”Vision is the art of seeing things invisible”
Jonathan Swift

to my parents

ISBN: 978-90-393-5442-1

Thesis printing was financially supported by UIPS.

Analysing macromolecular structures with native mass spectrometry: Insights to virus structure and assembly

Macromoleculaire structuren geanalyseerd met behulp van natieve massaspectrometrie:

Inzichten over de structuur en opbouw van virussen

(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. J.C. Stoof, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op woensdag 24 november 2010 des middags te 12.45 uur

door

Charlotte Uetrecht

geboren op 28 april 1982, te Ostercappeln, Duitsland

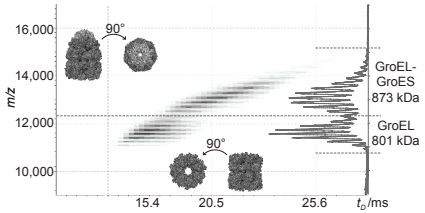
Promotor: Prof. dr. A.J.R. Heck

This work was financially supported by the Netherlands Proteomics Centre.

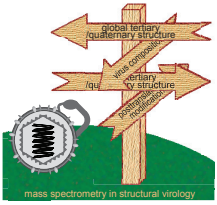
Table of contents

Chapter I **Introduction** **5**

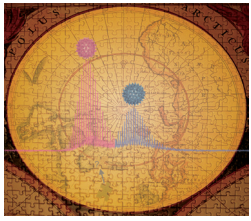
1. Native and ion mobility mass spectrometry 7



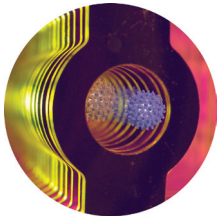
2. Analysing virus structure and assembly using mass spectrometry 44



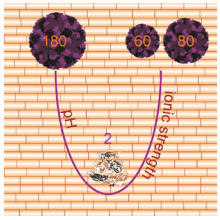
Chapter II **High-resolution mass spectrometry of viral assemblies: molecular composition and stability of dimorphic hepatitis B virus capsids** **73**



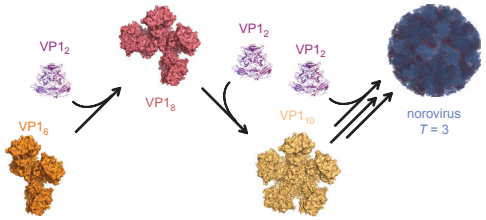
Chapter III **Stability and shape of hepatitis B virus capsids *in vacuo*** **83**



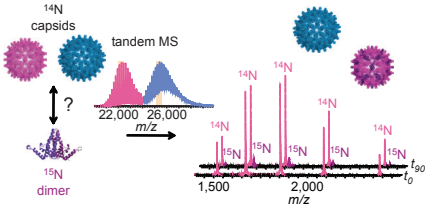
Chapter IV **Norwalk virus assembly and stability monitored by mass spectrometry** **91**



Chapter V **Interrogating viral capsid assembly with ion mobility mass spectrometry: from sheet- to sheath-like structures** **107**



Chapter VI **Subunit exchange rates in Hepatitis B virus capsids are geometry- and temperature-dependent** **121**



Chapter VII **Summary** **129**
1. Summarising discussion 131
2. Nederlandse samenvatting 134

Appendix I **Supplementary data** **137**
Chapter II 138
Chapter III 140
Chapter IV 143
Chapter V 146

Appendix II **About author** **151**
Curriculum vitae 152
Publication list 153
Acknowledgements 154

Chapter I

Introduction

In parts taken from:

Ion mobility mass spectrometry of proteins and protein assemblies

Charlotte Uetrecht, Rebecca J. Rose, Esther van Duijn, Kristina Lorenzen, Albert J.R. Heck

Chem Soc Rev **39**, 1633 - 1655 (2010)

Abbreviations: β_2m – β_2-microglobulin; CCMV – Cowpea Chlorotic Mottle Virus; CID – collision induced dissociation; cp – capsid protein; DC – direct current; ds – double-stranded; EHSS – exact hard sphere scattering; EM – electron microscopy; EMD – electrophoretic mobility diameter; ESI – electrospray ionisation; FAIMS – high field asymmetric waveform ion mobility spectrometry; GEMMA – gas phase electrophoretic mobility molecular analyser; Hb – haemoglobin; HBV – Hepatitis B virus; H/D – hydrogen/deuterium; IM – ion mobility; IMMS – ion mobility mass spectrometry; MALDI – matrix-assisted laser induced desorption/ionisation; MS – mass spectrometry; *m/z* – mass to charge ratio; NMR – nuclear magnetic resonance; NVLP – norovirus-like particle; Ω – collision cross section; ORF – open reading frame; PA – hard sphere projection approximation; Q-ToF – quadrupole time of flight; RF – radio frequency; SAS – small angle scattering; SRIG – stacked ring ion guide; ss – single-stranded; T-wave – travelling voltage wave; TMV – tobacco mosaic virus; ToF – time of flight; TRAP – tryptophan RNA binding attenuation protein; TTR – transthyretin; TWIMS – travelling voltage wave ion mobility (mass) spectrometry

1. Native and ion mobility mass spectrometry

Often, proteins are studied out of their cellular context to gain insights about their function in biochemical experiments. However, in the cell proteins operate in a dense network and prevalently form complexes. [Alberts, 1998] mass spectrometry (MS) based proteomics have elucidated the high interconnectivity in such protein networks. [Heck, 2008] On the other hand, the assembly process and complex stoichiometry is harder to investigate. Amongst the large protein complexes, viruses or their proteinous capsid shells are special since the number of different proteins that associate is limited. Studying virus structure is therefore of interest to understand protein assembly but also to bio- and nanotechnology. [Steinmetz *et al.*, 2009] Before we see how MS can assist the structural virologist, an overview of native MS in general and ion mobility MS (IMMS) in particular is provided.

Traditionally, MS has been a powerful analytical method enabling the structural analysis of small molecules, and later on peptides and proteins. With the advent of native MS, using a combination of electrospray ionisation (ESI) and Time of Flight (ToF) analysis, MS could also be applied to the mass determination of large protein complexes such as ribosomes and whole viruses. [Rostom *et al.*, 2000; Tito *et al.*, 2000] A particular advantage of native MS is the high sensitivity and speed allowing structural studies on endogenously expressed protein complexes identified in large scale proteomics approaches. Using the tandem affinity approach such complexes can be readily purified and subjected to native MS. [Lorenzen *et al.*, 2007a] Often, an initial identification of constituting subunits and their posttranslational modifications from the protease digested protein complexes is beneficial for the data interpretation. [Synowsky *et al.*, 2006] Moreover, careful selection of the tagged complex subunit enables the discrimination between complex variants specific to cellular compartments as shown for the nuclear and cytosolic yeast exosome. [Synowsky *et al.*, 2009] The mass of the assembly, provided by native MS, can be related to the complex stoichiometry utilising exact protein masses from MS of the denatured sample. In cases where several combinations are possible or the resolution is poor, tan-

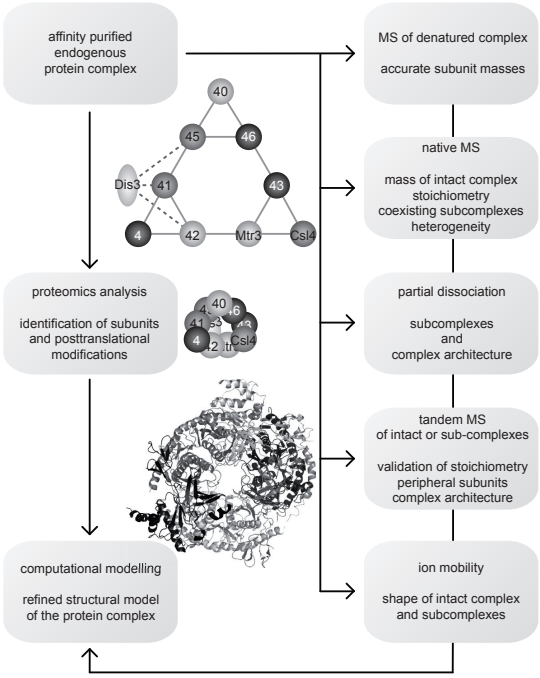


Fig. 1: Native MS workflow
A general workflow for structure determination of endogenously expressed protein complexes is depicted. After initial characterisation using proteomics and denatured mass spectra of the intact proteins, the subunit arrangement can be obtained from native MS and tandem MS experiments. An example for the proteins of the yeast exosome is shown in the center. In combination with global shape information from IMMS and modelling, atomic structures can be derived. Here, the homologous human exosome is displayed instead of the modelled yeast exosome. [Taverner *et al.*, 2008]

dem MS can assist the unambiguous assignment of complex stoichiometry. Especially, heterogeneous assemblies in overlapping mass to charge (*m/z*) regions can be distinguished.

Topological information on a protein complex can also be obtained from tandem MS, since small peripheral subunits are preferentially ejected upon collision induced dissociation (CID). Additionally, partial dissociation of the assembly with organic modifiers or pH variation can create subcomplexes which can reveal the entire architecture when combined with tandem MS. [Taverner *et al.*, 2008] A schematic native MS workflow is depicted in Fig. 1. Furthermore, the formation of dissociation products in solution and the gas phase can reflect the complex stability and strength of interactions. [Tahallah *et al.*, 2002; van Duijn *et al.*, 2006]

Even protein-protein and protein-ligand binding affinities have been estimated by native MS including the identification of potential cooperative effects. [Fu *et al.*, 2010; Kitova *et al.*, 2007; Lorenzen *et al.*, 2008] Whereas the complex architecture can be deduced from endogenously expressed samples, studies on binding properties and complex dynamics are more commonly applied to overexpressed proteins. Dynamics or reactions often can be monitored in real-time using label free approaches in native MS as opposed to NMR, fluorescence resonance energy transfer or radioactivity pulse-chase experiments. For example, the formation of mixed dodecamers of small heat shock proteins from different organisms has been studied revealing a structural sub-organisation into dimers. [Sobott *et al.*, 2002a] Moreover, transiently formed complexes, which are inaccessible to purification and therefore classical structural studies, could be detected in the folding cycle of the bacterial chaperone GroEL by native MS. [van Duijn *et al.*, 2005]

More recently, ion mobility spectrometry (IM) has been coupled to MS providing a new dimension in the analysis of biomolecules, with ion mobility separating ions according to differences in size and shape. In combination with native MS, IMMS opens up avenues for the detailed structural analysis of large and heterogeneous protein complexes, providing information on the stoichiometry, topology and cross section of these assemblies and their composite subunits. With these characteristics, IMMS offers a complementary tool in the context of structural biology. Native MS, its advantages and limitations have been extensively reviewed and the reader is referred to the following for more detailed information. [Benesch *et al.*, 2007; Bich & Zenobi, 2009; Heck & van den Heuvel, 2004; Heck, 2008; Sharon & Robinson, 2007; van Duijn, 2010] In the introduction of this thesis, I primarily focus on IMMS, which has recently been established for the analysis of larger proteins and protein assemblies. The development, instrumentation, approaches and applications of IMMS are introduced. In the second part of the introduction, I focus on general general concepts in virology and the possibilities mass spectrometric techniques hold to solve open questions in the structural biology of viruses.

1.1. General overview

As has been well documented, MS is an important and unique analytical tool for structural and biophysical studies of macromolecules. [Benesch *et al.*, 2007; Bich & Zenobi, 2009; Heck & van den Heuvel, 2004] The coupling of ion mobility to MS allows even greater insight into the properties of proteins and the complexes into which they assemble. Ion mobility has been in use for many decades, as a stand-alone technique as well as in combination with MS. However, the application of IMMS to intact proteins in their native-like conformation was not realised until the 1990s, and to large protein assemblies, only in the last few years.

This part of the introduction describes and summarises the use of native MS and IMMS in the study of proteins and protein assemblies, starting with the historical background of the technical developments, followed by the basic theory underpinning ion mobility and native mass spectrometry. Then, it leads towards a critical discussion of recent literature, focussing not only on experimental but also on computational aspects. I conclude with a view to the future prospects in the field of IMMS analysis of proteins and protein complexes.

1.1.1. Development of ion mobility and mass spectrometry

Ion mobility spectrometry and mass spectrometry share a similar theoretical background. When research into ions in gases began at the end of the 19th century, Rutherford, Thomson and Townsend observed that when X-rays pass through a gas, it becomes conductive. This finding initiated the use of gases for studying ions and their properties. Langevin laid the theoretical basis for the research which ultimately resulted in IM, [Langevin, 1903] whilst Thomson is seen today by many as the pioneer of MS. [Thomson, 1913] The historical development of these two methods and how they were combined is summarised in Fig. 2.

IM is a technique that separates ions based on their mobility, or the ability to move through a certain medium. This is highly dependent on the charge and shape of an ion. MS, by contrast, gives information on the *m/z*. Thus, if both techniques are coupled, ions can be identified by their mass, as derived from MS, and their overall structure can be determined simultaneously.

Langevin was one of the first to describe the nature of mobility, in particular, the interactions between ions and gaseous molecules, and the influence of the gas on the mobility of the ion. Based on Langevin's theory, ions can be separated by their characteristic velocity through a gas-filled electric field, the basic concept behind IM. Each ion usually drifts at a specific, constant velocity through such an electric field and all ions are thereby separated depending on their individual properties and those of the gas through which the ions drift. The nature of this technique does not allow the analysis of a continuous incoming flow of ions, but requires the introduction of discrete packets of ions into the ion mobility device, typically a 'drift tube', with each separation cycle being completed before the next begins. This major problem was solved by using pulsed injections of ions into the drift tube at the end of the 1920s by van de Graaf, Cravath, and through further

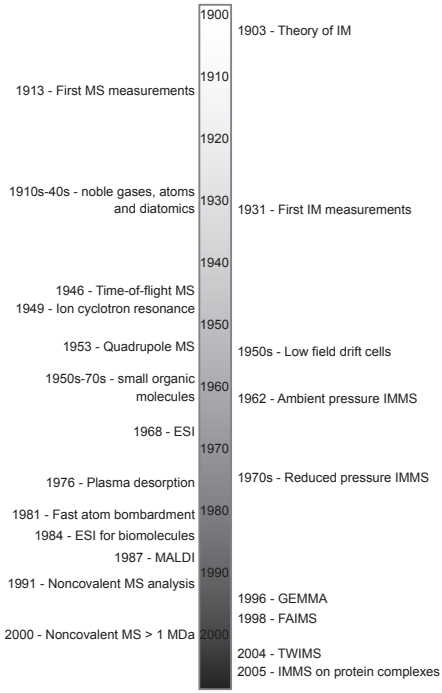


Fig. 2: Historical developments in MS and IM
Ion mobility (IM) and mass spectrometry (MS) have a common theoretical background and were developed in parallel; some of the most important events are shown in the timeline. Until the middle of the 20th century, analysis focussed on atomic species. New instrumentation in mass spectrometry and ion mobility enabled the detection of small organic molecules. In the last 25 years, the techniques were extended to study biomolecules and finally protein complexes. This progress was largely enabled by the development of ESI and new ion mobility devices.

modification by Bradbury. [Bradbury, 1931; Cravath, 1929; van de Graaf, 1928]

Early developments in MS followed a few years after Thomson's discovery, when Aston built the first prototype of what later became the basis of magnetic/sector field mass spectrometers. In contrast to IM, MS usually operates under vacuum conditions and uses distinctive properties of an ion to determine its exact mass. Although MS was used for the discovery and elucidation of isotopes as early as the start of the 20th century, [Aston, 1919] developing the basis of modern MS instrumentation took until the middle of that century. Throughout the years, both IM and MS continued to develop with the invention of new instrumentation. An interest in the behaviour of ions in air, due to research into pollution, warfare and space exploration, significantly aided the progression of IM. In the 1950s and '60s, McDaniel [McDaniel *et al.*, 1962] and Kebarle [Kebarle & Hogg, 1965] performed experiments with drift cells under a low electric field to study the separation and reactions of ions and small molecules in the gas phase. Shortly before McDaniel and Kebarle developed the predecessors of modern IM instrumentation, in the late 1940s and early '50s, the basis was laid for most of the mass analysers used in MS today. For example, the concept of ToF analysers, a detector that is commonly used in combination with IM today, was published by Stephens in 1946. [Stephens, 1946] The first prototype machine was reported in 1948 [Cameron & Eggers, 1948] and a couple of years later Wiley and McLaren came up with the design of a commercial linear ToF. [Wiley & McLaren, 1955] The coupling of IM with MS was first accomplished in the 1960s to study gas phase ion-molecule reactions. Since then, there have been combinations of IM with different types of MS analysers: initially magnetic sector instruments, [Barnes *et al.*, 1961; McDaniel *et al.*, 1962] then quadrupoles (Q), [Johnsen *et al.*, 1970; Thomson *et al.*, 1973] ToFs [McAfee *et al.*, 1967; McKnight *et al.*, 1967] and in recent years, ion traps and FTICRs as well. Different instrument types used in IMMS have lately been reviewed. [Kanu *et al.*, 2008]

The first experiments using IMMS focused on the investigation of properties of atoms or simple molecules, such as isotopes, gases and metal ions. Both the Bowers group [Kemper & Bowers, 1990] and the Jarrold group [Dugourd *et al.*, 1997] took the basis of an instrument developed by Böhringer and Arnold, [Bohringer & Arnold, 1982] and enabled new possibilities by implementing the ability to select a specific *m/z* prior to IMMS. Initially, this was used to investigate thermal reactions of mass-selected atomic ions and metal clusters (by Bowers and Jarrold, respectively). As will be discussed below, since these early studies, IMMS has also become tremendously useful for the much larger and more complex samples analysed today. For instance, to cover the protein contents of complete cells, the additional separating dimensions, such as provided by IMMS, are highly beneficial.

A relatively new application of IMMS is in structural biology, namely research of the structural properties of proteins and protein assemblies. The Jarrold and Clemmer groups did much of the pioneering work in this field analysing small proteins at the end of the 1990s. [Clemmer & Jarrold, 1997; Hudgins *et al.*, 1998] Their research initially focused on small organic and biological molecules, in particular peptides such as bradykinin and proteins like cytochrome C and ubiquitin, using a modified ion mobility-ToF mass spectrometer that enabled them to record mobilities and *m/z* ratios for multiple ions simultaneously. [Hoaglund *et al.*, 1998] A similar IMMS instrument was also designed by Bowers and co-workers. [Wyttenbach *et al.*, 2001] However, all of these instruments used for IMMS of proteins were developed in-house, and

dedicated IMMS instruments only became commercially available when Waters introduced the Synapt HDMS a couple of years ago. [Pringle *et al.*, 2007] Furthermore, the studies reported by Jarrold, Bowers and Clemmer did not deal with larger intact protein complexes. The retention of conformation, composition and stability of proteins and protein complexes in the gas phase has always been a major concern when dealing with native mass spectrometry of protein assemblies, [Heck & van den Heuvel, 2004; Loo, 1997] an issue that can now be addressed specifically by IMMS. In one of the first publications regarding the study of a functional protein complex by IMMS, Ruotolo *et al.* reported that the 11 member ring architecture of the RNA binding protein TRAP is largely maintained in the gas phase. [Ruotolo *et al.*, 2005] IMMS can thus contribute to the question of stability of proteins and their conformations in the gas phase. Moreover, IMMS can provide information about protein structure and conformational changes relatively quickly, requiring only a very low sample concentration and consumption. In the near future, IMMS will be able to provide, in conjunction with rapidly developing computational modelling tools, detailed information about the shape and conformation of a protein complex, its stability and its subunit arrangement.

1.1.2. General principle of ion mobility

The basic principles for IM in all applications are similar: a sample is introduced into the IM device after ionisation has taken place. Ion mobility separation depends on the characteristic behaviour of different ions in the gas phase, under the influence of an electric field and at either low vacuum or atmospheric pressure conditions, which result in a specific drift velocity. In general, this relies on the fact that properties such as size, shape and charge affect the mobility of an ion under given conditions. Therefore, this gas phase electrophoretic technique can be used analytically to deduce structural information regarding the ion of interest. The main advantages are the simplicity and speed of the measurement, the high sensitivity and selectivity, the fast generation of results and small size (*i.e.* portability) of the instrument. For these reasons, it has been used for many decades as a stand-alone method for the analysis of drug substances, explosives and other small molecule analytes. [St. Louis *et al.*, 1990]

An ion mobility spectrometer in its most basic form (a drift tube) consists of a gas-filled cell through which ions travel under the influence of a static electric field (see Fig. 5(a)). The velocity of the ion, v, is a product of the electric field, E, and the mobility of the ion, K, and is determined by measuring the time required, t_D, to traverse a drift cell of known dimensions, d. [Mesleh *et al.*, 1996]

$$v = K E = \frac{d}{t_D} \quad (1)$$

The reduced ion mobility, K_0, normalised with respect to pressure, P, and temperature, T, can thus be defined as:

$$K_0 = \frac{d}{E\, t_D} \frac{273}{T} \frac{P}{760} \quad (2)$$

This mobility constant can also be described in terms of the characteristics of the ion and the conditions of the drift cell, namely the ion's charge, $z\,e$, average collision cross section, Ω, the number density of the drift gas, N, and the reduced mass of the ion and buffer gas, μ.

$$K_0 = \frac{3\,z\,e}{16\,N} \frac{1}{\Omega} \left(\frac{2\,\pi}{\mu\, k_B\, T}\right)^{1/2} \quad (3)$$

The mass, charge and collision cross section of an ion thus determine its mobility under defined IM conditions. In coupling IM with mass spectrometry, the mass and charge become known factors, allowing the Ω of an ion to be calculated. This characteristic represents the orientationally averaged area of the ion which is able to interact with the buffer gas. For larger ions (*e.g.* proteins), the Ω can be approximated computationally using different, relatively simple models, such as hard sphere scattering (see section 1.4). [Creaser *et al.*, 2004; Hoaglund-Hyzer *et al.*, 1999] Ions with a larger apparent diameter undergo a greater number of collisions with the buffer gas, consequently their passage through the drift cell is retarded in comparison to smaller ions which experience less friction. This drift tube setup is the simplest version of the IM technique; there are, however, different variants and applications of IM, which will be discussed later.

1.1.3. General applications of ion mobility

The various IM techniques, especially in combination with MS, have found a wide range of applications during the years. Visible to everyone, but usually passed unnoticed, IM is employed at airports to scan for explosives and narcotics. IM is commonly used in a variety of industrial applications; examples include the semiconductor, pharmaceutical, medical and petrochemical industries, as well as chemical agent detection. Since the 1950s, military-based applications have been a major driving force behind the development of IM. Mainly applied as a detection system for chemical warfare agents and to monitor their destruction today, [Hill & Steiner, 2006] it was already in use to sense fumes and hazardous gases during the Second World War. Since then, the instrumentation has improved in various ways leading, for instance, to bench top IM and MS instruments small enough to be carried around easily. Further demonstrations of the power of IM include the separation of structural isomers, including chiral organic compounds and polyaromatic hydrocarbons, for which it has been used since the 1970s. [Griffin *et al.*, 1973; Hagen, 1979; Karpas *et al.*, 1986; Shen Nan *et al.*, 1974] Since the early nineties, IM has been a routine method for the detection of trace amounts of illicit drugs in forensics, notably heroin and cocaine, [Dussy *et al.*, 2008] with the detection limit for various narcotics being around 1 ng. In chemical industry, IM is often used to detect hazardous side products that can occur during production processes. This is often crucial in two aspects, one is of course the toxicity of these products for employees and the environment, but secondly, the ability to detect undesired by-products ensures product quality and enables a continuous process-

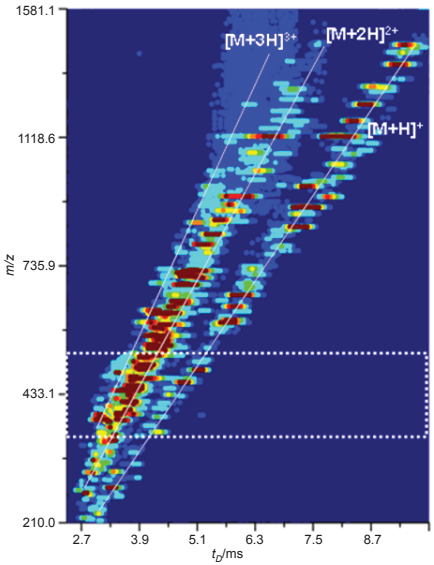


Fig. 3: IMMS separation of differentially charged peptides in proteomics
Differentially charged species are separated in ion mobility. This is especially useful in proteomics where the peptide charge can be directly assigned. Shown is a typical two-dimensional drift time (t_D) to *m/z* plot obtained for a tryptic digest of plasma. The singly, doubly, and triply charged peptide ion families are clearly separated and marked with the white lines. [Valentine *et al.*, 2006] (Reprint with permission from [Valentine *et al.*, 2006]. © 2006 American Chemical Society)

ing, and thus is also of economic interest. [Baumbach, 2006; Kolehmainen *et al.*, 2003] In the early days of IM, it was used by Arnold to monitor compounds in the stratosphere of our planet. [Arnold *et al.*, 1981] Today IM also finds a broad range of applications in environmental monitoring, from the detection of gases [Wu *et al.*, 1998] to microbiological organisms [Sheibani *et al.*, 2008].

In fundamental research, both IM and IMMS are highly significant and have a broad application range. IM can be used as a research tool for studying the structure of gas phase ions, or as a gas phase separation method. It has been used for structural studies of inorganic clusters (*e.g.* carbon and silicon), [Clemmer & Jarrold, 1997; Jarrold, 1995; von Helden *et al.*, 1993a] polymers such as poly(propylene glycol) and poly(methyl methacrylate), [Gidden *et al.*, 1999; Gidden *et al.*, 2000] DNA and oligonucleotides [Hoaglund *et al.*, 1997; Koomen *et al.*, 2002] as well as peptides and proteins. [Clemmer *et al.*, 1995; Clemmer & Jarrold, 1997; Creaser *et al.*, 2004; Hudgins *et al.*, 1999; Jarrold, 2000] As a separation technique, IM has proven powerful in the fields of metabolomics, glycomics and proteomics. [Dwivedi *et al.*, 2008; Jin *et al.*, 2005; Kanu *et al.*, 2008; McLean *et al.*, 2005; Plasencia *et al.*, 2008] The added dimension of separation it provides makes IMMS an attractive method for proteomic applications on complex samples. In addition to simply reducing the complexity of the sample, peptides are also separated according to their charge, facilitating peptide identifications (Fig. 3). [Valentine *et al.*, 2006] Before we focus on IMMS as a method for the structural elucidation of proteins and protein complexes, we now give a general overview of instrumentation used in native MS. Other techniques in native MS are introduced in conjunction with related case studies in IMMS.

1.1.4. Instrumentation in native mass spectrometry

In order to gain structural information on biomolecules with IMMS and MS, soft ionisation techniques are required to transfer the analytes into the gas phase without causing fragmentation. [Fenn *et al.*, 1989] ESI [Dole *et al.*, 1968; Fenn *et al.*, 1989; Whitehouse *et al.*, 1985] is favoured in MS of proteins and protein complexes over MALDI [Karas *et al.*, 1985; Koichi *et al.*, 1988] (matrix-assisted laser induced desorption/ionisation). [Bich & Zenobi, 2009] The main advantages in ESI are the creation of multiply charged ions from solution, which is essential for the study of noncovalent protein-protein interactions in native MS. [Bich & Zenobi, 2009; Heck & van den Heuvel, 2004; Heck, 2008; Sharon & Robinson, 2007] Singly charged ions of large complexes in MALDI appear at a high *m/z* in the mass analysers significantly reducing the mass accuracy, ion transmission and detection. [Bich & Zenobi, 2009] On the other hand, the spectra are simplified compared to ESI and the high salt tolerance and limited sample preparation make it in interesting technique for the analyses of protein complexes. Nonetheless, when looking at complexes in the MDa range, cross-linking is often required to covalently link protein assemblies.[Bich & Zenobi, 2009; Bolbach, 2005] Therefore, we focus on the ESI process (Fig. 4), which was applied throughout the research reported in this thesis.

Generally, in electrospray, high voltages are applied to a conductive capillary to create charged droplets of the analyte solution. Two mechanisms have been proposed for the transfer of ions into the gas phase, namely the ion emission [Iribarne & Thomson, 1976] and charged residue model. [Fernandez de la Mora, 2000] The latter one is now accepted for large biomolecules like protein complexes. [Cole, 2000; Fernandez de la Mora, 2000; Kebarle, 2000] Recently, Hogan *et al.* combined the two models to explain the ionisation of macromolecules.

[Hogan *et al.*, 2009] After formation of droplets from the Taylor cone and Coulombic fission, the protein resides in small droplets and ions in the solvent actually carry the excess charge; positive in most applications. According to the charged residue model the droplet fission continues until the Rayleigh limit charge is reached, which is defined by the size, *i.e.* conformation, of the protein. Finally, residual solvent evaporates and the charge is transferred to the desolvated analyte. However, some proteins show some discrepancy with this theory. It was proposed that neutral solvent evaporates from the droplet until a critical electrical field is reached and solvent ions evaporate from the droplet rather than the macromolecules as in the ion emission model. If enough energy is applied, the residual charge is eventually transferred to the completely desolvated macromolecule. In this adapted model, the charge of the protein is not only dependent on its size, but also the nature of the solvent ions explaining why the acquisition mode (positive or

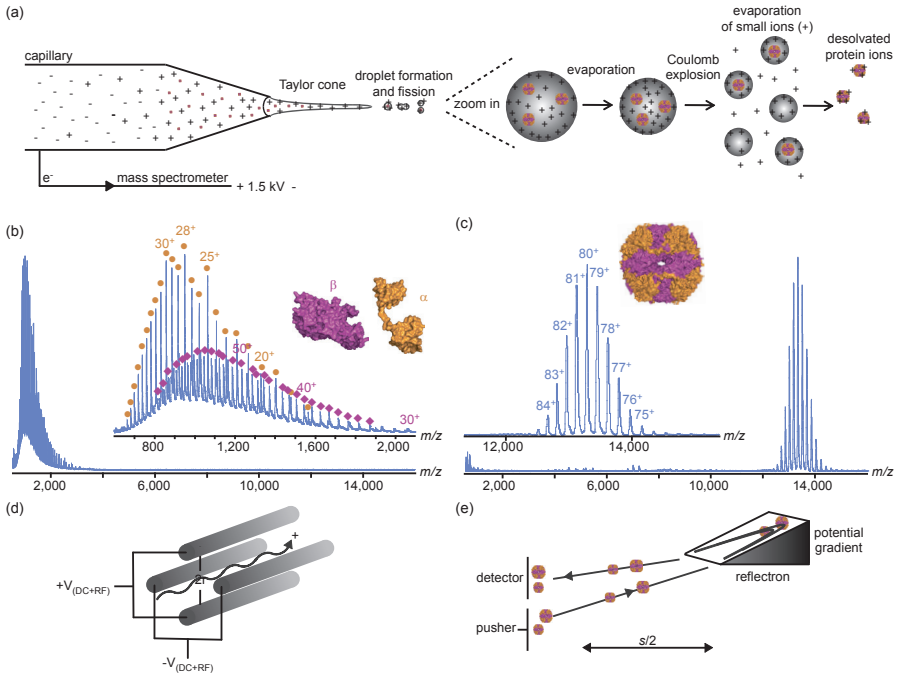


Fig. 4: Electrospray ionisation and mass analysers used in native MS
(a) Schematic representation of the nano-ESI process in positive ion mode. A high electric potential is applied to the conductive capillary containing the analyte solution. Initially, an elongated Taylor cone is formed, from which the droplets are released. Droplets undergo multiple fission events. The zoom in depicts steps in the combined charged residue and ion evaporation model. [Hogan *et al.*, 2009] After initial water evaporation, the surface charge increases until a certain energy is reached leading to a Coulomb explosion. The resulting tiny droplets keep shrinking *via* evaporation of small ions and water. Eventually, the analyte is completely desolvated carrying the residual charges. Depending on the solution conditions, individual subunits (b) or intact noncovalent complexes (c) are observed. The *helicobacter pylori* urease contains two subunits α and β, which dissociate and unfold in an acetonitrile-water solution. The corresponding highly charged ions appear below *m/z* 2,000. Upon exchange to ammonium acetate the signal of the intact 24-mer is observed at *m/z* 13,000 showing that noncovalent complexes can be preserved in the gas phase. [Pinkse *et al.*, 2003] (d) A quadrupole analyser consists of four parallel rods where opposing rods share the same potential. The combination of DC offset and RF determines the *m/z* at which ions enter a stable trajectory towards the detector. Ions of other *m/z* hit the rods. (e) Mass measurement in a ToF depends on the flight time. However, ions of the same mass albeit different kinetic energy (represented by size) arrive at dissimilar time points. In reflectron mode, ions become refocused in space and time thereby increasing the resolution. Low energy ions penetrate the opposing potential less and reach the detector at the same time as high energy ions.

negative ions), certain electrolytes or additions to the solution modify the protein charge. [Bagal *et al.*, 2009; Heck & van den Heuvel, 2004] To further increase the sensitivity, the solvent flow is usually reduced in nanoESI of proteins and protein complexes. [Bich & Zenobi, 2009; van den Heuvel & Heck, 2004] The specific requirements for noncovalent complexes are discussed later.

