Sanchez-Eugenia, G. Marti, J. Navaza, D. Guerin, and E. Neumann. 2013. Cryo-TEM reconstructions of Triatoma Virus particles: a clue to unravel genome delivery and capsid disassembly. *J. Gen. Virol.* 94:1058-1068.
37. Muscio, O. A., J. L. LaTorre, and E. A. Scodeller. 1987. Small nonoccluded viruses from triatomine bug *Triatoma infestans* (hemiptera: Reduviidae). *J. Invertebr. Pathol.* 49:218-220.
38. Kol, N., Y. Shi, M. Tsvitov, D. Barlam, R. Z. Shneck, M. S. Kay, and I. Rousso. 2007. A stiffness switch in human immunodeficiency virus. *Biophys. J.* 92:1777-1783.
39. Van Den Heuvel, R. H. H., E. Van Duijn, H. Mazon, S. A. Synowsky, K. Lorenzen, C. Versluis, S. J. J. Brouns, D. Langridge, J. Van Der Oost, J. Hoyes, and A. J. R. Heck. 2006. Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. *Anal. Chem.* 78:7473-7483.
40. Tahallah, N., M. Pinkse, C. S. Maier, and A. J. R. Heck. 2001. The effect of the source pressure on the abundance of ions of noncovalent protein assemblies in an electrospray ionization orthogonal time-of-flight instrument. *Rapid Communications in Mass Spectrometry.* 15:596-601.
41. Lorenzen, K., C. Versluis, E. van Duijn, R. H. H. van den Heuvel, and A. J. R. Heck. 2007. Optimizing macromolecular tandem mass spectrometry of large non-covalent complexes using heavy collision gases. *International Journal of Mass Spectrometry.* 268:198-206.
42. Pringle, S. D., K. Giles, J. L. Wildgoose, J. P. Williams, S. E. Slade, K. Thalassinou, R. H. Bateman, M. T. Bowers, and J. H. Scrivens. 2007. An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *International Journal of Mass Spectrometry.* 261:1-12.
43. Bush, M. F., Z. Hall, K. Giles, J. Hoyes, C. V. Robinson, and B. T. Ruotolo. 2010. Collision cross sections of proteins and their complexes: A calibration framework and database for gas-phase structural biology. *Anal. Chem.* 82:9557-9565.
44. Mesleh, M. F., J. M. Hunter, A. A. Shvartsburg, G. C. Schatz, and M. F. Jarrold. 1996. Structural information from ion mobility measurements: Effects of the long-range potential. *J. Phys. Chem.* 100:16082-16086.
45. Shvartsburg, A. A., and M. F. Jarrold. 1996. An exact hard-spheres scattering model for the mobilities of polyatomic ions. *Chemical Physics Letters.* 261:86-91.

46. Uetrecht, C., C. Versluis, N. R. Watts, P. T. Wingfield, A. C. Steven, and A. J. R. Heck. 2008. Stability and shape of hepatitis B virus capsids in vacuo. *Ange wandte Chemie - International Edition*. 47:6247-6251.
47. Roos, W. H. 2011. How to perform a nanoindentation experiment on a virus. *Methods in Molecular Biology*. 783:251-264.
48. Sader, J. E., J. W. M. Chon, and P. Mulvaney. 1999. Calibration of rectangular atomic force microscope cantilevers. *Rev. Sci. Instrum.* 70:3967-3969.
49. Snijder, J., I. L. Ivanovska, M. Baclayon, W. H. Roos, and G. J. L. Wuite. 2012. Probing the impact of loading rate on the mechanical properties of viral nanoparticles. *Micron*. 43:1343-1350.
50. Horcas, I., R. Fernández, J. M. Gómez-Rodríguez, J. Colchero, J. Gómez-Herrero, and A. M. Baro. 2007. WSXM: A software for scanning probe microscopy and a tool for nanotechnology. *Rev. Sci. Instrum.* 78:013705.

Chapter 6:

Cargo encapsulation triggers a stability switch in the bacterial nanocompartment encapsulin.

J. Snijder^{1,2,3}, O. Kononova^{4,5}, I.M. Barbu^{1,2}, C. Uetrecht^{1,2}, W.F. Rurup⁶, R.J. Burnley^{1,2}, M.S.T. Koay⁶, J.J.L.M. Cornellissen⁶, W.H. Roos³, V. Barsegov^{4,5}, G.J.L. Wuite³, A.J.R. Heck^{1,2}

¹ Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH, Utrecht, the Netherlands.

² Netherlands Proteomics Centre, Padualaan 8, 3584 CH, Utrecht, The Netherlands.

³ Natuur- en Sterrenkunde and LaserLab, Vrije Universiteit, De Boelelaan 1081, Amsterdam, The Netherlands

⁴ Department of Chemistry, University of Massachusetts, Lowell, Massachusetts 01854, United States

⁵ Moscow Institute of Physics and Technology, Moscow Region, Russia 141700

⁶ Department of Biomolecular Nanotechnology, MESA+ Institute, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

Abstract.

Bacterial nanocompartments that are structurally similar to virus capsids have recently been discovered. These compartments encapsulate specific enzymes, making them prospective carriers for applications in nanotechnology. Here, we combine native Ion Mobility Mass Spectrometry (IMMS), Atomic Force Microscopy (AFM) and high-performance computing to investigate the structural and energetic effects of cargo encapsulation in bacterial encapsulin. High precision native MS of cargo-loaded encapsulin revealed a mass of 1.966 MDa, corresponding to 60 copies of encapsulin and one copy of the hexameric cargo enzyme DyP. All-atom Molecular Dynamics simulations, IMMS and AFM nanoindentation *in vitro* and *in silico* reveal that cargo encapsulation leads to expansion and destabilization of the nanoshells. This weakening occurs through interactions of the cargo anchoring sequence with the shell interior. The data expose the delicate chemical and physical interplay between cargo and carrier in these nanocompartments.

Naturally occurring bacterial nanocompartments have recently been discovered (1-4). Much like bacterial micro-compartments, these nanocompartments are polyhedral protein cages that encapsulate enzymes to act as mini-organelles in prokaryotes (5-8). The subunits of bacterial nanocompartments adopt the HK97-fold, typical of the subunits of many virus and bacteriophage capsids (1-3). Bacteriophage capsids are notoriously resilient structures. They can withstand internal pressures of tens of atmospheres and have the rigidity of hard plastics (9-12). The use of engineered virus-like particles for nanotechnology is a rapidly growing field (13-16). The high degree of structural homology of the bacterial nanocompartments to bacteriophage capsids suggests their potential for similar applications.

Here, we have studied the bacterial nanocompartment encapsulin from *Brevibacterium linens* and *Thermotoga maritima*. Encapsulin is a sixty subunit nanocompartment with $T=1$ icosahedral symmetry that naturally occurs in a variety of prokaryotes. In *B. linens*, the native encapsulin is loaded with a dye-decolorizing peroxidase (DyP) whereas the *T. maritima* encapsulin accommodates a ferritin-like protein (Flp). A C-terminal anchoring sequence on the cargo enzyme targets the cargo of encapsulin to the shell interior. This naturally encoded mechanism for cargo loading makes encapsulin a particularly exciting scaffold as a nanocontainer/-reactor. It has been shown that a foreign cargo protein could indeed be targeted to encapsulin nanocompartments by utilizing the endogenous anchoring sequence for encapsulation (17-19).

Little is known about the interplay between the cargo and the physical and chemical properties of these nanocompartments. Revealing how cargo encapsulation influences the structure and stability of the shell would help making these nanocompartments into prospective platforms for applications in nanotechnology. Here we combine native Ion Mobility Mass Spectrometry (IMMS), Atomic Force Microscopy (AFM) nanoindentation and computational modeling to investigate the assembly of encapsulin and the effects of cargo encapsulation on the properties of the compartment. Native MS and AFM are strong tools to study the properties of viruses and their combined use has been shown to be a powerful approach to reveal the effects of cargo encapsulation on the stability of virus (-like) particles (10, 20, 21). This approach is complemented by extensive computer simulations, accelerated on Graphics Processing Units (GPUs) (22, 23). We performed all-atom Molecular Dynamics (MD) simulations of the complete encapsulin shell and used a simplified C_{α} -based Self-Organized Polymer (SOP) model to test the structure mechanically. The latter provides a detailed picture of the structure-stability relationship in AFM nanoindentation

(24). We demonstrate that encapsulin assembles uniquely with one copy of the native hexameric cargo-enzyme via a pathway that likely involves the addition of encapsulin dimers. Cargo encapsulation results in a radial expansion of the compartment. Cargo loading destabilizes encapsulin, which can be explained by the observed shape change and results via interactions of the anchoring sequence with the shell.

Results.

Cargo encapsulation and assembly of encapsulin.

The amount of native cargo carried in an encapsulin nanocompartment has not previously been quantified. In order to quantify the native cargo-load, we used native MS to analyze empty encapsulin shells from *Brevibacterium linens* and compared this to compartments loaded with the native DyP cargo enzyme (Figure 1). Native MS of empty encapsulin revealed one major particle distribution, with the expected mass of an intact 60mer $T=1$ shell (1.718 MDa), in addition to a minor population of 58mer partial assemblies, missing one dimer (an overview of experimental masses of all encapsulin assemblies can be found in Supplementary Table S1). Co-expression of the encapsulin protein with its native DyP cargo protein led to a distinct mass shift of approximately 248 kDa (to 1.966 MDa), corresponding to six copies of encapsulated DyP (theoretical mass 242 kDa). Electron microscopy reconstructions of free DyP revealed a hexameric structure of the enzyme (3), in agreement with the current native MS data (Supplementary Figure S1). Higher resolution MS analysis of DyP-loaded encapsulin on a modified Orbitrap for native MS (25, 26) confirmed that only compartments with one DyP hexamer are present, and additionally revealed the presence of a minor population of 58mer complexes, missing one dimer, but still containing a single DyP hexamer (Figure 1b). Thus, exclusively a single copy of the hexameric DyP cargo enzyme is contained within the encapsulin nanocompartment.

Closer inspection of the lower mass region in the spectra of both empty and DyP loaded encapsulin revealed that, in addition to the 60- and 58mer, more encapsulin sub-complexes were present. Monomers, dimers, higher order aggregates of dimers and half-shells (30mers) were detected (Figure 1c). These results indicate that the dimer is a particularly stable sub-structure and is likely to be the unit of assembly for encapsulin shells. In support of this suggestion, monomers appear to have particularly stable interactions across the 2-fold symmetry axis of the shell, owing to the extended E-loop of the subunits (3). The relatively high abundance of the 58mer is consistent with several computational studies show-

ing that near-complete shells can be particularly long-lived substructures during virus capsid assembly (27, 28). Our observation of the 58mer is some of the first direct experimental evidence for the stable/long-lived nature of such near-complete shells (29). Interestingly, the observation of DyP-loaded 58mers shows that this structural defect does not preclude cargo loading.

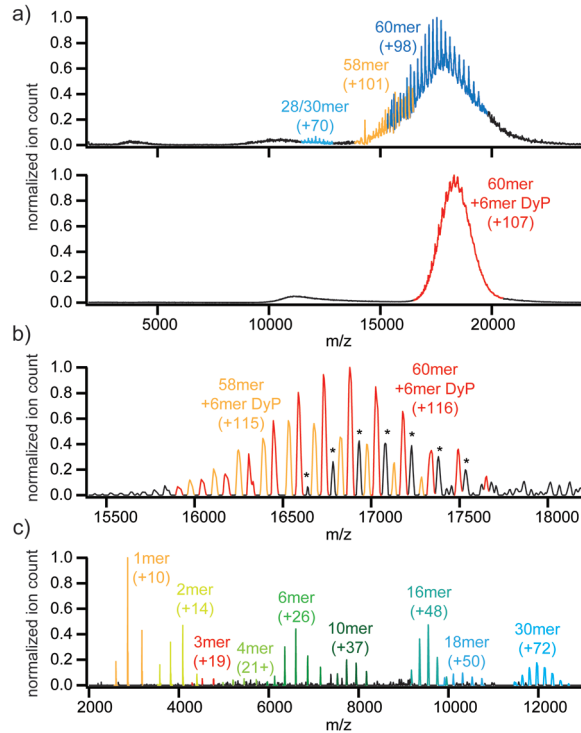


Figure 1. Encapsulin packages a single copy of a DyP hexamer. a) Native MS spectra (QTOF) of empty and DyP-loaded *B. linens* encapsulin. b) Native MS spectrum (Exactive plus EMR) of DyP-loaded *B. linens* encapsulin. In addition to the main species corresponding to 60mer+6mer DyP (1957.8 ± 0.5 kDa) and 58mer+6mer DyP (1901.0 ± 0.2 kDa), an unassigned species of 1963.8 ± 0.2 kDa is also detected, labelled with an asterisks (masses correspond to average \pm standard deviation across all charge states). c) Low mass region from a sample of empty encapsulin. Various smaller encapsulin oligomers are detected, including monomers, dimers and higher order multimers of dimers. In all panels main charge states are listed in parentheses.

Cargo encapsulation leads to expansion of the nanocompartment.

The structure of the related *T. maritima* encapsulin has been described to atomic-level resolution (3). It is worth noting that the reported structure corresponds to the enzyme-loaded compartment, and no such high-resolution structural data is currently available for the empty encapsulin shells. To test whether cargo encapsulation induces any structural changes in encapsulin, we first performed

native Ion Mobility MS (IMMS) on empty and DyP loaded compartments (see Figure 2a). Drift time measurements by IMMS can be used to calculate the gas-phase collisional cross section, Ω , of protein assemblies, a measure that is largely proportional to the rotationally averaged projection of the ion (30). Hence, IMMS yields shape information on the measured protein complex (21, 31, 32). IMMS revealed that cargo encapsulation results in a 9.8% increase in Ω , which corresponds to a 4.8% increase in radius, assuming that radius scales with Ω as $\Omega^{1/2}$.

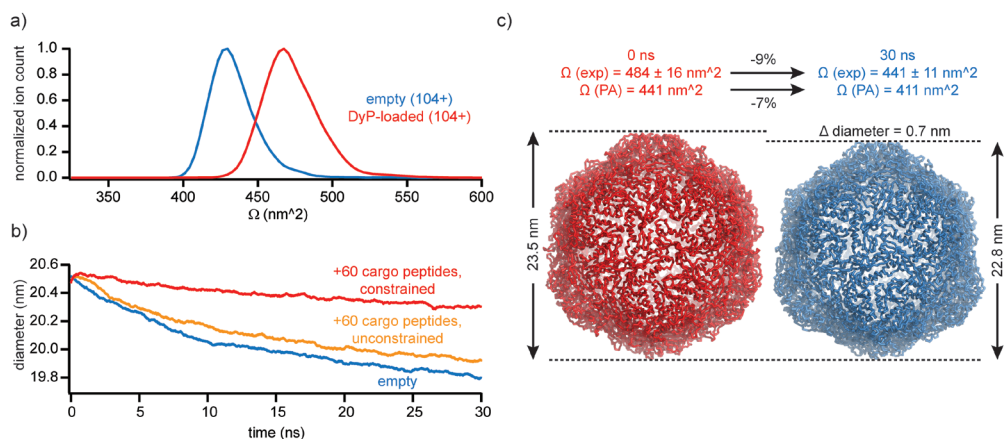


Figure 2. Cargo encapsulation triggers expansion of the encapsulin nanocompartment.

a) Collision Cross Section profiles of empty and DyP loaded encapsulin, illustrating the expansion of the filled particle. b) Diameter as a function of simulation time for the all-atom MD trajectories of *T. maritima* encapsulin shells with/without (un)constrained cargo elements, as indicated. c) Start- and end-structure from the all-atom MD simulations of the empty *T. maritima* encapsulin shell. The average experimental collisional cross section, Ω (exp), of empty and DyP loaded compartments (average \pm standard deviation, n=3) is shown vs. the rotationally averaged projections of the start- and end-structure of the MD simulation, Ω (PA).

To further test the effect of cargo encapsulation on the structure of encapsulin, we constructed a computer model for empty encapsulin shells, using the atomic structural model of filled encapsulin available from the Protein Data Bank (PDB) (3). We reconstructed complete shells of 60 subunits representing filled and empty encapsulin. To build the model for an empty shell, we removed all the cargo peptides from the PDB entry, which correspond to the anchoring sequences of the cargo enzyme. When constructing the model of the shell with the cargo peptides left in place, we constrained their centers of mass (COM) using a stiff harmonic spring. This mimics the presence of the cargo inside the shell. Next, we performed all-atom MD simulations of the entire 60 subunit nanocompartment in implicit solvent. The structure was allowed to equilibrate for 30

ns at 300 K. Over the course of 30 ns, the diameter of the empty compartment decreased by 3.6% (Figure 2b and Supplementary Figure S2). By contrast, for the shell containing cargo peptides with fixed COMes we observed a much smaller 0.8% decrease in diameter. When the COMes of the cargo peptide was left unconstrained, we still observed a 2.7% decrease in diameter, albeit at a slower rate compared to the empty compartment. Between the starting point (0 ns) and the 30 ns time point, the empty shell decreased by approximately 7% in its rotationally averaged projection (Figure 2c). Hence, our computer model of the empty encapsulin particle corroborates the finding by IMMS, namely that cargo encapsulation induces expansion of the shell.

Subsequent analysis of the dynamics of pentameric subunit of the empty compartment showed that there was a slight twisting motion around the 5-fold symmetry axis that was most pronounced in the E-loop of the subunits (Supplementary Figure S3). As the E-loop makes direct contact with the β -sheet region of the neighboring subunits across the 2-fold symmetry point, this twisting motion 'pulls' adjacent subunits closer together. With cargo peptides in the shell interior, the E-loop remains separated by ~ 1.0 - 1.5 nm distance from the β -sheet region of the neighboring subunits. Cargo was proposed to bind to the same beta-sheet of the encapsulin monomer as the inter-chain contacts of the E-loop(3), but on the opposite side, offering a possible explanation for the observed allosteric effect of cargo (un)binding.

In the original report on the structure of encapsulin, several pores were found in the shell that may be important for substrate flux in and out of the compartment (3). In our simulations, we observed that pores along the five-fold icosahedral symmetry axis were closed after initial equilibration of the PDB structure (Supplementary Figure S3). By contrast, many of the pores on the 2- and 3-fold axes remained open. However, these pores also close over the course of 30 ns at 300 K. When cargo peptides are present, the pores on the 3- and 5-fold symmetry points close during 30 ns of simulations. However, 7-8 out of 60 pores on the 2-fold symmetry points remain open. Therefore, our results suggest that cargo encapsulation leads to formation of an expanded state of the compartment with open pores, which is energetically less favorable in the absence of cargo binding.

Cargo encapsulation decreases the stability of the nanocompartment.

The MD simulations showed the total energy of the shell to decrease by 13.5%, indicating that there might be a difference in stability between the structures for empty and cargo-loaded compartments (see Supplementary Figure S2). To test

whether cargo encapsulation affects the stability of encapsulin, we performed AFM nanoindentation on empty shells vs. shells that were loaded with the natural DyP cargo. These experiments were complemented with nanoindentation *in silico* using Self-Organized Polymer (SOP) models of encapsulin shells (33). This has been proven to be successful at describing the mechanical properties of large biomolecules (24, 34-36). We used nanoindentation *in silico* to probe how the observed structural changes influence the integrity and mechanical properties of the nanocompartment (24, 36).

In the AFM experiments, the particles are imaged before nanoindentation to locate their center (Figure 3a). The contour of some particles appeared hexagonal or pentagonal, in agreement with their icosahedral symmetry. The height distribution of all imaged particles shows two maxima at 20 and 24 nm, which are consistent with particles deposited on their 2-/3-fold or 5-fold icosahedral symmetry axis, respectively (See Supplementary Figure S4). The orientation of the particle could not be unambiguously determined in most cases and we therefore considered the average over all particles for the analysis of the nanoindentation data.

For *in silico* nanoindentation, we employed the simplified C_{α} -based SOP models of encapsulin to represent empty or natural cargo loaded compartments (see Supplementary Methods). Briefly, the model for 'filled' encapsulin is based on the structure from the PDB but without cargo peptides. From this atomic model, a list of 'native' intra- and inter-chain contacts is compiled based on a distance cutoff between the C_{α} atoms. Each intra- or inter-chain contact is modeled with the same interaction potential, the strength of which ϵ_h is based on the average non-bonded energy per intra- or inter-chain contact. We calculated the value of ϵ_h for intra-/inter-chain contacts using the results of all-atom MD simulations (see coarse-graining procedure summarized in Supplementary Methods). Building on this same approach, the SOP model for empty encapsulin was constructed from the structure after 30 ns equilibration. Following the Langevin dynamics, *in silico* nanoindentation measurements were performed using a hard sphere of 15 nm radius to represent the AFM tip.

Encapsulin shells show a typical mechanical response to indentation. Stills from *in silico* nanoindentation in Figure 3b illustrates the deformation dynamics of the shells, which undergo I) initial indentation, followed by II) rupture of capsomer-capsomer interactions and III) subsequent structural collapse. Imaging after nanoindentation confirmed that the particles had collapsed under the applied forces (Figure 3c). In agreement with the experimental Force Distance

Curves (FDCs), the *in silico* nanoindentation also shows a linear increase in (deformation) force followed by a sharp discontinuity at higher forces (Figure 3d). Hence, the SOP model offers insight into the dynamic structural changes that define the mechanical response of encapsulin. This allows us to compare the force response for the different icosahedral symmetry types. Close examination of encapsulin structures revealed that rupture of the shell and structural collapse originate at strongly bent regions around the tip-particle contact area (Figure 3b).

When comparing between the different symmetry axes of the compartment (Supplementary Table S2.), it is interesting to note that, whereas indentation along the 2- and 3-fold symmetry axes resulted in a single-step linear indentation, followed by one discrete breaking event, indentation along the 5-fold symmetry axis displayed a two-step force-response. During the first discontinuity in the FDC, the vertex buckles inwards, but the shell does not rupture. Buckling is followed by a second linear force response that leads to collapse at higher forces (Supplementary Figure S5). Such buckling transitions have previously also been observed in *in silico* nanoindentation of HK97 (23). Whereas Finite Element Modeling (FEM) studies of AFM nanoindentation of virus capsids suggested that the buckling transition during indentation corresponds to mechanical failure of the shell, the SOP model shows that for encapsulin, the two structural transitions are not strictly the same (37). We observed that some particles in our AFM experiments display a similar two-step force response to our *in silico* indentations along the 5-fold symmetry axis (Figure S5).

The SOP model was based on the structure of *T. maritima* encapsulin, whereas we studied the effect of cargo encapsulation in compartments from *B. linens*. To validate our approach, we next compared the mechanical response for empty encapsulin from both species. These compartments were found to display indistinguishable spring constants and breaking forces, indicating that the *in silico* nanoindentation data for *T. maritima* encapsulin can be used to interpret the experimental results for *B. linens* encapsulin (Figure 4). As shown previously (38), we also determined that the difference in probe velocity between simulation (1000 nm/s) and experiment (30 nm/s) had only a very minor influence on the resulting spring constants and breaking forces (Supplementary Table S3). We found no evidence for a mechanically distinct population of particles that might be attributable to the 58mer complexes. *In silico* nanoindentation with models of 58mer shells revealed that they are indeed mechanically indistinguishable from complete 60mer compartments, despite the structural defect of missing a dimer (see Supplementary Figure S6).