After ionisation, the biomolecules are transferred into the vacuum stages of the mass spectrometer. The preferred analyser to separate large ions according to their *m/z* is the ToF because of the theoretically unlimited mass range. [Ens & Standing, 2005; Guilhaus *et al.*, 2000] The *m/z* of the ions can be easily determined from the flight time, *t*, as long as the flight length in the tube, *s*, and the applied pushing voltage, *V*, to accelerate the ions are known:

$$m/_z = 2 \times V \frac{t^2}{s^2} \quad (4)$$

Since the development of reflectron mode ToF instruments, the resolution has significantly increased and improved the detection of the charge state distributions from ESI (Fig. 4). From adjacent charge states (+/- one *z*) of a species, first the charge, which is always an integer, and then the mass, M_{ion}, can be deduced: [Mann *et al.*, 1989]

$$z = \frac{m/_z}{m/_{z+1} - m/_z} \quad (5)$$

$$M_{ion} = m/_z \times z - z \times H^+ \quad (6)$$

Especially for noncovalent protein complexes, charge state selection and CID in tandem MS are often used to increase the confidence in stoichiometry assignment and reveal information about the complex topology or stability. [Benesch *et al.*, 2006; van Duijn *et al.*, 2006] For this purpose, quadrupole analysers are placed in front of the ToF, which can transmit ions of a certain *m/z*. [Ens & Standing, 2005] Always, the opposing of four parallel rods are electrically connected. A combination of direct current (DC) and superimposed radio frequency (RF) voltage defines which ions enter a stable spiral trajectory or collide with the rods (Fig. 4). The quadrupole can also operate in RF only mode, thereby resolution is lost and a broad range of ions reaches the detector. [Douglas, 2009] In tandem MS instruments, the resolution is provided by the ToF analyser for monitoring the whole set of ions produced. [Ens & Standing, 2005] Standard quadrupoles perform best up to *m/z* 4,000, nevertheless the RF generator can be changed to lower frequencies to transmit ions of higher *m/z*. [Douglas, 2009; Sobott *et al.*, 2002b; van den Heuvel *et al.*, 2006]

We now move on to the general IM instrumentation as well as a basic introduction to travelling wave IM combined with MS (TWIMS), implemented in the first commercial instrument available for studying protein assemblies. Since the first publication using this technique about four years ago, [Ruotolo *et al.*, 2005] there has been ever-growing interest in the method and an increasing amount of publications accordingly. However, as the use of IMMS to study protein complexes is still in its infancy, it is necessary to gain deeper insight into the underlying fundamentals of the method and the possibilities it holds. Thus, we end the first part of the introduction with a discussion about the potential IMMS might offer in the near future, with a special emphasis on IMMS of protein complexes in combination with computational modelling and its position amongst traditional structural biology techniques in terms of the sample requirements, obtainable resolution, study of dynamics and differences to the physiological environment (sec-

tion 1.5.3).

1.2. Analysing individual proteins in the gas phase

Drift tube IM and high-field asymmetric waveform ion mobility spectrometry (FAIMS) have been extensively used to study the structure of single proteins and small protein aggregates. These methods and the instrumentation involved differ significantly, however, they share similar applications.

1.2.1. Drift tube ion mobility

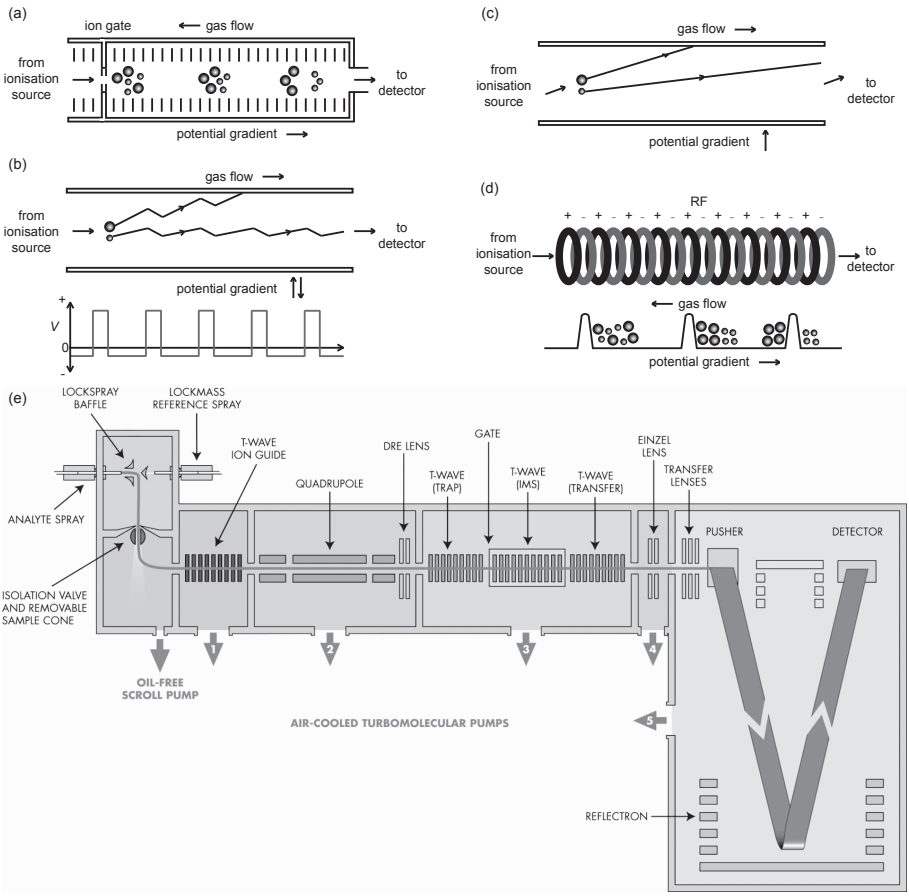
Drift tube ion mobility spectrometry, as mentioned above, is based on the temporal separation of ions of different size and shape under low electric field conditions of around 10 V cm^{-1} (for the general theory, see 1.1.2.). In the common setup, depicted schematically in Fig. 5(a), packets of ions are pulsed into the drift cell by an ion gate. Inside the cell, a neutral inert gas such as helium, argon or nitrogen is maintained at a constant pressure, generally in the range of 1-15 mbar. Ions traverse this cell under the influence of a uniform electric field, typically generated by a stack of ring electrodes (stacked ring ion guide or SRIG) connected by a series of resistors. [Creaser *et al.*, 2004; Fenn & McLean, 2008] The reported lengths of drift tubes range from a few centimetres to nearly a metre, and more recently, a cyclotron device of theoretically infinite length. [Baker *et al.*, 2007; McCullough *et al.*, 2008; Merenbloom *et al.*, 2009; Ruotolo *et al.*, 2004a] The length of the drift region must be accurately known in order to determine the precise velocity of the ions (eqn. (1)). The separation of ions in the tube typically occurs on a millisecond timescale. [Kanu *et al.*, 2008; Ruotolo *et al.*, 2002] Having traversed the ion mobility region, ions are directed towards the detector. In the case of IMMS, this involves focussing and transferring the ions into the high vacuum region of the mass analyser, whilst maintaining the mobility separation. Various combinations of drift tube IM with different mass analysers have been reported in which the mass analysers have been used solely to detect the *m/z* of separated ions, or to enable selection and/or storage of ions in quadrupoles or ion traps prior to the IM device. [Badman *et al.*, 2001; Clemmer *et al.*, 1995; Shelimov *et al.*, 1997]

1.2.2. High-field asymmetric waveform ion mobility spectrometry

In contrast to the simple drift tube setup described above, high drift fields are used in an alternative IM instrumentation, FAIMS. [Guevremont & Purves, 1999] In FAIMS, the ion mobility also becomes dependent on the strength of the applied electric field, since eqn. (1) breaks down at high ratios of E / N (*i.e.* greater than 2 x 10^{-21} V m^2). [Miller *et al.*, 2001; Purves & Guevremont, 1999] Here, it is effectively the dependence of mobility on the field strength rather than an absolute value of mobility that is measured. By applying an asymmetric waveform across two electrodes, ions alternately experience two distinct field strengths. The ions oscillate between the electrodes, and due to differential mobility in each field, will preferentially move towards one electrode (Fig. 5(b)). [Purves *et al.*, 1998] A compensation voltage is applied to correct this movement and enable ion transmission. Ions of different overall size and charge

Fig. 5: Characteristics of different ion mobility instrumentation
(a) Drift tube IM: In drift tube ion mobility, packets of ions traverse the cell along a potential gradient against a constant gas flow. Larger ions experience more collisions with the gas and therefore are retarded relative to smaller ions with a higher mobility. [Shvartsburg *et al.*, 2001] (b) FAIMS: The separation principle in FAIMSdiffers from the drift tube setup. The ions move with the gas flow and two alternating electric fields are applied perpendicular to the flow. Ions with different mobilities require distinct compensation voltages to transit the ion mobility cell (top). With the high fields used in FAIMS, the relation between mobility and compensation voltage is not trivial, preventing the direct determination of collision cross sections. An example of an asymmetric waveform showing the alternating voltages is shown (bottom). [Purves *et al.*, 1998]

require different compensation voltages to be transmitted through the FAIMS cell. Thus, the compensation voltage can be scanned to allow sequential detection of all ions present, or set at a pre-determined value to analyse a specific ion of interest (similar to a quadrupole mass analyser). For a given species, it is the compensation voltage, rather than the drift time, which is generally reported.



(c) GEMMA: The differential mobility analyser implemented in the GEMMA setup also uses an electric field perpendicular to the ion trajectory, with the latter in line with the gas flow. The field applied is constant and scanning a range of voltages leads to the transmission of different ions. From the required voltage the size of the ion can be determined. (d) TWIMS: In TWIMS alternating phases of RF voltage are applied to a stacked ring ion guide (top) on which a travelling wave is superimposed. Ions are pushed along in front of a potential wave in TWIMS (bottom). Due to the reverse gas flow, those ions with low mobilities experience the most friction and eventually roll over the crest of the wave and exit the cell last. [Giles *et al.*, 2004] (e) Schematic representation of a commercial TWIMS instrument. A travelling wave ion guide (TWIG) is used instead of a classical hexapole lens after the Z-electrospray ionisation source. A quadrupole mass analyser for *m/z* selection of ions is followed by three consecutive TWIGs, which function as a trapping or collision cell, an ion mobility cell and a transfer or additional collision cell. A different gas may be introduced in the trap and transfer region as the ion mobility device is in a separate chamber. Finally, ions are mass analysed in the ToF. [Pringle *et al.*, 2007] (Reprinted from [Pringle *et al.*, 2007], © 2007 with permission from Elsevier.)

Another difference in the operation of FAIMS is that the ions move with the gas flow under the influence of a perpendicular electric field, whereas in drift tube IM ions experience an axial field and move against the gas flow (Fig. 5). FAIMS is a continuous ion beam technique, and the electrodes cause focussing of the ions during separation, thus increasing sensitivity. However, whereas in FAIMS only one species can be analysed at a time, drift tube IM analyses all ions at once. Furthermore, the measurement of differential mobility by FAIMS does not allow direct calculation of average cross sections of the ions. [Purves *et al.*, 2000b] The exact characteristics controlling the mobility of an ion at high field strengths are not known precisely and appear to be independent of *m/z* and the low-field Ω. It has been suggested that structural flexibility is an important factor; if a structure is flexible, the orientation and distribution of charge on an ion can change relative to an applied field. [Purves *et al.*, 2000b] This would then lead to an apparent decrease in mobility under high fields compared to low fields that would not be observed for very rigid structures.

As yet, the theory of FAIMS separation is quite poorly understood and in most cases structural dimensions of the protein ions cannot be directly extracted from the data. Nevertheless, FAIMS in combination with MS has proven to be a valuable complementary technique for the analysis of proteins and peptides. Protein conformational studies have been reported for, amongst others, ubiquitin, [Purves *et al.*, 2000a; Purves *et al.*, 2000b; Purves *et al.*, 2001; Robinson & Williams, 2005] and to a lesser extent for cytochrome C [Purves & Guevremont, 1999; Shvartsburg *et al.*, 2006a]. Direct comparisons of FAIMS and drift tube IM data have indicated that both instrumental setups can provide similar results. [Covey & Douglas, 1993; Purves *et al.*, 2000a; Purves *et al.*, 2000b; Valentine *et al.*, 1997] Interestingly, in one study using FAIMS, increased waveform voltages and optimisation of the carrier gas allowed the separation of ubiquitin conformations with very subtle differences in their cross sections which were not resolved by drift tube IM. [Purves *et al.*, 2001] FAIMS has also been hyphenated with drift tube IM to enhance resolution, enabling the separation of more conformational ensembles of ubiquitin and cytochrome C compared to either drift tube IM or FAIMS alone. [Shvartsburg *et al.*, 2006b]

1.2.3. Early studies on gas phase protein conformation by IMMS

Initially, the application of IMMS to proteins focussed on the effects of desolvation and charge on the protein structure. In 1995, a pioneering study showed that different gas phase conformations of cytochrome C coexist, [Clemmer *et al.*, 1995] whereby the most compact structures were observed for low charge state ions (family A in Fig. 6(a)). At higher charge states, ions populated more extended species, a phenomenon which is likely due to repulsive Coulombic forces. Interestingly, several distinct conformations could be resolved for a particular ion, intermediate charge states in particular were found to exhibit high structural diversity ranging from compact and intermediate to extended conformational states. Some of the observed IMMS peaks, however, were seen to be relatively broad, which is likely attributable to the presence of even more similar and/or rapidly interconverting conformations (Fig. 6(b) left panel). In general, the Ω increased concurrently with the charge of the ion within a conformational 'family'. Comparison of the Ω estimated from available nuclear magnetic resonance (NMR) data and/or crystal structures of the folded protein, as well as from modelled extended random coil structures, suggested that the observed species were intermediate to these limiting conformational states and likely partially folded. Indeed, the different cross sectional 'families' that have been

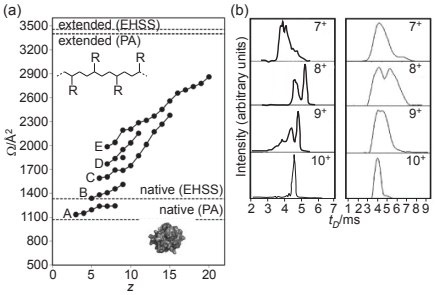


Fig. 6: Drift tube and TWIMS analysis of protein conformers
(a) Distinct conformational 'families' (A to E) can be separated in ion mobility over a range of charge states (z) for cytochrome C. At higher charge states, clear transitions between various conformations are observed. The dotted lines correspond to collision cross sections (Ω) calculated using two different models (PA – projection approximation, EHSS – exact hard sphere scattering) for the native crystal structure and an extended string-like conformation indicating that low charge states exhibit a native-like structure. [Badman *et al.*, 2001] (Reprint with permission from [Badman *et al.*, 2001]. © 2001 American Chemical Society) (b) The conformational separation is similar in drift tube ion mobility (left) [Badman *et al.*, 2001] and TWIMS (right) [Ruotolo *et al.*, 2005], despite the difference in resolution, as is apparent from the drift time (t_D) profiles for the 7^+ to 10^+ ions of cytochrome C. (From [Ruotolo *et al.*, 2005]. Reprinted with permission from AAAS)

detected for cytochrome C vary from close to the native structure to largely extended structures. [Badman *et al.*, 2001] The compact native-like state could also be observed for ions of intermediate numbers of charges. However, this population, when subjected to collisional heating (by increasing injection energy), underwent transitions to more unfolded states. [Shelimov *et al.*, 1997]

In contrast to these analyses, IM studies on the small protein bovine pancreatic trypsin inhibitor (BPTI) revealed that a conformation slightly smaller than the native structure was populated at low charge states (also predicted by molecular dynamics simulations). The Ω gradually increased towards higher charge states. However, there was no transition observed between distinct conformations, and collisional heating of the ions did not result in increased cross sections. Therefore, the native, folded state of BPTI is apparently much more stable, even being resistant to the Coulombic repulsion resulting from an increase in net charge. [Shelimov *et al.*, 1997]

In the case of the protein ubiquitin, which is only slightly larger than BPTI, activation at high injection energy resulted in the observation of a single elongated conformation, whereas at lower injection energies, compact and partially folded states were also seen, a similar behaviour to that described for cytochrome C. [Valentine *et al.*, 1997] The Ω of the compact structures were also slightly larger than estimated from the crystal structure. Here, the effect of the charge state on the observed drift time was examined further using proton-transfer reactions. It was found that the compact states were more likely to transfer a proton and thus decrease their charge. This observation was attributed to the higher Coulombic energy these states have; the repulsion decreases significantly upon charge loss. The elongated states of ubiquitin were not reactive, *i.e.* did not transfer a proton, and were not seen to undergo any folding event to become partially folded or compact. This indicates that the elongation of a protein ion decreases the Coulombic energy compared to a more compact structure. [Valentine *et al.*, 1997]

One approach which enhances the insight into gas phase conformations of proteins is the use of tandem or multidimensional IMMS. In these cases, the ions are transmitted through sequential drift tubes before entering a ToF analyser. [Koeniger *et al.*, 2006a; Koeniger *et al.*, 2006b; Merenbloom *et al.*, 2006] Collision cells between the different drift tubes allow experiments to be performed in which a specific conformation can be selected, activated and subsequently analysed in the next IM device to determine the resulting conformations. [Merenbloom *et al.*, 2006] By selecting and activating compact states of ubiquitin, a range of partially folded and extended conformations were formed. Interestingly, these retained some memory of their original struc-

ture, most likely in the form of secondary structure elements. [Koeniger *et al.*, 2006b] Similar to tandem MS, such tandem IMMS techniques can be useful *in situ*ations where broad peaks (consisting of mixed signals) are present or peaks are poorly resolved, since narrow sections of such peaks can also be isolated and further analysed. [Koeniger *et al.*, 2006a]

As can be inferred from the studies described so far, charge can have different and discrete effects on the mobility of an ion. Firstly, in accordance with eqn. (3), the mobility is directly proportional to charge. Assuming that the charge does not affect the conformation of the analyte at all, additional charges would increase the mobility, and thus decrease the drift time. However, increased charge can also cause longer drift times due to a gradual expansion of the structure of the protein ions. Furthermore, more charges can result in a sudden conversion to a completely different conformation. These last two structural effects are largely due to Coulombic repulsion. This is a phenomenon which is typically enhanced in the gas phase due to the lack of solvation, and concomitant decrease in dielectric constant. [Shelimov *et al.*, 1997; Wolynes, 1995] Of course, distinct conformations can already be present in solution as well, which may affect the charge state of the ions formed as is evident from studies on unfolding in acidic solutions.

1.2.4. Protein structure and folding

That protein ions can unfold in the gas phase due to Coulombic repulsions has become apparent through many IMMS experiments. In line with this, ions with fewer charges generally retain a more solution-like structure than their highly charged counterparts. However, the refolding of solution-denatured species can sometimes occur in the gas phase. Many studies over the last years have addressed the process of protein folding and unfolding in the gas phase. Such investigations allow intra-molecular interactions to be analysed in the absence of solvent thus, in principle, allowing the analysis of the intrinsic properties of the protein, providing a complementary contribution to the fundamental understanding of protein folding. [Jarrold, 1999; Jarrold, 2000]

Cytochrome C, analysed from an acidic solution in which it is denatured (apparent from the high charge states of the ions), mostly populates an extended conformation in the gas phase. [Shelimov & Jarrold, 1996] However, stripping some of the protons off in the gas phase was seen to generate species with lower charge states that had a shorter drift time, very similar to that expected for the native conformation. Therefore, charge reduction of a protein that was originally denatured can induce refolding to a more compact structure in the gas phase.

In the case of apomyoglobin, however, a significant activation barrier to complete gas phase refolding was reported. [Shelimov & Jarrold, 1997] Again, lower charge states of the ions were produced by proton stripping, resulting in a spontaneous collapse into a partially folded conformer. Interestingly, the structure collapsed further on collisional heating, forming a conformational state apparently slightly more compact than the native-like state. Thus for this protein, there exists an energy barrier which must be overcome in order to achieve the final gas phase folded state for these low charge states. This barrier is presumably due to a combination of Coulombic repulsive forces and the rearrangement of secondary structure elements. It is worth noting that the effect of collisional heating observed for apomyoglobin here (folding) is in contrast to the effect of collisional heating on partially folded/compact states of cytochrome C and ubiquitin (unfolding).

To monitor protein folding in the gas phase over prolonged periods of time (up to 30 s), an

instrument has been developed in which an ion trap is positioned before the IM device. [Badman *et al.*, 2001] In this setup, proteins are accumulated in the ion trap and released after specific periods of time, revealing complex unfolding/refolding activity. [Badman *et al.*, 2001; Badman *et al.*, 2005; Myung *et al.*, 2002] On the millisecond timescale, partially folded states were seen to unfold, whereas compact states often resulted in partially folded and elongated states with a variety of transition rates. [Myung *et al.*, 2002] Trapping proteins for longer timescales, up to 10 s, revealed very interesting folding behaviour. [Badman *et al.*, 2005] After a certain dwell time, the newly formed unfolded configurations underwent gas phase refolding events. This suggests that it is necessary for the ions to form relatively extended conformations before being able to sample other more stable and compact states. [Badman *et al.*, 2005] Evidently, ion mobility can be used to reveal protein folding behaviour on a timescale which is not easily accessible by other techniques.

The ability to detect different protein conformations makes IMMS attractive for analysing proteins involved in protein misfolding and amyloidogenic diseases, such as α-synuclein. This protein aggregates and can form amyloid fibrils, implicated in Parkinson's disease. At neutral pH, the protein is effectively unstructured, which largely hampers the use of traditional techniques, such as X-ray diffraction and NMR. IMMS has been used to obtain structural information about α-synuclein at different pH values. [Bernstein *et al.*, 2004] Surprisingly, compact species were detected at low charge states with a Ω similar to a theoretical globular form, whereas the structure elongates at higher charge states. In contrast to IMMS data on other proteins, [Badman *et al.*, 2001; Clemmer *et al.*, 1995; Shelimov *et al.*, 1997; Valentine *et al.*, 1997] there were no co-populated conformations of α-synuclein observed, but rather, a sudden increase in Ω of more than 50% occurred from charge state 8^- to 9^-. With regards to amyloidosis, the ligation of metal ions has been shown to increase the formation of amyloid fibrils, and the dimensions of these fibrils, indicated that a folded state of α-synuclein must be involved. From the IMMS data obtained, Bernstein *et al.* suggested that the ligation of the metal ions reduces the charge, and therefore causes collapse of the intrinsically unfolded structure into an aggregation-competent, compact state. [Bernstein *et al.*, 2004]

In the case of β_2-microglobulin (β_2m), which causes dialysis-related amyloidosis, the relative abundances of folded, partially folded and unfolded ensembles were compared at different pH values. [Borysik *et al.*, 2004] This is particularly important in this case, since the partially folded states are thought to be responsible for the aggregation of this protein into amyloid fibrils. This protein has since been subjected to further IMMS studies, using more recently developed technology, as discussed below in section 1.3.4.

The amount of information about protein structure obtained using IMMS can be enhanced by including extended computational approaches, such as in a study on the amyloidogenic protein involved in Alzheimer's disease, amyloid β-protein (Aβ42). [Bernstein *et al.*, 2005] In its monomeric form, observed as 3^- and 4^- charge states, ion mobility revealed two broad and overlapping peaks of similar intensity. The cross sections differed by approximately 10%, and evidently relate to distinct conformational ensembles that cannot interconvert on a millisecond timescale. Molecular dynamics studies, however, suggested that there may actually be more than two conformational states. [Baumketner *et al.*, 2006] Indeed, four 'families' of conformations and their relative populations were proposed, which is compatible with the broad peaks

observed. Thus, combining computational studies with experimental IMMS can aid and enhance confidence in data interpretation, as illustrated further in section 1.4.

As well as to study intact proteins, IMMS has been applied extensively to analyse the formation of secondary structure using synthetic peptides as model systems. [Hartings *et al.*, 2003; Kinnear *et al.*, 2001; Ruotolo *et al.*, 2004b; Zilch *et al.*, 2007] IMMS is an effective method for studying simple peptides, and in combination with modelling techniques (see section 1.4), which are typically more accurate for smaller systems, provides a greater understanding of the fundamentals of gas phase protein structure and folding. However, the relationship between gas phase and solution is complex as, for instance, the propensity to form helices can be quite different. [Jarrold, 1999] In an illustrative example, the peptide $AcA_{14}KG_3A_{14}K$, designed to form a helix-turn-helix motif, was analysed by IMMS and molecular dynamics simulations. [Zilch *et al.*, 2007] Indeed, both the ion mobility and simulation data showed a folded coiled-coil conformation. In addition, an extended form which is effectively unfolded and a metastable state were separated and identified by IMMS. The latter state relates to a misfolded ensemble of structures, apparently kinetically trapped, and transforming to the coiled-coil geometry only on increasing the temperature. These and other such analyses can provide valuable information on the stabilisation of protein secondary structure.

1.2.5. Studies on small noncovalent protein aggregates

A few relatively small noncovalent protein complexes have been studied by drift tube IMMS, particularly amyloidogenic proteins, in which aggregation is pronounced under certain solvent conditions. Identification of oligomeric states and their relative abundances by MS can sometimes be ambiguous, especially as the different copy numbers of homo-oligomers can give rise to peaks at the same *m/z*. Here, ion mobility can deconvolve such signals by analysing the relative size of the components. For instance, Bernstein *et al.*, studying Aβ42, analysed a single mass spectrometric peak by ion mobility and found that four different species contributed to this signal, namely a dimer (*z* charges), a tetramer (2 *z*), a hexamer (3 *z*) and a dodecamer (6 *z*). [Bernstein *et al.*, 2005] The relative proportions of these different oligomers could be altered using CID, which resulted in the decrease of the larger species and a predominance of the dimer. Furthermore, the monomer-dimer-tetramer interconversion was analysed as a function of temperature to determine the activation energies of dissociation. Thus, it is possible to deduce the relative populations of different oligomeric species, and even their related thermodynamic properties, using IMMS.

1.3. The analysis of protein complexes by ion mobility mass spectrometry

There is currently a significant gap between the generation of high throughput data on protein-protein interaction maps from, for instance, proteomics and yeast two-hybrid approaches, and the structural biology field determining the spatial arrangement of proteins within a specific functional macromolecular protein complex. In order to connect both research fields, there is a pressing demand for biophysical techniques that can monitor conformations of large proteins and intact protein assemblies. [Heck, 2008] IMMS may have an important role in such investigations. Since it was shown that different conformations of smaller proteins can be separated in the gas phase by IMMS based on their mass, charge and structure, as discussed above, [Clemmer *et al.*, 1995; Kanu *et al.*, 2008; Pringle *et al.*, 2007; Ruotolo *et al.*, 2008; Thalassinos *et al.*, 2004] this method has rapidly been gaining interest among structural chemists and biologists to

study additionally larger and more heterogeneous assemblies.

However, IMMS is not yet a high throughput method for the analysis of protein complexes to reveal their topology maps. Nevertheless, tremendous progress has been made in this field during the last few years, mirrored by many publications. [Alverdi *et al.*, 2008; Lorenzen *et al.*, 2008; Ruotolo *et al.*, 2005; Ruotolo *et al.*, 2007; Ruotolo *et al.*, 2008; Uetrecht *et al.*, 2008b; van Duijn *et al.*, 2009] The aim of most of these studies was to investigate the quaternary structure of protein complexes. Importantly, many of these early results reported at least the partial preservation of solution phase structures in the gas phase. [Lorenzen *et al.*, 2008; Ruotolo *et al.*, 2005; Ruotolo *et al.*, 2007; Uetrecht *et al.*, 2008b; van Duijn *et al.*, 2009]

There are certain considerations that need to be taken into account when obtaining mass spectrometric data of native, functional protein assemblies, or even entire molecular machineries. Firstly, modifications to the commercially available instrumentation are generally necessary. [Benesch *et al.*, 2007; Heck, 2008; van den Heuvel & Heck, 2004] In particular, protein complexes require higher pressures in the mass spectrometer to reduce their speed and thereby increase focussing and transmission. [Tahallah *et al.*, 2001] Heavy rare gases like xenon have also been beneficial for higher sensitivity and efficient complex dissociation in CID. [Lorenzen *et al.*, 2007b] Additionally, lower frequency quadrupoles as mentioned above and pushers with a low repetition rate in the ToF are preferred. [Sobott *et al.*, 2002b; van den Heuvel *et al.*, 2006] Secondly, noncovalent complexes usually require buffered solutions to retain their integrity. However to enable proper desolvation and therefore mass determination, non-volatile salts such as phosphate-based buffers have to be avoided and replaced by volatile buffer-like solutions; aqueous ammonium acetate typically provides the best compromise (Fig. 4(c)). Such non-covalent or native mass spectrometric methods in combination with IM can provide invaluable insights on the structure of protein complexes. [Bich & Zenobi, 2009; Heck, 2008; Sharon & Robinson, 2007]

To date, only one commercial IMMS instrument capable of analysing large protein complexes is on the market. [Giles *et al.*, 2004] Unlike conventional ion mobility instruments, a travelling voltage wave (T-wave) is used in the IM cell to propel the ions through this device and reduce their transit time. As a consequence, both sensitivity and analysis speed are significantly increased. [Giles *et al.*, 2004; Scarff *et al.*, 2008] In the following section, we focus on IMMS studies performed on T-wave instruments, however, before we do so, we briefly describe the GEMMA instrumentation, which has been used previously to obtain IM data on very large protein complexes.

1.3.1. GEMMA

GEMMA, or gas phase electrophoretic mobility molecular analyser, is a somewhat older variation of an IM instrument that may be used to study large protein assemblies. [Allmaier *et al.*, 2008; Kaddis & Loo, 2007] GEMMA allows the estimation of molecular mass, which is linearly correlated to the electrophoretic mobility diameter (EMD) for singly charged analytes. Therefore, GEMMA is not coupled with MS, although it generally utilises an ESI source for the generation of ions from protein complexes. The highly charged ions resulting from the electrospray process are introduced into a bipolar neutraliser which causes charge reduction, resulting in mostly neutral, some singly charged and a small amount of multiply charged species. This process is crucial for the determination of both the EMD and the mass. The charged particles

are then separated depending on their electrophoretic mobility in air (see Fig. 5(c) for the setup of the mobility analyser used in GEMMA). It is possible to analyse particles ranging from low nm to µm in diameter by this technique. Consequently, it has particularly been exploited in structural biology for the analysis of large protein complexes, such as the 20S proteasome and intact virus particles. [Kaddis *et al.*, 2007; Laschober *et al.*, 2008; Loo *et al.*, 2005] GEMMA has a much lower mass resolution than mass spectrometry, but does have the advantage that the analysis of particles with a molecular weight greater than 50 MDa is possible, and it is thus a useful contribution to the arsenal of techniques for gas phase structural analysis. [Kaddis & Loo, 2007]

1.3.2. Travelling voltage wave ion mobility mass spectrometry

Travelling voltage wave ion mobility mass spectrometry (TWIMS) is a recent development which has been incorporated into a commercial instrument consisting of an ion mobility unit integrated in a Q-ToF MS. [Giles *et al.*, 2004; Pringle *et al.*, 2007] RF ion guides are often used to stop the radial spread of ions in regions of mass spectrometers where the pressure is relatively high. In TWIMS, a SRIG is used, whereby opposite phases of RF voltage are applied to adjacent ring electrodes, resulting in a radially-confining effective potential well. [Giles *et al.*, 2004] A DC potential pulse is superimposed on the RF on sequential electrodes, producing a moving electric field or so-called T-wave, which propels ions axially through the centre of the electrodes. Ions 'surf' on the wave, moving down the potential gradient caused by the pulse into a lower field region. Then, when the pulse moves forward this is repeated, the ions are driven forward and transmitted through the cell (Fig. 5(d)). Since the ions are given an additional axial velocity by the T-wave, their transit time is reduced. Transmission can be optimised by altering the velocity of the wave and amplitude of the voltage pulse, as well as the gas pressure. [Giles *et al.*, 2004]

The T-wave SRIG can be used as a collision cell, a transport or storage device, for example to enhance the duty cycle of an orthogonal ToF mass analyser, or as an ion mobility device. The advantages of the T-wave and optimal settings for each of these uses have been extensively described by Giles *et al.* [Giles *et al.*, 2004] It is worth mentioning here that in the final instrumental setup, the so-called Tri-wave technology is implemented, meaning that three consecutive SRIGs, a collision and storage cell (trap), an ion mobility cell, and an ion delivery device or second collision cell (transfer), all operate with the T-wave principle. Fig. 5(e) shows a schematic representation of the T-wave as implemented in a Synapt HDMS instrument.

It is possible to use the T-wave SRIG as an ion mobility cell at pressures above approximately 0.2 mbar, and with optimisation of the wave properties for the specific analysis required. [Giles *et al.*, 2004] In this case, those ions with lower mobility cannot keep up with the velocity of the T-wave, and roll over the potential wave crest a number of times. The more mobile ions experience few of these 'roll-over' events, and thus have the fastest transit time through the TWIMS cell. The presence of one T-wave SRIG before and after the IM region allows CID to be carried out before or after IM (or both) to probe, for instance, the energetics of different protein conformations or the composition of protein complexes. [Giles *et al.*, 2004; Pringle *et al.*, 2007] If the T-wave SRIGs are required to function as a collision cell or ion delivery device, the gas pressure is normally around 3×10^{-2} mbar. [Lorenzen *et al.*, 2008; Ruotolo *et al.*, 2007; Ruotolo *et al.*, 2008; Uetrecht *et al.*, 2008b; van Duijn *et al.*, 2009] At this pressure, there is not enough

friction to separate ions according to their mobility. Instead, all the ions surf on the T-wave, reducing the transit time of the ions and thus allowing fast mass scanning. [Giles *et al.*, 2004]

In initial experiments to test the TWIMS ion mobility device, a mixture of gramicidin S and leucine enkephalin, peptides of 1,141 Da and 555 Da, respectively, was used. [Giles *et al.*, 2004] This study showed that the ion mobility separation efficiency was dramatically influenced by the gas pressure in the device. Effective separation was only achieved at relatively high pressures for these peptides. However, this introduced the necessity for a compromise in terms of conformational resolution, since high pressures were also seen to reduce the transmission of the ions. That TWIMS has the potential to separate singly charged monomers and multiply charged oligomers (of different stoichiometry) present in a sample was also clearly shown by the subsequent analysis of bradykinin, a nonapeptide of 1,060 Da.

One effect of applying a T-wave to the ion mobility cell is that the shape of a protein cannot be measured directly from the transit time of the ions. Instead, data calibration is required, which can be somewhat cumbersome since only a few calibrants are available (*i.e.* proteins or other species for which the Ω is accurately known). This will certainly improve as more Ω values are published on a broader range of samples. A detailed protocol for calibrating ion mobility data acquired on T-wave instruments has been reported by Ruotolo *et al.*, [Ruotolo *et al.*, 2008] and is currently the most general and widely used approach.

Shvartsburg and Smith [Shvartsburg & Smith, 2008] recently described the fundamentals of TWIMS, which bears relevance to the problem of calibration. A clear explanation and overview of the foundations of the T-wave technology were given, using a priori theory and numerical simulations of ion dynamics. As the authors stated, understanding of the T-wave method provides the basis for rational instrument design and operation, as well as more accurate and reliable IMMS data. This, of course, is essential for confident and correct ion structure characterisation. Finally, a relationship between drift time and Ω was introduced, to allow the Ω to be assessed directly from TWIMS data. Unfortunately, the authors ascertained that this relationship can only be applied to small analytes, meaning that the Ω of a protein or protein complex still has to be derived by calibration.

1.3.3. General applications of TWIMS

Since its development in 2004, applications of the Tri-wave technology have been explored widely. Thalassinos *et al.* [Thalassinos *et al.*, 2004] showed that TWIMS could be used to probe the Ω of proteins, providing results that were in good agreement with cross sections in the literature (*i.e.* those derived from other techniques). Various applications have been published to date, ranging from reducing the complexity of protein and peptide samples, to the structural analysis of large proteins and entire functional assemblies. Here, an illustrative selection of these articles is discussed, dealing primarily with protein structure elucidation, that clearly display some of the key advantages of IMMS over conventional mass spectrometry (and tandem mass spectrometry) studies.

In order to validate the TWIMS approach, several studies have compared the technique to conventional drift tube IM. The analysis of a peptide mixture by a TWIMS device revealed a similar pattern of separation to that obtained by a normal SRIG mobility device, however, the degree of separation actually increased significantly. [Pringle *et al.*, 2007] On the other hand, the resolution of different protein conformations can be lower when using TWIMS, but the

detected features are still comparable to drift tube IM data (Fig. 6(b)). An important quality of IMMS is the ability to separate isomeric species. In a normal mass spectrum this would not be possible, as the species appear at identical *m/z* values. Although the TWIMS instrument offers only moderate resolution, it does provide the potential to distinguish between the isomeric Gly-Arg-Gly-Asp-Ser and Ser-Asp-Gly-Arg-Gly peptides that differ by only 5% in Ω. [Pringle *et al.*, 2007] Overall, the mobility separation characteristics of the TWIG are similar to conventional drift tube devices. Even though the resolution of the ion mobility cell is relatively low, it has a high transmission efficiency and separation power. Taken together, these qualities result in an instrument that is very promising for the characterisation of biologically relevant systems.

An example of the importance of IMMS as a separation method is its use in conjunction with another popular method in mass spectrometry based structural biology, hydrogen deuterium (H/D) exchange (Box 1). H/D exchange MS has proved very valuable in the study of protein conformations and dynamics, and is often combined with proteolysis followed by liquid chromatography (LC) to separate the numerous peptides generated. [Sinz, 2007] However, compromises for the LC separation are essential to prevent high levels of deuterium back-exchange, a long-standing issue in H/D exchange MS. As a consequence, the LC separation is not optimal and peptides often co-elute. To circumvent this problem, H/D exchange has been successfully combined with an LC-IMMS approach to further reduce the complexity of the sample, by using IM as a gas phase separation method at the peptide level. [Iacob *et al.*, 2008] It was found that the incorporation of deuterium does not significantly affect the drift time of either proteins or peptides in TWIMS; additionally the level of back exchange was comparable with that observed using a standard Q-ToF instrument. The extra dimension provided by IM allows the efficient separation of overlapping isotopic patterns of co-eluting peptides present in a complex digest (see also Fig. 3). Thus, ion mobility separation increases the amount of information available from a single experiment in H/D exchange LC-MS.

The analysis of human haemoglobin (Hb) illustrates the benefits of the separation power of IMMS in relation to human health. Many variants of this protein are known to exist and are generally caused by a single base mutation in one of the corresponding genes. The result is

Box 1: H/D exchange MS

This approach is comparable to NMR experiments in which the exchange of deuteriums (D) to hydrogens (H) in a protein is monitored over time. Generally, hydrogens in the amino acid side chains are replaced very fast, therefore only effects on the exchange rates of amide hydrogens in the protein backbone are generally monitored. The speed of deuterium incorporation depends on the protein or protein complex structure. Amide hydrogens involved in secondary element or contact formation are protected against exchange due to hydrogen bonding relative to those in more flexible regions or exposed to the surface. [Sinz, 2007; Marcsisin & Engen, 2010]

A native protein or protein complex is diluted into deuterated water (D_2O) solution and the amount of deuterium can be determined as mass increase in MS at various points in time. Using native MS as readout provides information on major conformational changes, for example upon ligand binding. [Alverdi *et al.*, 2008]

To localise structural changes in the protein or protein complex, the samples from different points in time are diluted into acidic conditions (~pH 2.5) and kept close to 0 °C. This slows down the back exchange to hydrogen in the H_2O based solution. This also leads to unfolding of the protein increasing the accessibility towards proteases. However, few proteases can efficiently digest proteins at low pH. Pepsin has an optimum activity at low pH and is the preferred enzyme for H/D exchange. After digest, the obtained peptides are subjected to LC-MS analysis and identified by exact mass and fragmentation pattern. The mass shift from deuterium incorporation can then be monitored over time, elucidating which peptides are engaged in structure formation. To limit back exchange, the whole sample processing including LC separation has to be performed fast. Even though pepsin is regarded an unspecific protease, the detected peptides are highly reproducible. [Sinz, 2007]

either a structurally abnormal Hb or a reduced rate of globin chain synthesis. To date, over 1,000 Hb variants have been identified, many of which are disease related. [Hardison *et al.*, 2002; Williams *et al.*, 2008] Mutations in the globin chain can be identified by a mass spectrometric method. [Wild *et al.*, 2001] The screening approach is quite straightforward, but sometimes fails due to sample complexity. There is a high chance that one variant has (nearly) the same mass as another variant. Williams *et al.* [Williams *et al.*, 2008] showed that, in this case as well, TWIMS can reduce the sample complexity, thereby increasing the success rate of Hb variant identification. This achievement has significant clinical potential as it allows the unambiguous determination of Hb variants in a patient, which is essential to prescribe the right treatment.

The potential of IMMS for the analysis of proteins related to human diseases has also been clearly shown in studies relevant to glycosylation disorders. [Vakhrushev *et al.*, 2008] Glycoproteins and glycopeptides are engaged in many important biological processes, for instance cell differentiation and cell-cell interactions. The human urine glycome (glycourinome) is a very heterogeneous population, containing many different and often isomeric carbohydrate variants. [Vakhrushev *et al.*, 2006] Consequently, identification of oligosaccharides and glycopeptides from the glycourinome is complicated even with high-resolution mass spectrometry. Many ions overlap in *m/z*, hampering the chance of correct assignment. It has been shown that IMMS can be used to reduce the complexity of the diverse urine glycome mixtures, making sequence assignments of individual glycotargets in the sample possible. [Vakhrushev *et al.*, 2008] As structural profiling of glycotargets is important for biomarker discovery and diagnosis of human diseases, IMMS may become a helpful tool in this area of research.

Williams and Scrivens have described the coupling of desorption electrospray ionisation and neutral desorption/extractive electrospray ionisation to TWIMS for the direct accurate mass analysis of active ingredients formulated into pharmaceutical samples. [Williams & Scrivens, 2008] Here, the collision cross sections of poly(ethylene glycol) as determined by TWIMS were found to be in good agreement with earlier published data. This technique can also be used to facilitate the detection of trace impurities in drug products. [Eckers *et al.*, 2007] Although we will not elaborate any further on the application here, its importance to the pharmaceutical industry is obvious.

1.3.4. TWIMS in studying the conformational space sampled by single proteins

As TWIMS cannot yet be used to determine a Ω directly, the important question arises whether the value obtained from calibrating TWIMS data is comparable to the Ω determined by drift tube IM or derived from an X-ray or NMR structure. To address this question in detail, several groups have investigated native and unfolded proteins with TWIMS. [Pringle *et al.*, 2007; Thalassinos *et al.*, 2004]

As an example, TWIMS was used to study conformers of cytochrome C and the amyloidogenic β_2m, [Smith *et al.*, 2007] both of which had been the subject of IMMS studies previously. [Badman *et al.*, 2001; Borysik *et al.*, 2004; Mao *et al.*, 1999b] This former knowledge provided a good basis from which to illustrate the potential of TWIMS for unravelling protein folding pathways and protein misfolding processes. The data generated were in excellent agreement with previous studies; [Smith *et al.*, 2007] an example is given in Fig. 6. Again, the existence of multiple native-like conformations was found, and for cytochrome C, more extended structures as well.

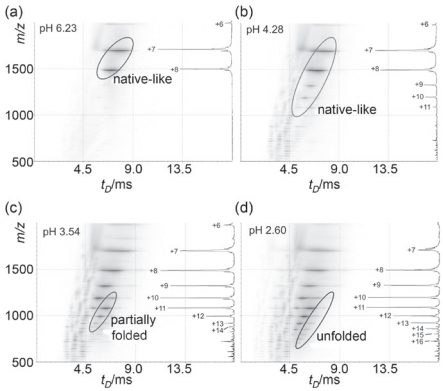


Fig. 7: Protein folding in solution and the gas phase TWIMS can monitor unfolding of proteins in solution, as shown here for β_2m upon acidification. At near neutral pH, only the native-like conformation is present in the drift time (t_D) to *m/z* contour plot (a) and the summed mass spectrum (inset, charge states are indicated). From pH 4.28 (b) to pH 2.60 (d), ions with a higher number of charges are observed, indicating unfolding of the protein. At pH 3.54, a species with a different structure, *i.e.* a slower drift time distribution, is populated corresponding to a partially folded intermediate (c). Unfolded β_2m forms at very low pH (d). These structural transitions have implications for amyloid formation, as (partial) unfolding is thought to be a prerequisite for aggregation. [Smith *et al.*, 2007] (Reprinted from [Smith *et al.*, 2007], with permission from Elsevier)

In the case of β_2m, earlier work had revealed the presence of three different protein structures, namely the native, partially folded and acid-unfolded β_2m, whose abundance could be influenced by changes in the pH. [Borysik *et al.*, 2004] TWIMS was therefore used to investigate the structural heterogeneity of β_2m under different conditions. [Smith *et al.*, 2007] At physiological pH, only fully folded β_2m was present, which is not able to form amyloid fibrils. Upon acidification, both the charge state distribution and the ion mobility plot were affected (Fig. 7). Higher charge states indicated first partial and then complete unfolding of β_2m, which was confirmed by the increased collision cross sections derived from the IMMS analysis. In addition, at moderately acidic pH (4.8), the ion mobility data revealed a second native-like conformer with very little difference in collision cross section, but with a higher propensity for protonation. In total, four conformer 'families' were detected: two native-like states, a partially unfolded and a fully unfolded β_2m species. Both the partially folded and unfolded β_2m may be involved in fibril formation and aggregation. The ability of TWIMS to resolve the different protein conformations of β_2m further illustrates the potential of this technique to provide insights into protein folding and heterogeneous assembly reactions in a manner comparable to conventional (drift tube) IMMS.

Evidently, the three dimensional structure of a protein is crucial for its biological function. In order to validate whether TWIMS can be applied to such studies with the same confidence as conventional IMMS, estimated Ω values for numerous proteins have been compared to values derived from published X-ray crystallography and NMR structures, and were in very good agreement. [Scarff *et al.*, 2008; Smith *et al.*, 2009] Therefore, this again underlines that TWIMS can provide information about three dimensional structures of biological relevance. However, for the most accurate collision cross sections, the TWIMS conditions require careful optimisation for each individual protein. A full calibration has to be performed for each set of instrumental parameters, reinforcing the fact that this is not yet a high throughput method for structural determination of proteins and protein complexes.

1.3.5. Structural and biophysical characteristics of protein complexes analysed by IMMS

Of all the applications of TWIMS, one of the most exciting and innovative is the determination of structural and biophysical characteristics of large protein complexes. Many native MS reports have strongly suggested that solution phase structures of protein complexes can be partially retained in the gas phase, but conclusive proof was still largely missing. [Benesch *et*

al., 2006; Benesch & Robinson, 2006; Synowsky *et al.*, 2006; van Duijn *et al.*, 2005; van Duijn *et al.*, 2006] Supporting evidence had come from such observations as the fact that trypsin and lysozyme ions can still be biologically active after soft-landing of the ions on a detector following their trajectory through the vacuum of a mass spectrometer. [Ouyang *et al.*, 2003; Takats *et al.*, 2004] Drift tube IM frequently detects compact structures of single proteins, with similar dimensions as the native state found in solution. In 2005, Robinson's group first used ion mobility with native mass spectrometry to study a large functional, multimeric protein assembly. [Ruotolo *et al.*, 2005] Most importantly, they provided one of the first pieces of direct evidence that the structure of protein complexes can remain native-like within the high vacuum of a mass spectrometer. Thus, T-wave IMMS provided definitive results that the quaternary topologies of protein complexes can be maintained in the absence of solvent. [Ruotolo *et al.*, 2005]

The oligomeric tryptophan RNA binding attenuation protein complex (TRAP) is an undecameric assembly of approximately 88 kDa, which forms a ring-like structure in solution. [Antson *et al.*, 1999] TRAP is a regulatory protein in bacteria and binds to RNA to facilitate the abortion of transcription. The affected gene products are involved in tryptophan biosynthesis, whereby tryptophan inhibits its own production by interacting with TRAP. The ring topology and oligomeric state of the TRAP complex made it an ideal assembly for analysing structural behaviour in the gas phase, and investigating the potential of protein complex TWIMS. Three different conformations of the TRAP complex were revealed, assigned to a native ring-like structure, a more compact intermediate assembly and a collapsed structure. Whereas the lower charge states mainly represented the native ring-like TRAP complex, at increasing charge states, compact intermediate and collapsed undecameric TRAP conformations were increasingly populated. Combining these observations with a molecular modelling approach, model structures for each of the TRAP complex conformations could be assigned. Tryptophan binding to the TRAP ring resulted in a Ω similar to the apo TRAP complex, with the binding of tryptophan increasing the stability of the ring-like conformation. The subsequent binding of RNA to the holo complex stabilised the TRAP ring even further, in good agreement with previous X-ray crystallography data. [Antson *et al.*, 1999; Ruotolo *et al.*, 2005]

Coulombic repulsion generally has less influence on protein ions with a low number of charges, compared to more highly charged ions. The most native or "compact" structures, therefore, are typically observed for protein ions with fewer charges, whereas higher charge states of the same protein generally take on more extended conformations. [Catalina *et al.*, 2005; Kaltashov & Abzalimov, 2008; Ruotolo *et al.*, 2007; Sun *et al.*, 2007] Interestingly, the TRAP ring exhibited more compact structures at higher charge states. A possible explanation for this is that at higher charge states, the TRAP complex will undergo more energetic collisions with neutral gas molecules, resulting in a higher internal energy of the ions. This relatively high internal energy can induce changes in the TRAP complex structure, whereby the ring topology can relax and collapse. This effect may be limited to certain systems, as it has since been shown that not all ring structures undergo a structural collapse in the gas phase at higher charge states. [Lorenzen *et al.*, 2008; van Duijn *et al.*, 2009] Overall, this pioneering work on the TRAP complex clearly demonstrates the ability of IMMS to determine the general topology of a protein complex and to monitor different structures simultaneously. Furthermore, it has provided direct evidence that the structure of an oligomeric protein assembly can remain intact in the gas phase. However, careful interpretation of IMMS data is always essential, as native and non-

native structures can co-exist depending on the ion's charge state and the instrumental settings. [Ruotolo *et al.*, 2005]

The largest structures analysed by TWIMS to date are those of viruses such as the Hepatitis B virus (HBV) capsids. [Uetrecht *et al.*, 2008b] This virus can cause severe liver diseases and cancer in humans. Additionally, there is high interest in applying viral capsids as vectors for drug delivery or cell transfection, as reaction chambers, and for other nanotechnology purposes. A precise biophysical characterisation of these large macromolecular particles is therefore of fundamental interest. Interestingly, HBV can form, both *in vivo* and *in vitro*, icosahedral capsids of different sizes, consisting of either 90 ($T = 3$) or 120 ($T = 4$) dimers, with total masses of 3 or 4 MDa, respectively. [Zlotnick *et al.*, 1996] Although these stoichiometries have long been assumed, only accurate mass determination from native mass spectrometry has confirmed them unambiguously (see chapter II). [Uetrecht *et al.*, 2008a] The shape of the HBV capsids in the gas phase was also studied by a TWIMS approach (see chapter III). Strikingly, these analyses revealed that both $T = 3$ and $T = 4$ capsids populated at least two distinct conformers (Fig. 8). [Uetrecht *et al.*, 2008b] As the ratio between the two conformers of both icosahedral capsids was not affected by the acceleration voltage or charge state, it was suggested that they originated from the ionisation process. Uetrecht *et al.* [Uetrecht *et al.*, 2008b] hypothesised that the capsids most likely adopt isoenergetic states prior to ionisation and become trapped in one of these upon desolvation, leading to two distinct conformations for both $T = 3$ and $T = 4$ capsids. The IMMS analysis, with calibration, could be used to determine the Ω of the capsids, which were in excellent agreement with radii determined from cryo-EM reconstructions. [Zlotnick *et al.*, 1996] This information suggests that the hollow capsid structures can be largely retained in the gas phase. Moreover, these studies also revealed that TWIMS is capable of transmitting large protein complexes of up to several Mega-Dalton.

Even before the introduction of IMMS to the field, native mass spectrometry enabled the determination of the stoichiometry, stability, and topology of a protein complex, and in some cases, also provided information on quaternary structure and protein complex dynamics. [Benesch *et al.*, 2007; Heck & van den Heuvel, 2004; Sharon & Robinson, 2007; van den Heuvel & Heck, 2004] The use of tandem MS in native MS has been especially important for its applications within the structural biology field. When subjecting a protein complex to tandem MS, the complex is energetically activated and eventually dissociated, typically by CID. The kinetic energy of the collision gas is transferred into internal energy of the protein complex in numerous collision events. When a certain threshold is attained, the internal energy is sufficient for the complex to fall apart, which typically occurs in an asymmetric manner in CID. [Benesch *et al.*, 2006; Benesch *et al.*, 2007; Benesch, 2009; Jurchen & Williams, 2003] From a multimeric assembly, single subunits tend to dissociate individually. The dissociating subunit obtains a high number of charges before expulsion from the assembly, leaving, by contrast, the remainder of the initial complex with relatively few charges.

It has been suggested that this asymmetric dissociation process is largely driven by Coulombic repulsion, [Benesch *et al.*, 2006; Benesch *et al.*, 2007; Jurchen & Williams, 2003] and that prior to dissociation, partial protein unfolding of the ejected monomer occurs. This has most clearly been shown by studies on cytochrome C and α-lactalbumin dimers. [Jurchen & Williams, 2003; Jurchen *et al.*, 2004] These protein dimers showed clear asymmetric charge state

distributions upon CID. However, when the conformation of the monomers within the complex was restricted by introducing disulfide cross-links, dissociation still occurred but now both monomers obtained an equal number of charges. [Jurchen & Williams, 2003; Jurchen *et al.*, 2004]

Obviously, IMMS provides unique opportunities to visualise the potential unfolding of a subunit within a complex before it is expelled from the assembly. Tetrameric transthyretin (TTR) was used to test the current model for protein complex dissociation. TTR is an abundant protein in human plasma, and can form fibrillar protein deposits associated with amyloidosis, implicated, for example, in neurodegeneration. In particular, two aggregation prone point mutants have been identified. Upon analysis by TWIMS, it was seen that the TTR complex, subjected to energetic activation, populated partially folded intermediate states, in which one or more subunits showed unfolding behaviour. [Ruotolo *et al.*, 2007] The IM drift time distributions were seen to broaden, towards larger tetrameric TTR structures, upon complex activation. Molecular modelling (see section 1.4) confirmed that partial unfolding of complex subunits was occurring before dissociation, in agreement with the current model for gas

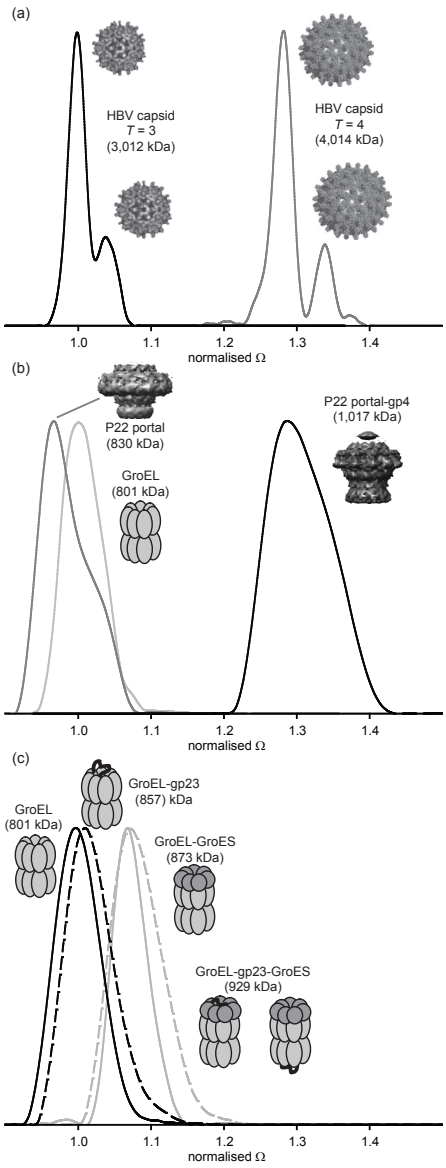


Fig. 8: Ion mobility on protein complexes
(a) Different conformations of protein complexes can be separated by ion mobility. The two geometries of the Hepatitis B capsid ($T = 3$ and $T = 4$) both show two distinct conformations in the gas phase. The collision cross sections (Ω) are normalised on the small $T = 3$ to highlight the size difference. Absolute Ω for the capsids were: 687 nm^2 and 717 nm^2 ($T = 3$), 875 nm^2 and 914 nm^2 ($T = 4$), suggesting that the viral capsid structure is retained in the gas phase. [Uetrecht *et al.*, 2008b] (b) Conformational changes upon binding of gp4 to the portal complex of bacteriophage P22 monitored by TWIMS. The portal complex (medium grey) and the chaperone GroEL (light grey) are of similar mass, but the portal has a significantly smaller Ω indicating that it is rather compact (traces are normalised on GroEL for better comparison with (c)). The binding of gp4 changes the Ω drastically (>30%, black), whereas the mass increases by only 22%. Absolute Ω for the complexes were: 230 nm^2 (GroEL), 221 nm^2 portal and 286 nm^2 portal-gp4 complex. [Lorenzen *et al.*, 2008] The gp4 induced conformational change could also be confirmed by an electron microscopy study (structures are depicted). [Zheng *et al.*, 2008] (c) Conformations can remain unchanged when a protein substrate is bound in a cavity, as in the case of GroEL. The chaperone with the substrate gp23 (black dashed) and GroEL alone (black solid) exhibit essentially the same Ω, whereas adding the capping cochaperone GroES in absence or presence of the substrate results in a similar increase in Ω (grey solid and dashed, respectively). Evidently, the substrate does not contribute to the Ω and is buried in the central cavity of the chaperone. Ω are normalised on GroEL. Absolute Ω were: 244 nm^2 (GroEL), 247 nm^2 (GroEL-gp23), 261 nm^2 (GroEL-GroES) and 262 nm^2 (GroEL-gp23-GroES). [van Duijn *et al.*, 2009]

phase dissociation of protein complexes.