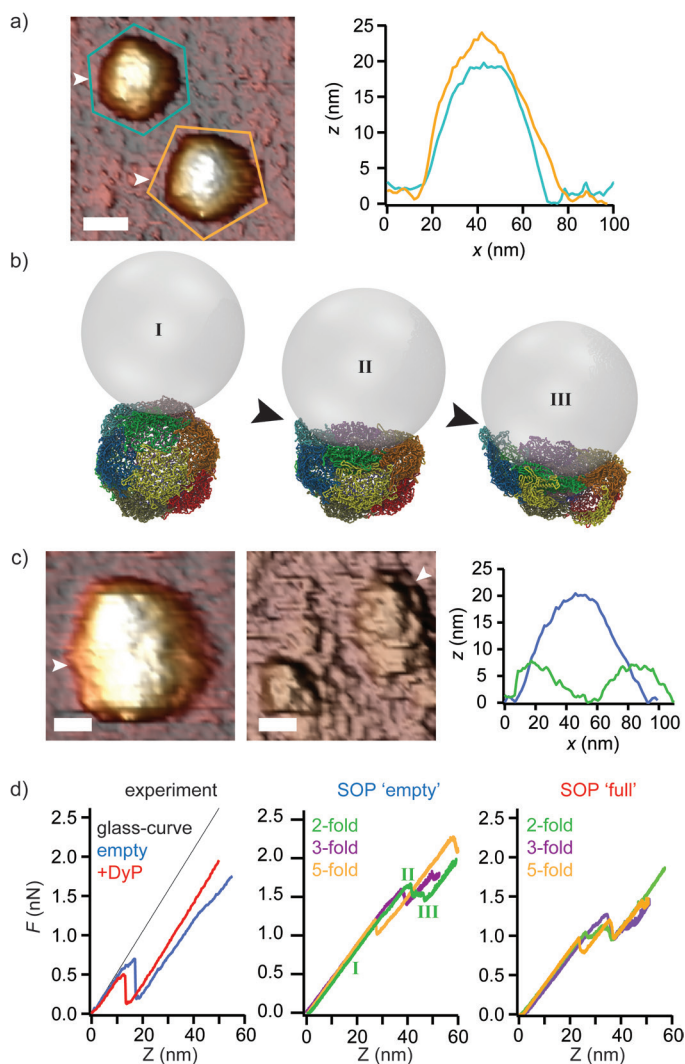


Figure 3. Comparison of experimental AFM nanoindentation of encapsulin nanocompartments with *in silico* indentation. a) AFM images of empty *B. linens* encapsulin. Scale bar is 30 nm, image is colored according to height, from dark brown (0 nm) to white (25 nm). Corresponding height profiles are taken along the lines indicated with white arrowheads. b) Snap-shots from *in silico* nanoindentation of 'empty' encapsulin along the 2-fold symmetry axis (see green curve in "d"), 'SOP empty'). The black arrowhead indicates the region of strong imposed curvature where the shell ruptures at the onset of structural collapse. c) AFM image of encapsulin nanocompartment before and after indentation. Scale bar is 20 nm. Corresponding height profile is taken along the line of the white arrowheads. d) Typical FDCs of encapsulin nanoindentation taken from experiment and simulations ("SOP", models for 'empty' and 'filled' shells). The corresponding parts of the FDC from which the images in "b)" were taken, are indicated on the graph ("SOP 'empty', green curve: I, II and III).

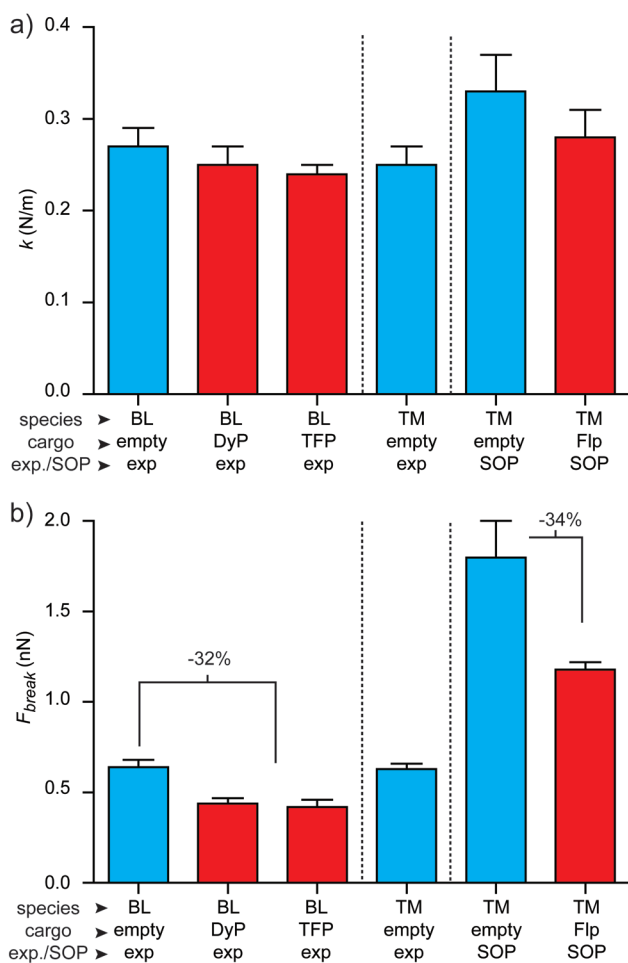


Figure 4. Cargo encapsulation destabilizes encapsulin nanocompartments. a) Spring constant. b) Breaking force. The corresponding species of the encapsulin probed is indicated: *B. linens* (BL) or *T. maritima* (TM). Cargo is indicated as follows: empty (blue), DyP, TFP or Flp (all three red, note that Flp is the native *T. maritima* enzyme). Experiments or simulation (labeled as “exp” and “SOP”, respectively) are reported as the average \pm SEM (see supplementary table S3 for particle numbers). For simulations, the average over the three icosahedral symmetry axes is shown.

Comparing empty and DyP-loaded encapsulin, we observe no change in the average spring constant of the compartment. Simulations also revealed similar values for the spring constant of the empty vs. filled compartments. There is no observed difference in stiffness, but there is a $\sim 30\%$ decrease in breaking force (peak force) for the filled compartment, showing that cargo loading destabilizes the encapsulin shell. The breaking forces are higher in the simulations than in experiments, but we did observe a similar 34% decrease in the breaking force for

filled vs. empty encapsulin. Hence, the shape change observed in simulations is sufficient to explain the destabilizing effect of cargo encapsulation. The free energy required to break encapsulin can be estimated from *in silico* nanoindentation (see Supplementary methods; Supplementary Table S2). Approximately 2.5-3.0 Mcal/mol more energy is required to break the ‘empty’ encapsulin shell, compared to the ‘filled’ shell, accounting for approximately 60% of the 5 Mcal/mol difference in the potential energy of the full vs. empty compartment. It should be noted that the difference in required work is mostly enthalpic and that it is highly dependent on the local symmetry at the point of indentation.

Interestingly, DyP loaded encapsulin also becomes more sensitive to material fatigue, as shown by repeated *in vitro* indentation at a lower force (see Supplementary Figure S7). Finally, we observe a decrease in breaking force upon loading the compartments with foreign material, similar to that observed when loading the encapsulin compartment with the natural DyP cargo. Given that the only structural element that is common between DyP and the foreign Teal Fluorescent Protein cargo is the C-terminal anchoring sequence, this result suggests that the observed destabilization is attributable to interactions of the C-terminal anchoring sequence with the shell.

Discussion.

Our experiments revealed that encapsulin nanocompartments assemble uniquely with a single copy of the native hexameric cargo enzyme (Figure 5). Our computational models suggest that cargo encapsulation stabilizes an expanded state of the nanocompartment (in which the shell exhibits open pores), which is energetically less favorable in the absence of cargo. For the empty encapsulin shell, our simulations indicate a translocation of the E-loop toward the cargo binding β -sheet region of the neighboring subunit, which results in “twisting” motion of the subunits, leading to shell shrinkage and pores closing. This effect almost disappears when cargo peptides are bound to the shell interior. In this respect, the destabilizing effect of the cargo might be a direct consequence of having to maintain a porous shell to allow in-/efflux of enzyme substrates.

The structural similarity of bacterial nanocompartments to virus capsids has raised questions about their origin. Given the high degree of structural homology, it is interesting to see that the nanocompartments are mechanically also similar to bacteriophage capsids. The Young’s Modulus of the shell material can be estimated from the spring constants obtained from the experiment and sim-

ulations (10). We obtain comparable values of ~ 1 GPa from experiment and simulations. The Young's Modulus of bacteriophage HK97 is estimated at essentially the same value (39), showing how the structural similarity between the systems translate into very similar material properties. Hence, encapsulin nanocompartments provide equally robust nanoshells as bacteriophages, with a built-in mechanism for cargo encapsulation. Our AFM-based data on the mechanical resilience of the nanocompartment and our MS-based findings of spontaneously occurring structural defects (i.e. the occurrence of 58mer compartments) provide important considerations for the use of encapsulin in all capacities as assembly scaffold, support structure or as nanoreactor or –container. Having established the natural loading capacity of the compartment provides a further guiding principle for the use of encapsulin as nanoreactor or –container. These novel insights into the assembly and physical properties of the shell thus serve as important guidelines in the rational design and engineering of encapsulin based nanomaterials.

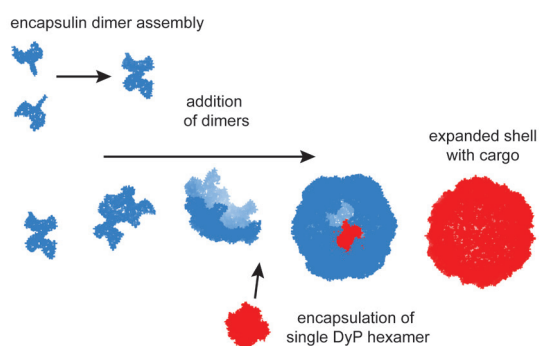


Figure 5. Schematic of encapsulin assembly and cargo encapsulation. We propose an assembly pathway that involves addition of dimers. A single copy of hexameric DyP is encapsulated, triggering expansion of the shell.

Methods.

Sample Preparation.

Empty and DyP or TFP filled encapsulin were expressed in *E. coli* cells and purified using previously described procedures (3, 17).

Native Mass Spectrometry.

Samples were buffer exchanged to 100-200 mM ammonium acetate pH 6.8 with five rounds of concentration-dilution, using Vivaspin 500 10 kDa MWCO cen-

trifugal filter units. Aliquots of 1-2 μL , at a final concentration of 5-10 μM , were loaded into gold-coated borosilicate capillaries (prepared in-house) for nanoelectrospray ionization. Samples were analyzed on a QTOF II instrument that was modified for optimal transmission of large ions (40). The source pressure was set at 10 mbar, collision cell pressure at 1.5×10^{-2} mbar, using xenon as collision gas (41, 42). Capillary voltage was set at 1300-1500 V and the sample cone voltage was set at 160 V. Higher mass ions were analyzed at a collision voltage of 100 V, the lower mass ions at 30-60 V. Additional measurements were performed on an Exactive plus EMR, modified for transmission of higher m/z ions, as described in Chapter 3 (25, 26). Capillary voltage was set at 1200 V, source fragmentation at 30 V. HCD collision voltage was set at 70 V, using xenon as collision gas. Spectra were acquired at a transient time of 32 ms using the transient averaging function. IMMS was performed on a Synapt HDMS instrument. Capillary and sample cone voltages were set at 1300 V and 160 V, respectively. TRAP/TRANSFER regions were operated at 1.5×10^{-2} mbar (xenon) at 30/60 V. The TWIMS cell was operated at 5.5×10^{-1} mbar, using a wave velocity of 200 m/s with a wave height of 2-12 V, ramped over 60% of the cycle. Collisional Cross Sections were calibrated using denatured Ubiquitin, Cytochrome C and Myoglobin (45/5% acetonitrile/formic acid), as well as native GroEL (50 mM ammonium acetate pH 6.8) (43, 44). The rotationally averaged projections of encapsulin models were calculated with the Driftscope PA algorithm.

Atomic Force Microscopy.

Encapsulin samples were prepared in 20 mM Tris pH 7.4, 20 mM magnesium dichloride, 150 mM sodium chloride, at a final concentration of 0.5-1 μM . We used silanized glass slides as substrates for AFM imaging and nanoindentation. The glass slides were prepared by cleaning them in a saturated solution of potassium hydroxide in 90% ethanol, followed by thorough rinsing with MilliQ water, drying and overnight incubation in a hexamethyl disilazane vapour, as previously described (45). A droplet of 100 μL of the encapsulin solution was incubated for 30 minutes on the glass slide, after which another 100 μL of buffer was added. We used Olympus OMCL-RC800PSA rectangular, silicon-nitride cantilevers with a nominal spring constant of 0.05 N/m and a nominal tip radius of 15 nm. Cantilevers were calibrated using the method of Sader *et al.* (46), giving an average value 0.0524 N/m. Measurements were performed with a Nanotec electronica AFM in jumping mode. Imaging was performed with an average maximum imaging force of approximately 30-50 pN. Nanoindentation was performed with a probe velocity of 30 nm/s, unless otherwise stated (38). Fatigue experiments were performed by placing the tip 35 nm above the substrate and performing consecutive approach-retraction cycles of 20-30 nm, which resulted in a maxi-

imum force of 0.2-0.3 nN on the nanocompartment. AFM images were analyzed with WSxM software and nanoindentation data was analyzed using a previously described home-built LabView application (38).

All-atom MD simulations.

The complete 60 subunit shell with $T = 1$ symmetry number was reconstructed from the crystal structure (PDB code 3DKT) (3). To mimic empty (cargo loaded) shell, we removed (retained) cargo peptides bound to the shell interior. To perform MD simulations accelerated on a GPU, we employed the Solvent Accessible Surface Area (SASA) model (47) of implicit solvation and the CHARMM19 force field. First, encapsulin structures were energy-minimized using the steepest-descent algorithm. These were then heated to 300 K and were equilibrated for 1 ns. The structures obtained were used to perform MD simulation runs. To account for the presence of cargo, the COMs of cargo peptides were constrained using a harmonic spring with the spring constant of 10 kcal/mol/Å². To eliminate potential artifacts associated with the SASA model of implicit solvation, we also performed MD simulations using the Generalized Born (GB) model (48). This model is based on a different philosophy for describing implicit solvation, with which we obtained similar results.

SOP-model based nanoindentations *in silico*.

Dynamic force measurements *in silico* were performed using Langevin Dynamics simulations accelerated on a GPU. We used the structure of encapsulin shell obtained before and after a 30 ns all-atom MD simulation run to model the “full” and “empty” particle, respectively. The SOP models of encapsulin particles were constructed as described in Supplementary Methods. We utilized a spherical tip of radius $R_{tip} = 15$ nm to compress the shell along the 2-, 3-, and 5-fold symmetry axis. The tip-shell interactions were described by the repulsive Lennard-Jones potential $V_{LJ}(r_i) = \epsilon[\sigma/(r_i - R_{tip})]^6$, where r_i is the i^{th} particle coordinate, $\epsilon = 4.18$ kJ/mol and $\sigma = 1.0$ Å. To prevent the encapsulin shell from rolling, we constrained the bottom portion of the shell by fixing six C_α-atoms. To mimic the experimental force measurements, the tip exerted the time-dependent force $f(t) = f(t)n$ in the direction n perpendicular to the surface of the shell. The force magnitude $f(t) = r_f t$ increased linearly in time t (force-ramp) with the force-loading rate $r_f = \kappa v_f$, where $v_f = 1.0$ μm/s is the cantilever velocity and $\kappa = 0.05$ N/m is the cantilever spring constant. The resisting force (F) from the encapsulin shell, which corresponds to the experimentally measured indentation force, is calculated using the energy output. In simulations, we controlled the piezo displacement Z (cantilever base), and the cantilever tip position X , both of which quantify the extent of indentation. The shell spring constant k_{enc} was obtained from a generated FZ

curve, using the slope K and the relationship $1/K = 1/\kappa + 1/k_{enc}$ for the combined system (encapsulin plus tip). We evaluated the total work of indentation w_{ind} by converting the FZ curves into the FX curves and by integrating the area under the curves. We repeated this procedure for the retraction curves to evaluate the reversible part of work w_{rev} (see Supplementary Methods). In the NPT-ensemble, the reversible work is $w_{rev} = \Delta G = \Delta H - T\Delta S$, where the thermodynamic state functions ΔG , ΔH , and ΔS are the Gibbs free energy change, enthalpy change and entropy change, respectively. Using the energy output from simulations, we estimated ΔH and $T\Delta S$.

Acknowledgements.

We kindly thank Markus Sutter and Nenad Ban (ETH, Zurich) for supplying plasmids for the expression of encapsulin and for initial guidance in the expression and assistance in the purification of the encapsulin. W.H.R acknowledges the support by a VIDI grant from the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO). A.J.R.H and G.J.L.W acknowledge the support of a Projectruimte grant from Fundamenteel Onderzoek der Materie (FOM). G.J.L.W. acknowledges support by NanoNextNL. V.B. acknowledges the support from the Russian Ministry of Education and Science (Grant 14.A18.21.1239)

Supplementary information.

Supplementary information can be found in the appendix.

References

1. Akita, F., K. T. Chong, H. Tanaka, E. Yamashita, N. Miyazaki, Y. Nakaishi, M. Suzuki, K. Namba, Y. Ono, T. Tsukihara, and A. Nakagawa. 2007. The Crystal Structure of a Virus-like Particle from the Hyperthermophilic Archaeon *Pyrococcus furiosus* Provides Insight into the Evolution of Viruses. *J. Mol. Biol.* 368:1469-1483.
2. McHugh, C. A., J. Fontana, D. Nemecek, N. Cheng, A. A. Aksyuk, J. B. Heymann, D. C. Winkler, A. S. Lam, J. S. Wall, A. C. Steven, and E. Hoiczyk. 2014. A virus capsid-like nanocompartment that stores iron and protects bacteria from oxidative stress. *EMBO J.* 33:1896-1911.
3. Sutter, M., D. Boehringer, S. Gutmann, S. Günther, D. Prangishvili, M. J. Loessner, K. O. Stetter, E. Weber-Ban, and N. Ban. 2008. Structural basis of enzyme encapsulation into a bacterial nanocompartment. *Nature Structural and Molecular Biology.* 15:939-947.
4. Contreras, H., M. S. Joens, L. M. McMath, V. P. Le, M. V. Tullius, J. M. Kimmey, N. Bionghi, M. A. Horwitz, J. A. J. Fitzpatrick, and C. W. Goulding. 2014. Characterization of a *Mycobacterium tuberculosis* nanocompartment and its potential cargo proteins. *J. Biol. Chem.* 289:18279-18289.
5. Yeates, T. O., C. S. Crowley, and S. Tanaka. 2010. Bacterial microcompartment organelles: Protein shell structure and evolution. *Annual Review of Biophysics.* 39:185-205.
6. Yeates, T. O., M. C. Thompson, and T. A. Bobik. 2011. The protein shells of bacterial microcompartment organelles. *Curr. Opin. Struct. Biol.* 21:223-231.
7. Kerfeld, C. A., S. Heinhorst, and G. C. Cannon. 2010. Bacterial microcompartments. *Annual Review of Microbiology.* 64:391-408.
8. Kang, S., and T. Douglas. 2010. Some enzymes just need a space of their own. *Science.* 327:42-43.
9. Ivanovska, I. L., P. J. De Pablo, B. Ibarra, G. Sgalari, F. C. MacKintosh, J. L. Carrascosa, C. F. Schmidt, and G. J. L. Wuite. 2004. Bacteriophage capsids: Tough nanoshells with complex elastic properties. *Proc. Natl. Acad. Sci. U. S. A.* 101:7600-7605.
10. Roos, W. H., R. Bruinsma, and G. J. L. Wuite. 2010. Physical virology. *Nature Physics.* 6:733-743.
11. Roos, W. H., I. L. Ivanovska, A. Evilevitch, and G. J. L. Wuite. 2007. Viral capsids: Mechanical characteristics, genome packaging and delivery mechanisms.

Cellular and Molecular Life Sciences. 64:1484-1497.

12. Hernando-Perez, M., R. Miranda, M. Aznar, J. L. Carrascosa, I. A. Schaap, D. Reguera, and P. J. de Pablo. 2012. Direct Measurement of Phage phi29 Stiffness Provides Evidence of Internal Pressure. *Small*. 8:2366-2370

13. O'Neil, A., C. Reichhardt, B. Johnson, P. E. Prevelige, and T. Douglas. 2011. Genetically programmed in vivo packaging of protein cargo and its controlled release from bacteriophage P22. *Angewandte Chemie - International Edition*. 50:7425-7428.

14. Patterson, D. P., P. E. Prevelige, and T. Douglas. 2012. Nanoreactors by programmed enzyme encapsulation inside the capsid of the bacteriophage P22. *ACS Nano*. 6:5000-5009.

15. Comellas-Aragonès, M., H. Engelkamp, V. I. Claessen, N. A. J. M. Sommerdijk, A. E. Rowan, P. C. M. Christianen, J. C. Maan, B. J. M. Verduin, J. J. L. M. Cornelissen, and R. J. M. Nolte. 2007. A virus-based single-enzyme nanoreactor. *Nature Nanotechnology*. 2:635-639.

16. Wörsdörfer, B., K. J. Woycechowsky, and D. Hilvert. 2011. Directed evolution of a protein container. *Science*. 331:589-592.

17. Rurup, W. F., J. Snijder, M. S. T. Koay, A. J. R. Heck, and J. J. L. M. Cornelissen. 2014. Self-sorting of foreign proteins in a bacterial nanocompartment. *J. Am. Chem. Soc.* 136:3828-3832.

18. Moon, H., J. Lee, J. Min, and S. Kang. 2014. Developing genetically engineered encapsulin protein cage nanoparticles as a targeted delivery nanoplat-form. *Biomacromolecules*. 15:3794-3801.

19. Tamura, A., Y. Fukutani, T. Takami, M. Fujii, Y. Nakaguchi, Y. Murakami, K. Noguchi, M. Yohda, and M. Odaka. 2014. Packaging guest proteins into the encapsulin nanocompartment from *Rhodococcus erythropolis* N771. *Biotechnol. Bioeng.* 112:13-20.

20. Uetrecht, C., and A. J. R. Heck. 2011. Modern Biomolecular Mass Spectrometry and its Role in Studying Virus Structure, Dynamics, and Assembly. *Angewandte Chemie International Edition*. 50:8248-8262.

21. Snijder, J., C. Uetrecht, R. J. Rose, R. Sanchez-Eugenía, G. A. Marti, J. Agirre, D. M. A. Guérin, G. J. L. Wuite, A. J. R. Heck, and W. H. Roos. 2013. Probing the biophysical interplay between a viral genome and its capsid. *Nature Chemistry*. 5:502-509.