Mutations in the amino acid sequence of a protein can affect its conformation and consequently the stability of any complex of which it is a part. For example, an aggregation prone TTR variant dissociates more readily in CID than the wild type equivalent. However, TWIMS experiments in conjunction with CID showed that it still exhibited the same unfolding pathway, but that the dissociation of the variant complex occurred at a lower degree of unfolding. Interestingly, the structural stabilisation upon ligand binding was increased in the mutant compared to the wild-type complex. Together, these data indicate that decreased intersubunit contacts cause destabilisation in the variant, enabling dissociation with limited unfolding, and the ligand partially compensates the lowered binding affinity. The combination of tandem MS and IMMS experiments can therefore elucidate how complex stability is influenced by mutations. [Hyung *et al.*, 2009] In another example, IMMS has been shown to distinguish the different conformations of normal and sickle cell haemoglobin, which result from point mutations. [Scarff *et al.*, 2009]

In a similar way to the TTR study, TWIMS can be applied to monitor conformational changes in a protein complex upon ligand binding. It is well known that the structure of a protein can be affected by its interaction with other proteins, ligands, nucleotides *etc.* [Benesch *et al.*, 2007; Scholten *et al.*, 2008] For example, the activity and moreover the structure of the cyclic guanosine 3´,5´-monophosphate (cGMP)-dependent protein kinase (PKG) is significantly affected by cGMP binding. Protein kinases are ubiquitous and involved in a variety of cell signalling pathways. Relatively little structural data is available for PKG. It is known to form a 150 kDa homodimer and requires the binding of four cGMP molecules for activation. Small-angle X-ray scattering data indicated an expansion of the PKG structure upon cGMP binding, and Fourier transform infrared spectroscopy also suggested a large movement of PKG structural domains. [Zhao *et al.*, 1997] In a multiplexed mass spectrometric approach, the implications of cGMP binding on the structure of PKG were investigated. [Alverdi *et al.*, 2008] Global H/D exchange measurements, native mass spectrometry, localised H/D exchange analysis and IMMS studies were performed. Using TWIMS, the gas phase ion mobility of inactive and active (cGMP bound) PKG was analysed. This clearly showed a small but significant increase in Ω for cGMP bound PKG as compared to inactive PKG, which is consistent with a structural rearrangement in PKG to a more open conformation induced by cGMP. The observed shift in charge state distributions in native mass spectrometry, as well as the H/D exchange measurements, added further support to this hypothesis. Combination of all these data allowed a new, detailed structural model of the cGMP-induced activation of PKG to be proposed. [Alverdi *et al.*, 2008]

In a similar manner, a major conformational change in the first step of bacteriophage P22 tail assembly could be exposed. [Lorenzen *et al.*, 2008] In order for DNA to be packed into the head of the bacteriophage, a protein complex called the portal, which connects the head to the tail, is essential. The working mechanism of this molecular motor is dependent on the binding of proteins that are involved in the tail formation. However, the exact assembly mechanism of the P22 tail is still not completely understood. Lorenzen *et al.* used a native mass spectrometric approach, including TWIMS, to determine the stoichiometry of the portal, as well as to reveal a major conformational change upon association to the first tail accessory protein (gp4). Native MS and IMMS showed that the unbound form of the portal adopted a relatively compact structure, but on binding of gp4 to the portal, a more open conformation was induced, as indicated

by a significant increase in Ω (Fig. 8). This conformational change in the portal is thought to be the starting point for a cascade of events assisting in the stabilisation of newly filled P22 particles, which marks the end of phage morphogenesis. In strong agreement with these findings, a concurrent cryo-electron microscopy (cryo-EM) analysis also visualised this drastic structural rearrangement in the P22 portal upon binding of gp4. [Zheng *et al.*, 2008]

The structure of proteins largely determines their function *in vivo*. In the cellular context, some proteins cannot fold into their "correct" structure, instead requiring assistance from chaperones. GroEL-GroES is one of the best studied chaperone systems from *Escherichia coli*, but still the substrate folding process remains intriguing. [van Duijn *et al.*, 2005] Since the folding cycle is a dynamic process, it is very challenging to study the structure of the transient intermediates. Recently, IMMS analysis has been used to study the gas phase conformations of different chaperonin complexes. [van Duijn *et al.*, 2009] Ω values were determined for different chaperonin complexes and, for the first time, the dimensions of a ternary complex between GroEL, a substrate and GroES could be determined. The data unambiguously showed substrate burial in the GroEL cavity, which was retained in the gas phase, since the overall conformations of the substrate bound complexes were similar to that of the empty GroEL or GroEL-GroES (Fig. 8). [van Duijn *et al.*, 2009] For two GroEL assemblies, crystal structures are available and the calculated Ω showed good agreement with the experimental values, emphasising the fact that IMMS can reveal details of biological processes.

An interesting issue addressed in the above study is the influence of charge reduction on the global shape of a protein complex. To probe this, the chaperonin complexes were analysed by IMMS using imidazole, a charge-stripping base, as an additive in the ESI solution. On average the Ω values of the significantly charge-reduced complexes were approximately 5% smaller, indicating a more compact structure. These results likely reflect the decrease in Coulombic repulsion upon charge reduction, a phenomenon which at higher charge states causes protein complex ions to adopt more extended structures with larger collision cross sections (section 1.2.3.). [Ruotolo *et al.*, 2007] The charge state reduction was found not to affect the relative differences observed in Ω for the chaperonin assemblies. Importantly, this study revealed that the resolution observed in the TWIMS data could be significantly enhanced by analysing charge reduced species, allowing the separation of closely related conformers.

1.4. Molecular modelling

As modelling peptide structures in the gas phase has become more accurate over recent years, so has the process of protein structure determination recently become more integrative, combining molecular modelling with common techniques for structure determination. [Johnson, 2008; Sharon & Robinson, 2007; Szymczyna *et al.*, 2009] The huge progress in available computing power has made structure prediction of large proteins and complexes possible. However, the accuracy of the models obtained is still highly dependent on the input data, in terms of quality as well as quantity (*e.g.* atomic resolution or amount of related structures and conformational restraints). [Lee *et al.*, 2007; Moult, 2006; Taverner *et al.*, 2008; van Gunsteren *et al.*, 2008; Zhang, 2008]

In relation to this, ion mobility data can add constraints to structural modelling. In particular, the determined Ω can discriminate between different predicted models. This is especially useful where several structures with similar energy minima are available. IMMS is not a high

resolution structural method, but nevertheless, has the great advantages of low sample consumption and relatively fast analysis. Therefore, modelling can create a bridge from IMMS derived data on protein complex topology and overall shape to high resolution structural data at atomic levels. Generally, two strategies have been applied for the combination of IMMS data with computational approaches. The ion mobility data can be used to support *in silico* protein folding models or, alternatively, models can be used to explain and extend the information gained from IMMS. [Arteca *et al.*, 2001; Daggett, 2006] In particular, the dynamics and conformational changes associated with protein complexes are of special interest for the integration of IMMS and structure prediction. [Alverdi *et al.*, 2008; Lorenzen *et al.*, 2008; Noe & Fischer, 2008; Ruotolo *et al.*, 2007; Sharon & Robinson, 2007]

1.4.1. Modelling approaches

In general, the sequence of a protein determines its structure and thus its function. [Moult, 2006; van Gunsteren *et al.*, 2008; Zhang, 2008] Consequently, the most common way to obtain a structural model is by comparative or homology modelling. Firstly, a database search provides a list of related proteins with different degrees of sequence homology. Such proteins may adopt a similar fold. Next, available crystal and NMR structures are sampled in order to obtain a consensus structure, which is used as a starting model. This is then adapted and structurally refined to the protein sequence of interest. For proteins with high sequence similarity (above 30%), calculated models are generally correctly aligned across the sequence. Where target and template protein share less than 20% sequence similarity, structures are often misaligned, but the overall fold can usually still be determined. The problem of misalignments increases exponentially with the size of the protein. Generally, the modelled structure cannot be of higher accuracy than the best input or template structure. In order to increase the accuracy in comparative modelling, new refinement methods are required. [Ginalski, 2006] The accuracy and correctness of modelling techniques are assessed by joint initiatives like the CASP project (Critical Assessment of Techniques for Protein Structure Prediction); an extensive overview on CASP7 has been reported, [Trapane & Lattman, 2007] and information on CASP8 is currently available online. [Moult, 2008]

An alternative to homology modelling is the more complex approach of *de novo* modelling of protein structures. [Moult, 2006; Zhang, 2008] So far, even though the force fields used are reasonable, it has rarely resulted in accurate alignments and correct three-dimensional structures due to the multitude of possibilities. Furthermore, *de novo* modelling requires substantial computing power. In both comparative (homology) and *de novo* modelling, the computational resources necessary depend on the protein size. Small structures can be modelled in all-atom representation, while large proteins are typically modelled after coarse-graining, in which several atoms, for example those comprising each individual amino acid, are clustered to decrease the degrees of freedom. [Colombo & Micheletti, 2006]

Finally, an intermediate approach between comparative and *de novo* modelling is based on the theory that most of the possible substructures of proteins are already known. [Moult, 2006; Zhang, 2008] These substructures may consist of just a few amino acids, a secondary structure unit or a particular arrangement of secondary structure elements. All currently known and any yet to be discovered protein folds are expected to be combinations of such substructures. The various possible combinations of these defined structural fragments result in a large set of pos-

sible structures. Particularly, in these cases, the obtained models can be very distinct, and IMMS can then provide the constraints necessary to identify the most reliable structures.

1.4.2. Calculating collision cross sections from available or modelled structures

Before computational models can be compared with a Ω value derived from IMMS analysis, the theoretical Ω has to be calculated for each structure. For this purpose, several algorithms have been developed using different theoretical assumptions, [Clemmer & Jarrold, 1997; Scarff *et al.*, 2008; Shvartsburg *et al.*, 2001; von Helden *et al.*, 1993b; Wyttenbach *et al.*, 1997; Wyttenbach *et al.*, 2000] three of which are commonly used as well as being implemented in the publicly available mobcal program. [Mesleh *et al.*, 1996; Shvartsburg & Jarrold, 1996] First of all, the most reliable results are estimated by trajectory calculations. [Shvartsburg *et al.*, 1997] However, these are computationally expensive, as they employ realistic potentials and take into account multiple scattering events and long-range interactions between ion and buffer gas. At the other end of the spectrum, the hard sphere projection approximation (PA) is the most reductive method. [Mesleh *et al.*, 1996] It assumes spherical ions possessing the same average projection areas as the input structure, and thereby simplifies the scattering process, as multiple collisions of a gas molecule do not occur. This approach can be especially problematic for molecules with concave surfaces, and results for convex faces can also deviate significantly from experimental values due to topological roughness of the exterior. Other assumptions that influence the accuracy of the method are the uniform distribution of charges across the surface and the absence of long-range interactions between ions and the buffer gas atoms.

The exact hard sphere scattering (EHSS) determines more realistic mobilities, because the surface structure is taken into account. [Shvartsburg & Jarrold, 1996] Only slightly more computation time is required as compared to the PA method. Nonetheless, it is not very suitable for small molecules like carbon clusters, where direct ion-buffer gas interactions become important, as these are also neglected in this algorithm. This compromise arises from the need for short computation times.

As stated above, both PA and EHSS ignore long-range interactions, which gain importance with increasing size of the molecules. [Mesleh *et al.*, 1996] Moreover, concave and rough surfaces are typically observed in large molecules such as proteins. Therefore, PA is the least suitable method for proteins, and it is generally advised that, where possible, the trajectory method should be used. [Shvartsburg *et al.*, 1997] For 'medium–sized' ions, it is highly recommended, especially if the experimentally-derived Ω is to be used to discriminate between different structural models of a particular molecule.

Notably, all three methods (trajectory, PA, EHSS) assess the Ω with helium as the *in silico* buffer gas, as this is the common gas used in experiments on smaller analytes. Today, for proteins and protein complexes, other gases, such as nitrogen or larger rare gases, are also infused into the mobility chamber. [Ruotolo *et al.*, 2005; Uetrecht *et al.*, 2008b] As mentioned above, ion-buffer gas interactions are important for the determination of the Ω. Thus, one might argue that the application of different gases exhibiting discrete polarisabilities hampers the comparison of a modelled and an experimentally derived Ω. However, a study by Chen *et al.* proved no substantial influence of ion-induced dipoles, at least for proteins. [Chen *et al.*, 1997] The measured Ω was nearly indistinguishable for neon, argon and krypton, with required polarisation forces decreasing sequentially from helium to krypton.

There are currently several publicly available programs for Ω calculation (see Table 1 for an overview). For example, Sigma from the Bowers group utilises both EHSS and PA calculations, although it excludes the trajectory method. [von Helden *et al.*, 1993b; Wyttenbach *et al.*, 1997; Wyttenbach *et al.*, 2000] The Jarrold group released the mobcal program, which handles all three of these theories. [Mesleh *et al.*, 1996; Shvartsburg & Jarrold, 1996] Mobcal is widely used in IMMS analyses of proteins. [Ruotolo *et al.*, 2008] It utilises data from the protein data bank (PDB) as input, although the structural coordinates must first be converted into the appropriate data/file format. Interestingly, the program requires and calculates electron densities from the atomic coordinates, however presently fails to incorporate electron microscopy data, essentially representing such electron distributions. Such an inclusion would be particularly helpful as relatively more EM than X-ray data is available for some of the larger heterogeneous protein assemblies currently being targeted by IMMS.

Scarff *et al.* (2008) recently showed that a protein Ω as determined with a T-wave instrument is usually intermediate to the PA and EHSS results calculated from published X-ray and NMR structures. [Scarff *et al.*, 2008] Clearly, this shows the applicability of computational modelling even in combination with an instrument that does not allow the direct measurement of Ω. These results have been confirmed by others. [Jarrold, 2000; Ruotolo *et al.*, 2005; Ruotolo *et al.*, 2007; van Duijn *et al.*, 2009] As previously described in section 1.3.5, a close agreement between IMMS-derived Ω and radii from EM could be established for HBV capsids. [Uetrecht *et al.*, 2008b] The IMMS results can be further validated by mobcal calculations based on the $T = 4$ HBV crystal structure, with a high correlation seen between molecular modelling and experimental data (907 nm^2 (PA) and 1258 nm^2 (EHSS) compared to 869 nm^2 and 909 nm^2 in IMMS). In this case, the PA value fits extraordinarily well with the IMMS results for the larger conformation, which is likely due to the spherical nature and relatively smooth surface of the capsids.

Recently, the use of Monte Carlo simulations was introduced to determine the Ω from a known structure. [Smith *et al.*, 2009] This "Leeds" algorithm applies a straightforward statistical approach to estimate the area in which the buffer gas could collide with protein ions, and thus is similar to the PA method. In addition to the low amount of computation time required, the program can handle PDB files directly without the necessity of data conversion, thereby speeding up the entire calculation process. From the data presented thus far, this new algorithm shows great potential, as in most cases the experimental Ω closely matched the calculated values.

Altogether, there is a wide range of computational methods for determining the cross section of either directly modelled protein structures, or those available from the PDB. In the next section, we move on to discuss how this repertoire of techniques can be exploited in conjunction with IMMS data, highlighting how protein structure determination can benefit in the future from integrating computational modelling and ion mobility.

Tab. 1: List of programs to calculate collision cross sections from atomic structures.

Name	Applied models	Authors	Source
Leeds algorithm	Monte Carlo algorithm similar to PA	Smith *et al.* [Smith *et al.*, 2009]	http://www.astbury.leeds.ac.uk/
Mobcal	PA, EHSS, trajectory method	M.F. Jarrold *et al.* [Mesleh *et al.*, 1996; Shvartsburg & Jarrold, 1996]	http://www.indiana.edu/~nano/
Sigma	PA	M.T. Bowers *et al.* [von Helden *et al.*, 1993a; Wyttenbach *et al.*, 1997; Wyttenbach *et al.*, 2000]	http://bowers.chem.ucsb.edu/

1.4.3 The integration of molecular modelling and IMMS

As described in section 1.2.3, the unfolding and refolding behaviour of proteins in the gas phase has been analysed in some detail by IMMS. For several systems, similar features could be observed upon acidic unfolding in solution and gas phase activated unfolding. [Arteca *et al.*, 2001; Jarrold, 2000] Molecular dynamics simulations have been employed in order to understand how the protein structure opens up upon unfolding. Instead of directly fitting the IMMS data, these calculations were run independently. Even so, the distinct features observed in IMMS could be reproduced by the modelled data, showing loss of tertiary structure until almost fully extended protein threads were formed. [Alonso & Daggett, 1998; Mao *et al.*, 1999a] Moreover, simulations have also reproduced the global collapse often observed in gas phase refolding of proteins, and additionally indicated that most of the contacts formed are likely to be non-native.

In general, proteins exhibit multiple conformations for a range of charge states upon unfolding. Typically, multiple and very distinct structures are present at a critical charge. This broad structural ensemble suggests a smooth minimum energy landscape around this specific charge with low energy or activation barriers, where the Coulombic repulsion is still mostly compensated by attractive interactions. [Arteca *et al.*, 2001] For larger proteins, IMMS peaks tend to be broad compared to those relating to peptides. This is likely to be due to the presence of ensembles of closely-related conformations, possibly caused by differences in the distribution of charges across the protein chain. The main advantages of gas phase folding studies above those in solution are the absence of solvent and the shorter time frame necessary for the transitions to occur. These two factors largely reduce complexity and thus enable all-atom simulations to be performed. Valuable information about protein folding can therefore be gained at relatively low expense.

Molecular dynamics simulations are limited with respect to the size of the protein or protein complex of interest. However, in combination with other techniques, IMMS can also be used to obtain a global picture of the native structure of large protein complexes. Structural information on large complexes is often difficult to obtain by common biophysical techniques such as X-ray crystallography and EM. However, the structural insight available from lower resolution EM data is greatly enhanced by knowing the exact stoichiometry, dimensions and approximate position of the individual subunits in the assembly. This information and more, for example, regarding the dynamics of protein complexes, can be provided by mass spectrometry and IMMS. [Heck & van den Heuvel, 2004; Heck, 2008; Sharon & Robinson, 2007]

IMMS can enhance our knowledge not only of the shape of an entire protein complex, but also of individual folded subunits when the purified proteins are analysed separately. Molecular replacement and docking procedures can then be used to build up a protein complex structure, applying the constraints determined by IMMS experiments. When high resolution structures of the components are not available, the correct homology models can be selected, as described above. Recently, the integration of molecular modelling, IMMS and EM enabled the deduction of the organisation of the 19S proteasome lid and the yeast exosome (the common workflow is depicted in Fig. 1). [Taverner *et al.*, 2008] However, the complete proteasome lid could not be modelled as there were no atomic structures of homologous proteins deposited in the PDB. Some of the more sophisticated modelling strategies presented, especially *de novo* modelling, may be useful in such situations, although this requires more computation power and combined

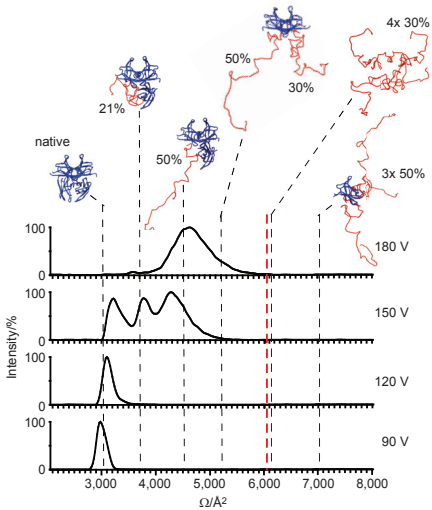


Fig. 9: Disassembly of protein complexes in the gas phase monitored by IMMS
Upon collision induced dissociation, individual subunits in a protein complex are thought to unfold prior to ejection. Indeed, ion mobility analysis of collisionally activated tetrameric transthyretin reveals an increase in collision cross section (Ω) when raising the activation voltage above 120 V. At 90 V the observed peak corresponds well to the Ω calculated for the native protein complex (blue, dashed line). The crystal structure is shown in blue. Several structures were modelled including one or more subunits with different degrees of unfolding (red) in the complex. The Ω peak at 180 V agrees well with one 50% unfolded subunit attached to the complex. The red dashed line, which also marks the end of the distribution at 180 V, corresponds to a completely unfolded monomer docked to the complex. From [Ruotolo *et al.*, 2007], © Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.

input from experimentalists and computational biologists.

Another field where IMMS and molecular modelling have come together is in the analysis of protein complex structures in the gas phase, for example upon collisional activation of the protein complex inside the mass spectrometer. [Ruotolo *et al.*, 2007] This problem is related in some respect to unfolding studies on single proteins. In 2005, it was shown that the ring topology of the TRAP complex is retained *in vacuo* for low charged ions, but higher charge states collapse, as discussed in detail above. [Ruotolo *et al.*, 2005] A structural collapse could also be observed upon collisional activation of the complex. In general, complexes dissociate *via* sequential subunit loss when sufficiently activated in the gas phase. This suggests that, in the case of TRAP, the applied energy first relaxes the structure leading to a collapsed conformation, which presumably involves non-native contacts and partially misfolded subunits. Upon further activation, the subunits start to unfold and are finally released from the complex, as has been shown for the tetrameric TTR complex. [Ruotolo *et al.*, 2007] The more globular TTR does not collapse prior to dissociation, but rather expands. Modelling partially unfolded subunits and docking them to the complex structure gave rise to the model that was in best agreement with the IMMS data. This suggested that one subunit unfolded almost completely, whereas the other three mostly retained their native fold, as illustrated in Fig. 9. [Ruotolo *et al.*, 2007] An interesting question is whether or not larger complexes behave similarly, especially the more heterogeneous protein assemblies. Here, the possibilities to distribute the energy are significantly increased and several subunits may start losing their native-like structure simultaneously. Clearly, more studies are required to fully understand gas phase activation and dissociation of protein assemblies.

In order to study the assembly of protein complexes or aggregation as observed in amyloids, the sensitivity of MS, in general, has been exploited for the identification of intermediates. [Bernstein *et al.*, 2005] A common problem in such analyses is the discrimination between on- and off-pathway oligomers. For this purpose, substructures of the complex can be modelled and compared with the Ω estimated by IMMS. Off pathway intermediates may have distinct structures which disagree with the models. In addition, some unspecific oligomers of certain proteins, *i.e.* non-functional aggregates, could potentially be identified in this fashion. However, it remains a question whether the resolution in IMMS is high enough to detect such differences.

It is by now well established that large conformational changes often occur upon ligand binding or complex formation, and these can be observed and identified by IMMS. [Alverdi *et al.*, 2008; Lorenzen *et al.*, 2008; van Duijn *et al.*, 2009] With respect to drug development in particular, it is important to precisely locate structural transitions in a protein. [Lee *et al.*, 2007; Totrov & Abagyan, 2008] IMMS provides constraints that could be used to discriminate between different models based on related structures, and thus define a conformational change. Other mass spectrometric techniques that can add further insights, especially into conformational changes and protein-protein interfaces for molecular modelling, are H/D-exchange and cross-linking; these can additionally be coupled with IMMS for multi-dimensional analyses. [Gingras *et al.*, 2007; Sinz, 2007] In summary, the inclusion of modelling and other computational techniques within IMMS studies can significantly enhance the amount of information that can be gained. Such combination allows the inference of high resolution structures from relatively low resolution ion mobility data. IMMS is thus a competitive structural biology technique, providing accurate, relevant and unique data.

1.5. Appraisement and future prospects

From the plethora of studies described in this introduction, IMMS can evidently be regarded as a valuable addition to the toolbox of structural biology. Data obtained from IMMS experiments allows the identification and analysis of co-populated conformational states, observation of folding behaviour in the absence of solvent and the analysis of oligomers and heterogeneous complexes. Structural information on larger macromolecular assemblies can be derived, particularly with the use of TWIMS technology. Despite the high efficiency in transmission of large ions, TWIMS has a lower resolving power than classic drift tube IM. Optimisation of the resolution of both TWIMS and drift tube instruments, and extending the mass range on drift tube instruments, will certainly increase the amount of valuable information it is possible to obtain for structural biology studies.

Due to the difficulty in predicting the structure of proteins in the gas phase, most research groups opt for computational modelling techniques to interpret their data. Ion mobility and structural modelling of peptides is at present in a more advanced state than for protein complexes. Indeed, databases of the cross sections of peptides have been published for peptides containing up to 24 residues, and the Ω of a protonated peptide can be predicted without the necessity for any structural modelling. [Mosier *et al.*, 2002; Tao *et al.*, 2007; Valentine *et al.*, 1999] Greater understanding of protein structures in the gas phase as well as a larger pool of standardised data would be beneficial for data interpretation. In the remainder of this section, some concluding remarks on major outstanding issues are provided that the community needs to address in order to further expand the credibility and applicability of IMMS in structural biology studies, as well as a discussion of the complementarity of IMMS and other biophysical techniques.

1.5.1. Structural measurements in the gas phase

Studying protein and protein complex conformation in the gas phase by IMMS has both pros and cons. On the one hand, IMMS can provide invaluable insights into the structure and cross sections of well-defined (mass-selected) protein complexes, seemingly of unlimited complexity and molecular weight, and with sensitivity unattainable by most other techniques. Gas phase studies can eliminate the complexity resulting from the presence of solvent, thus enabling

the analysis of the intrinsic properties of the protein. Moreover, transitions between conformational states can be monitored on a very short timescale. This makes IMMS an attractive tool for studying the inherent ability of proteins to adopt folded tertiary structures. Short-lived species, for example assembly intermediates or transient interactions, can become trapped and thus effectively preserved when the solvent layer is removed, providing snapshots of non-equilibrium conditions. Although partial desolvation of protein complexes would allow the effects of solvation on the gas phase structure to be analysed, mass accuracy strongly depends on complete desolvation of protein complexes before MS analysis. [Benesch *et al.*, 2007; Heck & van den Heuvel, 2004]

Conversely, a major problem with these gas phase studies is that the lack of solvent can affect the structure of the protein or complex in various ways such that the Ω obtained by IMMS does not necessarily resemble the Ω of the structure in solution. During the transition into the gas phase, very flexible ions or extended protein conformations can expand or even collapse. Experimental conditions in the ESI process, controlling the transfer from solution into the gas phase, need to be well-balanced to retain the integrity of the protein structure as much as possible, preventing undesired complex dissociation or unfolding induced by collisional activation or Coulombic repulsion. Fortunately, several IMMS studies have also shown that protein structure is often retained upon transfer into the gas phase, thus allowing biologically relevant information to be obtained, but critical assessment is always advised. [Ruotolo *et al.*, 2005; Ruotolo *et al.*, 2007; Uetrecht *et al.*, 2008b; van Duijn *et al.*, 2009]

1.5.2. Calibration and the measurement of absolute collision cross sections

As discussed in this introduction, the direct determination of the Ω of a protein is not possible with T-wave instruments, which are otherwise well suited to analyse macromolecular structures. [Ruotolo *et al.*, 2008] Calibrations must therefore be performed using species with known collision cross section. In order to improve the calibration and minimise systematic errors, standards should be analysed that cover the entire drift time and *m/z* range. So far, typically only small proteins far below 100 kDa in weight have been used for the calibration of TWIMS data in order to determine the absolute Ω, obviously too low in mass compared to the large assemblies currently under investigation. The Ω values of various charge states of these relatively small proteins were determined using a drift tube instrument and have been shown to be in fair agreement with TWIMS data. [Ruotolo *et al.*, 2008; Scarff *et al.*, 2008] However, one issue that can be encountered with this method of calibration is that proteins can unfold upon collisional activation in the gas phase, which leads to some ambiguity in assigning drift time, and thus Ω, correctly. The amount of collisional activation is related to several experimental settings on the instrument (*i.e.* accelerating and trapping voltages, gas pressure), which generally need to be tuned in the optimisation process, making inter-laboratory and -instrument comparisons of Ω more difficult. Techniques like drift tube IM, which are capable of direct Ω determination, have so far been restricted to proteins and complexes of several hundred kDa, mostly because of differences in ESI source designs and limited sensitivity when compared to TWIMS instruments. Expansion of the molecular weight range of drift tube instruments would therefore be very valuable in expanding the TWIMS calibration range.

One standard often used for *m/z* calibration in MS is an aqueous solution of CsI, [Heck & van den Heuvel, 2004] which can form large clusters across the relevant, large *m/z* range. Such

CsI clusters should be less affected by collisional activation than proteins in terms of structural change. [MaierBorst *et al.*, 1997] For these salt clusters, Ω values should be predictable, since the cluster geometries are quite well known. This could then lead to CsI acting as a calibration standard for TWIMS instruments, without the necessity for comparable drift tube results. Synthetic elongated polymers could provide another option to solve the problem of calibration, although it has recently become evident that these can also form equally elegant three-dimensional folded structures in the gas phase. [Trimpin *et al.*, 2007] More spherical structures are adopted by dendrimers, which are available across a broad mass range, presenting another potential calibrant. Shvartsburg *et al.* have recently shown that the drift times in a T-wave device can be predicted for small molecules, [Shvartsburg & Smith, 2008] however, calibration is still necessary for proteins. Improvement of such prediction methods could potentially remove the requirement for the laborious calibration of TWIMS data. Future work is required to establish which systems could be best used as a standard for calibration.

1.5.3. The role of IMMS in structural biology

We have discussed thus far not only the wide range of applicability of IMMS, but also some of the advantages and disadvantages of using it as a structural technique. The question now arises as to how it fits into the arsenal of tools already available and extensively used to study structures. Ideally, such a tool would provide not only an atomic resolution structure, but also information regarding dynamics, structural changes, co-existing conformations and different interaction partners, thereby using the least possible resources. In reality, of course, every technique has disadvantages and requires compromises to be made, and often, a combined approach is the best. Here, we discuss IMMS in terms of the requirements of the structural biology field, in the context of other commonly used methodologies.

The structural resolution possible using IMMS depends greatly on the size of the analyte. For small molecules, *e.g.* oligosaccharides, it is possible to determine the difference between constitutional isomers. [Plasencia *et al.*, 2008] For very large protein complexes, such as the HBV capsids, the uncertainty is approximately 10 nm^2 (or 1%). [Uetrecht *et al.*, 2008b] This level of resolution is evidently much lower than what can be achieved with X-ray crystallography and NMR. Although achieving the highest resolution possible is very desirable, both of these techniques have drawbacks making them inapplicable to or inappropriate for many studies. Firstly, both methods typically require milligrams of sample, significantly more than IMMS (in the nanogram to microgram range). Furthermore, the sample concentration has to be much higher, which can lead to such problems as protein aggregation, resulting in a loss of integrity of the protein structure. The second major restriction of the use of these high resolution techniques is that they are limited by the size of the protein or protein complex. To date, the largest proteins reported to have structures globally determined by NMR have molecular weights of approximately 100 kDa, although new developments in NMR allow some insights into local structure and dynamics of much larger protein complexes, such as the proteasome and GroEL. [Amero *et al.*, 2009] The molecular weight range for which X-ray crystallography has been used has extended considerably over the last decade, a famous example being the full structure of the 520 kDa heterogeneous polymerase II complex. [Cramer *et al.*, 2000; Cramer *et al.*, 2001] In contrast to this, IM coupled to a ToF analyser allows detection of ions over a theoretically unlimited mass range covering large as well as small species at once. Currently, the

largest structures in the gas phase defined by IMMS are, to the best of knowledge, the different forms of HBV capsids (3 MDa and 4 MDa). [Uetrecht *et al.*, 2008b]

Although X-ray crystallography gives high structural resolution for comparatively large proteins and protein complexes, the significance and accuracy of the results can sometimes be questionable. The crystal lattice is, like the gas phase, in principle an unnatural environment for a protein, and a major concern is the impact that crystal packing might have on biologically relevant structures and interactions. [Kobe *et al.*, 2008] In addition, local disorder in the crystal lattice, often indicative of protein dynamics, can reduce the average electron density such that local atomic positions cannot accurately be described. A further point worth noting is that most of the available crystal structures originate from recombinant overexpressed proteins, often modified to increase yield and stability or to decrease flexibility, due to the difficulties in purifying proteins expressed at endogenous levels in sufficient amounts required for the crystallisation process. In contrast, IMMS as MS studies can principally be performed on endogenously expressed proteins directly purified from their natural habitat. [Lorenzen *et al.*, 2007a; Synowsky *et al.*, 2006]

Alternative techniques that can provide structural information about large molecules such as protein complexes are atomic force microscopy (AFM), EM, small angle X-ray scattering (SAXS) and small angle neutron scattering (SANS). Data of intermediate structural resolution on the order of 1 nm can be obtained by AFM and EM, whereas small angle scattering (SAS) techniques are considered to be of low resolution, comparable to IMMS. Both AFM and EM require only low sample concentrations, and it is possible to obtain information about the dimensions of a protein or complex, sometimes even surface structure, on a relatively short timescale. [Roos & Wuite, 2009; Steven & Belnap, 2005] For these optical techniques, large macromolecular assemblies pose no problems. EM typically provides higher resolution, by averaging the low resolution images of many individual particles. AFM, by contrast, has the advantage that the features of a single particle can be analysed. In order to obtain a resolution of 1 nm by EM, 10,000 to 100,000 molecules have to be selected, classified if more than one family is identified, and aligned, a rather time-consuming process. [Rosenthal & Henderson, 2003] This is particularly difficult for proteins or particles of less than a few hundred kDa, and even more problematic if the object lacks symmetry, thus, EM is best suited for the structural analysis of large particles with high symmetry.

Since SAS methods provide only low resolution structural data, they are frequently combined with modelling. Over the last two years, the field of SAS has approached the determination of structures from biological systems with multiple components, such as protein complexes. [Neylon, 2008] One of the main advantages of SANS and SAXS is that they allow structural studies in an environment closer to physiological conditions than other structural methods, including IMMS. However, the experiments can cause problems due to the preparation, quality and quantity of samples required: approximately 200-500 µl solution with 1-10 mg/ml protein. With such high concentrations, the issue of potential protein aggregation arises, as with X-ray crystallography and NMR. Not only can this affect the identification of biologically relevant interaction sites between proteins, but can significantly affect the measurement, since SAS is more sensitive to larger particles.

Thus, many different structural techniques are available, allowing different types of infor-

mation to be gained on different samples. As has been mentioned, IMMS experiments have the distinct advantages of requiring very little sample, at low concentration, and enabling analysis over a very broad mass range. Additionally, the experimental approach can be considered much less cost and labour intensive than other methods, with IMMS data often being obtained and processed within a timescale of a few days or weeks. The integration of computational methods with IMMS can, evidently, increase this time substantially, dependent on the methodologies applied. The main disadvantages of IMMS are the relatively low structural resolution possible, as well as the non-physiological environment in which the samples are analysed. The low resolution means that it is only the average shape of an ensemble of ions that is measured. However, the distribution of drift times obtained can allow some conclusions to be drawn regarding the structural diversity of an ensemble, or the kinetics of interconverting conformational states. Data interpretation in native MS and IMMS is not always straightforward, however, more and more programs are currently becoming available to identify artefacts in MS, making these techniques more accessible to the broad scientific community. [Liepold *et al.*, 2009; McKay *et al.*, 2006; van Breukelen *et al.*, 2006] Another important property of IMMS, unattainable by at least the high resolution techniques, is its ability to simultaneously analyse different species within a mixture. Due to the two distinct dimensions of separation, this includes different conformational states of the same protein, both small interaction partners and large assemblies, and protein aggregates. IMMS is also especially suited to analyse the structural changes undergone by molecules [Alverdi *et al.*, 2008; Lorenzen *et al.*, 2008], a major advantage in comparison with other techniques.

In most cases, a structural biology question needs to be addressed by more than one method. Each technique used for resolving structural properties of proteins and protein complexes has benefits and disadvantages, and the data interpretation requires experienced analysts in order to draw justified conclusions. To continue improving the accuracy and certainty of the information gained, it becomes ever more necessary to verify and expand data obtained by one method with the application of other techniques. [Johnson, 2008; Szymczyna *et al.*, 2009] Its ability to complement other structural methods in niche applications makes IMMS a powerful new tool in structural biology.

2. Analysing virus structure and assembly using mass spectrometry

The work described in this thesis focuses on the structural analysis of viruses by native MS and IMMS, which I have introduced in the previous section. Viruses, in addition to being a threat for human health, are also cataclysmic for agriculture and fermentation processes. Furthermore, viruses appear to be good model systems to study the assembly of protein complexes, since the viral protein shells, capsids, often have the amazing ability to self-organise their folding and assembly. [Steinmetz *et al.*, 2009] Their natural capacity of encapsulating material, *i.e.* the viral genome, renders them an interesting target for nanotechnological applications that exceed far beyond drug delivery. [Singh *et al.*, 2006]

In order to effectively prevent viral infections and fully exploit the technological applicability, the precise biophysical characterisation of the virus assembly and maturation processes is crucial. [Mitragotri & Lahann, 2009] However, viral assemblies are large, which somewhat hampers the structural analysis by conventional techniques like X-ray and NMR as discussed in section 1.5.3. Another problem is the transient nature of assembly and maturation intermediates impeding their purification. Here, mass spectrometric techniques such as IMMS come into play enabling real-time studies even on low abundant, heterogeneous or transient species. [Morton *et al.*, 2008] Other aspects relevant to virology can be easily tackled by MS, for example mapping posttranslational modifications onto the virus structure. [Siuzdak, 1998]

After unveiling general concepts in virology, we outline how MS can assist in virus characterisation, especially focussing on structural aspects. In the last part, we introduce the viral systems investigated and the open questions related to their structure and assembly.

2.1. General overview

Viruses are small infectious agents that replicate intracellularly. [Pe'ery & Mathews, 2007] With sizes from nm to low µm, most of them are invisible under the light microscope. [Claverie & Abergel, 2009; Steven *et al.*, 2005a] The first evidence of their existence was found in an indirect fashion more than a hundred years ago, when Beijerinck showed that tobacco plants can be infected with sterile filtered cell extracts from diseased leaves. [Beijerinck, 1898] Despite their small size, viruses are ubiquitous, every organism has its viruses; from bacteria and archaea to man. [Ackermann & Kropinski, 2007; Lawrence *et al.*, 2009; Mercer *et al.*, 2010; Nelson & Citovsky, 2005; Pearson *et al.*, 2009; Yamada *et al.*, 2006] Estimates suggest that after prokaryotes, viruses account for the second largest amount of biomass on earth. [Suttle, 2005] Amongst the viruses, those hosting in bacteria are most abundant and are called phages. Generally, viruses lack a metabolism and therefore take control of the host cell to facilitate their own replication. The viral genome encodes all proteins necessary for replication in the adequate host, which range in number from a few to several hundreds. Due to their inability to reproduce independently, viruses are often not regarded as a form of life. [Moreira & Lopez-Garcia, 2009] However, the discussion is still ongoing and was further powered by the discovery of giant mimiviruses (microbe mimicking viruses). [Claverie & Abergel, 2009; La Scola *et al.*, 2008] These viruses of amoeba are larger in physical and genome size than the smallest known bacteria and have an unusually complex structure. Mimivirus can even carry a satellite virus or virophage in the virion, named Sputnik, which requires and modifies the mimivirus replication machinery for multiplication in the host. Furthermore, the evolutionary ancestry of viruses is still unclear: did they arise from pieces of nucleic acids replicating in cells; or reduce from cel-

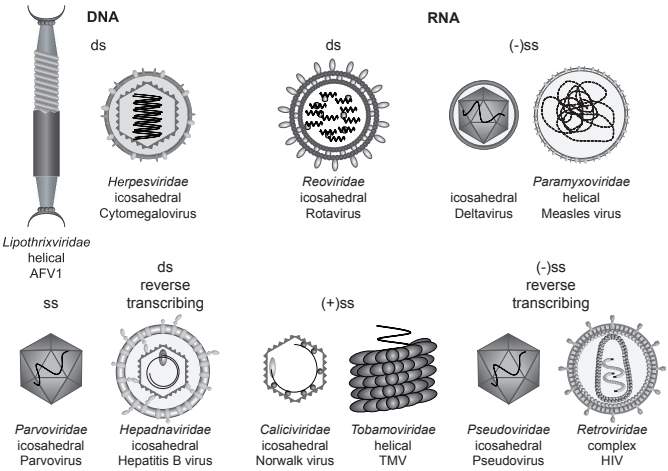


Fig. 10: Virus classification
Virus classification and taxonomy are based on the type of nucleic acid. Diverse capsid morphologies, including icosahedral, helical or complex are observed for the genome forms. A limited selection of virus families and the corresponding morphology are depicted. For more profound overview see http://www.ictvdb.org. [Condit, 2007]

lular organisms? [Forterre & Prangishvili, 2009; Koonin, 2009]

2.1.1. Classification

The ubiquity amongst species already foreshadows a broad diversity complicating viral classification. Generally, viruses are distinguished based on either their host organism or morphogenetic characteristics and genome organisation. [Condit, 2007; Fauquet & Fargette, 2005; Rohwer & Edwards, 2002; van Regenmortel & Mahy, 2004] The genome can be encoded by both, RNA or DNA, in single- or double-stranded (ss and ds) form. The information can be located on (+) or (–) strands and in some cases requires reverse transcription for replication. Viruses can contain single or multiple pieces of nucleic acid in linear or circular form (Fig. 10). The real evolutionary relationship between viruses is often difficult to obtain from one single feature, obstructing the extension of virus taxonomy beyond orders. The organisation into viral families and genera gets constantly rearranged, especially now as more viruses become sequenced and structurally characterised. The genome is enclosed in a protein shell forming the nucleocapsid. This capsid can be helical, icosahedral or complex. Already the nucleocapsid can facilitate host cell attachment and entry, but especially eukaryotic viruses are often enveloped with a lipid bilayer containing the adaptor proteins. [Harrison, 2007]

2.1.2. Lifecycle

In order to replicate, viruses have to identify the host and introduce their genome into the cell. Next, the protein and nucleic acid machinery of the cell is utilised for production of the viral constituents, finally followed by assembly and release of the infectious virus. After successful attachment to the host cell, the nucleocapsid enters the cytosol *via* various mechanisms, like membrane fusion or phagocytosis as in the case of eukaryotic viruses (Fig. 11). [Huiskonen & Butcher, 2007; Mercer *et al.*, 2010; White *et al.*, 2008] Bacteriophages generally inject their genome directly into the cell, [Grayson & Molineux, 2007; Marvin, 1998] whereas most other viruses release the genome from the capsid at the pore complexes of the cell nucleus. [Whittaker *et al.*, 2000] An exception are (+)ssRNA viruses, which also uncoat in the cytoplasm, but replicate their genome outside the nucleus in membrane proximity. [Denison, 2008] Some viruses get incorporated metastably into the host genome, for example retroviruses and lysogenic phages. Replication of such proviruses is then induced spontaneously or after a trigger.

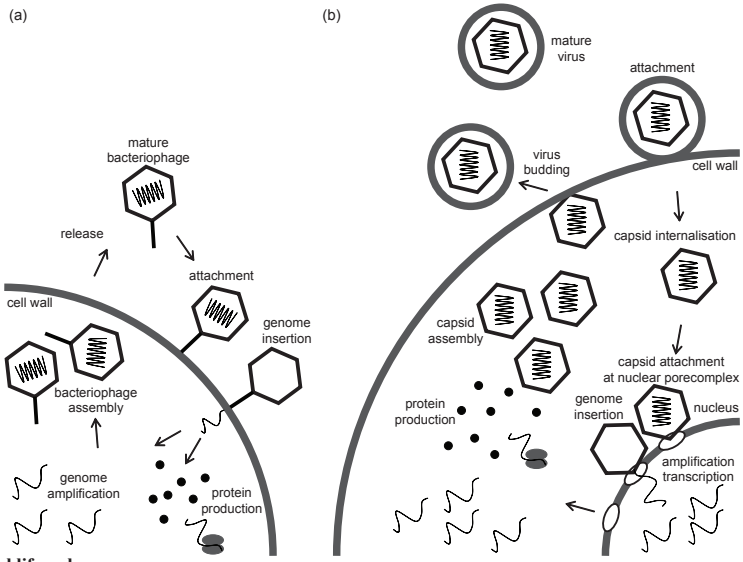


Fig. 11: Viral lifecycle
(a) Prokaryotic viruses (phages) attach to the host cell and directly inject their genome as shown here for a tailed phage. Then, the genome is amplified and proteins transcribed in the cytosol. In case of tailed phages, an empty capsid forms, then the genome is packed followed by maturation. The assembled phages accumulate in the cytoplasm until the cell ruptures. (b) Eukaryotic viruses transfer their capsid into the cytoplasm. Internalisation can occur by membrane fusion in case of enveloped viruses. At the nucleus the capsid disintegrates and releases the genome, which gets reproduced in the nucleus. Protein fabrication and assembly take place in the cytoplasm. For (+)ssRNA viruses the genome is also amplified in the cytoplasm. After assembly and possibly maturation the virus is released from the cell through budding or destructing the cell.

[Brussow *et al.*, 2004; Bushman *et al.*, 2005; Oppenheim *et al.*, 2005] Upon transcription and translation, the virus takes over the host metabolism for production of its proteins and genome. [Netherton *et al.*, 2007] At late stages of infection, the structural proteins either assemble around new copies of nucleic acid or assemble independently and package the nucleic acid afterwards. [Patient *et al.*, 2009; Patton *et al.*, 2007; Sun *et al.*, 2010] Eventually, mature virions are released by budding like vesicles or the cell can rupture as a consequence of the viral load. [Chen & Lamb, 2008; Wang *et al.*, 2000] The virus can then spread and infect new cells, beginning the cycle anew. Again, multiple ways of replication occur during the lifecycle of different viruses. An important step is usually maturation, signalling complete assembly of infectious particles that are ready for release. This process includes conformational changes triggered by the nucleic acid incorporated or attached proteins and posttranslational modifications such as phosphorylation, glycosylation, proteolysis and cross-linking. [Cutalo *et al.*, 2004; Duda *et al.*, 1995; Gertsman *et al.*, 2009; Lanman & Prevelige, 2005; Steven *et al.*, 2005b] Viruses represent a strong selective pressure on the host. Resistant cells have an evolutionary advantage by increased viability. As a consequence, viruses evolve by constantly altering their genome. RNA viruses, in particular, often show a high mutagenesis rate enforcing a constant struggle for survival. [Estes *et al.*, 2006] Another mechanism which results in an increased virulence and a switch in species specificity is the recombination of partial genomes from related viral strains, as is common in influenza viruses. [Casjens, 2005; Yen & Webster, 2009]

2.1.3. Capsid assembly and structure

As we have seen, the nucleocapsid is of eminent importance in the viral lifecycle since it encapsulates and protects the genome. It is common to all viruses, sometimes even mediating host cell attachment and infection. [Estes *et al.*, 2006] Additionally, the capsids are of high interest for other applications as we will see later on. [Singh *et al.*, 2006] One of the reasons is the exquisite interplay of simplicity and high efficiency of self-assembly viruses possess. Typically, the protein shell is formed by multiple copies of one or a few different proteins, [Caspar & Klug, 1962; Zlotnick, 2005] although decorations with other proteins are common in non-enveloped viruses, especially the lambdoid or tailed bacteriophages. [Johnson & Chiu, 2007] These attached proteins can increase the capsid stability, or serve a role during infection or genome packaging. The high copy number of a limited set of relatively small proteins decreases the size of nucleic acid to encode them, which therefore needs less space for encapsulation. [Caspar & Klug, 1962] This exemplifies the brilliant and efficient principles underlying virus structure.

Many capsid proteins (cp) can readily be produced in high quantities and represent ideal model systems to study protein (self-)assembly. [Steinmetz *et al.*, 2009] Even though, the building blocks in capsid assembly are different for certain viruses, their formation is generally in agreement with nucleation theory as was also proposed in amyloid assembly. [Wetzel, 2006; Zandi *et al.*, 2006; Zlotnick, 2005] First, the nucleus has to be formed, which may be an oligomeric assembly or just a conformational change within a monomer. After formation of the assembly nucleus, further building blocks attach to the nucleus until the capsid is completed (Fig. 12). Under conditions of efficient assembly, the nucleation is generally the rate limiting step and only a small fraction of the proteins will be in this state. The elongation rate is then much higher leading to immediate propagation of the nucleus to a capsid and to low concentrations of possible assembly intermediates. Therefore, the intermediate oligomeric species forming the nucleus or higher order assemblies are only present in trace amounts under assembly conditions. [Dokland, 2000; Hagan & Elrad, 2010; Zandi *et al.*, 2006; Zlotnick, 2005] For some viruses, it is possible to change the solution conditions favouring overnucleation. Intermediates then become kinetically trapped. Although competing processes, like misfolding and unspecific aggregation, can prevent the detection of intermediates. [Duda *et al.*, 1995; Stockley *et al.*, 2007; Zlotnick *et al.*, 2002; Zlotnick, 2007] In conclusion, a low nucleation rate ensures effective capsid assembly. Nucleation can be triggered by increasing concentration or posttranslational modifications to secure a sufficiently high titer of cp in the cell. Also, interactions with the nucleic acid can facilitate nucleus formation. [Hagan, 2009; Kivenson & Hagan, 2010]

The corresponding viral capsids are often highly stable towards changes in environment and can resist extreme pH values, high concentrations of denaturants and organic solvents, dilution to cp concentrations not facilitating assembly and even dehydration. [Flenniken *et al.*, 2009; Newman *et al.*, 2003; Roos & Wuite, 2009; Singh & Zlotnick, 2003; Siuzdak *et al.*, 1996] This effect is reflected by a strong hysteresis observed in dissociation experiments.

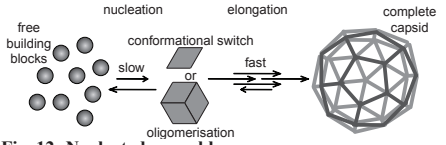


Fig. 12: Nucleated assembly
Conformational change or oligomer formation can result in nucleation. The nucleus formation is a slow reaction, whereas the subsequent addition of building blocks proceeds fast towards capsid completion. In cases where the nucleation is a fast process, assembly intermediates accumulate due to overnucleation and few capsids are formed. [Zlotnick, 2005]

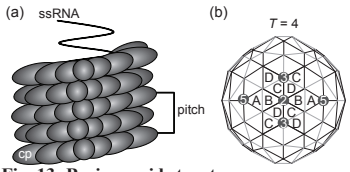


Fig. 13: Basic capsid structures
(a) Helical capsids form around a nucleic acid core. The cp subunits bind to the core in a helical fashion. The number of turns when building blocks appear in the same radial position is called pitch. This together with the diameter and number of subunits per turn is a general descriptor for helical capsids. (b) Icosahedral capsids contain 12 pentameric vertices with 5-fold symmetry. Additionnally, hexamers can be inserted, 30 in $T = 4$ capsids. The quasi-equivalent subunits (A, B, C, D) and symmetry axes (2-, 3-, 5-fold) are labelled. [Caspar & Klug, 1962]

Theoretical and experimental results suggest that the interaction between individual building blocks is rather weak and cannot account for the high apparent stability. [Ceres & Zlotnick, 2002; Zlotnick, 2005] However, in the capsid the binding energies of the subunits add up, explaining the pseudo-stability of the capsids under conditions where assembly does not occur. Nevertheless, theory suggests that in such a pseudo-equilibrium, there are always some free building blocks in solution and should exchange with proteins in the capsids. This process has been termed "capsid breathing" and some indirect evidence has been found. [Bothner *et al.*, 1998; Ceres & Zlotnick, 2002; Hilmer *et al.*, 2008; Parent *et al.*, 2007; Reisdorph *et al.*, 2003]

In general, two basic capsid structures are observed (Fig. 13), helical and icosahedral, which both allow the formation of a regular shell with multiple copies of a single cp due to the high symmetry. [Caspar & Klug, 1962] As laid out above, encoding few and low mass proteins is beneficial for the virus' size. The prototype of a helical virus is the tobacco mosaic virus (TMV). Here, cp monomers assemble around the RNA and the length of the genome defines the nucleocapsid size. [Kegel & van der Schoot, 2006; Klug, 1999] Helical viruses are defined by the number of proteins per turn and the pitch, which refers to the number of turns after which two monomers along the longitudinal axis realign in the same radial position. In this arrangement, capsids cannot be capped with the same protein. Alternatively, according to Caspar and Klug, icosahedral structures allow the complete closure of the shell using just one type of protein. [Caspar & Klug, 1962] Additionally, the almost spherical structure reduces the surface area relative to the volume. An icosahedron consists of 12 vertices, 20 faces and 30 edges corresponding to the different symmetry axis (5 , 3- and 2-fold, respectively). At least 30, generally dimeric, building blocks are required to form the smallest possible icosahedron, (resembling a Buckminster fullerene) where all proteins are accommodated in pentamers. Larger capsids are formed by introduction of hexamers. Only certain numbers of hexamers can be inserted to produce a perfect icosahedron reflected by the triangulation number (T): $T = h^2 + hk + k^2$ where h and k can be any integer and $T = 2$ is therefore not allowed. The amount of building blocks corresponds to $30T$. The surrounding contacts between subunits in hexamers and pentamers are different. However, the conformational changes to compensate this are marginal, yielding cp structures that are quasi-equivalent. In some viruses, one cp forms the pentamers and another the hexamers. [Vellinga *et al.*, 2005; Zhang *et al.*, 2010b] Additional hexamers may be introduced in a ring-like fashion leading to prolate capsids as in bacteriophage Phi29. [Choi *et al.*, 2006; Moody, 1999] Complex viruses often deviate from icosahedral symmetry, for example the HIV cp typically forms conical, but also rod-shaped, capsids. [Ganser-Pornillos *et al.*, 2004]

2.1.4. Applications in health, industry and nanotechnology

As is well-known, there is no cure for people infected with HIV. [Hokey & Weiner, 2006; Moreno *et al.*, 2010] In general, there are few effective antivirals that clear any virus completely after infection, in humans and other organisms. Thus, the treatment usually reduces the

symptoms, *i.e.* keeps the organism alive, until its immune system responds. [Althouse *et al.*, 2010; Lee *et al.*, 2009; Schinazi *et al.*, 2010] Furthermore, resistance arises against antivirals constantly, creating a demand for new drugs. In favourable cases, antisera maybe produced to treat patients. [Toovey, 2007] However, vaccination of a population is usually the method of choice to prevent the initial infection and also viral spread amongst non-vaccinated individuals. [Althouse *et al.*, 2010] Vaccines can consist of entire viruses inactivated by chemicals or heat. Additionally, less virulent strains from a different species or carrying mutations can be administered, so the immune system becomes acquainted to the virus. Where this is not possible, protein subunits exposed on the viral surface can be overexpressed and applied. [Grgacic & Anderson, 2006; Plotkin, 2009] However, high mutation rates prevent the conventional vaccine production for viruses like HIV, and other approaches, like DNA-based vaccination, are currently tested. [Hokey & Weiner, 2006] This is relying on the higher conservation of a certain protein and, therefore, certain genome stretches. Usually, the non-exposed parts of proteins are less prone to mutation and are a potential target for antivirals. This includes drugs preventing the capsid assembly, since the cp interaction site is highly conserved in a specific virus. [Lanman & Prevelige, 2005; Neira, 2009]

Next to drug and vaccine development, viruses or parts of these can be used directly to treat bacterial infections. Increasingly, bacteria acquire multi-resistance towards common antibiotics enforcing the search for new treatments. A phage therapy can help in such cases using the pathogen specific bacteriophage. [Housby & Mann, 2009] For applications in food production, parts of phages are currently under investigations to treat bacterial contamination. Furthermore, phages can be highly specific (narrow host range), sometimes infecting only a particular bacterial strain and can be used for sero-typing to identify bacterial infections. [O'Flaherty *et al.*, 2009]

Other applications use the fact that viruses can encapsulate material, especially nucleic acids, and deliver it into a cell. In gene therapy, viral vectors transfect the patient's cells with functional DNA. Dependent on the specificity, only a cell-type, organ or the whole organism can be treated. This has been tested for cancer and genetic diseases, so far with limited success. [Gillet *et al.*, 2009] Another option in anticancer therapy can be to direct cytotoxic drugs to the abnormal cells only, thereby massively reducing the side-effects observed in chemotherapy. In order to be cancer cell specific, biomarkers have to be identified. Then, the virus surface can be modified to recognise those. [Cattaneo *et al.*, 2008; Destito *et al.*, 2009; Singh *et al.*, 2006] Ligands or antibodies can be conjugated to lysine or cysteine residues. This also enables the use of viruses for biomedical imaging of target cells. [Manchester & Singh, 2006]

Viruses are also of interest in the field of nanotechnology, where there is much interest in such an organic material that reversibly assembles into well-defined stoichiometries. Moreover, both the inner and outer faces offer various functional groups for chemical modification. Displayed groups can be modified by inserting mutations and their spatial arrangement is highly controlled. This is difficult to achieve in other systems. [Flenniken *et al.*, 2009; Manchester & Singh, 2006; Singh *et al.*, 2006; Steinmetz *et al.*, 2009] The conjugated particles can then function as biosensors for ligands as seen above or display other material in an ordered array, like quantum dots or gold particles. Filamentous viruses can form nanowires when modified at both ends. Functionalised capsids have been shown to multiply the readout in detection arrays. At the

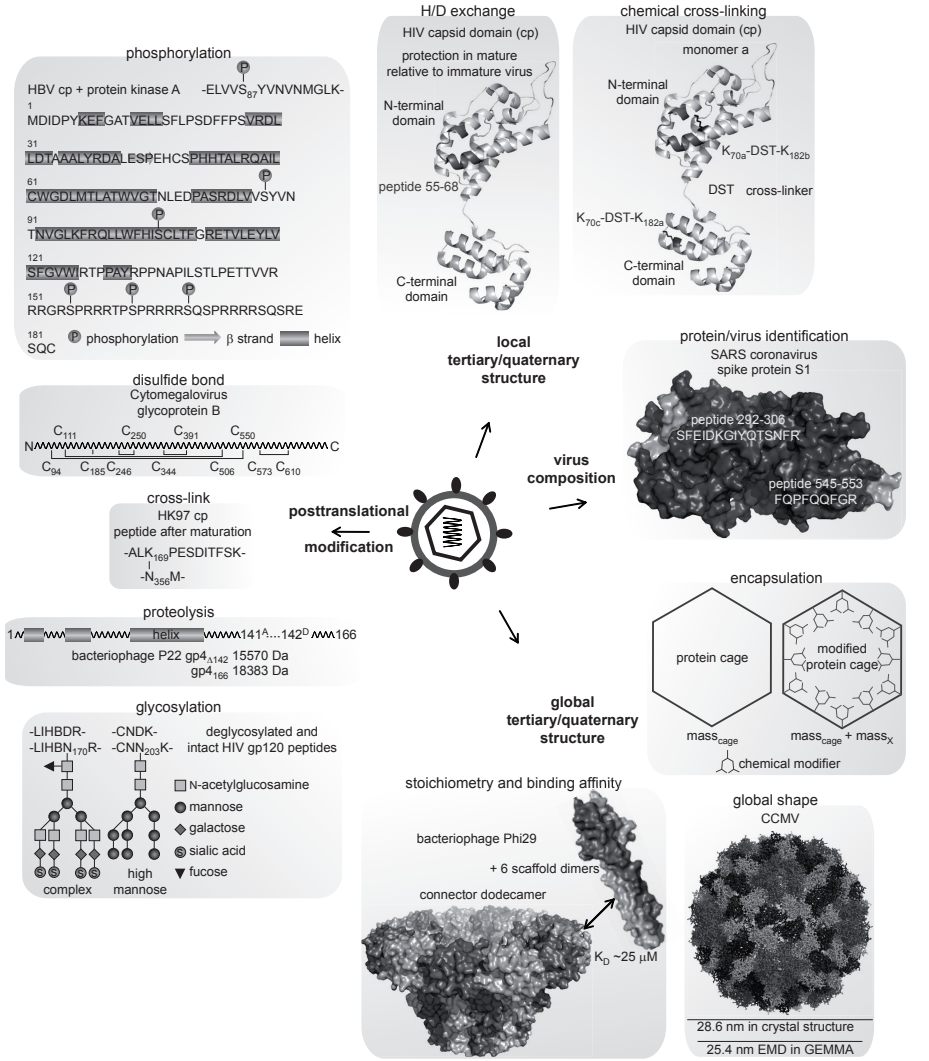


Fig. 14: Applications of MS in structural virology
Examples ranging from virus indentification, mapping of posttranslational modifications and structural analysis of viruses at different levels are illustrated, showing the broad applicability of MS. The protein compositon of a virus can be determined using proteomics approaches. For instance, peptides from the SARS coronavirus spike protein S1 found in the envelope could be identified (mapped onto the modelled structure 1Q4Z). [Zeng *et al.*, 2004] Posttranslational modifications can be mapped on such peptides and are sometimes even characterised upon analysing the denatured proteins as for the proteolytic cleavage site in bacteriophage P22 gp4. [Lorenzen *et al.*, 2008] The HBV cp carries multiple phosphorylation sites, some of which were found by MS analysis. [Kang *et al.*, 2006] Also, the location of cross-links like in HK97 maturation or more common disulfide bonds can be established. [Duda *et al.*, 1995; Lopper & Compton, 2002] MS even enabled the precise disposition of the heterogeneous glycosylation patterns in the major HIV surface protein. [Cutalo *et al.*, 2004] More structural information can be obtained on an (almost) atomic level applying H/D exchange and chemical cross-linking. Thereby, a crucial step in HIV capsid maturation was dissected. [Lanman & Prevelige, 2005] On a global level, native MS provides information on the stoichiometry of an assembly and in favourable cases the binding affinity. [Fu *et al.*, 2010] The encapsu-

same time, the background can be reduced if the dye is located inside the capsid. [Singh *et al.*, 2006]

Additionally, capsids provide a more or less closed reaction chamber. Enzymes are usually immobilised on surfaces, which often causes denaturation. The protein shell keeps them in the solution phase while still separating from the environment. This has been applied in single molecule studies on horseradish-peroxidase in Cowpea Chlorotic Mottle Virus (CCMV) capsids. Those are porous and both substrate and product diffuse in and out facilitating turnover. [Comellas-Aragones *et al.*, 2007] A smaller protein cage, ferritin, has been used as a scaffold to grow metal clusters. [Kang *et al.*, 2009a] The final particle size depends on the inner diameter of the protein shell. Accordingly, the wealth of capsids offers different scaffolds to produce well-defined clusters separately over a broad range. [Flenniken *et al.*, 2009]

2.2. Mass spectrometry in structural virology

Viruses are well-studied systems. However, multiple steps essential during the lifecycle including the assembly process and its regulation are not yet quite understood, which would be necessary to fully realise the potentials viruses offer in various scientific fields and assist drug development. [Mitragotri & Lahann, 2009] Furthermore, the emergence of ever new pathogens requires fast analysis to classify and characterise the viruses. [Cheng *et al.*, 2009] Since the 1980s, when intact biomolecules had been transferred into the mass spectrometer, a multitude of techniques have been applied to shed light on the protein composition, modifications and structure of viral proteins. [Chang, 2009; Kaddis & Loo, 2007; Lanman & Prevelige, 2005; Morton *et al.*, 2008; Siuzdak, 1998] We have already noted some of the advantages native MS and IMMS offer compared to other techniques, especially in structure elucidation. Nevertheless, other MS applications also deliver insights into virus structure at levels that were previously inaccessible or very laborious. Some examples are presented in the following section, starting with the search for the virus and its proteins (Fig. 14).