22. Zhmurov, A., R. I. Dima, Y. Kholodov, and V. Barsegov. 2010. SOP-GPU: Accelerating biomolecular simulations in the centisecond timescale using graphics

- processors. *Proteins: Structure, Function and Bioinformatics*. 78:2984-2999.
23. Zhmurov, A., K. Rybnikov, Y. Kholodov, and V. Barsegov. 2011. Generation of random numbers on graphics processors: Forced indentation in silico of the bacteriophage HK97. *J Phys Chem B*. 115:5278-5288.
24. Kononova, O., J. Snijder, M. Brasch, J. Cornelissen, R. I. Dima, K. A. Marx, G. J. L. Wuite, W. H. Roos, and V. Barsegov. 2013. Structural transitions and energy landscape for cowpea chlorotic mottle virus capsid mechanics from nanomanipulation in vitro and in silico. *Biophys. J*. 105:1893-1903.
25. Rose, R. J., E. Damoc, E. Denisov, A. Makarov, and A. J. R. Heck. 2012. High-sensitivity Orbitrap mass analysis of intact macromolecular assemblies. *Nature Methods*. 9:1084-1086.
26. Snijder, J., M. Van De Waterbeemd, E. Damoc, E. Denisov, D. Grinfeld, A. Bennett, M. Agbandje-Mckenna, A. Makarov, and A. J. R. Heck. 2014. Defining the stoichiometry and cargo load of viral and bacterial nanoparticles by orbitrap mass spectrometry. *J. Am. Chem. Soc.* 136:7295-7299.
27. Nguyen, H. D., V. S. Reddy, and C. L. Brooks III. 2007. Deciphering the kinetic mechanism of spontaneous self-assembly of icosahedral capsids. *Nano Letters*. 7:338-344.
28. Rapaport, D. C. 2012. Molecular dynamics simulation of reversibly self-assembling shells in solution using trapezoidal particles. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*. 86:051917.
29. Pierson, E. E., D. Z. Keifer, L. Selzer, L. S. Lee, N. C. Contino, J. C. -. Wang, A. Zlotnick, and M. F. Jarrold. 2014. Detection of Late Intermediates in Virus Capsid Assembly by Charge Detection Mass Spectrometry. *J. Am. Chem. Soc.* 136:3536-3541.
30. Uetrecht, C., R. J. Rose, E. Van Duijn, K. Lorenzen, and A. J. R. Heck. 2010. Ion mobility mass spectrometry of proteins and protein assemblies. *Chem. Soc. Rev.* 39:1633-1655.
31. Uetrecht, C., I. M. Barbu, G. K. Shoemaker, E. Van Duijn, and A. J. R. Heck. 2011. Interrogating viral capsid assembly with ion mobility-mass spectrometry. *Nature Chemistry*. 3:126-132.
32. Knapman, T. W., V. L. Morton, N. J. Stonehouse, P. G. Stockley, and A. E. Ashcroft. 2010. Determining the topology of virus assembly intermediates using ion mobility spectrometry-mass spectrometry. *Rapid Communications in Mass Spectrometry*. 24:3033-3042.
33. Hyeon, C., R. I. Dima, and D. Thirumalai. 2006. Pathways and Kinetic Bar-

riers in Mechanical Unfolding and Refolding of RNA and Proteins. *Structure*. 14:1633-1645.

34. Zhmurov, A., A. E. X. Brown, R. I. Litvinov, R. I. Dima, J. W. Weisel, and V. Barsegov. 2011. Mechanism of fibrin(ogen) forced unfolding. *Structure*. 19:1615-1624.

35. Duan, L., A. Zhmurov, V. Barsegov, and R. I. Dima. 2011. Exploring the mechanical stability of the C2 domains in human synaptotagmin 1. *J Phys Chem B*. 115:10133-10146.

36. Kononova, O., Y. Kholodov, K. E. Theisen, K. A. Marx, R. I. Dima, F. I. Ataul-lakhanov, E. L. Grishchuk, and V. Barsegov. 2014. Tubulin bond energies and microtubule biomechanics determined from nanoindentation in silico. *J. Am. Chem. Soc.* 136:17036-17045.

37. Klug, W. S., W. H. Roos, and G. J. L. Wuite. 2012. Unlocking internal prestress from protein nanoshells. *Phys. Rev. Lett.* 109:168104.

38. Snijder, J., I. L. Ivanovska, M. Baclayon, W. H. Roos, and G. J. L. Wuite. 2012. Probing the impact of loading rate on the mechanical properties of viral nanoparticles. *Micron*. 43:1343-1350.

39. Roos, W. H., I. Gertsman, E. R. May, C. L. Brooks III, J. E. Johnson, and G. J. L. Wuite. 2012. Mechanics of bacteriophage maturation. *Proc. Natl. Acad. Sci. U. S. A.* 109:2342-2347.

40. Van Den Heuvel, R. H. H., E. Van Duijn, H. Mazon, S. A. Synowsky, K. Lorenzen, C. Versluis, S. J. J. Brouns, D. Langridge, J. Van Der Oost, J. Hoyes, and A. J. R. Heck. 2006. Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. *Anal. Chem.* 78:7473-7483.

41. Tahallah, N., M. Pinkse, C. S. Maier, and A. J. R. Heck. 2001. The effect of the source pressure on the abundance of ions of noncovalent protein assemblies in an electrospray ionization orthogonal time-of-flight instrument. *Rapid Communications in Mass Spectrometry*. 15:596-601.

42. Lorenzen, K., C. Versluis, E. van Duijn, R. H. H. van den Heuvel, and A. J. R. Heck. 2007. Optimizing macromolecular tandem mass spectrometry of large non-covalent complexes using heavy collision gases. *International Journal of Mass Spectrometry*. 268:198-206.

43. Bush, M. F., Z. Hall, K. Giles, J. Hoyes, C. V. Robinson, and B. T. Ruotolo. 2010. Collision cross sections of proteins and their complexes: A calibration framework and database for gas-phase structural biology. *Anal. Chem.* 82:9557-9565.

44. Ruotolo, B. T., J. L. P. Benesch, A. M. Sandercock, S. - Hyung, and C. V. Robinson. 2008. Ion mobility-mass spectrometry analysis of large protein complexes. *Nature Protocols*. 3:1139-1152.
45. Roos, W. H. 2011. How to perform a nanoindentation experiment on a virus. *Methods in Molecular Biology*. 783:251-264.
46. Sader, J. E., J. W. M. Chon, and P. Mulvaney. 1999. Calibration of rectangular atomic force microscope cantilevers. *Rev. Sci. Instrum.* 70:3967-3969.
47. Ferrara, P., J. Apostolakis, and A. Caflisch. 2002. Evaluation of a fast implicit solvent model for molecular dynamics simulations. *Proteins: Structure, Function and Genetics*. 46:24-33.
48. Onufriev, A., D. Bashford, and D. A. Case. 2000. Modification of the generalized born model suitable for macromolecules. *J Phys Chem B*. 104:3712-3720.

Chapter 7:

Future outlook.

Summary.

Samenvatting in het Nederlands.

Acknowledgements.

Curriculum vitae.

List of publications.

Future outlook.

The work described in this thesis demonstrates how native mass spectrometry can be used to study protein complexes with masses up to several megadaltons. We have shown in Chapter 2 that on a modified QTOF mass spectrometer, virus capsids as big as 18 megadalton can be detected with resolved charge states, allowing mass determinations with a precision below 0.1%. Whereas improvements in sensitivity are always welcome for any analytical technique, it was revealed that on the QTOF platform not sensitivity, but the obtained peak widths are currently the major limiting factor, thereby identifying poor desolvation of the ions as a main bottleneck for the current practical mass limits in native MS. The new Orbitrap-based platform described in Chapter 3 is also capable of analyzing virus particles of a few megadaltons, with higher resolving power, but still exhibits problems with sensitivity towards very high m/z ions. The upper mass limit for protein complexes on this platform therefore seems currently limited to approximately 5 megadalton. The work described in subsequent chapters demonstrates how aspects of structure, assembly, composition and cargo encapsulation of virus(-like) particles can all be studied by native MS. These studies of large protein complexes present an interesting interaction between studying the biology of these proteins and further pushing the development of our analytical tools.

An ideal MS experiment would be able to yield an accurate and quantitative distribution of masses in a sample over a wide dynamic range, while simultaneously offering the possibility to dissect each component in that mass distribution from the top down to determine its composition, including localization of the subunits and post-translational modifications they may bear. We are currently still quite far removed from such a comprehensive mass spectrometry experiment. As discussed, TOF-based native MS experiments still lack resolving power and might benefit tremendously from improved desolvation of ions. If sharper peaks can be obtained, such that charge resolved spectra can be acquired for larger protein complexes as well, sensitivity is the next important bottleneck, as seems to be the case for the Orbitrap-based platform described in Chapter 3. For TOF analyzers, improved sensitivity at high m/z may be obtained by adapting new detectors that are being developed for applications in MALDI – where ions typically obtain a lower number of charges – such as cryo detection TOF analyzers, nanomembrane detectors and active pixel detectors (1-5).

We discussed strategies to improve transmission of high m/z species on the Orbitrap platform in Chapter 3, which involved changes to the vacuum stages and

the HCD cell. If ion detection with the Orbitrap could be further improved in terms of the signal-to-noise ratio, single particle charge detection MS would also appear feasible on an Orbitrap based platform. Following the strategy of the charge detection MS device of Jarrold and co-workers, trapping and detection of single ions at sufficient signal-to-noise ratio would allow the magnitude of the signal to be determined precisely and accurately enough to become a direct measure of charge state (6-9).

Though native MS has seen great success in studying relatively monodisperse protein complexes up to several megadalton, as well as smaller polydisperse assemblies, it is currently still incapable of providing meaningful mass information of polydisperse samples of larger assemblies such as cytoskeletal filaments, amyloid fibrils or structures with lipid bilayers such as small vesicles and enveloped viruses. A single particle mass spectrometry approach would circumvent current issues with highly heterogeneous samples. Resolving individual charge states is no longer necessary as both m/z and z are determined for every individual ion. Alternative single particle MS approaches have also been described using nano(electro) mechanical mass sensors, which circumvent the problems with charge determination altogether (10, 11). Though these approaches appear very promising, they still lack the precision and accuracy to confidently assign stoichiometries to protein assemblies, especially when they might consist of many different components of which some might be relatively small. To combine the precision of TOF and Orbitrap based native MS platforms with a single molecule experiment would arguably be a big step forward. The development of on-line separation methods of intact protein complexes such as size-exclusion chromatography or capillary electrophoresis with improved sensitivity and selectivity would also improve our capacity to deal with heterogeneity in a native MS experiment. In that regard it should be noted that current single-particle MS methods are not nearly fast enough yet to be compatible with on-line chromatography.

A next step would be to improve the possibility of top-down characterization on native MS instruments. Currently available fragmentation schemes yield only limited information about the composition of the protein complex. CID is strongly limited by the available number of charges and only yields sequence information in rare cases (12). The requirement for high precursor charge is also limiting the use of ETD in native MS. Developing new fragmentation schemes such as the use of surface induced dissociation, UV photo-dissociation and MS3 type experiments will improve top-down characterization of intact protein complexes to further contribute towards such a comprehensive native MS experi-

ment (13-17). The ability for top-down characterization is also somewhat limited by the current mass range of available quadrupole mass filters. In addition, selection with the quadrupole occurs before desolvation in most instruments, which limits our ability to separate components in this selection stage of the instrument. Improving the capacity for CID in the source or interface region of the instruments could be helpful in that regard.

There are thus plenty of options left to explore to improve the power of a native MS experiment. At some point, though, the field must run into some fundamental limitations of the approach or find some greater challenges ahead to say the least. We have talked about the problems associated with buffer adduct formation repeatedly. It being mostly an undesired artefact of electrospray ionization, it seems like a difficult problem to circumvent altogether. Seeing how electrospray ionization is such an important tool for native mass spectrometry, it is perhaps also ironic to what extent the field is limited by the restrictions posed by ESI. It is because of ESI that a native MS experiment is limited in concentration range, sensitivity and restricted in the use of buffer components. Also, unpredictable ionization efficiencies make quantitation inherently tedious in an ESI experiment.

Being an essential core technology in ESI-MS, electrospray ionization is also the biggest bottleneck in native mass spectrometry, as it is the main limiting factor of spectrum quality. For now, better mass analyzers can only compensate so much. The isotope distributions of protein complexes become increasingly broader with larger assemblies, until resolving the individual adducted species and assigning them by accurate mass will require resolutions in the 10^6 - 10^9 range and accuracies in the parts-per-billion range, respectively (18). I do not see those requirements being met in the near future; not across the entire relevant m/z range for native MS experiments. Also note that, as we focus on larger protein assemblies, consisting of a greater number of subunits, with a greater number of ligands bound, the relative precision in mass that is required to constrain all possible permutations of stoichiometries that fit an experimental mass becomes impractical, because of these issues with buffer adduct formation.

Native MS should be viewed as a valuable complement to the available analytical tools in structural biology, especially where it concerns large protein complexes. It can be used to study the ensemble of quaternary structures or stoichiometries of protein complexes, without a bias towards structured regions, as is common in tools like x-ray crystallography and cryo electron microscopy. Conversely, native MS offers little in terms of the spatial arrangement of components in a protein

complex. To get the bigger picture, a combination of complementary methods is thus required. Mass spectrometry based structural biology in a broader sense, including not only native MS but also cross-linking MS and hydrogen-deuterium exchange MS, is emerging or perhaps already established as a key focus of hybrid methods in structural biology. Alongside the recent revolution in cryo electron microscopy (19), the development of native MS and MS-based structural biology is offering structural details on a wider variety of protein complexes than ever before.

Recent years have shown more and more examples of MS-based structural biology techniques combined with other, perhaps more established analytical tools in structural biology (20-23). This hybrid approach of integrating structural constraints from many techniques is inherently more comprehensive and offers more confident models. As illustrated by the combination of native MS with atomic force microscopy in chapters 5 and 6 of this thesis, the added value of specificity offered by native MS also argues in favor of a hybrid approach in biophysics. All things considered, native MS is a powerful analytical tool to study the composition and stoichiometry of large protein complexes. With further development of TOF and Orbitrap instrumentation, single particle CD-MS and NEMS approaches and improved strategies for top-down characterization of protein complexes, I envisage an increasingly bigger role for native MS in fields like structural biology, biophysics and nanotechnology.

References

1. Aksenov, A. A., and M. E. Bier. 2008. The Analysis of Polystyrene and Polystyrene Aggregates into the Mega Dalton Mass Range by Cryodetection MALDI TOF MS. *J. Am. Soc. Mass Spectrom.* 19:219-230.
2. Ellis, S. R., J. H. Jungmann, D. F. Smith, J. Soltwisch, and R. M. A. Heeren. 2013. Enhanced detection of high-mass proteins by using an active pixel detector. *Angewandte Chemie - International Edition.* 52:11261-11264.
3. Park, J., Z. Aksamija, H. - Shin, H. Kim, and R. H. Blick. 2013. Phonon-assisted field emission in silicon nanomembranes for time-of-flight mass spectrometry of proteins. *Nano Letters.* 13:2698-2703.
4. Park, J., H. Qin, M. Scalf, R. T. Hilger, M. S. Westphall, L. M. Smith, and R. H. Blick. 2011. A mechanical nanomembrane detector for time-of-flight mass spectrometry. *Nano Letters.* 11:3681-3684.
5. Wenzel, R. J., U. Matter, L. Schultheis, and R. Zenobi. 2005. Analysis of megadalton ions using cryodetection MALDI time-of-flight mass spectrometry. *Anal. Chem.* 77:4329-4337.
6. Contino, N. C., and M. F. Jarrold. 2013. Charge detection mass spectrometry for single ions with a limit of detection of 30 charges. *International Journal of Mass Spectrometry.* 345-347:153-159.
7. Keifer, D. Z., E. E. Pierson, J. A. Hogan, G. J. Bedwell, P. E. Prevelige, and M. F. Jarrold. 2014. Charge detection mass spectrometry of bacteriophage P22 pro-capsid distributions above 20 MDa. *Rapid Communications in Mass Spectrometry.* 28:483-488.
8. Pierson, E. E., D. Z. Keifer, L. Selzer, L. S. Lee, N. C. Contino, J. C. - Wang, A. Zlotnick, and M. F. Jarrold. 2014. Detection of Late Intermediates in Virus Capsid Assembly by Charge Detection Mass Spectrometry. *J. Am. Chem. Soc.* 136:3536-3541.
9. Contino, N. C., E. E. Pierson, D. Z. Keifer, and M. F. Jarrold. 2013. Charge detection mass spectrometry with resolved charge states. *J. Am. Soc. Mass Spectrom.* 24:101-108.
10. Naik, A. K., M. S. Hanay, W. K. Hiebert, X. L. Feng, and M. L. Roukes. 2009. Towards single-molecule nanomechanical mass spectrometry. *Nature Nanotechnology.* 4:445-450.
11. Hanay, M. S., S. Kelber, A. K. Naik, D. Chi, S. Hentz, E. C. Bullard, E. Colinet, L. Duraffourg, and M. L. Roukes. 2012. Single-protein nanomechanical mass

spectrometry in real time. *Nature Nanotechnology*. 7:602-608.

12. Benesch, J. L. P. 2009. Collisional Activation of Protein Complexes: Picking Up the Pieces. *J. Am. Soc. Mass Spectrom.* 20:341-348.

13. Cannon, J. R., M. B. Cammarata, S. A. Robotham, V. C. Cotham, J. B. Shaw, R. T. Fellers, B. P. Early, P. M. Thomas, N. L. Kelleher, and J. S. Brodbelt. 2014. Ultraviolet photodissociation for characterization of whole proteins on a chromatographic time scale. *Anal. Chem.* 86:2185-2192.

14. Cammarata, M. B., and J. S. Brodbelt. 2015. Structural characterization of holo- and apo-myoglobin in the gas phase by ultraviolet photodissociation mass spectrometry. *Chemical Science*. 6:1324-1333.

15. Belov, M. E., E. Damoc, E. Denisov, P. D. Compton, S. Horning, A. A. Makarov, and N. L. Kelleher. 2013. From protein complexes to subunit backbone fragments: A multi-stage approach to native mass spectrometry. *Anal. Chem.* 85:11163-11173.

16. Wysocki, V. H., C. M. Jones, A. S. Galhena, and A. E. Blackwell. 2008. Surface-Induced Dissociation Shows Potential to Be More Informative Than Collision-Induced Dissociation for Structural Studies of Large Systems. *J. Am. Soc. Mass Spectrom.* 19:903-913.

17. Zhou, M., S. Dagan, and V. H. Wysocki. 2012. Protein subunits released by surface collisions of noncovalent complexes: Nativelike compact structures revealed by ion mobility mass spectrometry. *Angewandte Chemie - International Edition*. 51:4336-4339.

18. Lössl, P., J. Snijder, and A. J. R. Heck. 2014. Boundaries of mass resolution in native mass spectrometry. *J. Am. Soc. Mass Spectrom.* 25:906-917.

19. Bai, X. -, G. McMullan, and S. H. W. Scheres. 2014. How cryo-EM is revolutionizing structural biology. *Trends Biochem. Sci.* 40:49-57.

20. Geiger, S. R., K. Lorenzen, A. Schrieck, P. Hanecker, D. Kostrewa, A. J. R. Heck, and P. Cramer. 2010. RNA Polymerase I Contains a TFIIF-Related DNA-Binding Subcomplex. *Mol. Cell.* 39:583-594.

21. Greber, B. J., D. Boehringer, A. Leitner, P. Bieri, F. Voigts-Hoffmann, J. P. Erzberger, M. Leibundgut, R. Aebersold, and N. Ban. 2014. Architecture of the large subunit of the mammalian mitochondrial ribosome. *Nature*. 505:515-519.

22. Wiedenheft, B., E. Van Duijn, J. Bultema, S. Waghmare, K. Zhou, A. Barendregt, W. Westphal, A. Heck, E. Boekema, M. Dickman, and J. A. Doudna. 2011. RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc. Natl. Acad. Sci. U. S. A.*

108:10092-10097.

23. Jore, M. M., M. Lundgren, E. Van Duijn, J. B. Bultema, E. R. Westra, S. P. Waghmare, B. Wiedenheft, U. Pul, R. Wurm, R. Wagner, M. R. Beijer, A. Barendregt, K. Zhou, A. P. L. Snijders, M. J. Dickman, J. A. Doudna, E. J. Boekema, A. J. R. Heck, J. Van Der Oost, and S. J. J. Brouns. 2011. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nature Structural and Molecular Biology*. 18:529-536.

Acknowledgements.

After having spent one year in Albert Heck's lab as a master student, I got the opportunity to stay an additional four years to work on a PhD thesis. Now that the work is almost done, I want to thank Albert for the opportunity. I had a dozen interesting projects to work on in a great lab filled with the most incredible mass spectrometers. I collaborated with many great scientist from all over the world that took an interest because of the technology that Albert developed. I got every opportunity to present our work before the international mass spectrometry and virology communities. I learned a great deal about mass spectrometry, viruses and science in general. Thanks for taking an interest and for all the useful suggestions, encouragement and opportunities these past years. I am proud of the work of which I owe so much to Albert and the great group of people that he has gathered in Utrecht. When I started out in Albert's lab I worked under supervision of Rebecca Burnley. She has been great teaching me about the way of the mass spectrometrists. She was there with me from my first native MS spectrum to my first project to my first paper and has been incredibly helpful, patient and kind along the way. I owe a lot to the time, effort and attention that she invested in my projects as well.

Of course, a lab full of mass spectrometers doesn't run itself. It needs technicians taking care of business, dealing with moody instruments and moody scientists. So, I would like to thank Arjan (and his spirit animal Anja) for taking care of us in the native MS labs. Also thanks to Mirjam, Soenita, Harm and Jamila for their help and patience during my brief ventures into proteomics. A group of scientists also doesn't run itself. It needs a secretary taking care of business. Many thanks to Corine for helping out with all the paperwork, making arrangements for guests and students, all that stuff. Her help made life in the lab a lot easier. Thanks to all the past and present members of Albert's lab, especially the native MS group, for their useful comments and suggestions and their company in the lab. Thanks to MS Vision, especially Jan Commandeur and Jerre van der Horst for their help and discussion about our instruments. Thanks to Charlotte Utrecht, for laying the groundwork for many of my projects. Thanks to Michiel, for the work we did and for dealing with me on our trips to Bremen and Bloomington. Thanks to Hongtao, Vincent and Liana for their company in the office. Thanks to Chiara and Judith for the work and effort they put into their internships in the lab.

Thanks to the many collaborators I had the fortune of working with during this PhD work. Especially Wouter Roos and Gijs Wuite of the Vrije Universiteit Am-

sterdam. I enjoyed my internship in your lab as a master student, which turned out to be a great introduction to the field of physical virology and contributed a significant portion to the chapters of this thesis. Also thanks to Anika Wiegard and Ilka Axmann; David Veessler, Andrew Routh and Jack Johnson; Ruben Sanchez-Eugenía, Jon Agirre and Diego Guerin; Eugen Damoc, Eduard Denisov and Alexander Makarov; Rik Rurup, Melanie Brasch, Melissa Koay and Jeroen Cornelissen; Crystal Moyer and Glen Nemerow; Vijay Reddy; Ewa Szolajska and Jadwiga Chroboczek; Antonette Bennett and Mavis Agbandje-McKenna; Adrien Melquiond and Alexandre Bonvin, Olga Kononova and Valeri Barsegov; Dennis Thomas and Juergen Plitzko; Matthew Glover, Rachel Shi and David Clemmer; Matti Pronker and Bert Janssen; Remco Rodenburg and Piet Gros; Goran Milic, Manajit Hayer-Hartl and Franz-Ulrich Hartl.

Summary.

Native mass spectrometry (MS) is a powerful tool to study the composition and quaternary structure of protein complexes over a wide range of size and mass. As an analytical tool, native MS offers unmatched specificity and precision to pinpoint the stoichiometry of biomolecular complexes. It has been developed for analysis of larger protein complexes, such as intact virus particles, in recent years. The use of native MS was shown to be particularly convenient to characterize the assembly and composition of these large particles, offering important information on their role as pathogens and their use for nanotechnology and medicine. This thesis describes the development and application of native MS to analyze the structure and assembly of virus particles.

It is shown that virus particles as big as 18 megadalton can be analyzed with native MS on a modified quadrupole time-of-flight instrument. Analysis of the obtained peak shapes and theoretical considerations about the inherent spread in masses and the performance of the mass analyzer lead to the conclusion that the 18 megadalton particles are at the limit of what is feasible on the current generation of instruments, and shows that poor desolvation of the ions is currently the main limiting factor. Having established that 18 megadalton capsids can be analyzed by native MS, several maturation intermediates of bacteriophage HK97 were analyzed by native MS to learn about the maturation pathway of the virus.