2.2.1. Virus identification and composition

With existing technologies, it is quite easy to identify an unknown infectious pathogen by genotyping. [Belak *et al.*, 2009] However, essential information about the open reading frames (ORF), expression and function of the encoded proteins can be missing after genome sequencing. In proteomics experiments (Box 2), viral proteins can even be detected from complex samples like cell lysates and analysing the composition of the virion can reveal structural and accompanied proteins. [Zeng *et al.*, 2004] A prominent example is the human coronavirus causing SARS, which emerged in 2003 in Asia. All predicted structural proteins were identified using MS including some glycosylation and phosphorylation sites (Fig. 14). [Krokhin *et al.*, 2003; Ying *et al.*, 2004; Zeng *et al.*, 2004] Afterwards, such proteins may be over-expressed and analysed safely with classical structural techniques such as X-ray crystallography. [Bartlam *et al.*, 2007] In general, the production of purified material in large scale is also crucial for vaccine development as mentioned above. Additionally, isotopic labelling can be used to follow the time resolved expression of proteins, effects on the host or the protein stoichiometry in the assembled particles. [Go *et al.*, 2006; Jiang *et al.*, 2005] The latter may help to elucidate whether structural

lation of material in viral capsids may be monitored by following the mass increase relative to the unmodified protein cage. [Abedin *et al.*, 2009] The global shape of a viral protein complex can be investigated by ion mobility techniques such as GEMMA. [Kaddis *et al.*, 2007]

Box 2: Proteomics

The composition of protein complexes or even the proteome of an entire cell can be analysed by MS-based proteomics approaches. The sample of interest is therefore digested by a protease, trypsin in most applications, and the produced peptides are identified by MS. Separation techniques are crucial to cover most peptides from complex samples. Thus, electrophoresis at the protein and chromatography at the peptide level are generally applied and combined. Trypsin specifically creates peptides with a C-terminal lysine or arginine facilitating peptide identification. In tandem MS, fragments form around the peptide bonds in a way characteristic for the activation process, which is often CID. Multiple products can be formed from a selected peptide revealing the amino acid sequence. This is used in de novo sequencing of proteins. However, more common is the peptide sequence assignment using in silico digested proteins derived from large databases. Both, the exact mass of the peptide and its fragmentation pattern, are evaluated to obtain the most likely peptide and thereby identify the original protein. Furthermore, quantification of a protein or peptide is possible either on an absolute level employing an internal standard or relative comparing two conditions. For instance, two samples with differential isotope labels are mixed in defined amounts and simultaneously analysed by MS. This ensures that signal intensities of the resulting double peaks can be compared to obtain the relative amount of the proteins/peptides in the samples. [Heck & Krijgsveld, 2004]

proteins exhibit a copy number that is consistent with the capsid symmetry, thus indicating their likely location in the virus. Such information is beneficial for the interpretation of high resolution data from large assemblies. By mischance, the exact stoichiometry is difficult to obtain in absence of related and refined X-ray or EM structures. [Estrozi *et al.*, 2008; Steven & Belnap, 2005] Isotope labelling strategies in combination with proteomics provided information on the cellular changes upon SARS infection and allowed identification of host cell factors putatively involved in virus replication. [Jiang *et al.*, 2005; Zhang *et al.*, 2010a] Accordingly, drug design benefits from the identification of potential target sites on viral proteins as well as inside the host cell.

2.2.2. Virus maturation and modification

Since the viral assembly and maturation process is less prone to mutational changes, it offers a wealth of potential interference sites and is therefore of high interest to virologists. Therefore, we now want to focus on the impact of MS in elucidating the virus structure and its changes throughout the lifecycle. During virus assembly and maturation, and also upon infection, multiple modification and rearrangement steps can occur. Traditionally, conformational changes and protein interactions have been located using fluorescent labels or globally monitored by spectroscopic techniques. [Lakowicz, 1980; Masi *et al.*, 2010] Multiple degradation steps and/or mutagenesis studies have been employed to characterise and map posttranslational modifications. [Duda *et al.*, 1995; Kang *et al.*, 2008a] To identify amino acids crucial for interactions and structural changes, mutagenesis such as alanine scanning has proved invaluable. [Fu *et al.*, 2010; Moreira *et al.*, 2007] Unfortunately, these techniques often provide either little localised information or are very laborious, whereas MS can identify conformational changes and modifications in a fast manner. Also, after initial MS screening, specific sites may be selected and mutated to validate the results.

MS techniques have provided considerable assistance in unveiling the processes in capsid formation and maturation of the icosahedral bacteriophage HK97. This lambdoid phage stores its dsDNA under high pressure and thus demands a stable capsid structure and irreversible assembly to ensure protection of the genome. [Duda *et al.*, 2009; Steven *et al.*, 2005b] Initially, 420 cp subunits organise into pentamers and hexamers. These capsomers then build the spherical, thick-walled prohead I (Fig. 15). The N-terminal Δ-domain functions as a scaffold and is cleaved off leading to prohead II. Such examples of proteolytic cleavage have been observed in many viruses and the exact location can be identified *via* the resulting fragment masses or

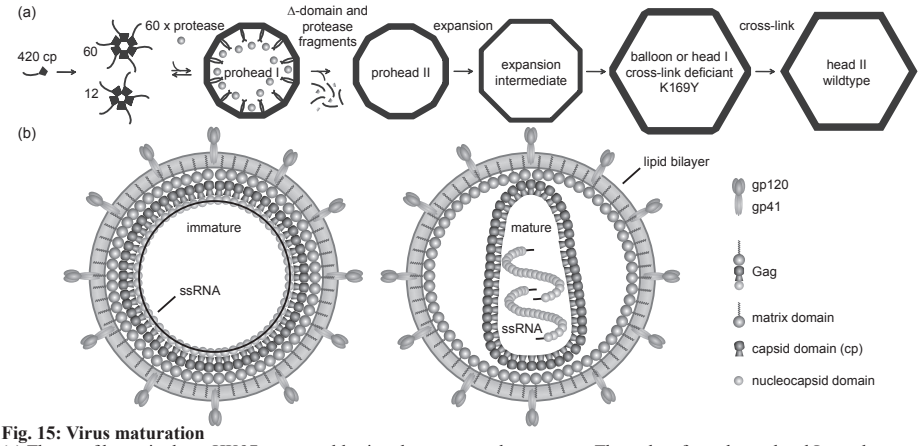


Fig. 15: Virus maturation
(a) The cp of bacteriophage HK97 preassembles into hexamers and pentamers. These then form the prohead I together with the viral protease. Assembly is facilitated by the cp Δ-domain, which functions as a scaffold. The protease cleaves the Δ-domain and itself leading to prohead II marking the first irreversible step in capsid formation. Under certain conditions expansion occurs followed by the second irreversible step, auto-catalytic cross-linking of cp subunits. The head I of a cross-link deficient mutant resembles the mature head II closely. [Gertsman *et al.*, 2009] (b) In HIV, the immature capsid of the Gag polyprotein encloses the (+)ssRNA and becomes enveloped. After release from the cell, the viral protease cleaves the Gag into three major proteins and some small peptides. The matrix domain stays bound to the lipid envelope via a myristoyl residue and the nucleocapsid domain is associated with the ssRNA. A conformational change in cp leads to a collapse of the spherical towards a conical capsid. At this stage the virus is infective. Other proteins, including the protease and peptide fragments, are omitted for clarity. [Lanman & Prevelige, 2005]

new terminal peptides observed in MS. [Pepinsky *et al.*, 1996] Based on such findings, mutants of the tail accessory protein gp4 from bacteriophage P22 could be constructed lacking the proteolytic cleavage site, thereby enabling EM and native MS structural investigations (see section 1.3.5. and Fig. 14). [Lorenzen *et al.*, 2008; Zheng *et al.*, 2008] A critical step in HK97 maturation is the conformational change leading to a thin-walled icosehedral capsid or phage-head. This expansion is usually accompanied by DNA packaging, but can be triggered *in vitro* in absence of DNA and the packaging machinery. Major structural changes associated with this transition were recognised in EM and crystallography studies. However, the achieved resolution was limited for some intermediate structures. H/D exchange MS on the different intermediates in head assembly, from capsomers to the nearly mature head I, could precisely demonstrate which amino acid stretches remained unchanged during maturation and which ones accounted for the observed global changes (see also Box 1). [Gertsman *et al.*, 2009; Gertsman *et al.*, 2010a; Gertsman *et al.*, 2010b] After expansion, the mature conformation is fixed by cross-linking of the cp subunits. Surprisingly, this is an auto-catalytic reaction boosted by the cp rearrangement. MS on trypsin-digested HK97 proheads and expanded heads detected the specific peptide carrying a lysine-asparagine cross-link. [Duda *et al.*, 1995]

However, disulfide bonds are more common in extracellular proteins and mature viruses. Stabilisation of the Cytomegalovirus glycoprotein B involves extensive disulfide bonding, which has been mapped with MS and is apparently common amongst *Herpesviridae* (Fig. 14). [Lopper & Compton, 2002] Another approach commonly applied to identify structural changes or binding surfaces is the introduction of artificial cross-links between adjacent lysines (Box 3). [Fu *et al.*, 2010; Kang *et al.*, 2009b; Lanman *et al.*, 2003] As noted above, the HIV capsid exhib-

its a pleiomorphic structure rendering classical structural analysis difficult, since high resolution generally relies on averaging multiple particles. [Ganser-Pornillos *et al.*, 2004] Nevertheless, the different capsid appearances share a common cp arrangement that could be revealed in H/D exchange and cross-linking experiments. [Lam *et al.*, 2002; Lanman *et al.*, 2003; Lanman *et al.*, 2004; Lanman & Prevelige, 2005] In HIV, the Gag polyprotein assembles around the (+)ssRNA at the plasma membrane. After budding of the enveloped virus, the viral protease residing in the spherical immature capsid cleaves the Gag protein. The released domains reorganise leading to mature virions. Then, the nucleocapsid domain is associated with the RNA in the now collapsed and conical core formed by the capsid domain (from here on cp), whereas the matrix domain stays bound to the envelope *via* an N-terminal myristoylation (Fig. 15). This myristoyl residue was also observed in MS analysis of the matrix domain in virus-like particles proving their similarity to the fully infectious virus. [Lanman *et al.*, 2004] Alanine scanning already identified certain sites in the N- and C-terminal domain of the cp that are involved in intersubunit binding between homotypic domains. However, the conformational changes and precise contact points accounting for maturation where anonymous. In H/D exchange, a previously unknown site in the N-terminal region of cp was found to be highly protected for *in vitro* assembled and mature particles, whereas immature cp showed similar behaviour to the free subunit. Interestingly, other cp peptides showed little differences between immature and mature capsids. [Lam *et al.*, 2002; Lanman *et al.*, 2004] Despite the power of H/D exchange in locating conformational changes and emerging interaction sites, this technique could not be used to determine the binding partner. Therefore, chemical cross-linking of lysines was carried out (Fig. 14). Cross-links can be incomplete, inter- and intramolecular, and thus complicate the analysis. The latter two provide useful information on monomer structure and subunit arrangement. Careful experimental design allowed filtering the data for peptides carrying intersubunit cross-links. [Lanman *et al.*, 2003] Thereby, a new interaction between the N- and C-terminal domains of adjacent monomers was established. This contact is likely driving maturation and is probably protected or inhibited in the Gag polyprotein. Since maturation is essential for infectivity of HIV, this interface is a potential target site for small drugs. [Lanman & Prevelige, 2005]

2.2.3. Virus-host interactions

Surface exposed proteins or domains usually mediate the attachment of the virus to the host. For enveloped viruses, such as HIV, those are located in the lipid bilayer and often are highly glycosylated at least in mammals. For instance, in the HIV envelope protein gp120, approximately 50% of the mass stems from glycans. Those have been suggested to be involved in antigenicity, shielding the virus from the immune system. The sites and types of glycosylation in particular are heavily affected by mutations, hampering vaccine development against HIV.

Box 3: Chemical cross-linking

Here, intra- and intermolecular interactions can be identified in a protein or protein complex. Primarily, chemical cross-linkers modifying lysines are applied. Residues that are adjacent in the native fold or assembly can become connected. Which residues are close enough in proximity is defined by the length of the cross-linker, thereby providing distance restraints for computational modelling. After modification of the proteins, residual reagent is removed or inactivated. Next, protein digest and MS analysis of the peptides are performed. The results are screened for peptides specific to the cross-linked proteins, *i.e.* those peptides that deviate in mass from the prediction or the unmodified sample. Since the chemical reaction is incomplete, dead end cross-linkers are frequently observed on peptides complicating the data interpretation. Already the exact mass can hint at the two peptides that are cross-linked. Validation and identification of the specific lysines that are modified stems from the fragmentation pattern in tandem MS. [Gingras *et al.*, 2007]

Box 4: Posttranslational modifications

Often posttranslational modifications are already evident from the mass spectrum of a denatured protein because of the characteristic mass increments. For example, glycosylation reactions are often incomplete resulting in peaks deviating by one building block. A mass difference of 162 Da would infer a conjugated hexose. Another frequently observed feature is proteolytic cleavage or unspecific degradation of a protein. The mass difference can sometimes be mapped to the protein sequence elucidating the cleavage site. [Lorenzen *et al.*, 2008]

For more thorough evaluation of the modifications, the proteins are usually digested and where possible the peptides of interest are enriched. Both the affected amino acid and the modification are identified from their fragmentation pattern in tandem MS. For instance, phosphorylations result in a characteristic neutral loss of 80 or 98 Da from the original peptide mass. [Boersema *et al.*, 2009] Fortunately, the number of different monosaccharides observed in protein glycosylations is very limited. Only certain arrangements are observed and are well characterised for asparagine or N-linked glycans (see also Fig. 14). Therefore, the fragmentation pattern unveils even the structure of the attached carbohydrates. For the initial identification usually, peptide digests are split and one sample is deglycosylated by endoglycosidase H. Hence, glycosylated peptides can easily be assigned and the size of the glycan moiety determined from the mass difference. [An *et al.*, 2009]

[Vigerust & Shepherd, 2007] Mass spectrometric studies revealed major changes in glycosylation pattern between different viral strains and expression systems (Box 4). [Go *et al.*, 2009; Liedtke *et al.*, 1997; Zhu *et al.*, 2000] After peptide digestion, the mass discrepancy between glycosylated and enzymatically deglycosylated samples determines the size of the carbohydrate. In tandem MS, the modified amino acid and the composition of the glycans may be determined. [Cutalo *et al.*, 2004; Zhu *et al.*, 2000] Stepwise enzymatic digests are another method to elucidate the glycan composition in combination with MS or other size-based separation techniques. [Liedtke *et al.*, 1997] The carbohydrates attached vary in length and in type. All three structural classes, high mannose, complex and hybrid, have been observed for gp120 (Fig. 14). Owing to the high structural flexibility and heterogeneity of carbohydrates, only a deglycosylated mutant of gp120 could be crystallised. To gain further information on distribution and conformation of the glycans identified in MS, molecular modelling was applied indicating that high mannose and complex carbohydrate moieties cluster in two separate areas. [Zhu *et al.*, 2000] It has been suggested that these clusters may increase interaction with the host cell glycans. Furthermore, binding of CD4 and chemokine receptors on human T-cells to gp120 is apparently influenced by the glycosylation pattern. However, the hypothesised stoichiometries of gp120-CD4 complexes were controversial. MALDI-MS of noncovalent and cross-linked complexes was used to address this question and clearly showed a 1:1 binding. [Borchers & Tomer, 1999]

The cellular receptors of viruses represent the most obvious virus-host interaction besides genome replication and protein translation to the non-specialist. However, the interplay is much more complex and the cellular machinery can also regulate the assembly process. For example, phosphorylation of HBV cp enhances the formation of DNA from RNA in the assembling capsid (for more detailed information on HBV see also 2.3.2.). [Melegari *et al.*, 2005] This is likely connected to the enhanced capsid assembly and encapsulation of RNA observed after phosphorylation of multiple C-terminal and a couple of central amino acid residues (Fig. 14). [Gazina *et al.*, 2000; Kang *et al.*, 2008a; Kang *et al.*, 2006] Protein kinase A or C each phosphorylate one of the two central serine residues, as was confirmed by MS and mutagenesis studies. [Kang *et al.*, 2008a; Kang *et al.*, 2006]

2.2.4. Capsids, assembly intermediates and encapsulation

Above we highlighted how MS can assist in the elucidation of various steps in viral lifecycles, but is the direct assessment of capsids or viruses and their assembly process also possible?

Indeed, the successful transfer of TMV and other viruses into the gas phase using ESI was shown already in 1996. [Siuzdak *et al.*, 1996] In these initial studies, the mass determination was impossible due the mass limitation of the employed quadrupole analyser. However, collection of the electrosprayed particles and subsequent EM analysis disclosed that the structures were retained. The TMV harvested after MS was even infective. Since then, multiple instrumental setups have been applied to estimate the mass of virus particles. Nevertheless, the mass accuracy obtained is still limited. The combination of *m/z* and charge detection in a ToF resulted in a mass of ~40 MDa for TMV. Despite this significant success, the mass distributions were broad and the charge determination was error prone, resulting in a substantial uncertainty in mass (± 15%). [Fuerstenau *et al.*, 2001] In GEMMA, masses are estimated from the EMD (electrophoretic mobility diameter) under the assumption that singly charged particles are present. The obtained values are in reasonable agreement with the theoretical masses of viral particles. [Kaddis *et al.*, 2007; Thomas *et al.*, 2004] Additionally, the determined virus diameter is usually within 10% of EM measurements, providing further evidence of the structural integrity after ESI as depicted for CCMV in Fig. 14. Early on, a relatively simple ESI-ToF facilitated analysis of the bacteriophage MS2, showing slightly resolved charge states. [Tito *et al.*, 2000] The observed charge states, however, were difficult to assign and further desolvation stages could assist the mass determination. Recently, another ionisation technique, laser induced acoustic desorption (LIAD), was coupled to charge and optical detection to establish the masses of viruses and even cells. This system is suitable for much larger analytes than ESI and has a great potential for particle identification, although the mass accuracy of several % does not allow detailed structural studies. [Chang, 2009; Nie *et al.*, 2006; Peng *et al.*, 2006]

For nanotechnology applications, not only the integrity of the capsids is important, but also their dynamics like “breathing” and reversible assembly. These are important for particle modification and to support reactions of encapsulated materials. For instance, limited proteolysis on intact capsids of several viruses followed by MS indicated that cp regions located at the inner face become transiently exposed to the exterior in flock house virus and CCMV. [Bothner *et al.*, 1998; Liepold *et al.*, 2005] Furthermore, reagents are able to penetrate the particles and modify the encapsulated nucleic acid. [Broo *et al.*, 2001] In all cases, the overall integrity of the capsids was not affected as validated by EM. Whether this is due to transient dissociation of individual cp subunits or channelling of flexible regions through pores is not entirely clear. However, capsids virtually identical in high resolution structures can deviate in their susceptibility to proteolysis, indicating that the native genome affected the dynamics more than other polynucleic acids. [Bothner *et al.*, 1999] Antivirals and antibodies can also modulate the dynamic properties, thus possibly offering the potential to open or close viral particles according to specific needs. [Reisdorph *et al.*, 2003] The incorporation of material and modification of capsids and other protein cages utilises these dynamic abilities. Additionally, the mass shift upon functionalising particles or creation of mixed assemblies can be readily detected with native MS providing a tool to monitor the reaction and control the quality (Fig. 14). [Abedin *et al.*, 2009; Kang *et al.*, 2008b; Kang *et al.*, 2009a] Dendrimers covalently attached to the inner face of a protein cage can be nurtured to subsequential generations. Thereby, the particle stability increases tremendously enabling applications in more extreme environments. The possibility to specifically modify the enclosed dendrimers further increases the utility in imaging, for example. [Abedin *et al.*, 2009]

Controlling the assembly reactions of viral capsids *in vitro* is a significant problem, as is the detection of assembly intermediates. As mentioned earlier, the nucleated character of the assembly process renders intermediates very low abundant. Nevertheless, a few studies of noncovalent complexes related to assembly processes have been performed by MS and will be described here.

Tailed bacteriophages contain a connector or portal complex, which functions as DNA conduit and forms the interface between the capsid and the tail. The latter likely seals the genome carrying capsid and facilitates attachment and penetration of the host cell. Interestingly, the connector cannot incorporate into existing capsids, instead the cp proteins assemble around it. The question arises how the interaction is mediated and ensured that only one connector is present per phage. *In vitro* assembly reactions suggested that the scaffold protein involved in arrangement of the cp also interacts with the connector. [Fu & Prevelige, 2009] In conclusion, the scaffold protein could be the missing link since it contacts both the capsid and connector. Native MS unambiguously proved the existence and stoichiometry of scaffold connector complexes (Fig. 14). Investigations on the concentration dependence of the assembly showed no cooperative effect and allowed the estimation of the binding affinity. In conjunction with cross-linking analysis of the binding interface, a model of the complex structure was established, marking the first step in bacteriophage assembly. [Fu *et al.*, 2010]

Since capsids bind and encapsulate nucleic acid, it is not surprising that the genome can represent a scaffold for assembly as well. For example, the complex of the HIV nucleocapsid protein with RNA-stemloops exhibits a 1:1 stoichiometry. In addition, complex formation was shown to require Zn^{2+} and the binding affinity was estimated. [Fabris *et al.*, 2007; Hagan & Fabris, 2003; Loo *et al.*, 1998] Also in bacteriophage MS2, RNA-stemloops interact with the cp subunits and are necessary for efficient *in vitro* assembly. [Stockley *et al.*, 2007] The RNA induces a conformational switch resulting in quasi-equivalence of RNA-associated and independent dimers. At appropriate mixing ratios of RNA and cp, the assembly reaction proceeds to completion and the time-course could be monitored by native MS. After several hours, the capsid is observed in addition to a few other, larger oligomers. Isotopic labelling in reassociation experiments could show that these oligomers are on pathway to capsid formation and grow *via* addition of dimers.

In conclusion, we have seen that viruses can be studied by various techniques in MS. Even intact particles can survive the transfer into the gas phase and assembly pathways may be elucidated with native MS. Next, the specific viruses analysed during this research project will be introduced and open questions related to their structure and assembly are discussed.

2.3. Scope of thesis

A variety of crucial steps occur during the viral lifecycle. Of these, capsid assembly is common to basically all viruses and essential to form an infectious virion. Therefore, this initial stage represents a potential drug target. Despite its key role, little is known about the assembly process itself and even less about capsid dissociation. Icosahedral capsids are a well defined system in terms of size, which render them ideal for native MS studies. Nevertheless, capsids have masses in the MDa range requiring instrumental adjustments. Furthermore, lipid envelopes and nucleic acids are often heterogeneous, challenging the mass spectrometric analysis with increased mass and also increased mass heterogeneity.

2.3.1. Open questions in capsid assembly

First of all, the stoichiometry of icosahedral capsids is theoretically well defined according to Caspar and Klug, who related the symmetry to the number of subunits required. [Caspar & Klug, 1962] Many viruses have been studied by EM and X-ray crystallography, ratifying this theory. These techniques rely on particle or rotational averaging to obtain featured structures. Upon inspection of a capsid structure, tight cp packing is observed, raising the question whether sterical hindrance could hamper the insertion of the last building block. The question of completion has not been addressed experimentally. [Zlotnick, 2007] On the other hand, incomplete assembly has been observed for mutants of HK97 cp, which lack all pentameric vertices. [Li *et al.*, 2005] The averaging process in high resolution structure determination masks the difference, and other mass measurements are not accurate enough to distinguish between almost and entirely complete capsids. [Wingfield *et al.*, 1995] In light of the encouraging results by Tito *et al.* and the improved instrumentation, [Lorenzen *et al.*, 2007b; Sobott *et al.*, 2002b; Tito *et al.*, 2000; van den Heuvel *et al.*, 2006] exact capsid masses may be determined to resolve the completion problem (see chapter II).

The next step would be to investigate the influence of the *T* number on capsid stability *in vacuo* and in solution. Several studies have indicated that not only the stoichiometry but also virus and capsid structures can be retained in the gas phase. [Kaddis *et al.*, 2007; Siuzdak *et al.*, 1996; Thomas *et al.*, 2004] The retention of structure is essential, otherwise potential stability differences *in vacuo* are unrelated to the icosahedral geometry. Hence, the conformation of capsids in the mass spectrometer should be disclosed by IMMS. Nowadays, the TWIMS instrumentation enables the investigation of large protein complexes. [Lorenzen *et al.*, 2008; van Duijn *et al.*, 2009] Some viral cp possess the ability to assemble into capsids with different *T* numbers. [Choi *et al.*, 2006; Crowther *et al.*, 1994; Driedonks *et al.*, 1978] Since the building block is identical, the stability should largely reflect geometric effects. The dissection of such effects can be attempted in the gas phase and in solution. An interesting point is whether the stabilities in the two environments are related (see chapter III).

Moreover, the assembly and dissociation pathways can be investigated. Major questions include the identity of formed products and intermediates using the possibility to detect low abundant species in MS; and the potential reversibility of the process. IMMS can be applied to elucidate the gas phase structure of the observed oligomers. Additionally, molecular modelling could assist in disclosing whether the subunits are arranged in a tightly packed, *i.e.* globular, planar or spherical fashion (see chapter IV and V).

Last but not least, capsids are dynamic entities, which are susceptible to limited proteolysis. [Bothner *et al.*, 1998] Theoretical approaches suggest that capsids exhibit “breathing”, in other words exchange subunits with free ones in solution. [Ceres & Zlotnick, 2002] A method has to be developed to monitor the incorporation of isotopically labelled subunits into preformed capsids and estimate the exchange rate (see chapter VI).

To address these questions, the capsids of two major human pathogens were investigated in the work described here, namely HBV and norovirus. We will now have a brief look at biological aspects including the viral lifecycle and then focus on structural features of each of these two viruses.

2.3.2. Hepatitis B virus

Even though vaccines against HBV are existing for almost 30 years, [Zanetti *et al.*, 2008] several hundred million people worldwide still suffer from chronic HBV infection. [Lai & Liaw, 2010; McMahon, 2009] Accordingly, HBV is a major cause of liver cirrhosis and cancer that eventually leads to death. The virus is transmitted through body fluids and was initially identified from patients after receiving blood or plasma transfusions. [Dane *et al.*, 1970] Akin viruses have been identified in avians and mammals forming the family *Hepadnaviridae*. Those are enveloped and reverse transcribing dsDNA viruses. [Schaefer, 2007] Notably, Hepatitis A and C viruses are completely unrelated to HBV and belong to distinct virus families. [Howard, 2002]

Initially, the ~3.2 kb genome in the mature HBV is a partially ds circular DNA enclosed by the spherical, icosahedral capsid. [Nassal, 2008; Steven *et al.*, 2005a] The surrounding envelope contains a viral glycoprotein in three length variants (S,M and L). Effective virus replication occurs in hepatocytes suggesting a specific receptor. While it is known that the L protein is required for infection and several cellular binding partners were identified, the receptor is still unassigned. [Xie *et al.*, 2010] The possibility that several proteins are involved in host cell attachment can not be excluded. It has also been hypothesised that the virus binds an unspecific surface factor and effective genome replication limits the reproduction to hepatocytes. [Seeger & Mason, 2000] However, somehow the virus attaches followed by endocytosis and membrane fusion. [Cooper *et al.*, 2003] The released capsid migrates to the nucleus where it disintegrates at the nuclear pore complex (Fig. 16). [Rabe *et al.*, 2009; Weigand *et al.*, 2010] Then, the partially ds circular DNA enters the nucleus, where it is completed and covalently closed. [Nassal, 2008] From this DNA, two forms of RNA get transcribed: the pregenomic RNA for generation of new virus particles and the subgenomic RNAs for protein synthesis. The HBV genome contains multiple overlapping ORFs, some of which are shifted, with separate initiation sites, but all end at the same polyadenylation site. Thus, multiple subgenomic RNAs are produced explaining the three length versions of the surface protein, which differ by one domain each.

In addition to the structural proteins, a reverse transcriptase and the x protein are encoded in the genome. HBx has a regulatory function affecting many signalling pathways. [Nguyen *et al.*, 2008] During protein synthesis, the cp accumulates in the cytoplasm, whereas the related e antigen of unknown function is secreted as a dimer. This protein lacks the C-terminal domain and instead is N-terminally extended. [Steven *et al.*, 2005a] At sufficient concentrations of cp in the cytosol, the capsid assembles around a copy of pregenomic RNA bound to the reverse transcriptase. Directly after capsid closure, the RNA is converted to the mature partially ds circular DNA by the transcriptase. Hence, the capsid functions as

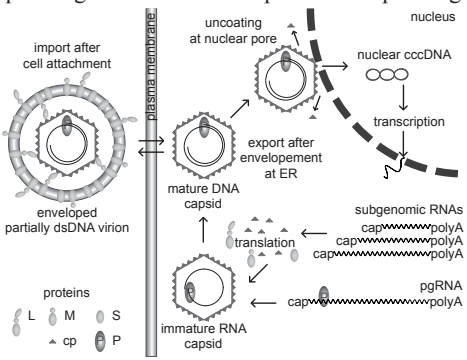


Fig. 16: HBV lifecycle
The lifecycle of HBV is shown schematically. Cell entry and uncoating at the nuclear pore complex are followed by covalent closure of the circular DNA genome (cccDNA) in the nucleus. Pregenomic (pg) and subgenomic RNAs are produced and enter the cytoplasm. The cp assembles around the pgRNA associated to the reverse transcriptase (P protein). After reverse transcription to circular partially dsDNA, the mature capsid becomes enveloped and released. The three envelope proteins (S, M, L) facilitate capsid binding and host cell attachment. [Nassal, 2008]

a microcompartment for this synthesis. Pores in the protein shell allow the diffusion of nucleotides. [Nassal, 2008; Patient *et al.*, 2009] The capsids associate with the envelope proteins, which have been translocated to the endoplasmatic reticulum after translation. Here, the capsid becomes enveloped and internalised exclusively after completion of DNA synthesis. The DNA likely induces a conformational change in the cp, which is supported by slight structural differences observed between RNA and DNA hosting capsids. [Roseman *et al.*, 2005] Then, mature virions are transported in vesicles to the cell membrane and secreted. [Patient *et al.*, 2009]

Remarkably, the HBV capsid occurs in two geometries both *in vivo* and *in vitro*. Either geometry can become enveloped and incorporate nucleic acid. [Crowther *et al.*, 1994; Dryden *et al.*, 2006; Seitz *et al.*, 2007] The larger $T = 4$ capsids (240 cp subunits) form the vast majority and likely represent the infectious virions. In the case of the smaller $T = 3$ particles (180 cp subunits), it is unclear whether they are infectious or by-products. Nevertheless, this dimorphism even *in vivo* is quite exceptional and only few other examples have been reported. [Choi *et al.*, 2006; Driedonks *et al.*, 1978] The capsid geometry is highly conserved amongst viruses of a particular family. [Steven *et al.*, 2005a] In line with this, capsid dimorphism was also identified in the avian *Hepadnavirus* duck Hepatitis B virus. [Kenney *et al.*, 1995] The HBV capsids are formed by dimeric building blocks of the cp, which is 183 amino acids in length and 21 kDa in mass. Wherein the N-terminal 140 residues form the assembly domain and the highly basic C-terminal "protamine domain" facilitates nucleic acid binding. Both are connected *via* a 10 amino acid long linker. [Lee *et al.*, 2003; Steven *et al.*, 2005a] Whereas the C-terminal region and the linker are largely unstructured in absence of nucleic acid, [Roseman *et al.*, 2005] the assembly domain exhibits a primarily α-helical fold. Two central helices of adjacent cp monomers form a four-helix bundle, which appears as a spike on the capsid in EM, stabilising the dimer. [Bottcher *et al.*, 1997; Conway *et al.*, 1997] Additionally, the conserved cysteine 61 can oxidise to an interdimer disulfide bridge. However, cysteines are not required for capsid assembly or integrity and can readily be replaced by alanines. [Zlotnick *et al.*, 1996]

The cp can be overexpressed and self-assembles in *Escherichia coli*. [Wingfield *et al.*, 1995] Interestingly, the purified particles resemble the native capsids indicating no major structural changes. [Kenney *et al.*, 1995; Roseman *et al.*, 2005] For most *in vitro* studies on HBV capsids, constructs lacking the C-terminal domain have been used, which also permitted X-ray crystallography. Of the 149 residue constructs, the first 142 were resolved in the $T = 4$ structure (Fig. 17). [Wynne *et al.*, 1999] Regrettably, a corresponding $T = 3$ structure is not available and few comparisons of capsid geometries have been reported by EM. [Bottcher *et al.*, 2006; Dryden *et al.*, 2006; Kenney *et al.*, 1995] Interestingly, reducing the length of the linker affects the ratio of the two geometries. The shorter the linker becomes, the more $T = 3$ is produced until the bare N-terminal domain loses the assembly competence below 140 amino acids. [Zlotnick *et al.*, 1996] For the efficient *in vitro* formation of capsids, high ionic strength and bivalent metal ions are beneficial. [Choi *et al.*, 2005; Stray *et al.*, 2004; Wingfield *et al.*, 1995] Below a critical cp concentration no assembly occurs, but preformed capsids are stable under dilute conditions. [Singh & Zlotnick, 2003] Theoretical analyses suggest that a conformational switch leads to formation of the nucleus (trimer of dimers) [Zlotnick *et al.*, 1999], to which dimers are rapidly added. [Zandi *et al.*, 2006; Zlotnick, 2005] Even though individual dimers interact weakly, the complete capsid could be stable due to the multiplicity of contacts and apparent increased local concentration. This would also explain the strong hysteresis effect for capsid dissociation.

[Ceres & Zlotnick, 2002; van der Schoot & Zandi, 2007] Significant dimer reformation only occurs at extremely basic pH and high denaturant concentrations. [Singh & Zlotnick, 2003; Wingfield *et al.*, 1995] Supporting data for the pseudo-stability of capsids and the nucleation process have been reported. [Newman *et al.*, 2003; Packianathan *et al.*, 2010; Singh & Zlotnick, 2003]

2.3.3. Norovirus

Noroviruses are the main cause of viral gastroenteritis in humans amongst all age groups. Currently, neither vaccination nor proper drug treatment are available. Even though healthy individuals usually recover from the infection after a few days, young children, elderly and patients with a weakened immune system can suffer from severe disease. The virus spreads easily *via* contaminated food or aerosols. The high stability of the virion against several common disinfectants and the high contagiousness propagate epidemic infections. Frequently, norovirus outbreaks are reported on cruise-ships, in hospitals and schools. The symptoms arise after 1 or 2 days in approximately 50% of the individuals in contact with the agent. Mutation and recombination rates of the various norovirus strains are high resulting in only short term immunity after infection and excluding classical vaccination strategies. Noroviruses cause substantial decreases in productivity and consecutively financial losses every year. Therefore, the development of fast detection techniques, antiviral drugs and vaccines is highly desirable despite the low mortality rate. [Donaldson *et al.*, 2010; Estes *et al.*, 2006; Glass *et al.*, 2009; Harris *et al.*, 2010]

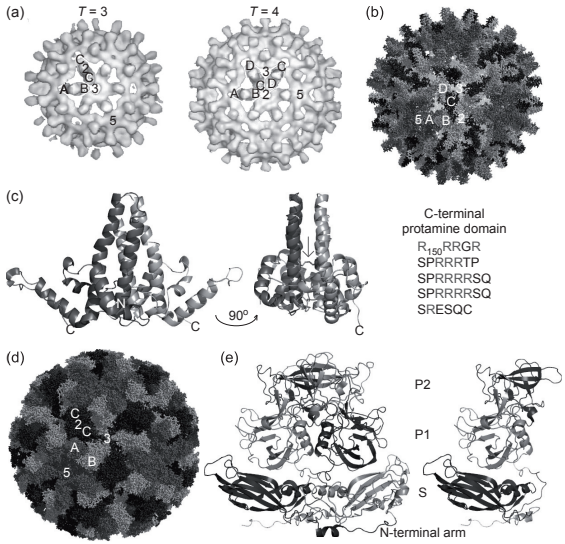


Fig. 17: Structure of HBV and NVLP
(a) EM reconstructions of the $T = 3$ and $T = 4$ HBV capsids formed by the N-terminal assembly domain. (b) Corresponding crystal structure of the $T = 4$ capsid. The spikes on the capsid surface are composed of four-helix bundles. These enable dimerisation of the cp. Two views in cartoon representation of the largely α-helical dimer are shown in (c). A cysteine bridge is formed between the two monomers at the base of the spike (grey arrow). The sequence of the C-terminal protamine domain missing in the crystallised construct is given on the right. The basic arginines are highlighted and the sequence repeats are separated. [Wynne *et al.*, 1999; Zlotnick *et al.*, 1996] (d) A much smoother surface is observed on the crystal structure of the $T = 3$ NVLP. (e) The dimer (left) and monomer (right) extracted from the capsid structure clearly show the high β-sheet content and the close interaction between the different cp domains. The N-terminal arm faces the capsid interior. Towards the outside face follow the shell (S) and the two protruding domains (P1 and P2). P2 is responsible for binding of the histoblood group antigen on the host cells. [Prasad *et al.*, 1999] Symmetry axes and pseudo-equivalent subunits are indicated on the capsid structures.

First identified around 1970, the Norwalk virus is the prototype strain of human noroviruses. [Kapikian *et al.*, 1972] These belong to the genus *norovirus* of the *Caliciviridae*, which are nonenveloped (+)ssRNA viruses. [Green *et al.*, 2000] We will, however, focus on the lifecycle of Norwalk virus from here on only referred to as norovirus. Due to a lacking cell culture system for norovirus replication, little is known about the exact replication mechanisms. Nevertheless, histoblood group antigens presented by gut epithelial cells were identified to

bind norovirus. [Marionneau *et al.*, 2002] The mechanism of cell entry is still subject to discussion, for example *caliciviruses* from other genera like feline calicivirus rely on a clathrin-based internalisation. [Makino *et al.*, 2006] Lately, studies on murine norovirus indicated that this may not be the case for *noroviruses*, instead proposing a dynamin II and cholesterol dependent mechanism. [Gerondopoulos *et al.*, 2010; Perry & Wobus, 2010] Whether this applies also for the human virus still has to be elucidated. Next, the virus has to uncoat and the RNA is directly released into the cytoplasm. The 7.7 kb RNA is bound to a viral protein, VPg, which is likely involved in translation initiation from the genomic RNA. Of the three encoded ORFs, the first is translated encoding a polyprotein. All non-structural proteins are located in this region, including p48, NTPase, p22, VPg, CL^{pro} and RdRp. [Hardy, 2005; Xi *et al.*, 1990] In general, RNA viruses replicate at intracellular membranes. [Denison, 2008; Hyde *et al.*, 2009] The membrane attachment of replication complexes is likely mediated by p48, which is apparently an integral membrane protein and may affect cellular transport. p22 may assist in this process. Even though, the functions of p48, NTPase and p22 are not clarified. NTPase shares similarities with helicases. Clearly, VPg binds the genomic RNA and was suggested to participate in RNA synthesis and translation. CL^{pro} represents the viral protease cleaving the polyprotein and releasing the other non-structural proteins to fulfil their function. Amongst the others, the RNA dependent RNA polymerase, RdRp, is released. This enzyme catalyses the synthesis of the (-)strand, which functions as template for replication of genomic and subgenomic RNAs. The latter ones function then as mRNA to produce the encoded proteins. This includes the structural proteins on ORFs 2 and 3, the major (VP1) and minor capsid protein (VP2), respectively. VP2 is required for infection and is present in low copy numbers in the virions. VP1 forms the capsid around the VPg bound genomic RNA, afterwards the virus is released. [Hardy, 2005] Little is known about virus release and maturation. [Katpally *et al.*, 2008]

Most investigations focus on recombinant norovirus-like particles (NVLP), which can be produced in an insect cell-baculovirus expression system. Despite the cotransfection of VP1 and VP2, the minor protein could not be structurally characterised in NVLPs. [Glass *et al.*, 2000; Jiang *et al.*, 1992] The particles exhibit a $T = 3$ icosahedral symmetry, as in the mature virus. [Prasad *et al.*, 1994] NVLPs are composed of 90 VP1 dimers and have been crystallised (Fig. 17). From inspection of the crystal structure, a pentamer of dimers was proposed as assembly nucleus whereupon elongation proceeds *via* addition of dimeric building blocks. [Prasad *et al.*, 1999] The particles are highly stable across a pH range of 3 - 7, however, they dissociate or reorganise at basic pH. In addition, smaller spherical particles have been observed under alkaline treatment, which could resemble $T = 1$ capsids. [Ausar *et al.*, 2006; White *et al.*, 1997]

In contrast to HBV, the norovirus cp VP1 contains primarily β-sheets. Consistent with the increased number of tasks, including genome encapsidation, virus integrity and host cell attachment, compared to the enveloped HBV, the cp is much larger comprising 520 amino acids (56 kDa). Four domains, the N-terminal arm, shell (S) and two protruding (P) domains are observed in the cp. The P domains face the exterior, whereby the outermost represents a sequence insertion into the inner domain. [Prasad *et al.*, 1994; Prasad *et al.*, 1999] Even though the P domain increases dimer stability and forms dimers on its own, it is not a prerequisite for assembly. [Bertolotti-Ciarlet *et al.*, 2002] Binding of the histoblood group carbohydrates have been mapped to this region and complex structures have been published recently. [Cao *et al.*, 2007; Choi *et al.*, 2008; Tan *et al.*, 2006; White *et al.*, 1996] The N-terminus faces the interior

and likely interacts through mediation of VP2 with the RNA. [Glass *et al.*, 2003; Prasad *et al.*, 1999] At least in presence of the N-terminal arm, the S domain is fully assembly competent. The resulting particles have the same geometry, but possess a smaller outer diameter and a smooth surface. [Bertolotti-Ciarlet *et al.*, 2002] Besides the pH stability of the capsids, very little is known about the assembly and disassembly. Together with the small size and the fact that only the genome is missing compared to mature particles, this renders them an interesting and ideal model system for native MS studies.

References

Abedin, M. J., Liepold, L., Suci, P., Young, M. & Douglas, T. (2009). Synthesis of a cross-linked branched polymer network in the interior of a protein cage. *J Am Chem Soc* **131**, 4346-4354.

Ackermann, H. W. & Kropinski, A. M. (2007). Curated list of prokaryote viruses with fully sequenced genomes. *Res Microbiol* **158**, 555-566.

Alberts, B. (1998). The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* **92**, 291-294.

Allmaier, G., Laschober, C. & Szymanski, W. W. (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.

Alonso, D. O. & Daggett, V. (1998). Molecular dynamics simulations of hydrophobic collapse of ubiquitin. *Protein Sci* **7**, 860-874.

Althouse, B. M., Bergstrom, T. C. & Bergstrom, C. T. (2010). Evolution in health and medicine Sackler colloquium: a public choice framework for controlling transmissible and evolving diseases. *Proc Natl Acad Sci USA* **107 Suppl 1**, 1696-1701.

Alverdi, V., Mazon, H., Versluis, C., Hemrika, W., Esposito, G., van den Heuvel, R., Scholten, A. & Heck, A. J. (2008). cGMP-binding prepares PKG for substrate binding by disclosing the C-terminal domain. *J Mol Biol* **375**, 1380-1393.

Amero, C., Schanda, P., Dura, M. A., Ayala, I., Marion, D., Franzetti, B., Brutscher, B. & Boisbouvier, J. (2009). Fast two-dimensional NMR spectroscopy of high molecular weight protein assemblies. *J Am Chem Soc* **131**, 3448-3449.

An, H. J., Froehlich, J. W. & Lebrilla, C. B. (2009). Determination of glycosylation sites and site-specific heterogeneity in glycoproteins. *Curr Opin Chem Biol* **13**, 421-426.

Antson, A. A., Dodson, E. J., Dodson, G., Greaves, R. B., Chen, X. & Gollnick, P. (1999). Structure of the trp RNA-binding attenuation protein, TRAP, bound to RNA. *Nature* **401**, 235-242.

Arnold, F., Henschen, G. & Ferguson, E. E. (1981). Mass spectrometric measurements of fractional ion abundances in the stratosphere--Positive ions. *Planet Space Sci* **29**, 185-193.

Arteca, G. A., Reimann, C. T. & Tapia, O. (2001). Proteins *in vacuo*: Denaturing and folding mechanisms studied with computer-simulated molecular dynamics. *Mass Spectrom Rev* **20**, 402-422.

Aston, F. W. (1919). The Mass-Spectra of Chemical Elements. *Phil Mag* **48**, 707-723.

Ausar, S. F., Foubert, T. R., Hudson, M. H., Vedvick, T. S. & Middaugh, C. R. (2006). Conformational stability and disassembly of Norwalk virus-like particles. Effect of pH and temperature. *J Biol Chem* **281**, 19478-19488.

Badman, E. R., Hoaglund-Hyzer, C. S. & Clemmer, D. E. (2001). Monitoring structural changes of proteins in an ion trap over approximately 10-200 ms: unfolding transitions in cytochrome c ions. *Anal Chem* **73**, 6000-6007.

Badman, E. R., Myung, S. & Clemmer, D. E. (2005). Evidence for unfolding and refolding of gas-phase cytochrome C ions in a Paul trap. *J Am Soc Mass Spectrom* **16**, 1493-1497.

Bagal, D., Kitova, E. N., Liu, L., El-Hawiet, A., Schnier, P. D. & Klassen, J. S. (2009). Gas phase stabilization of noncovalent protein complexes formed by electrospray ionization. *Anal Chem* **81**, 7801-7806.

Baker, E. S., Clowers, B. H., Li, F., Tang, K., Tolmachev, A. V., Prior, D. C., Belov, M. E. & Smith, R. D. (2007). Ion mobility spectrometry-mass spectrometry performance using electrodynamic ion funnels and elevated drift gas pressures. *J Am Soc Mass Spectrom* **18**, 1176-1187.

Barnes, W. S., Martin, D. W. & McDaniel, E. W. (1961). Mass Spectrographic Identification of the Ion Observed in Hydrogen Mobility Experiments. *Phys Rev Lett* **6**, 110-111.

Bartlam, M., Xu, Y. & Rao, Z. (2007). Structural proteomics of the SARS coronavirus: a model response to emerging infectious diseases. *J Struct Funct Genomics* **8**, 85-97.

Baumbach, J. (2006). Process analysis using ion mobility spectrometry. *Anal Bioanal Chem* **384**, 1059-1070.

Baumketner, A., Bernstein, S. L., Wyttenbach, T., Bitan, G., Teplow, D. B., Bowers, M. T. & Shea, J. E. (2006). Amyloid beta-protein monomer structure: a computational and experimental study. *Protein Sci* **15**, 420-428.

Beijerinck, M. W. (1898). Über ein Contagium vivum fluidum als Ursache der Fleckenkrankheit der Tabaksblätter. *Verh Akad Wetensch Afd Natuurkunde* **6**, 3-21.

Belak, S., Thoren, P., LeBlanc, N. & Viljoen, G. (2009). Advances in viral disease diagnostic and molecular epidemiological technologies. *Expert Rev Mol Diagn* **9**, 367-381.

Benesch, J. L., Aquilina, J. A., Ruotolo, B. T., Sobott, F. & Robinson, C. V. (2006). Tandem mass spectrometry reveals the quaternary organization of macromolecular assemblies. *Chem Biol* **13**, 597-605.

Benesch, J. L. & Robinson, C. V. (2006). Mass spectrometry of macromolecular assemblies: preservation and dissociation. *Curr Opin Struct Biol* **16**, 245-251.

Benesch, J. L., Ruotolo, B. T., Simmons, D. A. & Robinson, C. V. (2007). Protein complexes in the gas phase: technology for structural genomics and proteomics. *Chem Rev* **107**, 3544-3567.

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

Bernstein, S. L., Liu, D., Wyttenbach, T., Bowers, M. T., Lee, J. C., Gray, H. B. & Winkler, J. R. (2004). Alpha-synuclein: stable compact and extended monomeric structures and pH dependence of dimer formation. *J Am Soc Mass Spectrom* **15**, 1435-1443.

Bernstein, S. L., Wyttenbach, T., Baumketner, A., Shea, J. E., Bitan, G., Teplow, D. B. & Bowers, M. T. (2005). Amyloid beta-protein: monomer structure and early aggregation states of Abeta42 and its Pro19 alloform. *J Am Chem Soc* **127**, 2075-2084.

Bertolotti-Ciarlet, A., White, L. J., Chen, R., Prasad, B. V. & Estes, M. K. (2002). Structural requirements for the assembly of Norwalk virus-like particles. *J Virol* **76**, 4044-4055.

Bich, C. & Zenobi, R. (2009). Mass spectrometry of large complexes. *Curr Opin Struct Biol* **19**, 632-639.

Boersema, P. J., Mohammed, S. & Heck, A. J. (2009). Phosphopeptide fragmentation and analysis by mass spectrometry. *J Mass Spectrom* **44**, 861-878.

Bohringer, H. & Arnold, F. (1982). Temperature dependence of three-body association reactions from 45 to 400 K. The reactions $N^{+2}+2N_2 \rightarrow N^{+4}+N_2$ and $O^{+2}+2O_2 \rightarrow O^{+4}+O_2$. *J Chem Phys* **77**, 5534-5541.

Bolbach, G. (2005). Matrix-assisted laser desorption/ionization analysis of non-covalent complexes: fundamentals and applications. *Curr Pharm Des* **11**, 2535-2557.

Borchers, C. & Tomer, K. B. (1999). Characterization of the noncovalent complex of human immunodeficiency virus glycoprotein 120 with its cellular receptor CD4 by matrix-assisted laser desorption/ionization mass spectrometry. *Biochemistry* **38**, 11734-11740.
Borysik, A. J., Read, P., Little, D. R., Bateman, R. H., Radford, S. E. & Ashcroft, A. E. (2004). Separation of beta2-microglobulin conformers by high-field asymmetric waveform ion mobility spectrometry (FAIMS) coupled to electrospray ionisation mass spectrometry. *Rapid Commun Mass Spectrom* **18**, 2229-2234.
Bothner, B., Dong, X. F., Bibbs, L., Johnson, J. E. & Siuzdak, G. (1998). Evidence of viral capsid dynamics using limited proteolysis and mass spectrometry. *J Biol Chem* **273**, 673-676.
Bothner, B., Schneemann, A., Marshall, D., Reddy, V., Johnson, J. E. & Siuzdak, G. (1999). Crystallographically identical virus capsids display different properties in solution. *Nat Struct Biol* **6**, 114-116.
Bottcher, B., Wynne, S. A. & Crowther, R. A. (1997). Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. *Nature* **386**, 88-91.
Bottcher, B., Vogel, M., Ploss, M. & Nassal, M. (2006). High plasticity of the hepatitis B virus capsid revealed by conformational stress. *J Mol Biol* **356**, 812-822.
Bradbury, N. E. (1931). The Mobility of Aged Ions in Air in Relation to the Nature of Gaseous Ions. *Phys Rev* **37**, 1311-1319.
Broo, K., Wei, J., Marshall, D., Brown, F., Smith, T. J., Johnson, J. E., Schneemann, A. & Siuzdak, G. (2001). Viral capsid mobility: a dynamic conduit for inactivation. *Proc Natl Acad Sci USA* **98**, 2274-2277.
Brussow, H., Canchaya, C. & Hardt, W. D. (2004). Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev* **68**, 560-602.
Bushman, F., Lewinski, M., Ciuffi, A., Barr, S., Leipzig, J., Hannenhalli, S. & Hoffmann, C. (2005). Genome-wide analysis of retroviral DNA integration. *Nat Rev Microbiol* **3**, 848-858.
Cameron, A. E. & Eggers, D. F. (1948). An Ion "Velocitron". *Rev Sci Inst* **19**, 605-607.
Cao, S., Lou, Z., Tan, M. & other authors (2007). Structural basis for the recognition of blood group trisaccharides by norovirus. *J Virol* **81**, 5949-5957.
Casjens, S. R. (2005). Comparative genomics and evolution of the tailed-bacteriophages. *Curr Opin Microbiol* **8**, 451-458.
Caspar, D. L. & Klug, A. (1962). Physical principles in the construction of regular viruses. *Cold Spring Harb Symp Quant Biol* **27**, 1-24.
Catalina, M. I., van den Heuvel, R. H., van Duijn, E. & Heck, A. J. (2005). Decharging of globular proteins and protein complexes in electrospray. *Chemistry* **11**, 960-968.
Cattaneo, R., Miest, T., Shashkova, E. V. & Barry, M. A. (2008). Reprogrammed viruses as cancer therapeutics: targeted, armed and shielded. *Nat Rev Microbiol* **6**, 529-540.
Ceres, P. & Zlotnick, A. (2002). Weak protein-protein interactions are sufficient to drive assembly of hepatitis B virus capsids. *Biochemistry* **41**, 11525-11531.
Chang, H. C. (2009). Ultrahigh-mass mass spectrometry of single biomolecules and bioparticles. *Annu Rev Anal Chem (Palo Alto Calif)* **2**, 169-185.
Chen, B. J. & Lamb, R. A. (2008). Mechanisms for enveloped virus budding: can some viruses do without an ESCRT? *Virology* **372**, 221-232.
Chen, Y. L., Collings, B. A. & Douglas, D. J. (1997). Collision cross sections of myoglobin and cytochrome c ions with Ne, Ar, and Kr. *J Am Soc Mass Spectrom* **8**, 681-687.
Cheng, X., Chen, G. & Rodriguez, W. R. (2009). Micro- and nanotechnology for viral detection. *Anal Bioanal Chem* **393**, 487-501.
Choi, J. M., Hutson, A. M., Estes, M. K. & Prasad, B. V. (2008). Atomic resolution structural characterization of recognition of histo-blood group antigens by Norwalk virus. *Proc Natl Acad Sci USA* **105**, 9175-9180.
Choi, K. H., Morais, M. C., Anderson, D. L. & Rossmann, M. G. (2006). Determinants of bacteriophage phi29 head morphology. *Structure* **14**, 1723-1727.
Choi, Y., Gyoo Park, S., Yoo, J. H. & Jung, G. (2005). Calcium ions affect the hepatitis B virus core assembly. *Virology* **332**, 454-463.
Claverie, J. M. & Abergel, C. (2009). Mimivirus and its virophage. *Annu Rev Genet* **43**, 49-66.
Clemmer, D. E., Hudgins, R. R. & Jarrold, M. F. (1995). Naked protein conformations: cytochrome c in the gas phase. *J Am Chem Soc* **117**, 10141-10142.
Clemmer, D. E. & Jarrold, M. F. (1997). Ion mobility measurements and their applications to clusters and biomolecules. *J Mass Spectrom* **32**, 577-592.
Cole, R. B. (2000). Some tenets pertaining to electrospray ionization mass spectrometry. *J Mass Spectrom* **35**, 763-772.
Colombo, G. & Micheletti, C. (2006). Protein folding simulations: combining coarse-grained models and all-atom molecular dynamics. *Theor Chem Acc* **116**, 75-86.
Comellas-Aragones, M., Engelkamp, H., Claessen, V. I. & other authors (2007). A virus-based single-enzyme nanoreactor. *Nat Nanotechnol* **2**, 635-639.
Condit, R. C. (2007). Principles of Virology. In *Fields Virology*. Edited by D. M. Knipe, P. M. Howley, D. E. Griffin, R. A. Lamb, M. A. Martin, B. Roizman & S. E. Straus. Philadelphia: Lippincott Williams & Wilkins.
Conway, J. F., Cheng, N., Zlotnick, A., Wingfield, P. T., Stahl, S. J. & Steven, A. C. (1997). Visualization of a 4-helix bundle in the hepatitis B virus capsid by cryo-electron microscopy. *Nature* **386**, 91-94.
Cooper, A., Paran, N. & Shaul, Y. (2003). The earliest steps in hepatitis B virus infection. *Biochim Biophys Acta* **1614**, 89-96.
Covey, T. & Douglas, D. J. (1993). Collision cross sections for protein ions. *J Am Soc Mass Spectrom* **4**, 616-623.
Cramer, P., Bushnell, D. A., Fu, J. & other authors (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* **288**, 640-649.
Cramer, P., Bushnell, D. A. & Kornberg, R. D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292**, 1863-1876.
Cravath, A. M. (1929). The Rate of Formation of Negative Ions by Electron Attachment. *Phys Rev* **33**, 605-613.
Creaser, C. S., Griffiths, J. R., Bramwell, C. J., Noreen, S., Hill, C. A. & Thomas, C. L. P. (2004). Ion mobility spectrometry: a review. Part 1. Structural analysis by mobility measurement. *Analyst* **129**, 984-994.
Crowther, R. A., Kiselev, N. A., Bottcher, B., Berriman, J. A., Borisova, G. P., Ose, V. & Pumpens, P. (1994). Three-dimensional structure of hepatitis B virus core particles determined by electron cryomicroscopy. *Cell* **77**, 943-950.
Cutalo, J. M., Deterding, L. J. & Tomer, K. B. (2004). Characterization of glycopeptides from HIV-I(SF2) gp120 by liquid chromatography mass spectrometry. *J Am Soc Mass Spectrom* **15**, 1545-1555.
Daggett, V. (2006). Protein folding-simulation. *Chem Rev* **106**, 1898-1916.
Dane, D. S., Cameron, C. H. & Briggs, M. (1970). Virus-like particles in serum of patients with Australia-antigen-associated hepatitis. *Lancet* **1**, 695-698.
Denison, M. R. (2008). Seeking membranes: positive-strand RNA virus replication complexes. *PLoS Biol* **6**, e270.
Destito, G., Schneemann, A. & Manchester, M. (2009). Biomedical nanotechnology using virus-based nanoparticles. *Curr Top Microbiol Immunol* **327**, 95-122.
Dokland, T. (2000). Freedom and restraint: themes in virus capsid assembly. *Structure* **8**, R157-162.
Dole, M., Mack, L. L. & Hines, R. L. (1968). Molecular Beams of Macroions. *J Chem Phys* **49**, 2240-&.
Donaldson, E. F., Lindesmith, L. C., Lobue, A. D. & Baric, R. S. (2010). Viral shape-shifting: norovirus evasion of the human immune system. *Nat Rev Microbiol* **8**, 231-241.
Douglas, D. J. (2009). Linear quadrupoles in mass spectrometry. *Mass Spectrom Rev* **28**, 937-960.
Driedonks, R. A., Krijgsman, P. C. & Mellema, J. E. (1978). Characterization of alfalfa-mosaic-virus protein polymerization in the presence

of nucleic acid. *Eur J Biochem* **82**, 405-417.
Dryden, K. A., Wieland, S. F., Whitten-Bauer, C., Gerin, J. L., Chisari, F. V. & Yeager, M. (2006). Native hepatitis B virions and capsids visualized by electron cryomicroscopy. *Mol Cell* **22**, 843-850.
Duda, R. L., Hempel, J., Michel, H., Shabanowitz, J., Hunt, D. & Hendrix, R. W. (1995). Structural transitions during bacteriophage HK97 head assembly. *J Mol Biol* **247**, 618-635.
Duda, R. L., Ross, P. D., Cheng, N., Firek, B. A., Hendrix, R. W., Conway, J. F. & Steven, A. C. (2009). Structure and energetics of encapsidated DNA in bacteriophage HK97 studied by scanning calorimetry and cryo-electron microscopy. *J Mol Biol* **391**, 471-483.
Dugourd, P., Hudgins, R. R., Clemmer, D. E. & Jarrold, M. F. (1997). High-resolution ion mobility measurements. *Rev Sci Inst* **68**, 1122-1129.
Dussy, F. E., Berchtold, C., Briellmann, T. A., Lang, C., Steiger, R. & Bovens, M. (2008). Validation of an ion mobility spectrometry (IMS) method for the detection of heroin and cocaine on incriminated material. *Forensic Sci Int* **177**, 105-111.
Dwivedi, P., Wu, P., Klopsch, S. J., Puzon, G. J., Xun, L. & Hill, H. H. (2008). Metabolic profiling by ion mobility mass spectrometry (IMMS) *Metabolomics* **4**, 63-80.
Eckers, C., Laures, A. M., Giles, K., Major, H. & Pringle, S. (2007). Evaluating the utility of ion mobility separation in combination with high-pressure liquid chromatography/mass spectrometry to facilitate detection of trace impurities in formulated drug products. *Rapid Commun Mass Spectrom* **21**, 1255-1263.
Ens, W. & Standing, K. G. (2005). Hybrid quadrupole/time-of-flight mass spectrometers for analysis of biomolecules. *Methods Enzymol* **402**, 49-78.
Estes, M. K., Prasad, B. V. & Atmar, R. L. (2006). Noroviruses everywhere: has something changed? *Curr Opin Infect Dis* **19**, 467-474.
Estrozi, L. F., Neumann, E., Squires, G., Rozas-Dennis, G., Costabel, M., Rey, F. A., Guerin, D. M. & Navaza, J. (2008). Phasing of the Triatoma virus diffraction data using a cryo-electron microscopy reconstruction. *Virology* **375**, 85-93.
Fabris, D., Chaudhari, P., Hagan, N. & Turner, K. (2007). Functional investigations of retroviral proteinribonucleic acid complexes by nanospray Fourier transform ion cyclotron resonance mass spectrometry. *Eur J Mass Spectrom (Chichester, Eng)* **13**, 29-33.
Fauquet, C. M. & Fargette, D. (2005). International Committee on Taxonomy of Viruses and the 3,142 unassigned species. *Virol J* **2**, 64.
Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64-71.
Fenn, L. S. & McLean, J. A. (2008). Biomolecular structural separations by ion mobility-mass spectrometry. *Anal Bioanal Chem* **391**, 905-909.
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.
Flenniken, M. L., Uchida, M., Liepold, L. O., Kang, S., Young, M. J. & Douglas, T. (2009). A library of protein cage architectures as nanomaterials. *Curr Top Microbiol Immunol* **327**, 71-93.
Forterre, P. & Prangishvili, D. (2009). The origin of viruses. *Res Microbiol* **160**, 466-472.
Fu, C., Uetrecht, C., Kang, S., Morais, M., Heck, A. J., Walter, M. R. & Prevelige, P. E. (2010). A Docking Model Based on Mass Spectrometric and Biochemical Data Describes Phage Packaging Motor Incorporation. *Mol Cell Proteomics* **9**, 1764-1773.
Fu, C. Y. & Prevelige, P. E., Jr. (2009). In vitro incorporation of the phage Phi29 connector complex. *Virology*.
Fuerstenau, S. D., Benner, W. H., Thomas, J. J., Brugidou, C., Bothner, B. & Siuzdak, G. (2001). Mass spectrometry of an intact virus. *Angew Chem Int Ed Engl* **40**, 542-544.
Ganser-Pornillos, B. K., von Schwedler, U. K., Stray, K. M., Aiken, C. & Sundquist, W. I. (2004). Assembly properties of the human immunodeficiency virus type 1 CA protein. *J Virol* **78**, 2545-2552.
Gazina, E. V., Fielding, J. E., Lin, B. & Anderson, D. A. (2000). Core protein phosphorylation modulates pregenomic RNA encapsidation to different extents in human and duck hepatitis B viruses. *J Virol* **74**, 4721-4728.
Gerondopoulos, A., Jackson, T., Monaghan, P., Doyle, N. & Roberts, L. O. (2010). Murine norovirus-1 cell entry is mediated through a non-clathrin-, non-caveolae-, dynamin- and cholesterol-dependent pathway. *J Gen Virol* **91**, 1428-1438.
Gertsman, I., Gan, L., Guttman, M., Lee, K., Speir, J. A., Duda, R. L., Hendrix, R. W., Komives, E. A. & Johnson, J. E. (2009). An unexpected twist in viral capsid maturation. *Nature* **458**, 646-650.
Gertsman, I., Fu, C. Y., Huang, R., Komives, E. & Johnson, J. E. (2010a). Critical salt bridges guide capsid assembly, stability, and maturation behavior in Bacteriophage HK97. *Mol Cell Proteomics* **9**, 1752-1763.
Gertsman, I., Komives, E. A. & Johnson, J. E. (2010b). HK97 maturation studied by crystallography and H/^{2}H exchange reveals the structural basis for exothermic particle transitions. *J Mol Biol* **397**, 560-574.
Gidden, J., Jackson, A. T., Scrivens, J. H. & Bowers, M. T. (1999). Gas phase conformations of synthetic polymers: poly (methyl methacrylate) oligomers cationized by sodium ions. *Int J Mass Spectrom* **188**, 121-130.
Gidden, J., Wyttenbach, T., Jackson, A. T., Scrivens, J. H. & Bowers, M. T. (2000). Gas-Phase Conformations of Synthetic Polymers: Poly(ethylene glycol), Poly(propylene glycol), and Poly(tetramethylene glycol). *J Am Chem Soc* **122**, 4692-4699.
Giles, K., Pringle, S. D., Worthington, K. R., Little, D., Wildgoose, J. L. & Bateman, R. H. (2004). Applications of a travelling wave-based radio-frequency-only stacked ring ion guide. *Rapid Commun Mass Spectrom* **18**, 2401-2414.
Gillet, J. P., Macadangdang, B., Fathke, R. L., Gottesman, M. M. & Kimchi-Sarfaty, C. (2009). The development of gene therapy: from monogenic recessive disorders to complex diseases such as cancer. *Methods Mol Biol* **542**, 5-54.
Ginalski, K. (2006). Comparative modeling for protein structure prediction. *Curr Opin Struct Biol* **16**, 172-177.
Gingras, A. C., Gstaiger, M., Raught, B. & Aebersold, R. (2007). Analysis of protein complexes using mass spectrometry. *Nat Rev Mol Cell Biol* **8**, 645-654.
Glass, P. J., White, L. J., Ball, J. M., Leparc-Goffart, I., Hardy, M. E. & Estes, M. K. (2000). Norwalk virus open reading frame 3 encodes a minor structural protein. *J Virol* **74**, 6581-6591.
Glass, P. J., Zeng, C. Q. & Estes, M. K. (2003). Two nonoverlapping domains on the Norwalk virus open reading frame 3 (ORF3) protein are involved in the formation of the phosphorylated 35K protein and in ORF3-capsid protein interactions. *J Virol* **77**, 3569-3577.
Glass, R. I., Parashar, U. D. & Estes, M. K. (2009). Norovirus gastroenteritis. *N Engl J Med* **361**, 1776-1785.
Go, E. P., Wikoff, W. R., Shen, Z. & other authors (2006). Mass spectrometry reveals specific and global molecular transformations during viral infection. *J Proteome Res* **5**, 2405-2416.
Go, E. P., Chang, Q., Liao, H. X., Sutherland, L. L., Alam, S. M., Haynes, B. F. & Desaire, H. (2009). Glycosylation site-specific analysis of clade C HIV-1 envelope proteins. *J Proteome Res* **8**, 4231-4242.
Grayson, P. & Molineux, I. J. (2007). Is phage DNA 'injected' into cells--biologists and physicists can agree. *Curr Opin Microbiol* **10**, 401-409.
Green, K. Y., Ando, T., Balayan, M. S. & other authors (2000). Taxonomy of the caliciviruses. *J Infect Dis* **181 Suppl 2**, S322-330.
Grgacic, E. V. & Anderson, D. A. (2006). Virus-like particles: passport to immune recognition. *Methods* **40**, 60-65.
Griffin, G. W., Dzidic, I., Carroll, D. I., Stillwell, R. N. & Horning, E. C. (1973). Ion mass assignments based on mobility measuremets. Validity of plasma chromatographic mass mobility correlations. *Anal Chem* **45**, 1204-1209.
Guevremont, R. & Purves, R. W. (1999). Atmospheric pressure ion focusing in a high-field asymmetric waveform ion mobility spectrometer *Rev Sci Inst* **70**, 1370-1383.
Guilhaus, M., Selby, D. & Mlynski, V. (2000). Orthogonal acceleration time-of-flight mass spectrometry. *Mass Spectrom Rev* **19**, 65-107.
Hagan, M. F. (2009). A theory for viral capsid assembly around electrostatic cores. *J Chem Phys* **130**, 114902.
Hagan, M. F. & Elrad, O. M. (2010). Understanding the concentration dependence of viral capsid assembly kinetics--the origin of the lag time and identifying the critical nucleus size. *Biophys J* **98**, 1065-1074.