It is furthermore shown how we extended the mass range of a novel type of Orbitrap mass analyzer for native MS to allow analysis of virus particles of several megadaltons. The transmission at high mass-to-charge ratio is improved by modifying the ion guides of the instrument. Sensitivity is improved 5-fold in the high m/z range to allow analysis of virus particles with improved resolution compared to the time-of-flight based platforms. It is demonstrated that the Orbitrap with extended mass range can be used to quantify cargo loading in a virus-like particle for nanotechnology and that it can be used to determine the stoichiometry of capsids in heterogeneous mixtures.

The use of both the quadrupole time-of-flight and the Orbitrap-based instrument is demonstrated in this thesis in three applications to virus structure and assembly. First, a combination of native MS with proteomics and hydrogen-deuterium exchange MS is used to uncover a previously unknown component of mature adenovirus capsids. It is shown that a product of cement protein pVI is associated with the main capsid protein as a specific site on the capsid interior. Next, it is shown how native MS and atomic force microscopy can be used

to study the interactions between a capsid and its genome. As complementary technologies they are used to uncover a mechanism for pH-triggered disassembly of triatoma virus and establish a model for the assembly of the capsid, based on structural information from native MS. Finally, native MS is used again in combination with atomic force microscopy to study the effect of cargo encapsulation in a recently discovered bacterial nanocompartment named encapsulin. The assembly pathway of the encapsulin particle is studied by native MS and cargo loading is shown to result in a structural change of the compartment that destabilizes the shell.

This thesis thereby illustrates the great potential of native MS to study virus capsid structure and assembly. Several potential improvements of current native MS instrumentation as well as possible fundamental limitations of the technique are discussed to establish the current and future role of native MS in the analytical toolbox to study protein complex structure and assembly.

Samenvatting in het Nederlands.

Natieve massa spectrometrie (MS) is een waardevolle techniek om de samenstelling en quaternaire structuur van eiwitcomplexen over een breed massabereik op te kunnen helderen. Onder de beschikbare analysetechnieken heeft natieve MS ongeëvenaarde specificiteit en precisie in de bepaling van de stoichiometrie van een biomoleculair complex. De techniek is in voorbijgaande jaren ook ontwikkeld voor de analyse van grote eiwitcomplexen zoals intacte virusdeeltjes. Natieve MS is vooral heel bruikbaar om de assemblage en samenstelling van deze deeltjes te bepalen. Daardoor levert het nuttige informatie op over de virusdeeltjes in hun hoedanigheid als ziekteverwekkers, maar ook in hun toepassing in nanotechnologie en geneeskunde. Dit proefschrift beschrijft de verdere ontwikkeling en toepassing van natieve MS om de structuur en assemblage van virusdeeltjes te analyseren.

Er wordt beschreven dat virusdeeltjes van 18 megadalton nog met natieve MS geanalyseerd kunnen worden op een gangbaar quadrupole time-of-flight instrument. Op basis van de bekende prestaties van het instrument, de gemeten piekvormen en theoretische overwegingen aangaande de inherente spreiding van massa wordt geconcludeerd dat 18 megadalton tot de hoogst haalbare massa's behoort op de huidige generatie instrumenten. Op basis van diezelfde overwegingen wordt ook duidelijk dat incomplete desolvatie daarin de belangrijkste limiterende factor is. Na vastgesteld te hebben dat natieve MS bruikbaar is voor de analyse van virusdeeltjes van 18 megadalton wordt de techniek gebruikt om enkele tussenstappen in de rijping van bacteriofaag HK97 te bekijken om zodoende de rijping van de bacteriofaag beter te begrijpen.

Er wordt ook beschreven hoe het m/z bereik van een zeer populair type massa analysator, de zogenaamde Orbitrap, verbeterd kan worden voor natieve MS. Het verbeterde massabereik van het nieuwe instrument maakt nu de analyse van virusdeeltjes van enkele megadaltons mogelijk. Door de ionlenzen van het instrument te manipuleren wordt het transport van ionen met hoge massaladingverhouding verbeterd. Het resulteert in een vijfvoudige toename in de gevoeligheid van het instrument. Daardoor is het nu mogelijk om virusdeeltjes van enkele megadaltons te analyseren met betere resolutie in vergelijking met de analyse op het quadrupole time-of-flight instrument. Het verbeterde Orbitrap instrument is nu geschikt om de vracht van een nanotechnologisch virusdeeltje te kwantificeren en om de stoichiometrie van virusdeeltjes in heterogene mengsels te bepalen.

De toepassing van zowel het quadrupole time-of-flight als het Orbitrap instrument wordt in dit proefschrift aan de hand van drie illustratieve voorbeelden beschreven. Ten eerste wordt een combinatie van natieve MS met proteomics en waterstof-deuterium-uitwisseling MS gebruikt om een onbekend bestanddeel van rijpe adenovirusdeeltjes te karakteriseren. Er wordt beschreven dat een rijpingsproduct van het cementeiwit pVI op een specifieke plaats aan een van de belangrijkste capsidewitten aan de binnenkant van adenovirus bindt. Vervolgens wordt beschreven hoe een combinatie van natieve MS met atomaire krachtmicroscopie gebruikt kan worden om de interactie tussen een virusdeeltje en het virale genoom te karakteriseren. Er wordt beschreven hoe de pH het uiteenvallen van triatoma virus deeltjes kan veroorzaken en op basis van structuurinformatie uit natieve MS experimenten een model voor de assemblage van het deeltje voorgesteld. Tenslotte wordt er opnieuw een combinatie van natieve MS en atomaire krachtmicroscopie experimenten beschreven om het effect van de vracht op een bacterieel nanocompartiment te bepalen. De assemblage van het nanocompartiment met de vracht wordt met natieve MS bestudeerd en er wordt gekeken hoe de vracht een structuurverandering teweegbrengt die het compartiment destabiliseert.

Dit proefschrift illustreert daarmee de volle mogelijkheden van natieve MS om de structuur en assemblage van virusdeeltjes te bestuderen. Uiteindelijk worden enkele mogelijke verbeterpunten voor de huidige instrumenten voorgesteld en enkele fundamentele grenzen van de techniek besproken. Daarmee wordt vastgesteld wat de huidige en toekomstige toepassing van natieve MS is in de context van andere analytische technieken voor de studie van eiwitcomplexen.

Curriculum vitae.

Joost Snijder was born on the 17th of September 1988 in Asperen, The Netherlands. He started his studies in biology at Utrecht University in 2006, obtaining his bachelor's degree in 2009. As part of the research master 'Biomolecular Sciences' of the UU, he spent the next year in the Biomolecular Mass Spectrometry and Proteomics Group of Dr. Albert Heck to learn native mass spectrometry. During his internship he characterized proteins involved in calcium signaling as well as proteins involved in circadian timing in cyanobacteria. In 2010 he moved on to the lab of Dr. Wouter Roos and Dr. Gijs Wuite at the Vrije Universiteit Amsterdam to learn Atomic Force Microscopy (AFM) of virus particles. During this internship, AFM was used to image and manipulate virus capsids with the ultimate aim to understand their mechanical properties. After obtaining his master's degree in 2011, he returned the same year to the group of Dr. Albert Heck as a doctoral candidate. Here, he continued to work on virus capsids, using and developing native mass spectrometry as the method of characterization. The results from this work are presented in this thesis.

List of publications.

1. Snijder, J., M. van de Waterbeemd, M. S. Glover, L. Shi, D. E. Clemmer, and A. J. Heck. 2015. Conformational landscape and pathway of disulfide bond reduction of human alpha defensin. *Protein Sci.* doi: 10.1002/pro.2694.
2. Krishnamurthy, S., D. Veesler, R. Khayat, J. Snijder, R. K. Huang, A. J. R. Heck, J. E. Johnson, and G. S. Anand. 2014. Distinguishing direct binding interactions from allosteric effects in the protease–HK97 prohead I δ domain complex by amide H/D exchange mass spectrometry. *Bacteriophage.* 4:e959816.
3. Snijder, J., and A. J. R. Heck. 2014. Analytical approaches for size and mass analysis of large protein assemblies. *Annual Review of Analytical Chemistry.* 7:43-64.
4. Snijder, J., M. Van De Waterbeemd, E. Damoc, E. Denisov, D. Grinfeld, A. Bennett, M. Agbandje-Mckenna, A. Makarov, and A. J. R. Heck. 2014. Defining the stoichiometry and cargo load of viral and bacterial nanoparticles by orbitrap mass spectrometry. *J. Am. Chem. Soc.* 136:7295-7299.
5. Lössl, P., J. Snijder, and A. J. R. Heck. 2014. Boundaries of mass resolution in native mass spectrometry. *J. Am. Soc. Mass Spectrom.* 25:906-917.
6. Snijder, J., M. Benevento, C. L. Moyer, V. Reddy, G. R. Nemerow, and A. J. R. Heck. 2014. The cleaved N-Terminus of pVI binds peripentonal hexons in mature adenovirus. *J. Mol. Biol.* 426:1971-1979.
7. Benevento, M., S. Di Palma, J. Snijder, C. L. Moyer, V. S. Reddy, G. R. Nemerow, and A. J. Heck. 2014. Adenovirus Composition, Proteolysis and Disassembly Studied by in-depth Qualitative and Quantitative Proteomics. *J. Biol. Chem.* 289:11421-11430.
8. Snijder, J., R. J. Burnley, A. Wiegard, A. S. J. Melquiond, A. M. J. J. Bonvin, I. M. Axmann, and A. J. R. Heck. 2014. Insight into cyanobacterial circadian timing from structural details of the KaiB-KaiC interaction. *Proc. Natl. Acad. Sci. U. S. A.* 111:1379-1384.
9. Veesler, D., R. Khayat, S. Krishnamurthy, J. Snijder, R. K. Huang, A. J. R. Heck, G. S. Anand, and J. E. Johnson. 2014. Architecture of a dsDNA viral capsid in complex with its maturation protease. *Structure.* 22:230-237.
10. Rurup, W. F., J. Snijder, M. S. T. Koay, A. J. R. Heck, and J. J. L. M. Cornelissen. 2014. Self-sorting of foreign proteins in a bacterial nanocompartment. *J. Am. Chem. Soc.* 136:3828-3832.
11. Luque, D., A. De La Escosura, J. Snijder, M. Brasch, R. J. Burnley, M. S. T.

- Koay, J. L. Carrascosa, G. J. L. Wuite, W. H. Roos, A. J. R. Heck, J. J. L. M. Cornelissen, T. Torres, and J. R. Castón. 2014. Self-assembly and characterization of small and monodisperse dye nanospheres in a protein cage. *Chemical Science*. 5:575-581.
12. Snijder, J., C. Uetrecht, R. J. Rose, R. Sanchez-Eugenía, G. A. Martí, J. Agirre, D. M. A. Guérin, G. J. L. Wuite, A. J. R. Heck, and W. H. Roos. 2013. Probing the biophysical interplay between a viral genome and its capsid. *Nature Chemistry*. 5:502-509.
13. Snijder, J., R. J. Rose, D. Veesler, J. E. Johnson, and A. J. R. Heck. 2013. Studying 18 MDa virus assemblies with native mass spectrometry. *Angewandte Chemie - International Edition*. 52:4020-4023.
14. Snijder, J., V. S. Reddy, E. R. May, W. H. Roos, G. R. Nemerow, and G. J. L. Wuite. 2013. Integrin and defensin modulate the mechanical properties of adenovirus. *J. Virol.* 87:2756-2766.
15. Kononova, O., J. Snijder, M. Brasch, J. Cornelissen, R. Dima, K. Marx, G. L. Wuite, W. Roos, and V. Barsegov. 2013. Structural Transitions and Energy Landscape for Cowpea Chlorotic Mottle Virus Capsid Mechanics from Nanomanipulation in Vitro and in Silico. *Biophys. J.* 105:1893-1903.
16. Snijder, J., I. L. Ivanovska, M. Baclayon, W. H. Roos, and G. J. L. Wuite. 2012. Probing the impact of loading rate on the mechanical properties of viral nanoparticles. *Micron*. 43:1343-1350.
17. Snijder, J., R. J. Rose, R. Raijmakers, and A. J. R. Heck. 2011. Site-specific methionine oxidation in calmodulin affects structural integrity and interaction with Ca²⁺/calmodulin-dependent protein kinase II. *J. Struct. Biol.* 174:187-195.

Appendix:

Supplementary information.

Supplementary information to Chapter 2.

Challenging the current upper mass limit of native MS.

Joost Snijder, Rebecca J. Burnley, David Veessler, Jack E. Johnson, Albert J.R.
Heck.

Supplementary figures.

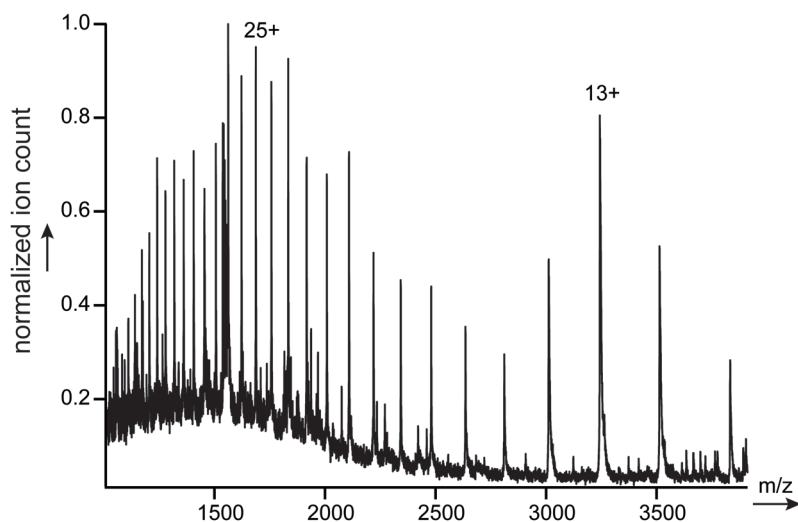


Figure S1. Spectrum of partially denatured capsid subunit monomer, generated by analyzing the Prohead-1^{gp5} samples in 50mM ammonium acetate buffer at pH 9.5. Charge states are as indicated, the average mass determined from this analysis is 42151 ± 1 Da.

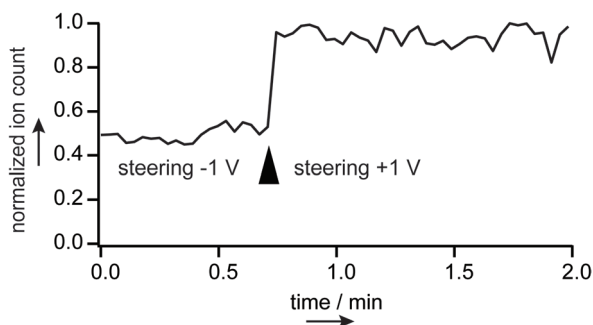


Figure S2. The effect of the 'steering' parameter on transmission of the Prohead-1^{gp5} particle. Shown is the normalized ion count per 2 s scan over a two minute acquisition window. At the start, the steering parameter is set at an optimal value for smaller protein complexes. The black arrowhead indicates where the parameter was changed to improve transmission of Prohead-1^{gp5}. Tuning of the steering parameter resulted in an approximate 2-fold gain in signal.

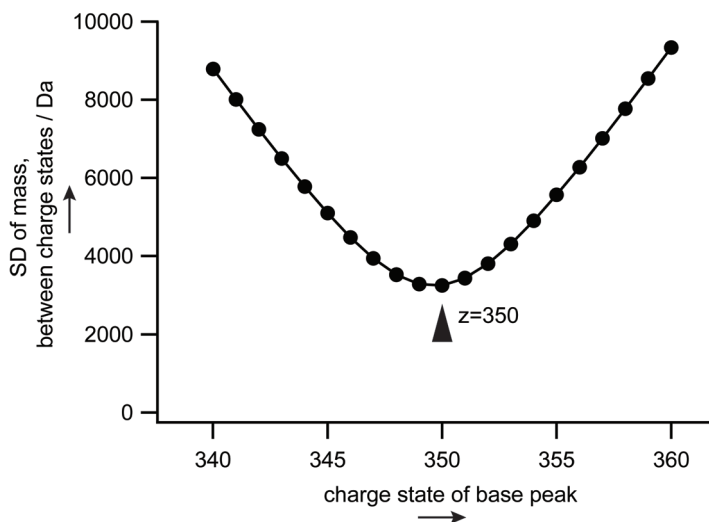


Figure S3. Charge state assignment strategy for Prohead-1^{BP5} spectra. Shown is the variation in calculated mass between charge states from the same spectrum (expressed as the standard deviation in units of Da) as a function of charge state assignment of the base peak. The minimum along this curve, at $z=350$, is considered the correct assignment. Note that the minimum is not very sharply defined: the accuracy of the assignment is assessed from replicate analyses.

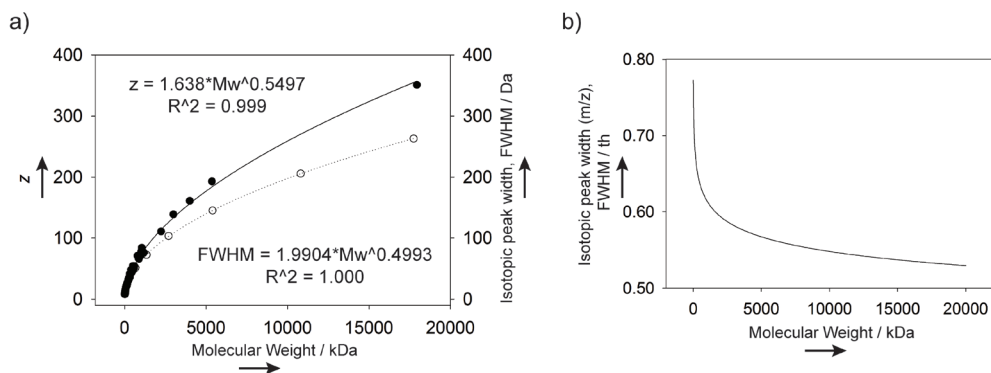


Figure S4. Isotope distribution and charging determine the inherent peak width. a) The experimentally determined base-peak charge states z (closed circles, left y-axis) and the simulated width of the isotope distribution (open circles, right y-axis) are shown as a function of molecular weight. b) The inherent width of protein complexes in nESI is equal to FWHM/z , giving the width of the isotope distribution in the m/z -dimension. As charging scales slightly stronger with molecular weight than FWHM does, the width of the peaks in m/z get increasingly narrow with increasing molecular weight.

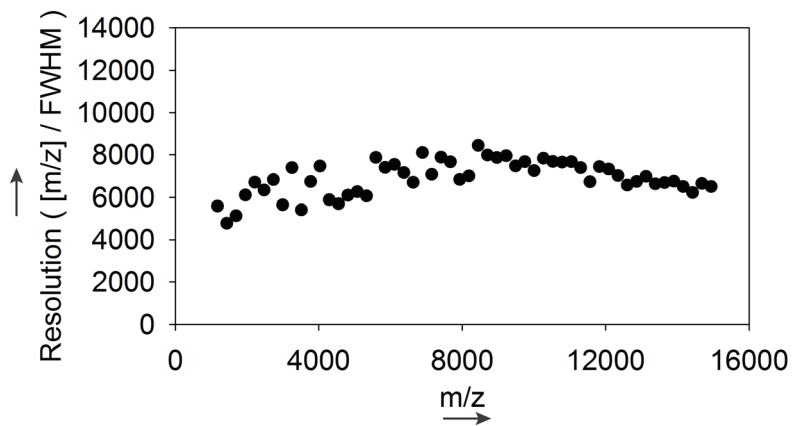


Figure S5. Experimentally determined resolution of the modified QTOF instrument used here. The peak width (FWHM) was determined from a spectrum of 25 mg/mL cesium iodide.

Supplementary Information to Chapter 3.

A new platform with improved resolution for native MS of virus particles.

Joost Snijder, Michiel van de Waterbeemd, Eugen Damoc, Eduard Denisov, Dmitry Grinfeld, Antonette Bennett, Mavis Agbandje-McKenna, Alexander Makarov, Albert J.R. Heck.

Supplementary text.

Limiting factors in transmission efficiency of high mass ions.

Before setting out to make further modifications, in order to extend the accessible mass range beyond what we reported previously, we investigated which factors would influence high mass ion transmission and detection on the Orbitrap mass analyzer. As the effective potential well in the Orbitrap analyzer is independent of mass, there is no fundamental limit on m/z to be detected. However, ion transport from the ion source as well as trapping of ions prior to injection into the analyzer, utilizes RF-only multipoles, which provide focusing in highly m/z -dependent RF quasi-potential wells. For example, in a quadrupole of inscribed radius r_0 with an RF voltage of amplitude V_{RF} applied to the rods (0-peak), the quasi-potential (in Volts) is given by:

$$\varphi(r) \approx (eV_{RF}^2) / ((m/z)(2\pi f)^2 r_0^2) * (r/r_0)^2 \quad (1)$$

and therefore is inversely proportional to m/z . A similar formula is also applicable for other types of RF-only devices, such as multipoles, ion tunnels, funnels, carpets, etc.- just the value of r_0 changes. For typical $f \approx 3.2$ MHz, $r_0 \approx 2.87$ mm (which corresponds to the standard C-trap settings in this instrument), $V_{RF} = 500$ V (0-p), $m/z = 10000$, the RF pseudo-potential wall that prevents the ions from impinging upon the quadrupole electrodes, amounts approximately to 0.7 V. Therefore, any process that supplies ions with radial energy in excess of this amount will simultaneously doom these ions to be lost for analysis. For example, if thermalized ions (with a radial kinetic energy at room temperature of about $kT \approx 0.025$ eV) enter as a parallel beam of radius R into an ion-optical lens with focal distance F followed by an RF-multipole, then after the focal point they will form a beam with angular divergence $\gamma = R/F$. If R is large enough so that for a final acceleration U of ions into a multipole

$$\varphi(r_0) < U * (R/F)^2 \quad (2)$$

then such ions are lost from the multipole. Even for slightly smaller R , with the beam approaching the RF rods, efficiency of ion transport is already drastically reduced. Typically, such lenses are just thin apertures that are used to separate differentially pumped regions and have RF-multipoles on one or both sides. Generally, we could represent their focal distance as

$$F = r_0 * C(U_0, U) \quad (3)$$

where U_0 is the ion energy per elementary charge prior to the lens and coefficient C is determined by the particular geometry. Typically, $U_0 \ll U$ and then $C \approx C(U)$.

The most efficient countermeasure to loss of ions would be to establish a sufficiently high pressure of residual gas in the corresponding multipole, so that any excess radial energy is damped in collisions, preferably over the length comparable to or smaller than F . In addition to this, an increase of RF amplitude or a decrease of frequency or radius of the multipole could be used for more efficient focusing of ions. It should be noted that an excessive increase of pressure becomes detrimental for RF focusing. Substantial damping by collisions takes place in the injection flatapole of the instrument used in this work (Fig. 1), and collisional cooling is also the likely reason why the pressure in the HCD cell needs to be raised to >0.1 mbar for trapping of high-mass ions and their fragments (see below). However, the pressure remains significantly lower in other parts of the instrument: i.e. the bent flatapole, transport octapole and the C-trap. While the pressure in the first two RF multipoles could be elevated relatively easily, it would be more difficult for the C-trap because it directly relates to the residual pressure in the Orbitrap analyzer (with a coefficient of approximately 3×10^{-6}) and the stopping length always exceeds F in the C-trap.

To estimate the limit on m/z , the best-case assumption for an incoming ion beam would include its complete thermalization, so that most of the beam ($\approx 68\%$) is contained within radius R_T where

$$\varphi(R_T) \approx kT \quad (4)$$

Substituting R_T from (4) into (2)-(3), we obtain the upper limit for transmitted m/z :

$$(m/z)_{max} \leq (C(U_0, U)) / ((2\pi f)^2 r_0^2) * (e V_{RF}^2) / \sqrt{U * kT} \quad (5)$$

To estimate the m/z limit for the C-trap, we need to take into account that capture of ions into the trap is accompanied by reflection in the retarding DC potential created by end apertures on both sides of the RF rods. Penetration of the DC potential into a multipole (see Supplementary Figure S1) depends on the distance from the aperture as $\exp(\pm x/h)$ where $1/h^2$ is the eigenvalue of the Laplace operator in the multipole's cross-section. It could be shown that at a sufficient distance from the aperture (of about r_0) the focal distance of the exponential retarding field depends weakly on the ion energy and is $F \approx 0.41 * h \approx 0.2 * r_0$. Therefore, for this field it holds that $C(U, U_0) \approx 0.2$. For a maximum RF voltage of $V_{RF} = 1500$ V

(0-p) and a residual energy from the original jet expansion corresponding to 5-8 V acceleration, this corresponds to $(m/z)_{max} \approx 29000-36000$, which roughly fits with the highest m/z experimentally observed so far in this mass spectrometer (see below). Due to the spread of initial velocities and coordinates, this cut-off is never sharp and well-defined, but rather manifests itself as an increasingly unstable and intensity-suppressed signal at higher m/z .

For acceleration lenses in the bent flatapole, transport octapole and on the entrance to the HCD cell, the focal distance F is much greater, of the order of $1.5 \cdot r_0$ or higher and therefore even lower RF amplitudes are sufficient to retain ions of such m/z at energies up to 10-20 V. However, ions do need to lose almost all their kinetic energy by the time they reach the back of the HCD cell to avoid catastrophic scattering in the exponential retarding field at the turning point. Indeed, formula (5) could be re-written as a limit on maximum kinetic energy of ions as they approach the end of a multipole:

$$U_{max} \leq C(U_0 U)^2 / ((2\pi f)^4 r_0^4) * (e^2 V_{RF}^4) / (kT (m/z)^2) \quad (6)$$

i.e. a change of frequency from e.g. 3.2 MHz to 2.8 MHz increases the allowed energy by a factor 1.7x, while in the exponential retarding field it plummets by a factor of 50x relatively to transmission lenses (see Supplementary Figure S2).

Supplementary tables.

Table S1. Peak assignments for HCD spectra of GroEL.

GroEL 14-mer					
m/z	FWHM	M/ Δ M, FWHM	z	mass (Da)	
10539.5	7.1	1478	76	800928	
10679.9	7.8	1363	75	800921	
10822.5	7.6	1433	74	800792	
10973.4	7.4	1479	73	800982	
11124.3	7.4	1495	72	800878	
11281.0	8.1	1392	71	800882	
11442.0	6.5	1760	70	800871	
				average	standard deviation
				800893	59

GroEL 13-mer				
m/z	FWHM	M/ Δ M, FWHM	z	mass (Da)
13047.1	8.7	1501	57	743628
13278.9	9.2	1451	56	743562
13522.3	9.5	1422	55	743674
13772.2	10.3	1332	54	743647
14032.5	9.8	1427	53	743670
14302.0	10.1	1414	52	743651
14583.0	10.2	1426	51	743683
14874.8	10.5	1422	50	743692
15179.1	10.9	1391	49	743727
15494.8	11.6	1335	48	743702
15825.0	11.8	1335	47	743726
16168.5	12.2	1329	46	743705
16527.8	12.5	1324	45	743705
16904.3	12.3	1374	44	743745
17297.0	12.8	1347	43	743728
17709.4	13.0	1361	42	743754

GroEL 13-mer					
m/z	FWHM	M/ Δ M, FWHM	z	mass (Da)	
18141.9	13.3	1360	41	743778	
18595.4	13.9	1341	40	743777	
19071.8	14.4	1325	39	743760	
19574.8	14.5	1351	38	743804	
20102.9	16.0	1258	37	743770	
20662.3	16.1	1281	36	743808	
21256.4	19.8	1074	35	743939	
				average	standard deviation
				743723	76

GroEL 12-mer					
m/z	FWHM	M/ Δ M, FWHM	z	mass (Da)	
19612.8	16.7	1171	35	686414	
20190.0	17.4	1160	34	686427	
20802.2	18.6	1120	33	686439	
21453.6	18.3	1172	32	686484	
22144.3	18.8	1175	31	686442	
22883.2	20.2	1135	30	686466	
23671.8	21.2	1119	29	686453	
24516.7	22.0	1116	28	686439	
25425.4	24.0	1058	27	686458	
26403.2	25.3	1042	26	686456	
27460.3	27.5	997	25	686481	
28606.6	28.9	989	24	686533	
29850.4	30.3	984	23	686535	
31208.5	31.4	993	22	686566	
32693.2	34.1	958	21	686536	
34327.8	36.1	952	20	686536	
36134.3	36.5	991	19	686533	
38143.8	40.9	934	18	686571	
				average	standard deviation
				686487	51

Table S2. Peak assignments of encapsulin, Dd, AAV1 and CCMV spectra.

Encapsulin + TFP					
m/z	FWHM	M/ Δ M, FWHM	z	mass (Da)	
16144.1	46.3	349	127	2050179	
16277.2	46.5	350	126	2050800	
16404.6	44.4	369	125	2050448	
16540.0	43.4	381	124	2050836	
16675.3	41.8	399	123	2050944	
16813.6	40.4	416	122	2051136	
16951.3	39.6	428	121	2050990	
17092.1	39.5	433	120	2050931	
17235.2	39.1	440	119	2050872	
17378.7	39.6	438	118	2050565	
17522.7	40.7	431	117	2050033	
17666.6	42.4	417	116	2049211	
17814.8	42.7	417	115	2048587	
				average	standard deviation
				2050425	761

Adenovirus 3 Dodecahedron				
m/z	FWHM	M/ Δ M, FWHM	z	mass (Da)
20038.2	43.7	459	176	3526545
20149.5	40.8	493	175	3525989
20263.2	47.9	423	174	3525623
20382.3	45.7	446	173	3525956
20501.1	45.2	454	172	3526009
20623.5	49.1	420	171	3526453
20745.0	48.3	429	170	3526480
20868.7	48.9	426	169	3526636
20992.9	48.0	437	168	3526639
21122.1	50.9	415	167	3527229
21246.6	50.9	417	166	3526776
21377.3	49.6	431	165	3527093
21509.8	54.1	397	164	3527446
21643.9	51.4	421	163	3527794

Adenovirus 3 Dodecahedron					
m/z	FWHM	M/ Δ M, FWHM	z	mass (Da)	
21779.3	54.2	402	162	3528077	
21920.3	48.4	453	161	3529009	
22056.0	51.2	431	160	3528803	
22200.9	55.1	403	159	3529776	
				average	standard deviation
				3527130	1159

Adeno Associated Virus serotype 1, series 1					
m/z	FWHM	M/ Δ M, FWHM	z	mass (Da)	
20578.6	15.4	1336	178	3662809	
20702.4	20.8	995	177	3664151	
20824.0	18.9	1102	176	3664844	
20943.5	21.5	975	175	3664936	
21062.4	21.9	960	174	3664684	
21185.8	23.4	904	173	3664972	
21308.7	23.3	913	172	3664923	
21433.1	23.1	929	171	3664891	
21559.8	20.6	1048	170	3665001	
21686.5	21.0	1035	169	3664846	
21812.8	22.3	976	168	3664384	
21946.8	23.4	939	167	3664944	
22073.1	20.5	1079	166	3663962	
22210.8	24.1	922	165	3664624	
				average	standard deviation
				3664569	601

Adeno Associated Virus serotype 1, series 2				
m/z	FWHM	M/ Δ M, FWHM	z	mass (Da)
20621.4	24.9	829	178	3670429
20742.5	21.2	978	177	3671247
20861.5	21.4	976	176	3671439

Adeno Associated Virus serotype 1, series 2					
m/z	FWHM	M/ Δ M, FWHM	z	mass (Da)	
20980.4	19.7	1063	175	3671388	
21103.9	21.7	974	174	3671906	
21226.7	22.2	957	173	3672048	
21348.0	20.6	1037	172	3671682	
21475.8	21.7	990	171	3672186	
21601.4	21.6	1001	170	3672068	
21727.6	22.0	987	169	3671795	
21856.0	22.2	986	168	3671645	
21989.5	20.0	1102	167	3672078	
22120.9	21.7	1022	166	3671895	
22249.7	19.6	1134	165	3671027	
				average	standard deviation
				3671631	488

Adeno Associated Virus serotype 1, series 3					
m/z	FWHM	M/ Δ M, FWHM	z	mass (Da)	
20658.9	17.8	1163	177	3656452	
20783.7	19.1	1088	176	3657755	
20900.5	20.7	1011	175	3657409	
21022.0	21.0	1001	174	3657656	
21143.6	20.4	1035	173	3657666	
21267.1	20.6	1035	172	3657762	
21391.8	21.8	982	171	3657830	
21518.6	22.2	969	170	3657985	
21643.1	22.9	943	169	3657512	
21771.5	21.8	997	168	3657444	
21903.9	21.6	1016	167	3657781	
22030.2	24.3	906	166	3656851	
22162.5	26.5	835	165	3656652	
				average	standard deviation
				3657443	485

Cowpea Chlorotic Mottle Virus					
m/z	FWHM	M/ Δ M, FWHM	z	mass (Da)	
22623.7	35.9	631	197	4456672	
22737.7	41.7	545	196	4456393	
22850.2	44.5	514	195	4455594	
22972.6	47.9	479	194	4456490	
23089.1	49.7	464	193	4456003	
23208.6	52.5	442	192	4455859	
23331.6	54.6	428	191	4456145	
23452.5	57.8	406	190	4455785	
23578.0	61.5	383	189	4456053	
23705.5	61.2	387	188	4456446	
23835.4	61.8	386	187	4457033	
23967.3	60.0	399	186	4457732	
24094.4	57.4	420	185	4457279	
24223.6	50.2	482	184	4456958	
24349.7	50.3	484	183	4455812	
				average	standard deviation
				4456417	618

Table S3. Quantification of TFP loading in encapsulin from HCD spectra of the loaded nanocompartment. The TFP-mass is calculated from the average determined mass, minus 59 (1st product ion) or 58 (2nd product ion) times the encapsulin monomer mass (28594 Da). The number of TFP (#TFP) is calculated by dividing the TFP-mass by the mass of one TFP monomer (34667 Da).

Assignment (#TFP)	mass (Da)		TFP-component		
	1st product ion	average	standard deviation	TFP-mass	#TFP
8		1964580	626	277534	8.01
9		1999406	197	312360	9.01
10		2034081	232	347035	10.01
11		2068801	200	381755	11.01
12		2103584	378	416538	12.02
2nd product ion	average	standard deviation	TFP-mass	#TFP	
9	1970968	161	312516	9.01	
10	2005560	391	347108	10.01	
11	2040093	275	381641	11.01	
12	2074680	495	416228	12.01	

Table S4. Stoichiometry of AAV1 capsids. Using theoretical VP1/VP2/VP3 masses, all stoichiometry's with theoretical masses that are within two standard deviations of the experimental mass are listed.

peak series	z = 164-177					z = 165-178				
	mass (kDa)		stoichiometry			mass (kDa)		stoichiometry		
	avg	stdev	VP1	VP2	VP3	avg	stdev	VP1	VP2	VP3
1	3643	0.7	0	10	50	3665	0.6	1	10	49
								4	0	56
2	3650	0.5	0	11	49	3672	0.5	1	11	48
			1	8	51					
3	3657	0.5	1	9	50	3679	0.6	2	9	49

Supplementary Figures.

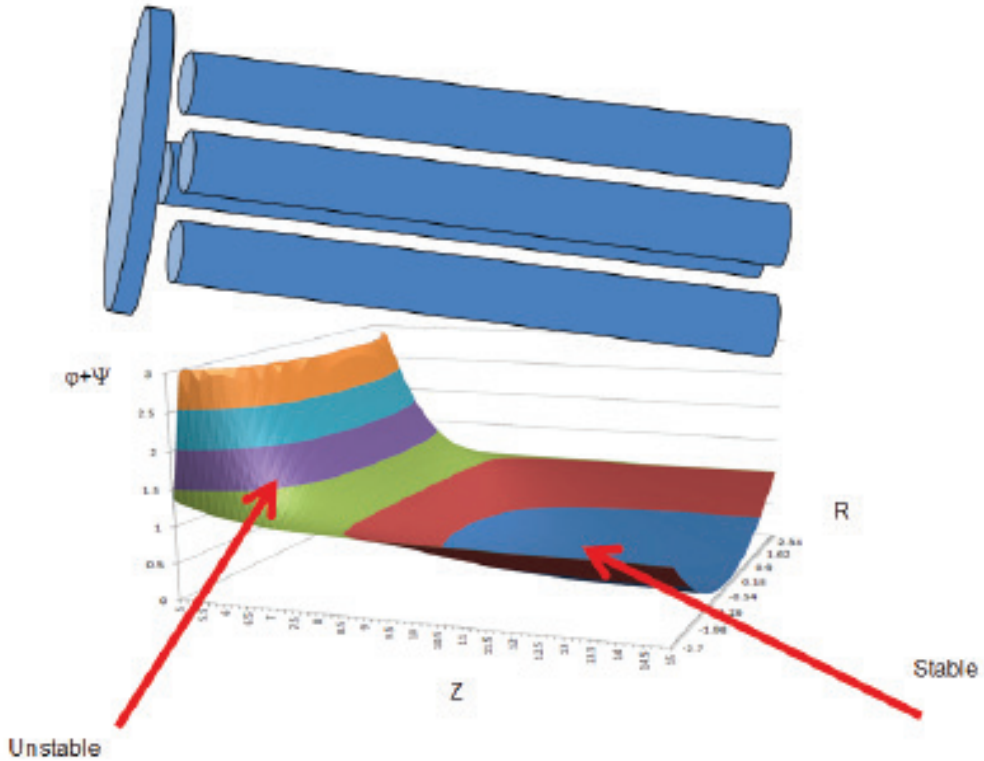


Figure S1. Ion instability near the end aperture of a RF-only field occurs when diverging radial force created by DC potential ψ overpowers focusing by RF quasi-potential ϕ .

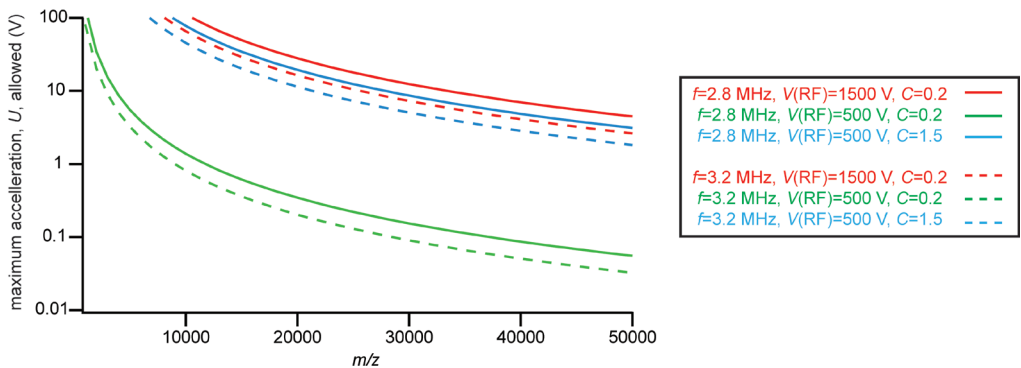


Figure S2. The maximum allowed acceleration of ions (for efficient transmission) as a function of m/z for different f , V_{RF} and C , according to equation 6 of the main text. The curves illustrate how reduced RF frequency (f), higher RF amplitude (V_{RF}) and increased focal distance (via increase in C) are beneficial for transmission of higher m/z ions.

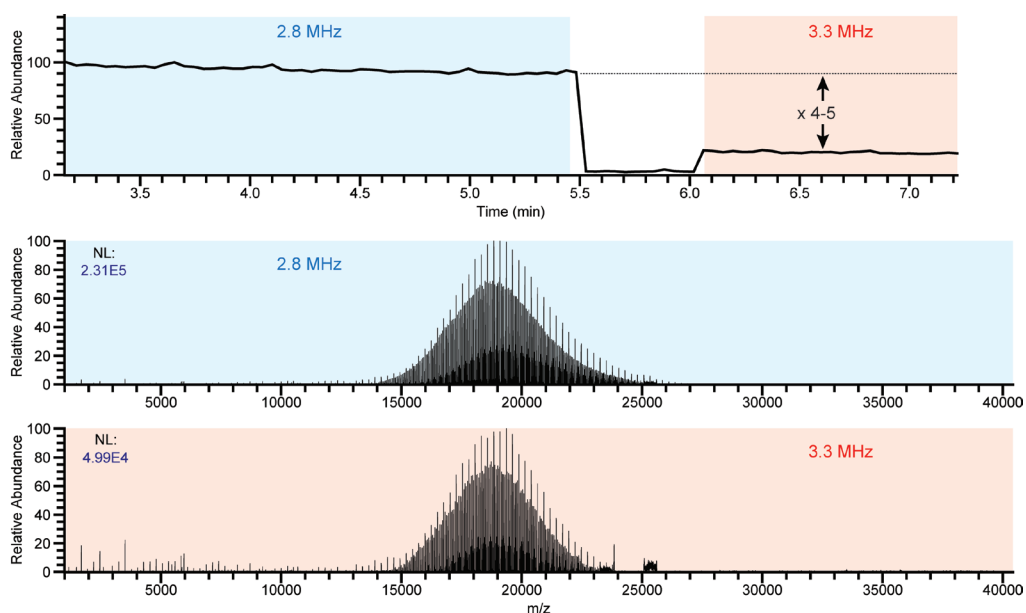


Figure S3. Effect of reduced RF frequency on the injection flatpole and bent flatpole on the transmission of CsI clusters at 20000 m/z . CsI spectra were acquired starting with the lowered RF frequency (2.8 MHz, highlighted in blue). Mid-acquisition, while maintaining flow through the nanoelectrospray capillary, we switched back to the original RF frequency of the ion guides (3.3 MHz, highlighted in red), thereby demonstrating a 4-5 fold gain in transmission due to the lowered RF frequency on these two particular ion guides.

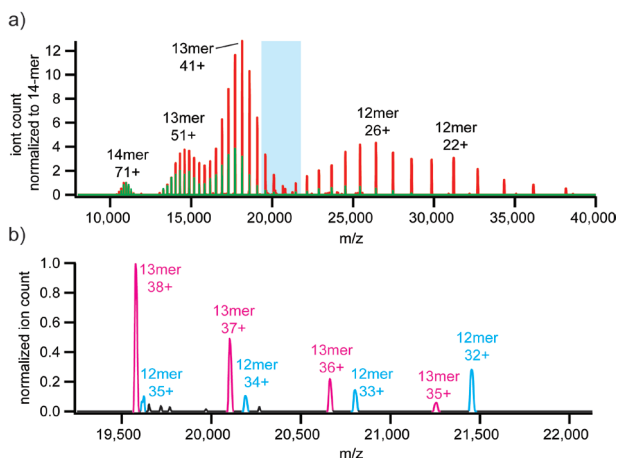


Figure S4. Effect of reduced RF frequency on HCD spectrum of GroEL. a) HCD spectrum of GroEL. Signals are normalized to the most intense 14-mer charge state precursor ion. The acquired spectrum before the modification is shown in green, that after modification in red. b) Zoom-in of the blue highlighted m/z region in “a)” where 13-mer and 12-mer ions overlap. The closely spaced 13/12-mer peaks are still well resolved. For detailed peak assignments see Supplementary Table S1.

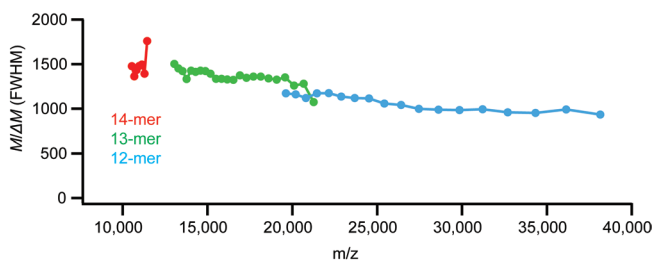


Figure S5. Effective resolution as a function of m/z for intact GroEL 14-mer and 13-/12-mer product ions on the modified Exactive Plus with lowered RF frequency.

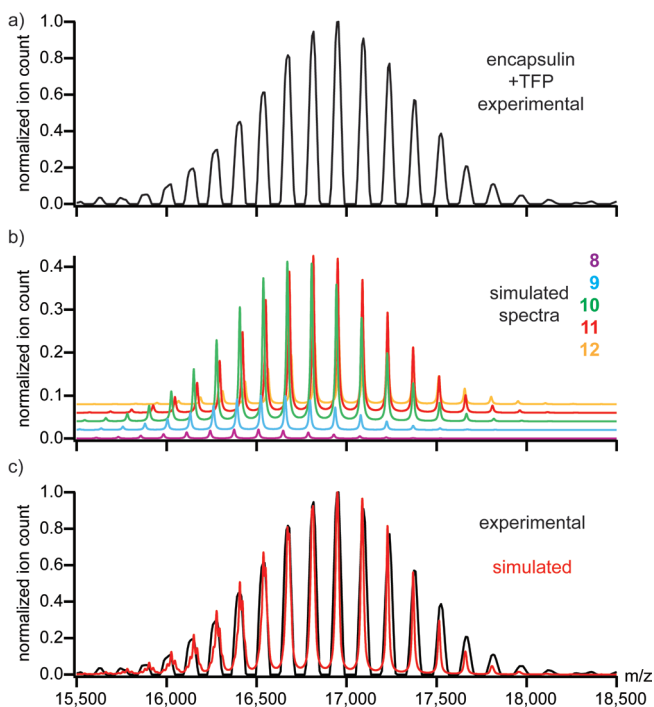


Figure S6. Validating quantitation of cargo encapsulation in encapsulin VLP. a) Experimental spectrum of intact encapsulin VLP. b) Simulated spectra of the identified encapsulin-cargo complexes. Note that each consecutive stoichiometry is offset in y . c) Overlay of the experimental spectrum with the sum of spectra in “b”).

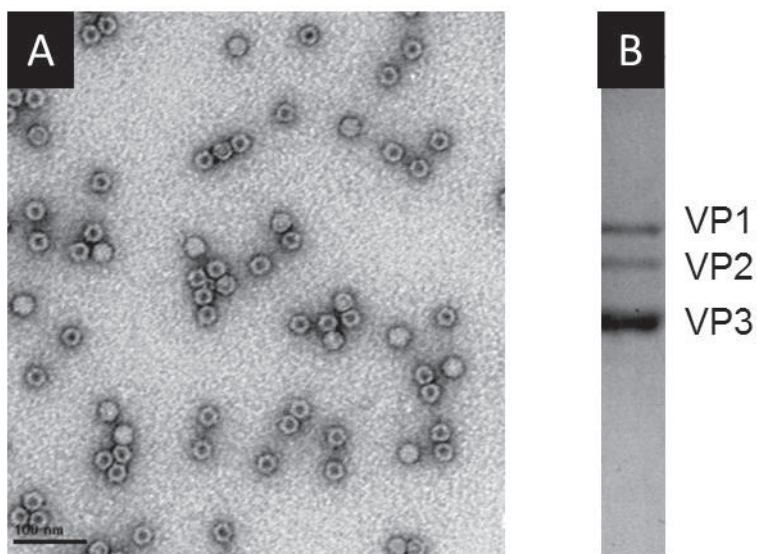


Figure S7. AAV1 capsid gross morphology and VP composition. A) 5 μ l of AAV1 capsids were stained with 2% uranyl acetate on holey carbon grids and imaged on a FEI spirit TEM. B) Coomassie stained SDS PAGE of denatured AAV1 capsids showing VP1,VP2 and VP3.

Supplementary Information to Chapter 4.

Mass spectrometry uncovers a novel component of mature adenovirus capsids.

Joost Snijder, Marco Benevento, Crystal L. Moyer, Vijay Reddy, Glen R.
Nemerow, Albert J.R. Heck.

Supplementary tables.

Table S1. Peptide identifications of pVIn from LC-MS/MS analysis of mature HAdV virions. The confidence of the identification is given by the IonScore. The higher the score, the more confident the identification. Scores of >20 are considered confident.

start	end	sequence	modifications	score
1	12	mEDINFASLAPR	N-Term(Acetyl)	105
1	12	mEDINFASLAPR	N-Term(Acetyl); M1(Oxidation)	97
13	33	HGSRPFmGNWQDIGTSNmSGG	M7(Oxidation); M18(Oxidation)	55
13	33	HGSRPFmGNWQDIGTSNMSGG	M7(Oxidation)	62
13	33	HGSRPFmGNWQDIGTSNMSGG		46

Table S2. PSM based quantitation of pVIn release upon heating. The fraction of released pVIn is calculated as the average \pm standard deviation from triplicate experiments, where for each experiment the number of PSM's, normalized to the PSM's for pVII, is divided by the average pVII-normalized number of pVIn PSM's in untreated virus.

condition	protein	number of PSMs		
		average	standard dev.	n
untreated	pVII	82.3	5.0	6
	pVIn	23.5	2.7	
heated	pVII	104	1.7	3
	pVIn	5.7	0.6	
		fraction released		
		average	standard dev.	
		0.81	0.02	

Table S3. Hexon-pVIn masses obtained from native MS analysis.

	mass (kDa)		
	average	standard deviation	error % (compared to theoretical mass)
hexon ₃	323.62	0.01	0.024
hexon ₃ pVIn ₁	327.24	0.02	0.021
hexon ₃ pVIn ₂	330.87	0.01	0.022
hexon ₃ pVIn ₃	334.49	0.01	0.022
pVIn	3.624	0.00	0.002

Supplementary Figures.

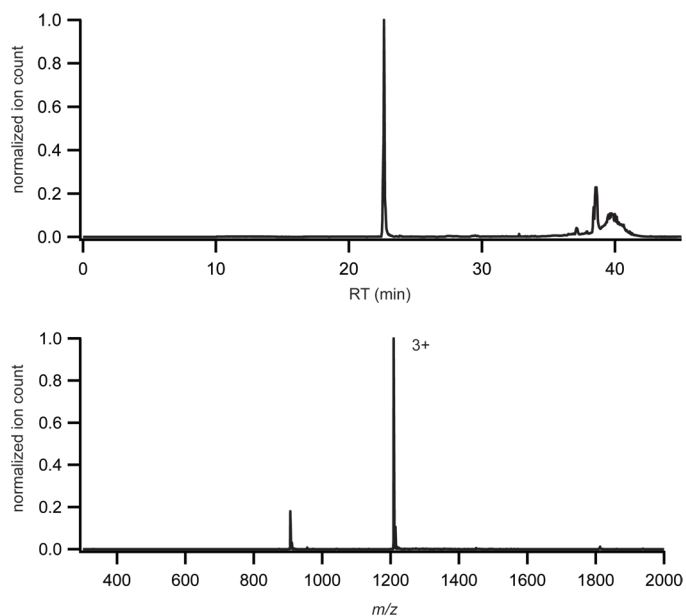


Figure S1. LC-MS analysis of isolated peptide from heat-released hexon trimers. A single peptide is recovered from the isolation procedure as evidenced from the base-peak chromatogram (top). The isolated peptide has a mass of 3624.09 Da, compared to a theoretical mass of 3624.01 for pVIn.

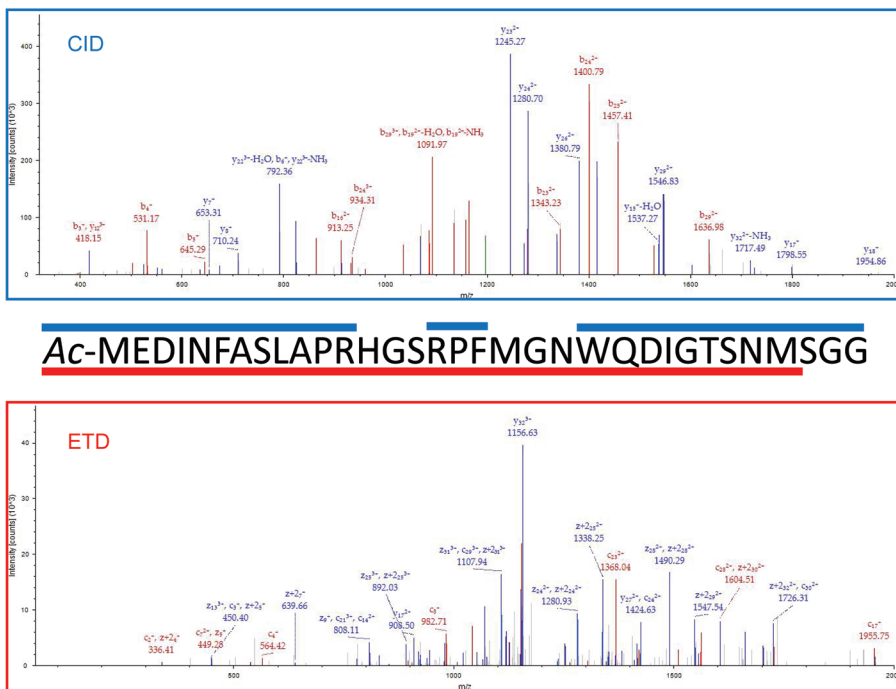


Figure S2. MS/MS analysis of the peptide isolated from the hexon complex. A combination of CID and ETD fragmentation covers the full sequence of the peptide and confirms that the peptide is pVIn.

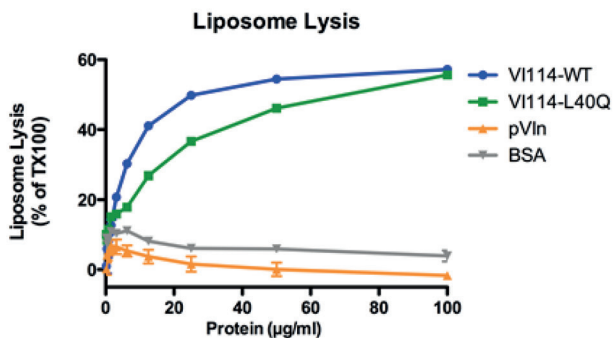


Figure S3. pVIn has no membrane lytic activity. Fluorescence measurements of SulfoB release from liposomes. Lytic activity is expressed as % of total fluorescence for Triton-X100 disrupted liposomes. Points represent the average \pm SEM from triplicate measurements.

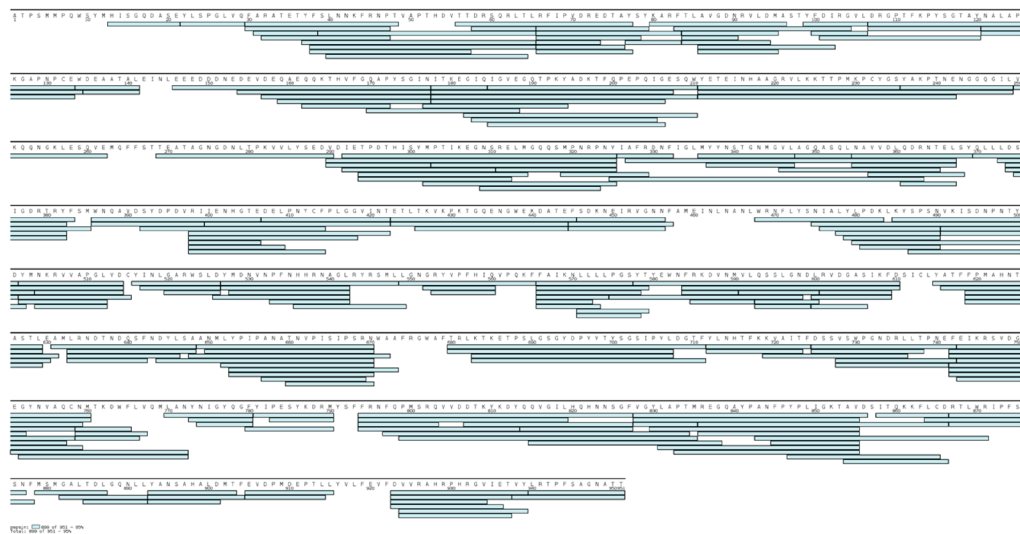
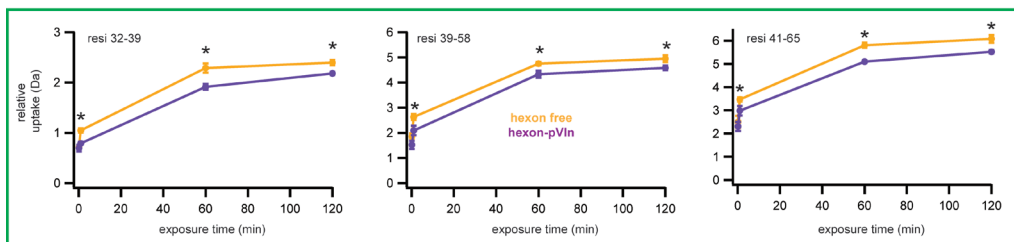


Figure S4. Sequence coverage of hexon from peptide ID's of the HDX-MS experiment.

affected by pVIn binding



selected representative peptides, not affected by pVIn binding

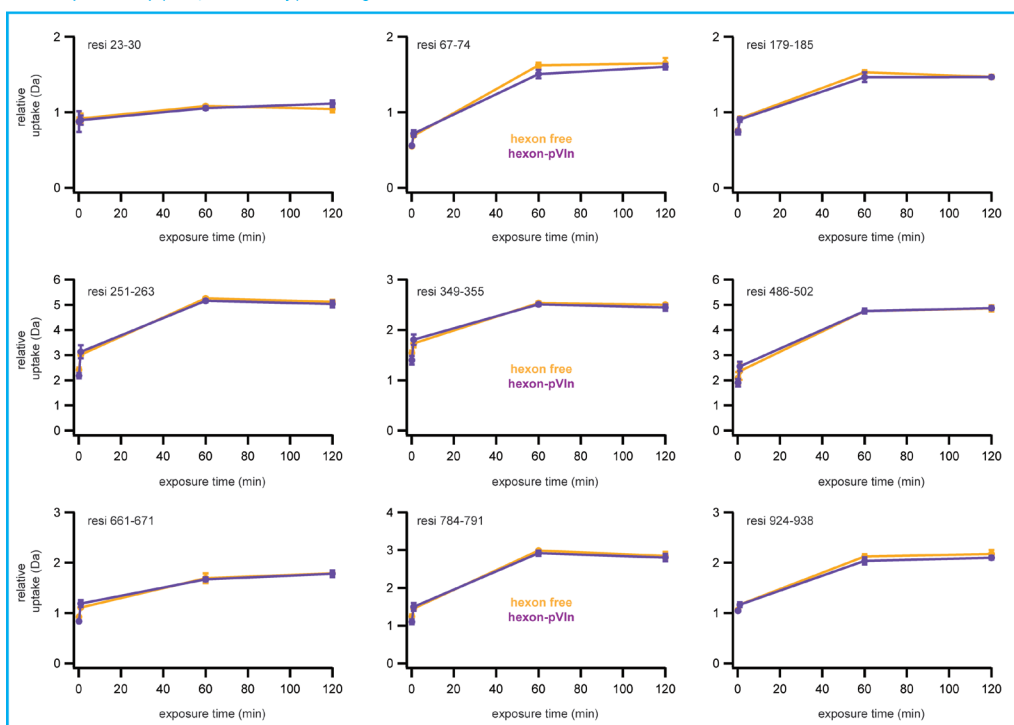


Figure S5. Comparison of peptides that exhibit protection for deuterium uptake upon pVIn binding (green box) with selected representative peptides that were unaffected by pVIn binding (blue box). Asterisks indicate that $p < 0.05$ in unpaired, two-tailed Student's t-tests.

Supplementary Information to Chapter 5.

Probing the biophysical interplay between a viral genome and its capsid.

J. Snijder, C. Uetrecht, R.J. Burnley, R. Sanchez-Eugenia, G.A. Marti, J. Agirre,
D.M.A. Guérin, G.J.L. Wuite, A.J.R. Heck, W.H. Roos.

Supplementary methods.

Autofluorescence.

We performed tryptophan fluorescence measurements by exciting at 295 nm and recording between 305 and 420 nm in a FluoroMax fluorimeter (Horiba Jobin-Yvon). TrV samples were at 0.1 mg/mL at the indicated pH.

Native-PAGE.

Native polyacrylamide gels prepared at 5% were run with no sample loaded for 30 minutes at 140 V and 4°C in order to equilibrate them with the running buffer (2.5 mM Tris-HCl pH 8.2 and 19 mM glycine). 20 μ L of TrV samples at 0.1 mg/mL were mixed with 5 μ L of native loading buffer (0.25 M Tris-HCl pH 6.8, 50% glycerol, 60 mM EDTA and 0.0125% (w/v) bromophenol blue). Loaded samples and molecular weight markers (High Molecular Weight Kit, GE Healthcare) were run for 4 hours at 100 V and 4°C. Finally gels were stained with staining solution (50% methanol, 10% acetic acid and 0.25% w/v R-250 Coomassie Blue) for 20 minutes and unstained overnight with 10% acetic acid solution.

Computed electrostatic binding energies.

Adaptive Poisson-Boltzmann Solver (APBS) software was used to calculate the binding energy between VP1/2/3 protomers within and between pentons. The binding free energy between protomer A and the rest of the complex B (either the other 4 penton protomers or the adjacent penton) was calculated taking into account the unfavorable process of desolvation of both proteins and the likely favorable process of electrostatic interaction between them: $\Delta G_{\text{binding}} = \Delta G_{\text{desolA}} + \Delta G_{\text{desolB}} + E_{\text{elec.AB}}$. The desolvation energies of protomers correspond to the loss of electrostatic interaction energy between the solvent and the protomer upon binding, and was calculated for each protomer by first calculating the electrostatic energy of the complex A-B with B protomer charges set to 0 (E1) and calculating the electrostatic energy of A (E2). Then, the desolvation energy of protomer A was obtained by: $\Delta G_{\text{desolA}} = E1 - E2$. Finally, the electrostatic energy was calculated as the electrostatic potential generated by protomer A at the position of atomic charges of protomer B. APBS input settings were: i) Linearized Poisson-Boltzmann equation was solved. ii) The number of grid points was 321 x 353 x 417, the length of the coarse grid was 214.977 x 262.910 x 312.543 and of the fine grid 146.457 x 174.653 x 203.849. iii) Protein dielectric constant was set to 2 and the solvent dielectric constant was set to 78.54. iv) Dissolvated ions were taken into account as follows: ion charge +1 at a concentration of 0.150 M with a radius of 2.0 Å and ion charge -1 at a concentration of 0.150 M with the same radius.

Genome Packing Density.

The packing density of the genomes in TrV and phage λ were calculated by dividing the total number of nucleotides by the inner volume of the capsid. Capsids were approximated as spheres to calculate the inner volumes. TrV was taken to have an inner radius of 11.7 nm, giving an inner volume of 6709 nm^3 . The radius of phage λ was taken as 28.2 nm, giving an inner volume of 93937 nm^3 . The total number of nucleotides was taken as 9010 for TrV and as 2×48502 for phage λ .

Supplementary tables.

Table S1. Experimental vs. sequence derived masses of TrV. The stated errors are the standard deviations of masses derived from different charge states within one spectrum, except in the case of “r-empty”, which is the standard deviation over five replicate analyses.

complex	experimental mass		theoretical mass	
	average	stdev	sequence ^a	exp ^b
penton	444.8	0.3	446.1	446.5
decamer	888.4	0.3	892.1	893.1
r-empty capsid	5362	2.5	5353	5359
virion	No resolved charge states		8252	8258
VP1	29.774	0.004	29.719	
VP2	28.405	0.0009	28.401	
VP3	31.129	0.0009	31.093	
VP4	5.504	0.0002	5.503	

^a based on sequence, verified by genomic sequencing and LC-MS/MS

^b based on experimental VP monomer masses

Table S2. Cross sections (Ω) of TrV disassembly products measured by IMS. Stated errors are standard deviations over triplicate experiments.

complex	origin		experimental Ω (nm ²) ^a		theoretical Ω (nm ²) ^a	
			average	standard deviation	from PDB code 3NAP ^b	as globular ^c
penton	pH		169	3	158	138
	urea	low z	165	1		
		high z	170	0.2		
			183	9		
			203	16		
penton dimer	urea		250	3	282	225

^aAverage over all charge states and triplicate analysis

^bFrom projection approximation (PA) calculations, note that residues are missing from the pdb coordinates, resulting in underestimation of the cross section.

^cThe expected cross section of a globular protein as estimated with the experimentally obtained fit of globular proteins presented in the main text.

Table S3. Particle diameters as determined by AFM. Heights were corrected by the estimated indentation during imaging, typically well below 1 nm. Stated errors are Standard Error of the Mean (SEM).

particle type	pH	corrected particle height (nm)		
		avg	SEM	n
n-empty	6.8	34.0	0.2	32
	8	33.7	0.2	36
	9	33.6	0.3	27
r-empty	6.8	34.3	0.2	14
	9	34.3	0.3	29
virion	6.8	34.1	0.2	25
	8	34.0	0.2	8

Table S4. Penton dimensions as determined from AFM imaging. The ‘ppd’ denotes the peak-to-peak distance: the lateral separation between adjacent five-fold related subunits as can be distinguished in the AFM image. Stated errors are Standard Error of the Mean (SEM). The height of the penton in the crystal structure of TrV (PDB code: 3NAP) is approximately 7 nm and adjacent VP1 protrusions are separated by approximately 6 nm.

type	height (nm)			ppd (nm)		
	avg	SEM	n	avg	SEM	n
force-induced	6.7	0.03	143	6.3	0.1	69
pH-induced	6.6	0.05	58	6.5	0.08	77

Table S5. Mechanical properties of TrV particles determined by AFM nanoindentation. For virion particles, the indicated subclassification into intact virion, uncoating intermediate and r-empty capsid was based on the density plot presented in the main text. ‘r-empty’ was defined as all particles with $F_{break} < 2$ nN, ‘uncoating intermediate’ as $F_{break} > 2$ nN and $k < 1$ N/m, all remaining particles were classified as ‘intact virion’. Stated errors are Standard Error of the Mean (SEM).

particle type		pH	k (N/m)			F_{break} (nN)		
			avg	SEM	n	avg	SEM	n
virion	intact virion	7	1.47	0.06	34	3.62	0.15	34
		8	1.34	0.14	6	3.81	0.17	6
		8->7	1.47	0.06	19	3.80	0.08	19
		combined	1.46	0.04	59	3.69	0.09	59
	uncoating intermediate	7	0.63	0.08	12	3.15	0.26	12
		8	0.49	0.07	12	3.24	0.16	12
		8->7	0.67	0.08	7	3.04	0.19	7
		combined	0.58	0.05	31	3.16	0.12	31
	r-empty	7	0.50	0.09	11	0.86	0.09	16
		8	0.42	0.12	14	0.77	0.11	14
		9	0.37	0.07	27	0.92	0.06	29
		9->7	0.46	0.12	8	1.35	0.25	8
		8->7	0.46	0.05	14	0.87	0.07	14
		combined	0.43	0.04	74	0.92	0.05	81
	n-empty	7	0.52	0.07	27	0.81	0.04	32
8		0.52	0.05	31	0.91	0.06	35	
9		0.37	0.05	26	0.90	0.06	26	
combined		0.47	0.03	84	0.87	0.03	93	

Table S6. Particle diameters as determined by Ion Mobility Spectrometry. There is no significant difference (> 2 Standard Deviation, SD) between empty capsids and the virion.

particle type	diameter (nm)	
	average	standard deviation
n-empty	33.3	1.5
r-empty	34	2
virion	36.5	0.4

Supplementary Figures.

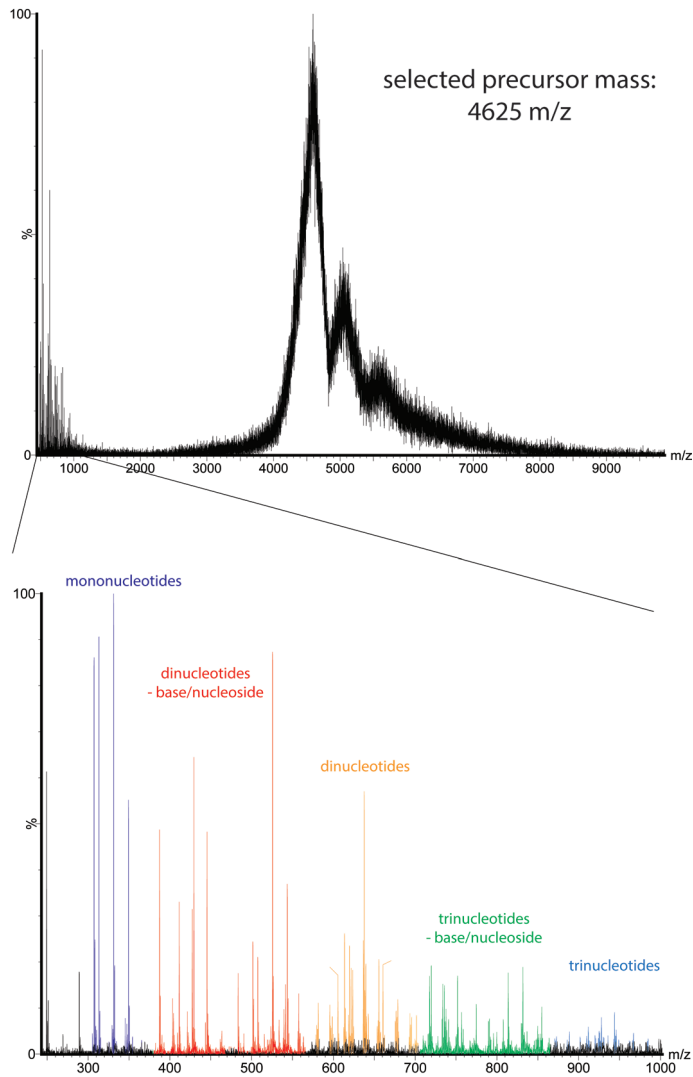


Figure S1. Tandem MS of unresolved baseline signal in the virion at pH >8. Top: tandem MS of 4625 m/z precursor ions, showing both extensive fragmentation and neutral loss. Bottom: zoom-in of ribonucleotide fragments of the 4625 m/z precursor ions.

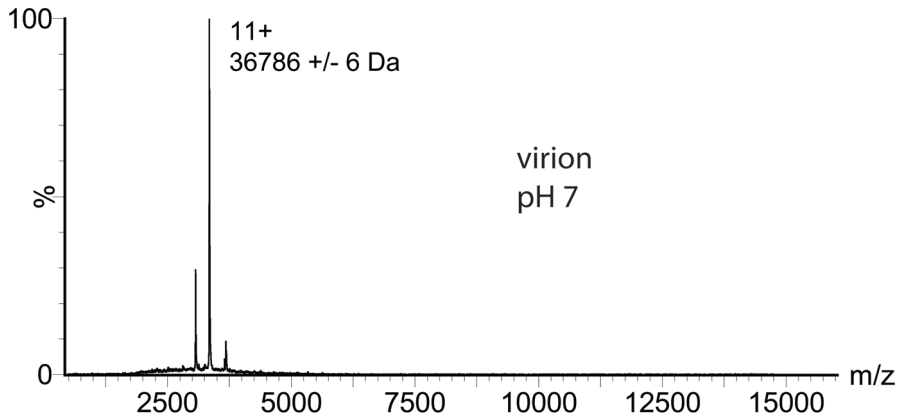


Figure S2. Low m/z region of the virion at neutral pH. There is no signal for ssRNA, VP4 and pentons. There is a distinguishable charge state distribution for this unknown 36 kDa component referred to in the main text, showing that it is unrelated to alkaline-triggered uncoating.

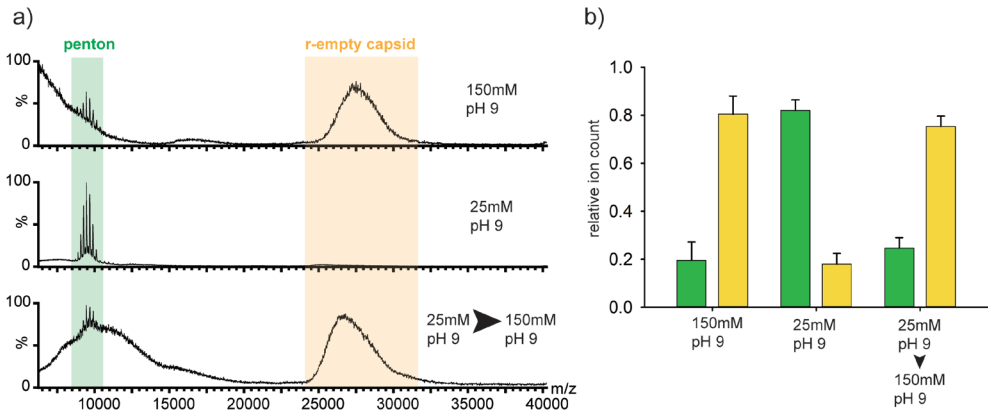


Figure S3. Pentons generated from the virion under alkaline conditions are assembly competent. a) Mass spectra of virion samples incubated for one day at 150 mM pH 9 or 25 mM pH 9 and after one day 25mM, one day 150 mM pH 9. Both r-empty capsids and penton are observed at 150mM ionic strength. More penton is formed under low ionic strength conditions. When the 25 mM pH 9 sample is subsequently transferred to 150 mM and incubated for one additional day, more r-empty capsid is formed. Hence, pentons can reassemble into empty capsids. b) Average relative ion count for penton (green) and capsid (yellow) from duplicate experiments. Error bars are SD.

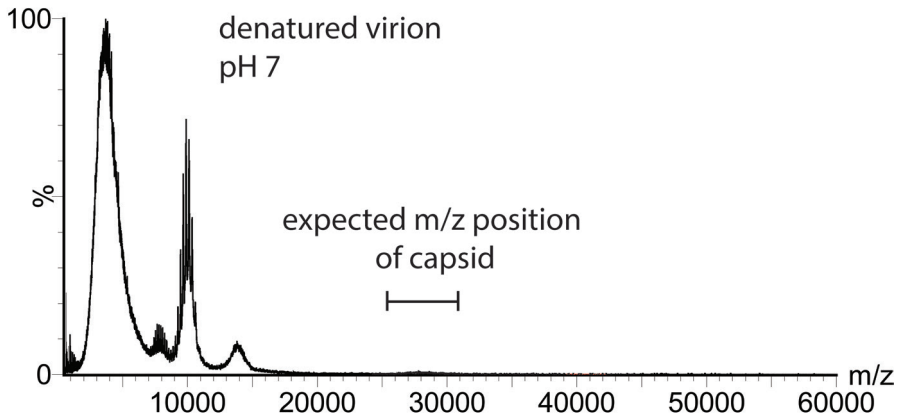


Figure S4. Overview scan of a urea denatured virion sample. There is no substantial signal for the empty capsid, showing that reassembly does not occur with denatured pentons/decamers.

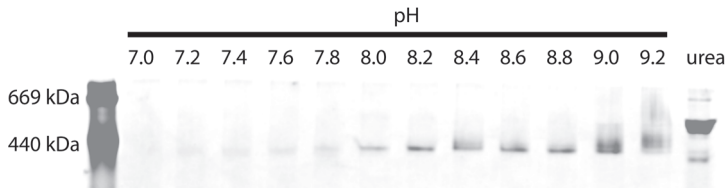


Figure S5. Native gel of virion at increasing pH and after urea denaturation. There is an increasing amount of ~440 kDa product on gel with increasing pH. This is consistent with the appearance of pentons, confirming that alkaline-triggered uncoating occurs via disassembly of the capsid.

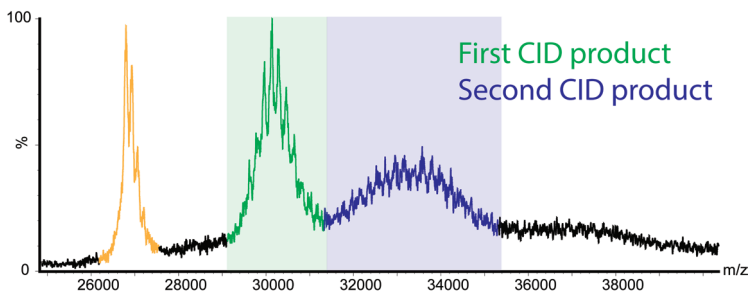


Figure S6. Tandem MS of r-empty capsids confirms the mass assignment, thus confirming release of ssRNA and VP4. Ions of the r-empty capsid (yellow) are selected in the instrument and subjected to high collision voltage (260 V with xenon as collision gas). This results in sequential dissociation of VP subunits. The large product ions indicated in green and blue correspond to the loss of one and two VP subunits, respectively. Their respective masses are calculated at 5332 ± 2 and 5306 ± 1 kDa, corresponding to the loss of first 29.5 and then another 26.5 kDa (stated errors are SD). This matches the expected masses of VP2/3 to within the error margin, confirming that the precursor is $60 \times \text{VP}_{1/2/3}$ without any discernible inclusion of ssRNA or VP4.

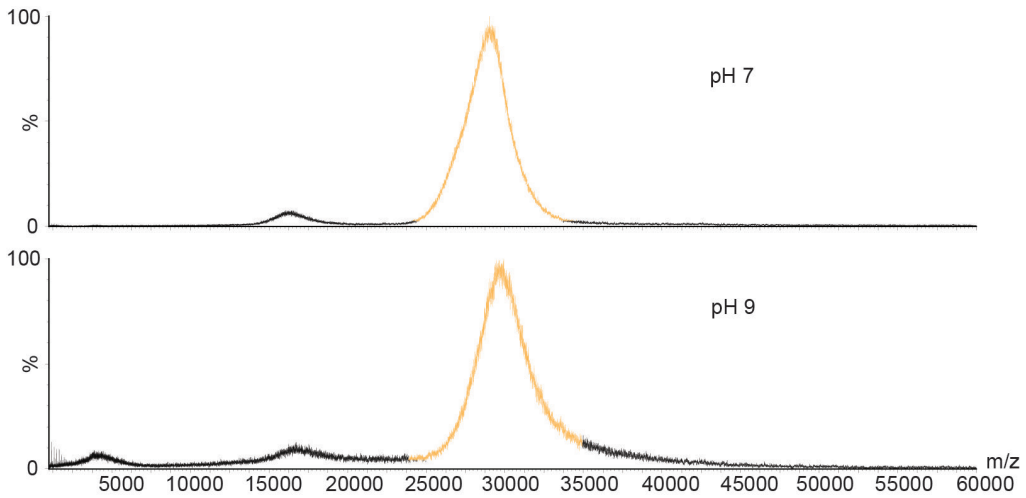


Figure S7. Native MS of n-empty capsids. There are no distinguishable disassembly products at alkaline pH and there is no notable shift in peak-position of the n-empty capsid under alkaline pH. This shows that there is no conformational change to the effect of a larger chargeable surface area under alkaline conditions. Also, it supports the data demonstrating that virion disassembly at alkaline pH is triggered by the ssRNA.

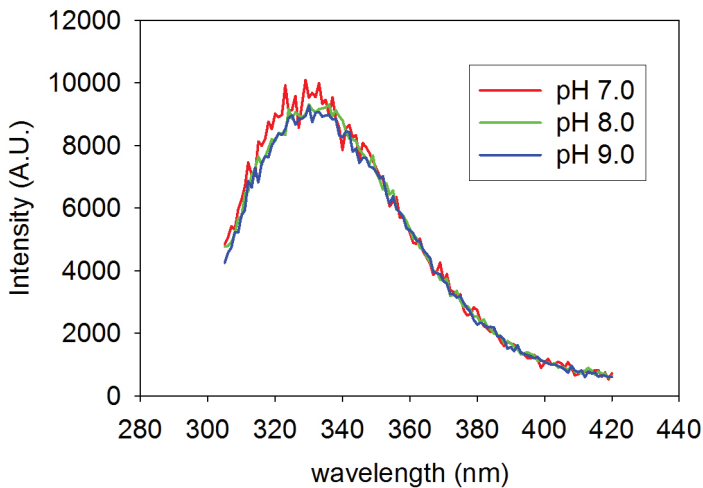


Figure S8. Auto-fluorescence of TrV virions at increasing pH. There is only a marginal pH-induced change in the auto-fluorescence spectra of TrV that likely relates to deprotonation of tyrosine residues. Therefore, there are no major conformational changes in the vicinity of aromatic residues. Note that TrV contains in total five tryptophan residues, none of which is exposed in either the full capsid or free penton (PDB code: 3NAP).

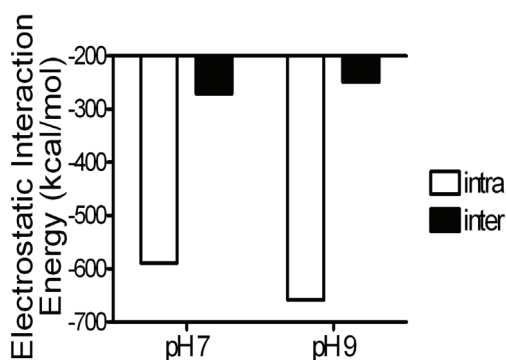


Figure S9. APBS calculations of the electrostatic interaction potential of VP1/2/3 protomers within (intra) and between (inter) pentons (see methods). The interaction is stronger within pentons than between pentons. There is no substantial difference between pH 7 and pH 9. This explains the occurrence of pentons as disassembly products and supports the hypothesis that the ssRNA genome triggers alkaline-induced disassembly in the virion.

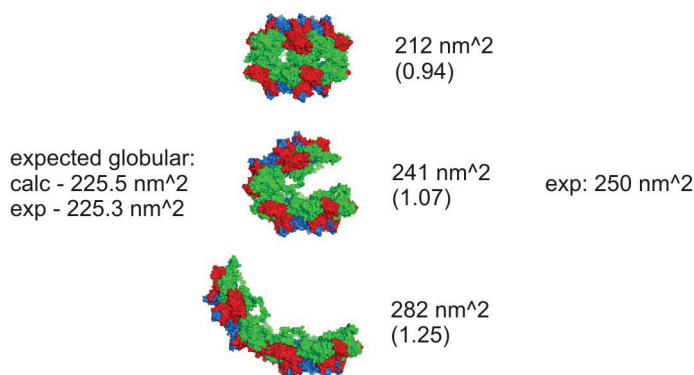


Figure S10. Assembly incompetent penton-dimers hinge on the penton-penton interface. Collision cross section calculations of the penton-dimers. ‘Expected globular’ refers to the expected cross section when entering the molecular weight of the assembly into the obtained fit of all globular assemblies. ‘calc’ denotes the expected cross section from the observed trend of MobCal PA calculations of globular proteins, ‘exp’ denotes the expected cross section from the observed experimental trend. The structures represent a penton-dimers model with an increasing angle between pentons. The bottom structure represents the angle as observed in the complete capsid of TrV (PDB code: 3NAP). The values correspond to the PA cross section of that model, the values in parentheses are the relative cross section compared to ‘expected globular’. These calculations suggest that the pentons hinge on their interface, thus explaining the relatively low cross section of the penton-dimers compared to the pentons.



Figure S11. AFM imaging of alkaline-triggered disassembly products. This shows that pentons are present in the virion sample at pH 9. Left image is 100x100 nm in xy and approximately 10 nm in z.

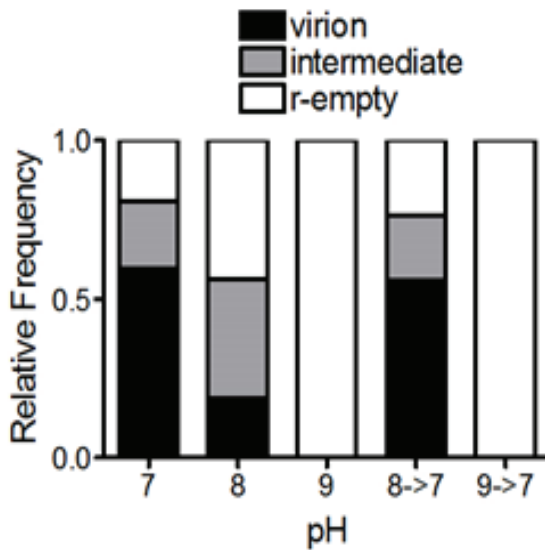


Figure S12. AFM data with the relative frequencies of virion, pre-disassemble state and r-empty capsid and reversibility of the transitions. There is a 2-fold increase of the uncoating intermediate (pre-disassemble state) at pH 8, and the transition is reversible. Reversing the pH from 9 to 7 does not result in reassembly of the virion (encapsidation of the ssRNA). Note that the increased frequency of the r-empty capsid at pH 8 is related to the time-scale. Samples at pH 8 are stored for several days over the course of the analysis, in the experiment of pH 8 -> 7, the sample is incubated overnight before reversing the pH back to 7.

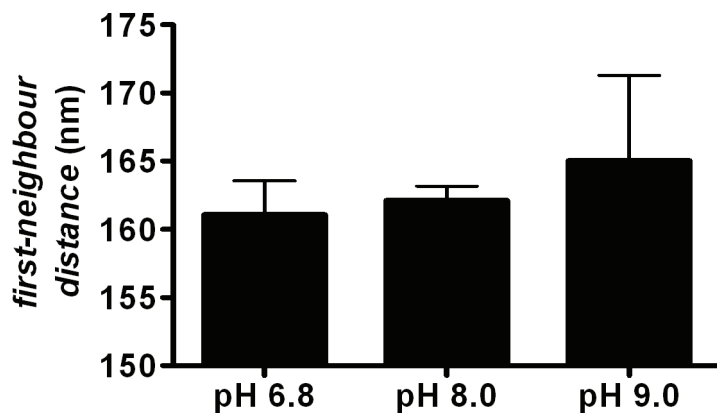


Figure S13. Particle density in AFM imaging as a function of pH. The particle density was determined from several scans of $2 \times 2 \mu\text{m}$ as the average first-neighbour distance. The particle density is constant over the tested pH range, confirming that the observed reassembly of virions into empty capsids does not arise from a sampling error.

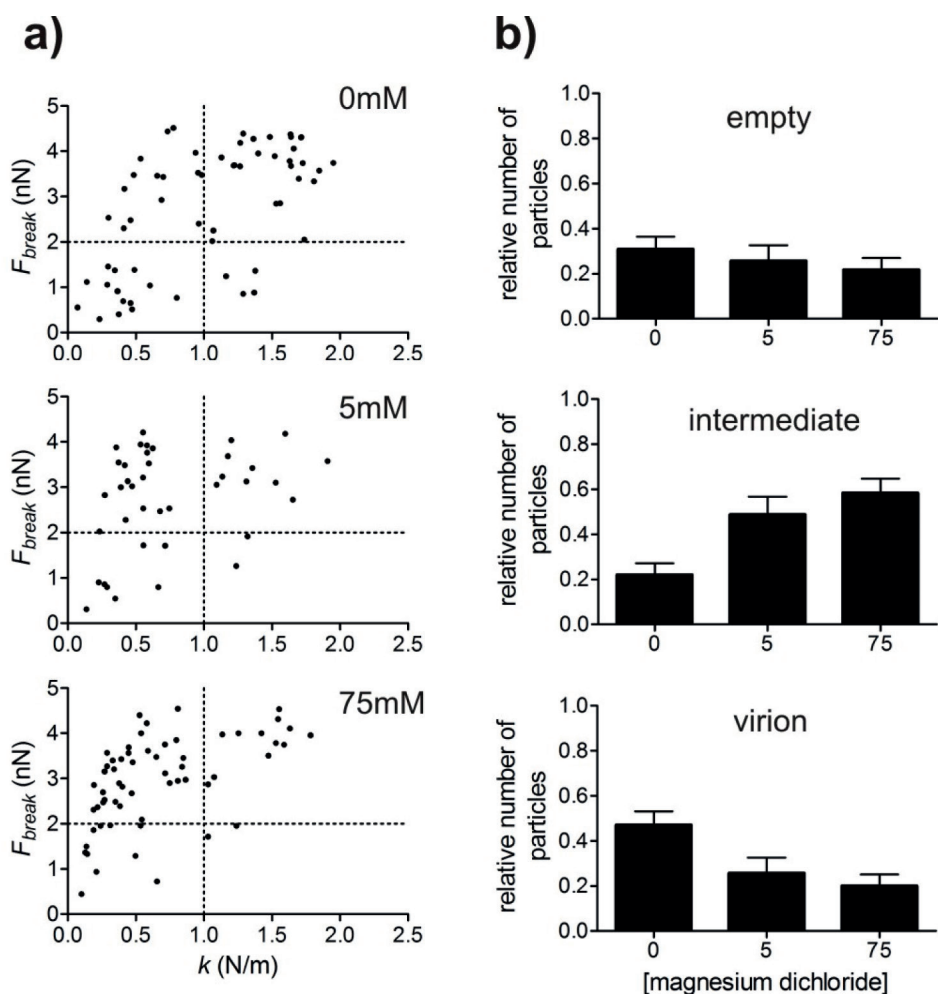


Figure S14. AFM nanoindentation of TrV in the presence/absence of magnesium dichloride at pH 6.8. a) 2D-plots of spring constant vs. breaking force of TrV virion samples without magnesium dichloride and with 5 or 75 mM magnesium dichloride. The boundaries used to classify different types of particles are indicated by dashed line and are the same as in the alkaline-triggered uncoating experiment: empty ($F_{break} < 2$ nN), ‘intermediate’ ($F_{break} > 2$ nN and $k < 1$ N/m) and virion ($F_{break} > 2$ nN and $k > 1$ N/m). b) Relative number of particles per particle-type, as a function of magnesium dichloride concentration. The error bars represent the relative SD as estimated assuming binomial distributions for each particle type (e.g. either ‘empty’ or ‘not empty’): the errors were calculated as $((Npq)^{0.5})/N$, where N is the total number of particles, p is the relative frequency of a particular particle-type and $q = 1 - p$. Adding magnesium dichloride results in a shift to lower spring constant for ssRNA-filled particles, as apparent from the increased frequency of ‘intermediates’. We also performed a control experiment with r-empty capsids to test the effect of 75 mM magnesium dichloride on the mechanical response of the capsid alone at pH 6.8 and found no significant difference in spring constant or breaking force compared to the r-empty capsid without magnesium dichloride ($k = 0.48 \pm 0.06$ N/m, $F_{break} = 0.67 \pm 0.03$ nN, $n = 16$). This shows that the magnesium dichloride does not affect the protein-protein interactions, instead it must act on the ssRNA.

Supplementary Information to Chapter 6.

Cargo encapsulation triggers a stability switch in the bacterial nanocompartment encapsulin.

J. Snijder, O. Kononova, I.M. Barbu, C. Uetrecht, W.F. Rurup, R.J. Burnley, M.S.T. Koay, J.J.L.M. Cornellissen, W.H. Roos, V. Barsegov, G.J.L. Wuite, A.J.R. Heck.

Supplementary methods.

Self-Organized Polymer (SOP) model.

In the topology-based SOP model, each residue is described by a single interaction center (C_α -atom). The potential energy function of the protein conformation,

$$U_{SOP} = U_{FENE} + U_{NB}^{ATT} + U_{NB}^{REP},$$

is a function of the coordinates r_1, r_2, \dots, r_N , where N is the total number of residues. The finite extensible nonlinear elastic (FENE) potential,

$$U_{FENE} = -\sum_{i=1}^{N-1} \frac{k}{2} R_0^2 \log \left(1 - \frac{(r_{i,i+1} - r_{i,i+1}^0)^2}{R_0^2} \right),$$

with the spring constant $k = 14$ N/m and the tolerance in the change of the covalent bond distance $R_0 = 2$ Å describes the backbone chain connectivity. The distance between residues i and $i+1$, is $r_{i,i+1}$, and $r_{i,i+1}^0$ is its value in the native structure. The Lennard-Jones potential,

$$U_{NB}^{ATT} = \sum_{i=1}^{N-3} \sum_{j=i+3}^N \varepsilon_n \left[\left(\frac{r_{ij}^0}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{ij}^0}{r_{ij}} \right)^6 \right] \Delta_{ij},$$

accounts for the non-bonded interactions that stabilize the native state (summation is performed over all the native contacts). If the residues i and j ($|i-j| > 2$) are within the cut-off distance $R_C = 8.5$ Å in the native state, then $\Delta_{ij} = 1$, and zero otherwise. The value of ε_n quantifies the average strength of the non-bonded interactions. We used $\varepsilon_n = \varepsilon_{inter} = 1.3$ kcal/mol and $\varepsilon_n = \varepsilon_{intra} = 1.1$ kcal/mol for the inter-chain contacts and intra-chain contacts, respectively (see next section). The non-native interactions are treated as repulsive in the potential

$$U_{NB}^{REP} = \sum_{i=1}^{N-2} \varepsilon_l \left(\frac{\sigma_l}{r_{i,i+1}} \right)^6 + \sum_{i=1}^{N-3} \sum_{j=i+3}^N \varepsilon_l \left(\frac{r_{ij}^0}{r_{ij}} \right)^6 (1 - \Delta_{ij})$$

(summation is over all the contacts between residues separated by the distance $> R_C$. To describe self-avoidance, a constraint is imposed on the bond angle formed

by residues i , $i+1$, and $i+2$ by including the repulsive potential with parameters ϵ_r , $= 1$ kcal/mol and $\sigma = 3.8$ Å, which determine, respectively, the strength and the range of the repulsion.

Parameterization of SOP-model for encapsulin shells.

The values of parameter ϵ_h of the SOP model, which contains the atomic-level details (number of binary contacts between amino acids and their energies), were calculated directly using the MD simulations of an atomic structure model of encapsulin shell at $T = 300$ K. Once determined, this parameter was ported to the SOP model of the encapsulin shell. All the native contacts were divided into two major groups which define the contact types: (1) the intra-chain contacts within each encapsulin subunit and (2) the inter-chain contacts at the interfaces formed by encapsulin proteins. We assumed that a pair of residues forms a contact if the distance between their C_α -atoms in the native state does not exceed the cut-off distance R_c . We used a standard choice of the cut-off distance $R_c = 8$ Å. The 30 ns MD simulation run was performed to calculate for each group of contacts (i) the average non-bonded energy (E_{intra} and E_{inter}), given by the sum of the van-der-Waals energy (Lennard-Jones potential) and the electrostatic energy (Coulomb potential), and (ii) the average number of binary contacts between amino acids (N_{intra} and N_{inter}) that stabilize the native encapsulin structure (native contacts). To carry out the all-atom MD simulations, we employed the Solvent Accessible Surface Area (SASA) model of implicit solvation (1), which is based on the CHARMM19 force-field (2). We used the output from SASA model based simulations (coordinate files and energy files) to calculate the values of E_{nb} and N_{nb} for the two contact groups. On average, there were $N_{intra} \approx 1,050$ intra-chain contacts and $N_{inter} \approx 90$ inter-chain contacts per protein monomer, and $E_{intra} \approx 1,350$ kcal/mol for the intra-chain contacts per protein monomer and $E_{inter} \approx 95$ kcal/mol for the inter-chain contacts between two monomers. Next, we divided the two numbers for each contact group to obtain $\epsilon_{intra} = E_{intra}/N_{intra} = 1.28$ kcal/mol and $\epsilon_{inter} = E_{inter}/N_{inter} = 1.05$ kcal/mol, which were then used in the SOP model based Langevin simulations of the mechanical indentation of encapsulin. We used a similar approach to perform the Langevin simulations of nanoindentation of CCMV shell (3).

Langevin simulations of encapsulin nanoindentation.

The indentation dynamics were obtained by integrating the Langevin equations for each particle position r_i in the overdamped limit, $\eta dr_i/dt = -\partial U_{tot}/\partial r_i + g_i(t)$. Here, $U_{tot} = U_{SOP} + U_{tip}$ is the total potential energy, which accounts for the contribution from the capsid conformation U_{SOP} and for the interaction of the i -th particle with the spherical tip $U_{tip}(r_i)$ (see Materials and Methods in the main

text). Also, $g_i(t)$ is the Gaussian distributed random force and η is the friction coefficient. The Langevin equations were propagated with the time step $\Delta t = 0.08\tau_H = 20$ ps, where $\tau_H = \zeta\varepsilon_h\tau_L/k_B T$. Here, $\tau_L = (ma^2/\varepsilon_h)^{1/2} = 3$ ps, $\zeta = 50.0$ is the dimensionless friction constant for a residue in water ($\eta = \zeta m/\tau_L$), and $m \approx 3 \times 10^{-22}$ g is the residue mass (4,5). Simulations of nanoindentation of encapsulin particles were carried out at (constant) room temperature using the bulk water viscosity, which corresponds to the friction coefficient $\eta = 7.0 \times 10^5$ pN ps/nm. In the course of simulations, the water viscosity and friction coefficient were kept constant, which corresponds to the condition of constant pressure.

Calculation of thermodynamic quantities.

We analyzed the energy output from simulations (potential energy U_{SOP}) to estimate the enthalpy change associated with particle deformation ΔH . The total work of deformation w can be obtained by converting the FZ curve into the FX curve and by integrating the area under the FX curve for the forward indentation. This procedure can then be repeated but for the retraction curve to resolve the reversible part of work $w_{rev} = \Delta G = \Delta H - T\Delta S$ and the entropic contribution $T\Delta S = \Delta H - w_{rev}$. The amount of dissipated energy and w_{rev} was estimated using the Crooks theorem (see next section).

Calculation of reversible work w_{rev} .

Crooks showed theoretically (6) and Bustamante and coworkers verified experimentally (7) that in a driven unfolding process from the initial point $X = X_{in}$ to some final point $X = X_{fin}$ and a reverse refolding process from $X = X_{fin}$ back to $X = X_{in}$, the probability distributions of the values of work for the forward process $p_{unf}(w)$ and reverse process $p_{ref}(-w)$ are connected via the relationship:

$$p_{unf}(w) / p_{ref}(-w) = \exp[-(w - \Delta G_{eq}) / k_B T]$$

The point at which $p_{unf}(w)$ and $p_{ref}(-w)$ intersect corresponds to the equilibrium work $w = \Delta G_{eq}$. Here, we adopted this formalism to describe the process of forced indentation followed by force-quenched retraction. In our simulations, we are limited to a few (3-5) trajectories for each indentation point. We can still use the Crooks relationship if we notice that $p_{unf}(w)$ and $p_{ref}(-w)$ intersect at their tails, which correspond to the extreme observations: the minimum work ($w_{min,forw}$) for the forward process and the maximum work ($w_{max,rev}$) for the reverse process ($w_{max,ref} < w_{min,forw}$). Hence, the reversible work is between the minimum work for the forward process and the maximum work for the reverse process, i.e. $w_{max,rev} < w < w_{min,forw}$. In the context of our indentation experiments *in vitro* or *in silico*, this corresponds to the minimum work $w_{min,ind} = \min\{w_{1,ind}, w_{2,ind}, \dots\}$ for the forward forced indentation ($p_{ind}(w)$) and the maximum work $w_{max,ret} = \max\{w_{1,ret}, w_{2,ret}, \dots\}$ for the reverse force-quenched retraction ($p_{ret}(-w)$).

$\{w_{1,ret}, w_{2,ret}, \dots\}$ for the backward force-quenched retraction ($p_{ret}(w)$), i.e. $w_{max,ret} < w < w_{min,ind}$. We estimated the reversible part of work $w_{rev} = \Delta G_{eq}$ by taking the arithmetic mean of $w_{max,ret}$ and $w_{min,ind}$, i.e. $w \approx (w_{min,ind} + w_{max,ret})/2$. We successfully used this approach to resolve the thermodynamic characteristics of the tubulin-tubulin interactions in from *in silico* nanoindentation measurements on microtubule polymers (8).

Supplementary references.

1. Ferrara, P., Apostolakis, J. & Caflisch, A. Evaluation of a fast implicit solvent model for molecular dynamics simulations. *Proteins: Structure, Function and Genetics* 46, 24-33 (2002).
2. MacKerell, A.D. et al. All-Atom empirical potential for molecular modeling and dynamics studies of proteins *J. Phys. Chem. B* 102, 3586–3616 (1998).
3. Kononova O., et al. Structural Transitions and Energy Landscape for Cowpea Chlorotic Mottle Virus Capsid mechanics from nanomanipulation in vitro and in silico. *Biophys. J.* 105, 1893-1903 (2013).
4. Barsegov, V., D. Klimov, & D. Thirumalai. Mapping the energy landscape of biomolecules using single molecule force correlation spectroscopy (FCS): Theory and applications, *Biophys. J.* 90, 3827-3841 (2006).
5. Kononova, O., L. Jones, & V. Barsegov. Order statistics inference for describing topological coupling and mechanical symmetry breaking in multidomain proteins. *J. Chem. Phys.* 139: 121913-121925 (2013).
6. Crooks, G. E. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences *Phys. Rev. E.* 60, 2721-2726 (1999).
7. Collin, D. et al. Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies. *Nature*, 437, 231-234 (2005).
8. Kononova, O. et al. Tubulin bond energies and microtubule biomechanics determined from nanoindentation in silico, *J. Am. Chem. Soc.* 136, 17036-17045 (2014).

Supplementary tables.

Table S1. Overview of encapsulin masses determined by native MS on the QTOFII instrument.

assigned species		mass (kDa)			
intact		average	standard deviation	theory	error (%)
60mer	empty	1718.2	0.307	1715.6	0.15
60mer	DyP-loaded	1966.0	0.339	1957.2	0.45
subcomplexes		average	standard deviation	theory	error (%)
58mer		1660.6	0.518	1658.5	0.13
30mer		860.0	0.340	857.8	0.26
18mer		516.4	0.184	514.7	0.33
16mer		458.4	0.259	457.5	0.19
10mer		286.1	0.032	285.9	0.04
6mer		171.6	0.006	171.6	0.04
4mer		114.4	0.001	114.4	0.04
2mer		57.2	0.003	57.2	0.04
1mer		28.6	0.002	28.6	0.07

Table S2. Summary of the mechanical properties and thermodynamics of encapsulin from *in silico* nanoindentation: F^* breaking force, Z^* cantilever displacement at breaking force, X^* tip displacement at breaking force, k_{sp} spring constant, and E Young's modulus.

Symmetry	F^* (nN)	Z^* (nm)	X^* (nm)	k_{sp} (N/m)	E (GPa)	ΔH (Mcal/mol)	ΔG (Mcal/mol)	$T\Delta S$ (Mcal/mol)
full encapsulin shell								
2-fold	1.1	26	4	0.26	1.07	2.95	1.93	1.02
3-fold	1.25	34	6.3	0.26	1.07	2.66	1.72	0.94
5-fold	1.17	33.8	9.8	0.37	1.52	2.6	1.70	0.90
empty encapsulin shell								
2-fold	1.55	40	7	0.31	1.29	4.02	2.95	1.07
3-fold	1.5	37	5	0.31	1.29	3.11	1.66	1.45
5-fold	2.2	57.7	11.3	0.45	1.74	4.43	3.38	0.05

Table S3. Overview of AFM nanoindentation results.

sample			k (N/m)		F_{break} (nN)		n
species	type	probe velocity (nm/s)	avg	SEM	avg	SEM	
<i>B. linens</i>	empty	30	0.27	0.02	0.64	0.04	46
		1000	0.36	0.01	0.82	0.05	22
	DyP-filled	30	0.25	0.02	0.44	0.03	54
	TFP-filled	30	0.24	0.01	0.42	0.04	26
<i>T. maritima</i>	empty	30	0.25	0.02	0.63	0.03	56

Supplementary figures.

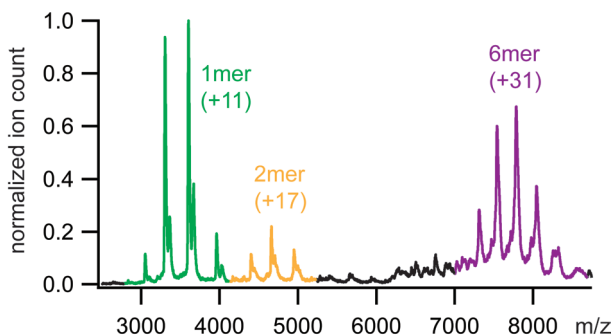


Figure S1. Native MS of *B. linens* DyP. DyP monomers (39.6 ± 0.002 kDa), dimers (79.2 ± 0.03 kDa) and hexamers (241.3 ± 0.03 kDa) are observed with native MS. The theoretical mass for monomeric DyP without heme cofactor is 39.6 kDa, the theoretical mass for hexameric DyP is 241.6 kDa, including cofactor. Main charge states are indicated in parentheses.

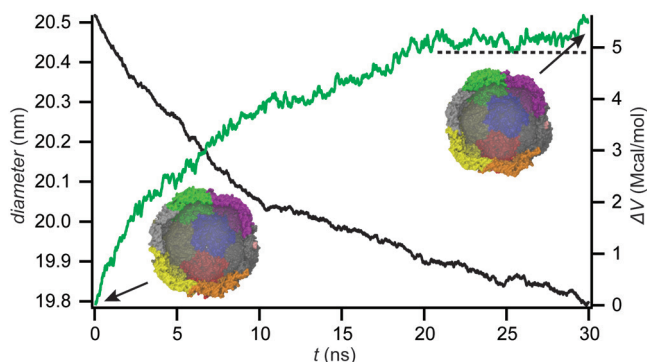


Figure S2. Diameter (black curve) and potential energy change (green curve) as a function of time from the all-atom MD simulation trajectory for *T. maritima* encapsulin shells at 300K. Also displayed are two snapshots of encapsulin, representing the shell structure at the beginning (0 ns) and at the end (30 ns) of a simulation run, showing the slight shrinkage and faceting of the shell.

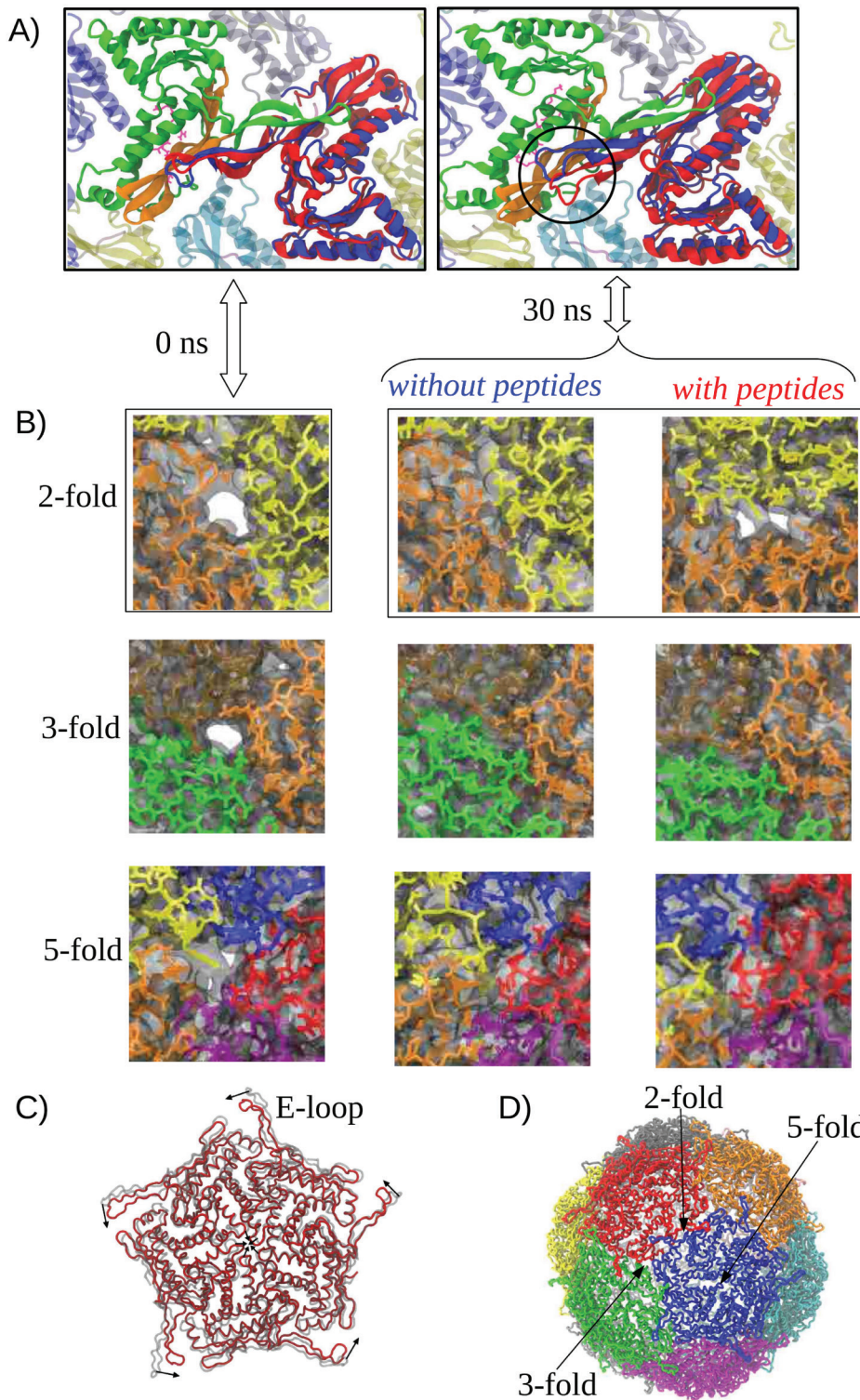


Figure S3. Detailed view of structural changes occurring in encapsulin shell with/without bound synthetic peptides during 30 ns of MD simulations. A) Position of E-loop in the protein subunit with/without bound synthetic peptides (red/blue) relatively to β -sheet region (orange) of neighboring subunit (green) across the 2-fold symmetry point at 0 ns (left) and after 30 ns (right) of equilibrium simulation. Circled is 1.5 nm difference in E-loop position for the structures with and without synthetic peptides. The observed twisting motion ‘pulls’ subunits closer together. The putative binding site for cargo anchoring sequence (magenta) on the shell interior is part of the same β -sheet at the E-loop from the neighboring subunit, suggesting a possible allosteric ‘pathway’ for the effect of cargo (un)binding. B) Close-up view of the pores in encapsulin shells with/without bound synthetic peptides as observed at the 0 ns and 30 ns time points of an MD simulation run at 300 K, viewed from the particle exterior. C) Essential Dynamics decomposition of encapsulin during 30 ns equilibration, shown for 5-fold symmetry related subunits. Start-structure is shown in gray, end-structure is in red. There is a slight twisting motion along the 5-fold symmetry axis, in particular of the E-loops. D) The SOP model of complete structure at 0 ns is shown as reference. Pores are closed along 5-fold axis, in both models. The pores on the 3-fold axes close after 30 ns at 300K, the pores around 2-fold axis close after 30 ns if synthetic peptides are not bound to the shell exterior.

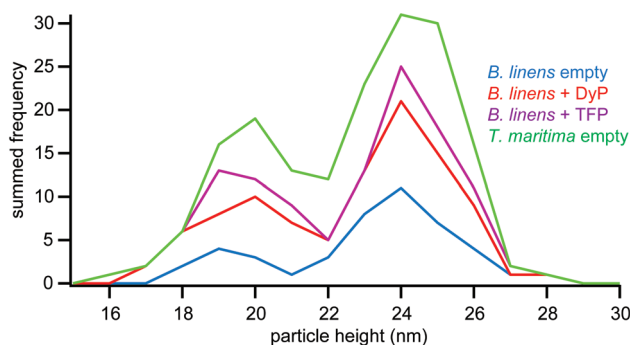


Figure S4. Height distribution of encapsulin nanocompartments as observed in AFM imaging. Sum frequency of height distributions of all encapsulin samples analyzed in this study. The height distribution is bimodal, which we attribute to particles being deposited on the different icosahedral symmetry axes.

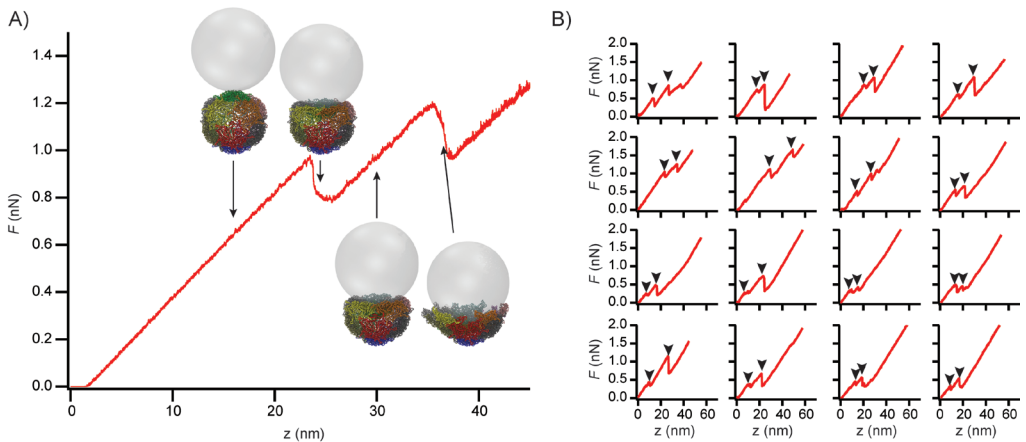


Figure S5. Langevin Dynamics simulations of *in silico* nanoindentation reveal a distinct buckling transitions before mechanical failure of the nanocompartment when probed along the 5-fold icosahedral symmetry axis. A) Typical FZ curve from the simulated indentation along the 5-fold axis. Corresponding structures before/after buckling (first force peak) and before/after mechanical failure (second force peak) are presented on the graph. B) Experimental FZ curves from AFM nanoindentation experiments that exhibit similar two-stage force responses. The main discontinuities are indicated on the curves. The corresponding AFM images were not conclusive as to the orientation of the particles.

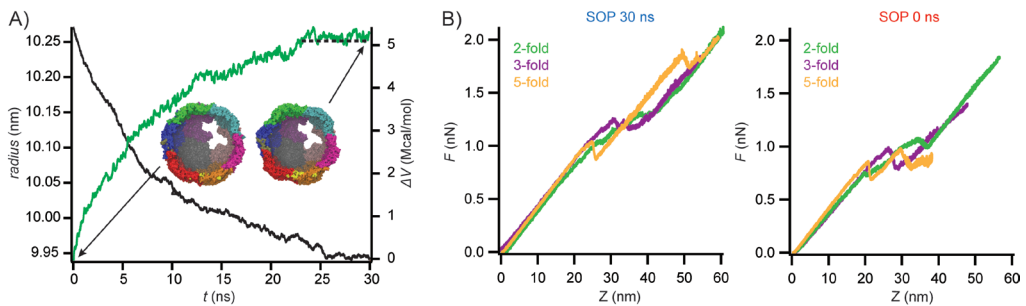


Figure S6. *In silico* nanoindentation of 58mer complexes of encapsulin. A) Diameter (black curve) and potential energy change (green curve) as a function of time from the all-atom MD simulation. One 2-fold symmetry related dimer was removed from the structure, after which all-atom MD equilibration was carried out using the same procedures as for the full 60mer complex. A similar decrease in radius and increase in potential energy was observed as for the complete shell. B) The start- and end structures were used to construct SOP models, labeled “SOP 0 ns” and “SOP 30 ns” respectively. *In silico* nanoindentation was performed on these models along the three icosahedral symmetry axes. The SOP models of 58mer complexes show a very similar force-response as complete 60mer shells. Average spring constant (across the three icosahedral symmetry axes) is 0.23 N/m for ‘SOP 0 ns’ and 0.30 N/m for ‘SOP 30ns’. Average breaking force is 1.0 nN for ‘SOP 0 ns’ and 1.4 nN for ‘SOP 30 ns’. These results indicate that possible 58-mer complexes of encapsulin might indeed be mechanically indistinguishable from complete shells. Given the rigidity of encapsulin nanocompartments, this demonstrates that assembly defects do not necessarily affect the mechanical resilience of the compartment.

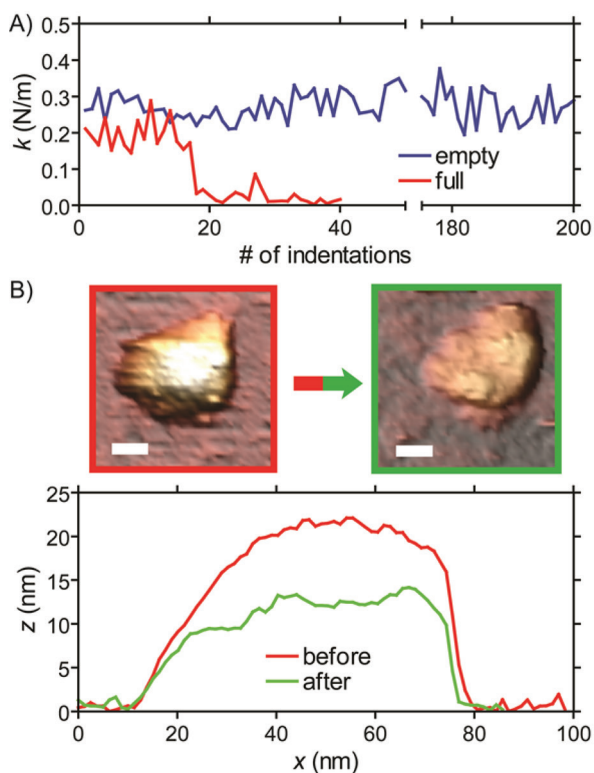


Figure S7. Cargo encapsulation makes encapsulin nanocompartments sensitive to material fatigue. A) Spring constant (k) is shown over repeated consecutive indentations at lower force ($F_{max} \sim 0.2$ nN). No decrease in spring constant is observed with the empty compartments, even after 200 consecutive approach-retraction cycles ($n=5$). DyP loaded compartments break after between 17-61 indentation cycles, with an average of 35 cycles ($n=4$). B) AFM imaging of a DyP loaded nanocompartment before and after the fatigue experiments. The decreased particle height and altered morphology show that the particle underwent a complete structural collapse after repeated indentation at low force. Scale bars are 20 nm. Images are colored according to height, from dark brown (0 nm) to white (25 nm).