Hagan, N. & Fabris, D. (2003). Direct mass spectrometric determination of the stoichiometry and binding affinity of the complexes between nucleocapsid protein and RNA stem-loop hairpins of the HIV-1 Psi-recognition element. *Biochemistry* **42**, 10736-10745.
Hagen, D. F. (1979). Characterization of isomeric compounds by gas and plasma chromatography. *Anal Chem* **51**, 870-874.
Hardison, R. C., Chui, D. H., Giardine, B., Riemer, C., Patrinos, G. P., Anagnou, N., Miller, W. & Wajcman, H. (2002). HbVar: A relational database of human hemoglobin variants and thalassemia mutations at the globin gene server. *Hum Mutat* **19**, 225-233.
Hardy, M. E. (2005). Norovirus protein structure and function. *FEMS Microbiol Lett* **253**, 1-8.
Harris, J. P., Lopman, B. A. & O'Brien, S. J. (2010). Infection control measures for norovirus: a systematic review of outbreaks in semi-enclosed settings. *J Hosp Infect* **74**, 1-9.
Harrison, S. C. (2007). Principles of Virus Structure. In *Fields Virology*. Edited by D. M. Knipe, P. M. Howley, D. E. Griffin, R. A. Lamb, M. A. Martin, B. Roizman & S. E. Straus. Philadelphia: Lippincott Williams & Wilkins.
Hartings, M. R., Kinnear, B. S. & Jarrold, M. F. (2003). The energy landscape of unsolvated peptides: the role of context in the stability of alanine/glycine helices. *J Am Chem Soc* **125**, 3941-3947.
Heck, A. J. & Krijgsveld, J. (2004). Mass spectrometry-based quantitative proteomics. *Expert Rev Proteomics* **1**, 317-326.
Heck, A. J. & van den Heuvel, R. H. (2004). Investigation of intact protein complexes by mass spectrometry. *Mass Spectrom Rev* **23**, 368-389.
Heck, A. J. (2008). Native mass spectrometry: a bridge between interactomics and structural biology. *Nat Methods* **5**, 927-933.
Hill, H. & Steiner, W. (2006). Ion Mobility Spectrometry for Monitoring the Destruction of Chemical Warfare Agents. In *Ecological Risks Associated with the Destruction of Chemical Weapons*, pp. 157-166.
Hilmer, J. K., Zlotnick, A. & Bothner, B. (2008). Conformational equilibria and rates of localized motion within hepatitis B virus capsids. *J Mol Biol* **375**, 581-594.
Hoaglund-Hyzer, C. S., Counterman, A. E. & Clemmer, D. E. (1999). Anhydrous protein ions. *Chem Rev* **99**, 3037-3080.
Hoaglund, C. S., Liu, Y., Ellington, A. D., Pagel, M. & Clemmer, D. E. (1997). Gas-Phase DNA: Oligothymidine Ion Conformers. *J Am Chem Soc* **119**, 9051-9052.
Hoaglund, C. S., Valentine, S. J., Sporleder, C. R., Reilly, J. P. & Clemmer, D. E. (1998). Three-dimensional ion mobility/TOFMS analysis of electrosprayed biomolecules. *Anal Chem* **70**, 2236-2242.
Hogan, C. J., Jr., Carroll, J. A., Rohrs, H. W., Biswas, P. & Gross, M. L. (2009). Combined charged residue-field emission model of macromolecular electrospray ionization. *Anal Chem* **81**, 369-377.
Hokey, D. A. & Weiner, D. B. (2006). DNA vaccines for HIV: challenges and opportunities. *Springer Semin Immunopathol* **28**, 267-279.
Housby, J. N. & Mann, N. H. (2009). Phage therapy. *Drug Discov Today* **14**, 536-540.
Howard, C. R. (2002). Hepatitis viruses: a pandora's box? *J Gastroenterol Hepatol* **17 Suppl**, S464-467.
Hudgins, R. R., Ratner, M. A. & Jarrold, M. F. (1998). Design of Helices That Are Stable in Vacuo. *J Am Chem Soc* **120**, 12974-12975.
Hudgins, R. R., Mao, Y., Ratner, M. A. & Jarrold, M. F. (1999). Conformations of Gly_nH^+ and Ala_nH^+ peptides in the gas phase. *Biophys J* **76**, 1591-1597.
Huiskonen, J. T. & Butcher, S. J. (2007). Membrane-containing viruses with icosahedrally symmetric capsids. *Curr Opin Struct Biol* **17**, 229-236.
Hyde, J. L., Sosnovtsev, S. V., Green, K. Y., Wobus, C., Virgin, H. W. & Mackenzie, J. M. (2009). Mouse norovirus replication is associated with virus-induced vesicle clusters originating from membranes derived from the secretory pathway. *J Virol* **83**, 9709-9719.
Hyung, S. J., Robinson, C. V. & Ruotolo, B. T. (2009). Gas-phase unfolding and disassembly reveals stability differences in ligand-bound multiprotein complexes. *Chem Biol* **16**, 382-390.
Iacob, R. E., Murphy, J. P., 3rd & Engen, J. R. (2008). Ion mobility adds an additional dimension to mass spectrometric analysis of solution-phase hydrogen/deuterium exchange. *Rapid Commun Mass Spectrom* **22**, 2898-2904.
Iribarne, J. V. & Thomson, B. A. (1976). Evaporation of Small Ions from Charged Droplets. *J Chem Phys* **64**, 2287-2294.
Jarrold, M. F. (1995). Drift tube studies of atomic clusters. *J Phys Chem* **99**, 11-21.
Jarrold, M. F. (1999). Unfolding, refolding, and hydration of proteins in the gas phase. *Acc Chem Res* **32**, 360-367.
Jarrold, M. F. (2000). Peptides and proteins in the vapor phase. *Annu Rev Phys Chem* **51**, 179-207.
Jiang, X., Wang, M., Graham, D. Y. & Estes, M. K. (1992). Expression, self-assembly, and antigenicity of the Norwalk virus capsid protein. *J Virol* **66**, 6527-6532.
Jiang, X. S., Tang, L. Y., Dai, J., Zhou, H., Li, S. J., Xia, Q. C., Wu, J. R. & Zeng, R. (2005). Quantitative analysis of severe acute respiratory syndrome (SARS)-associated coronavirus-infected cells using proteomic approaches: implications for cellular responses to virus infection. *Mol Cell Proteomics* **4**, 902-913.
Jin, L., Barran, P. E., Deakin, J. A., Lyon, M. & Uhrin, D. (2005). Conformation of glycosaminoglycans by ion mobility mass spectrometry and molecular modelling. *Phys Chem Chem Phys* **7**, 3464-3471.
Johnsen, R., Brown, H. L. & Manfred, A. B. (1970). Ion--Molecule Reactions Involving N_2^+, N^+, O_2^+, and O^+ Ions from 300°K to ~ 1 eV. *J Chem Phys* **52**, 5080-5084.
Johnson, J. E. & Chiu, W. (2007). DNA packaging and delivery machines in tailed bacteriophages. *Curr Opin Struct Biol* **17**, 237-243.
Johnson, J. E. (2008). Multi-disciplinary studies of viruses: the role of structure in shaping the questions and answers. *J Struct Biol* **163**, 246-253.
Jurchen, J. C. & Williams, E. R. (2003). Origin of asymmetric charge partitioning in the dissociation of gas-phase protein homodimers. *J Am Chem Soc* **125**, 2817-2826.
Jurchen, J. C., Garcia, D. E. & Williams, E. R. (2004). Further studies on the origins of asymmetric charge partitioning in protein homodimers. *J Am Soc Mass Spectrom* **15**, 1408-1415.
Kaddis, C. S., Lomeli, S. H., Yin, S., Berhane, B., Apostol, M. I., Kickhoefer, V. A., Rome, L. H. & Loo, J. A. (2007). Sizing large proteins and protein complexes by electrospray ionization mass spectrometry and ion mobility. *J Am Soc Mass Spectrom* **18**, 1206-1216.
Kaddis, C. S. & Loo, J. A. (2007). Native protein MS and ion mobility large flying proteins with ESI. *Anal Chem* **79**, 1778-1784.
Kaltashov, I. A. & Abzalimov, R. R. (2008). Do ionic charges in ESI MS provide useful information on macromolecular structure? *J Am Soc Mass Spectrom* **19**, 1239-1246.
Kang, H., Yu, J. & Jung, G. (2008a). Phosphorylation of hepatitis B virus core C-terminally truncated protein (Cp149) by PKC increases capsid assembly and stability. *Biochem J* **416**, 47-54.
Kang, H. Y., Lee, S., Park, S. G., Yu, J., Kim, Y. & Jung, G. (2006). Phosphorylation of hepatitis B virus Cp at Ser87 facilitates core assembly. *Biochem J* **398**, 311-317.
Kang, S., Oltrogge, L. M., Broomell, C. C., Liepold, L. O., Prevelige, P. E., Young, M. & Douglas, T. (2008b). Controlled assembly of bifunctional chimeric protein cages and composition analysis using noncovalent mass spectrometry. *J Am Chem Soc* **130**, 16527-16529.
Kang, S., Jolley, C. C., Liepold, L. O., Young, M. & Douglas, T. (2009a). From metal binding to nanoparticle formation: monitoring biomimetic iron oxide synthesis within protein cages using mass spectrometry. *Angew Chem Int Ed Engl* **48**, 4772-4776.
Kang, S., Mou, L., Lanman, J., Velu, S., Brouillette, W. J. & Prevelige, P. E., Jr. (2009b). Synthesis of biotin-tagged chemical cross-linkers and their applications for mass spectrometry. *Rapid Commun Mass Spectrom* **23**, 1719-1726.
Kanu, A. B., Dwivedi, P., Tam, M., Matz, L. & Hill, H. H., Jr. (2008). Ion mobility-mass spectrometry. *J Mass Spectrom* **43**, 1-22.
Kapikian, A. Z., Wyatt, R. G., Dolin, R., Thornhill, T. S., Kalica, A. R. & Chanock, R. M. (1972). Visualization by immune electron microscopy of a 27-nm particle associated with acute infectious nonbacterial gastroenteritis. *J Virol* **10**, 1075-1081.
Karas, M., Bachmann, D. & Hillenkamp, F. (1985). Influence of the Wavelength in High-Irradiance Ultraviolet-Laser Desorption Mass-Spectrometry of Organic-Molecules. *Anal Chem* **57**, 2935-2939.
Karpas, Z., Cohen, M. J., Stimac, R. M. & Wernlund, R. F. (1986). On the effects of structure and charge distribution on the mobility of ions.

Int J Mass Spectrom Ion Proc **74**, 153-159.
Katpally, U., Wobus, C. E., Dryden, K., Virgin, H. W. t. & Smith, T. J. (2008). Structure of antibody-neutralized murine norovirus and unexpected differences from viruslike particles. *J Virol* **82**, 2079-2088.
Kebarle, P. & Hogg, A. M. (1965). Mass-Spectrometric Study of Ions at Near Atmospheric Pressures. I. The Ionic Polymerization of Ethylene. *J Chem Phys* **42**, 668-674.
Kebarle, P. (2000). A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry. *J Mass Spectrom* **35**, 804-817.
Kegel, W. K. & van der Schoot, P. (2006). Physical regulation of the self-assembly of tobacco mosaic virus coat protein. *Biophys J* **91**, 1501-1512.
Kemper, P. R. & Bowers, M. T. (1990). A hybrid double-focusing mass spectrometer--high-pressure drift reaction cell to study thermal energy reactions of mass-selected ions. *J Am Soc Mass Spectrom* **1**, 197-207.
Kenney, J. M., von Bonsdorff, C. H., Nassal, M. & Fuller, S. D. (1995). Evolutionary conservation in the hepatitis B virus core structure: comparison of human and duck cores. *Structure* **3**, 1009-1019.
Kinnear, B. S., Hartings, M. R. & Jarrold, M. F. (2001). Helix unfolding in unsolvated peptides. *J Am Chem Soc* **123**, 5660-5667.
Kitova, E. N., Kitov, P. I., Paszkiewicz, E., Kim, J., Mulvey, G. L., Armstrong, G. D., Bundle, D. R. & Klassen, J. S. (2007). Affinities of Shiga toxins 1 and 2 for univalent and oligovalent Pk-trisaccharide analogs measured by electrospray ionization mass spectrometry. *Glycobiology* **17**, 1127-1137.
Kivenson, A. & Hagan, M. F. (2010). Mechanisms of capsid assembly around a polymer. *Biophys J* **99**, 619-628.
Klug, A. (1999). The tobacco mosaic virus particle: structure and assembly. *Philos Trans R Soc Lond B Biol Sci* **354**, 531-535.
Kobe, B., Guncar, G., Buchholz, R., Huber, T., Maco, B., Cowieson, N., Martin, J. L., Marfori, M. & Forwood, J. K. (2008). Crystallography and protein-protein interactions: biological interfaces and crystal contacts. *Biochem Soc Trans* **036**, 1438-1441.
Koeniger, S. L., Merenbloom, S. I. & Clemmer, D. E. (2006a). Evidence for many resolvable structures within conformation types of electrosprayed ubiquitin ions. *J Phys Chem B* **110**, 7017-7021.
Koeniger, S. L., Merenbloom, S. I., Sevugarajan, S. & Clemmer, D. E. (2006b). Transfer of structural elements from compact to extended states in unsolvated ubiquitin. *J Am Chem Soc* **128**, 11713-11719.
Koichi, T., Hiroaki, W., Yutaka, I., Satoshi, A., Yoshikazu, Y., Tamio, Y. & Matsuo, T. (1988). Protein and polymer analyses up to *m/z* 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* **2**, 151-153.
Kolehmainen, M., Rönkkö, P. & Raatikainen, O. (2003). Monitoring of yeast fermentation by ion mobility spectrometry measurement and data visualisation with Self-Organizing Maps. *Anal Chim Acta* **484**, 93-100.
Koomen, J. M., Ruotolo, B. T., Gillig, K. J. & other authors (2002). Oligonucleotide analysis with MALDI-ion-mobility-TOFMS *Anal Bioanal Chem* **373**, 312-317.
Koonin, E. V. (2009). On the origin of cells and viruses: primordial virus world scenario. *Ann NY Acad Sci* **1178**, 47-64.
Krokhin, O., Li, Y., Andonov, A. & other authors (2003). Mass spectrometric characterization of proteins from the SARS virus: a preliminary report. *Mol Cell Proteomics* **2**, 346-356.
La Scola, B., Desnues, C., Pagnier, I. & other authors (2008). The virophage as a unique parasite of the giant mimivirus. *Nature* **455**, 100-104.
Lai, M. & Liaw, Y. F. (2010). Chronic Hepatitis B: Past, Present, and Future. *Clin Liver Dis* **14**, 531-546.
Lakowicz, J. R. (1980). Fluorescence spectroscopic investigations of the dynamic properties of proteins, membranes and nucleic acids. *J Biochem Biophys Methods* **2**, 91-119.
Lam, T. T., Lanman, J. K., Emmett, M. R., Hendrickson, C. L., Marshall, A. G. & Prevelige, P. E. (2002). Mapping of protein:protein contact surfaces by hydrogen/deuterium exchange, followed by on-line high-performance liquid chromatography-electrospray ionization Fourier-transform ion-cyclotron-resonance mass analysis. *J Chromatogr A* **982**, 85-95.
Langevin, P. (1903). l'ionisition des gaz. *Ann Chim Phys* **28**, 289-384.
Lanman, J., Lam, T. T., Barnes, S., Sakalian, M., Emmett, M. R., Marshall, A. G. & Prevelige, P. E., Jr. (2003). Identification of novel interactions in HIV-1 capsid protein assembly by high-resolution mass spectrometry. *J Mol Biol* **325**, 759-772.
Lanman, J., Lam, T. T., Emmett, M. R., Marshall, A. G., Sakalian, M. & Prevelige, P. E., Jr. (2004). Key interactions in HIV-1 maturation identified by hydrogen-deuterium exchange. *Nat Struct Mol Biol* **11**, 676-677.
Lanman, J. & Prevelige, P. E., Jr. (2005). Kinetic and mass spectrometry-based investigation of human immunodeficiency virus type 1 assembly and maturation. *Adv Virus Res* **64**, 285-309.
Laschober, C., Wruss, J., Blaas, D., Szymanski, W. W. & Allmaier, G. (2008). Gas-phase electrophoretic molecular mobility analysis of size and stoichiometry of complexes of a common cold virus with antibody and soluble receptor molecules. *Anal Chem* **80**, 2261-2264.
Lawrence, C. M., Menon, S., Eilers, B. J., Bothner, B., Khayat, R., Douglas, T. & Young, M. J. (2009). Structural and functional studies of archaeal viruses. *J Biol Chem* **284**, 12599-12603.
Lee, D., Redfern, O. & Orengo, C. (2007). Predicting protein function from sequence and structure. *Nat Rev Mol Cell Biol* **8**, 995-1005.
Lee, N., Chan, P. K., Hui, D. S. & other authors (2009). Viral loads and duration of viral shedding in adult patients hospitalized with influenza. *J Infect Dis* **200**, 492-500.
Lee, S. M., Park, S. G., Park, E., Lee, J. Y. & Jung, G. (2003). The 113th and 117th charged amino acids in the 5th alpha-helix of the HBV core protein are necessary for pgRNA encapsidation. *Virus Genes* **27**, 227-235.
Li, Y., Conway, J. F., Cheng, N., Steven, A. C., Hendrix, R. W. & Duda, R. L. (2005). Control of virus assembly: HK97 "Whiffleball" mutant capsids without pentons. *J Mol Biol* **348**, 167-182.
Liedtke, S., Geyer, R. & Geyer, H. (1997). Host-cell-specific glycosylation of HIV-2 envelope glycoprotein. *Glycoconj J* **14**, 785-793.
Liepold, L., Oltrogge, L. M., Suci, P. A., Young, M. J. & Douglas, T. (2009). Correct charge state assignment of native electrospray spectra of protein complexes. *J Am Soc Mass Spectrom* **20**, 435-442.
Liepold, L. O., Revis, J., Allen, M., Oltrogge, L., Young, M. & Douglas, T. (2005). Structural transitions in Cowpea chlorotic mottle virus (CCMV). *Phys Biol* **2**, S166-172.
Loo, J. A. (1997). Studying noncovalent protein complexes by electrospray ionization mass spectrometry. *Mass Spectrom Rev* **16**, 1-23.
Loo, J. A., Holler, T. P., Foltin, S. K., McConnell, P., Banotai, C. A., Horne, N. M., Mueller, W. T., Stevenson, T. I. & Mack, D. P. (1998). Application of electrospray ionization mass spectrometry for studying human immunodeficiency virus protein complexes. *Proteins* **Suppl 2**, 28-37.
Loo, J. A., Berhane, B., Kaddis, C. S., Wooding, K. M., Xie, Y., Kaufman, S. L. & Chernushevich, I. V. (2005). Electrospray ionization mass spectrometry and ion mobility analysis of the 20S proteasome complex. *J Am Soc Mass Spectrom* **16**, 998-1008.
Lopper, M. & Compton, T. (2002). Disulfide bond configuration of human cytomegalovirus glycoprotein B. *J Virol* **76**, 6073-6082.
Lorenzen, K., Vannini, A., Cramer, P. & Heck, A. J. (2007a). Structural biology of RNA polymerase III: mass spectrometry elucidates subcomplex architecture. *Structure* **15**, 1237-1245.
Lorenzen, K., Versluis, C., van Duijn, E., van den Heuvel, R. H. H. & Heck, A. J. R. (2007b). Optimizing macromolecular tandem mass spectrometry of large non-covalent complexes using heavy collision gases. *Int J Mass Spectrom* **268**, 198-206.
Lorenzen, K., Olia, A. S., Uetrecht, C., Cingolani, G. & Heck, A. J. (2008). Determination of stoichiometry and conformational changes in the first step of the P22 tail assembly. *J Mol Biol* **379**, 385-396.
MaierBorst, M., Loffler, P., Petry, J. & Kreisle, D. (1997). Ion mobility measurements of metal halide clusters. *Z Phys D: At, Mol Clusters* **40**, 476-478.
Makino, A., Shimojima, M., Miyazawa, T., Kato, K., Tohya, Y. & Akashi, H. (2006). Junctional adhesion molecule 1 is a functional receptor

for feline calicivirus. *J Virol* **80**, 4482-4490.
Manchester, M. & Singh, P. (2006). Virus-based nanoparticles (VNPs): platform technologies for diagnostic imaging. *Adv Drug Deliv Rev* **58**, 1505-1522.
Mann, M., Meng, C. K. & Fenn, J. B. (1989). Interpreting Mass-Spectra of Multiply Charged Ions. *Anal Chem* **61**, 1702-1708.
Mao, Y., Ratner, M. A. & Jarrold, M. F. (1999a). Molecular dynamics simulations of the charge-induced unfolding and refolding of unsolvated cytochrome c. *J Phys Chem B* **103**, 10017-10021.
Mao, Y., Woenckhaus, J., Kolafa, J., Ratner, M. A. & Jarrold, M. F. (1999b). Thermal unfolding of unsolvated cytochrome c: Experiment and molecular dynamics simulations. *J Am Chem Soc* **121**, 2712-2721.
Marcsisin, S. R. & Engen, J. R. (2010). Hydrogen exchange mass spectrometry: what is it and what can it tell us? *Anal Bioanal Chem* **397**, 967-972.
Marionneau, S., Ruvoen, N., Le Moullac-Vaidye, B., Clement, M., Cailleau-Thomas, A., Ruiz-Palacois, G., Huang, P., Jiang, X. & Le Pendu, J. (2002). Norwalk virus binds to histo-blood group antigens present on gastroduodenal epithelial cells of secretor individuals. *Gastroenterology* **122**, 1967-1977.
Marvin, D. A. (1998). Filamentous phage structure, infection and assembly. *Curr Opin Struct Biol* **8**, 150-158.
Masi, A., Cicchi, R., Carloni, A., Pavone, F. S. & Arcangeli, A. (2010). Optical methods in the study of protein-protein interactions. *Adv Exp Med Biol* **674**, 33-42.
McAfee, K. B., Sipler, D. & Edelson, D. (1967). Mobilities and Reactions of Ions in Argon. *Phys Rev* **160**, 130-135.
McCullough, B. J., Kalapothakis, J., Eastwood, H., Kemper, P., MacMillan, D., Taylor, K., Dorin, J. & Barran, P. E. (2008). Development of an ion mobility quadrupole time of flight mass spectrometer. *Anal Chem* **80**, 6336-6344.
McDaniel, E. W., Martin, D. W. & Barnes, W. S. (1962). Drift Tube-Mass Spectrometer for Studies of Low-Energy Ion-Molecule Reactions. *Rev Sci Inst* **33**, 2-7.
McKay, A. R., Ruotolo, B. T., Ilag, L. L. & Robinson, C. V. (2006). Mass measurements of increased accuracy resolve heterogeneous populations of intact ribosomes. *J Am Chem Soc* **128**, 11433-11442.
McKnight, L. G., McAfee, K. B. & Sipler, D. P. (1967). Low-Field Drift Velocities and Reactions of Nitrogen Ions in Nitrogen. *Phys Rev* **164**, 62-70.
McLean, J. A., Ruotolo, B. T., Gillig, K. J. & Russell, D. H. (2005). Ion mobility–mass spectrometry: a new paradigm for proteomics. *Int J Mass Spectrom* **240**, 301-315.
McMahon, B. J. (2009). The natural history of chronic hepatitis B virus infection. *Hepatology* **49**, S45-55.
Melegari, M., Wolf, S. K. & Schneider, R. J. (2005). Hepatitis B virus DNA replication is coordinated by core protein serine phosphorylation and HBx expression. *J Virol* **79**, 9810-9820.
Mercer, J., Schelhaas, M. & Helenius, A. (2010). Virus entry by endocytosis. *Annu Rev Biochem* **79**, 803-833.
Merenbloom, S. I., Koeniger, S. L., Valentine, S. J., Plasencia, M. D. & Clemmer, D. E. (2006). IMS-IMS and IMS-IMS-IMS/MS for separating peptide and protein fragment ions. *Anal Chem* **78**, 2802-2809.
Merenbloom, S. I., Glaskin, R. S., Henson, Z. B. & Clemmer, D. E. (2009). High-resolution ion cyclotron mobility spectrometry. *Anal Chem* **81**, 1482-1487.
Mesleh, M. F., Hunter, J. M., Shvartsburg, A. A., Schatz, G. C. & Jarrold, M. F. (1996). Structural information from ion mobility measurements: Effects of the long-range potential. *J Phys Chem* **100**, 16082-16086.
Miller, R. A., Nazarov, E. G., Eiceman, G. A. & King, A. T. (2001). A MEMS radio-frequency ion mobility spectrometer for chemical vapor detection. *Sens Actuators A Phys* **91**, 301-312.
Mitragotri, S. & Lahann, J. (2009). Physical approaches to biomaterial design. *Nat Mater* **8**, 15-23.
Moody, M. F. (1999). Geometry of phage head construction. *J Mol Biol* **293**, 401-433.
Moreira, D. & Lopez-Garcia, P. (2009). Ten reasons to exclude viruses from the tree of life. *Nat Rev Microbiol* **7**, 306-311.
Moreira, I. S., Fernandes, P. A. & Ramos, M. J. (2007). Hot spots--a review of the protein-protein interface determinant amino-acid residues. *Proteins* **68**, 803-812.
Moreno, S., Lopez Aldeguer, J., Arribas, J. R., Domingo, P., Iribarren, J. A., Ribera, E., Rivero, A. & Pulido, F. (2010). The future of antiretroviral therapy: challenges and needs. *J Antimicrob Chemother* **65**, 827-835.
Morton, V. L., Stockley, P. G., Stonehouse, N. J. & Ashcroft, A. E. (2008). Insights into virus capsid assembly from non-covalent mass spectrometry. *Mass Spectrom Rev* **27**, 575-595.
Mosier, P. D., Counterman, A. E., Jurs, P. C. & Clemmer, D. E. (2002). Prediction of peptide ion collision cross sections from topological molecular structure and amino acid parameters. *Anal Chem* **74**, 1360-1370.
Moult, J. (2006). Rigorous performance evaluation in protein structure modelling and implications for computational biology. *Philos Trans R Soc Lond B Biol Sci* **361**, 453-458.
Moult, J. (2008). http://predictioncenter.org/casp8/.
Myung, S., Badman, E. R., Lee, Y. J. & Clemmer, D. E. (2002). Structural transitions of electrosprayed ubiquitin ions stored in an ion trap over similar to 10 ms to 30 s. *J Phys Chem A* **106**, 9976-9982.
Nassal, M. (2008). Hepatitis B viruses: reverse transcription a different way. *Virus Res* **134**, 235-249.
Neira, J. L. (2009). The capsid protein of human immunodeficiency virus: designing inhibitors of capsid assembly. *FEBS J* **276**, 6110-6117.
Nelson, R. S. & Citovsky, V. (2005). Plant viruses. Invaders of cells and pirates of cellular pathways. *Plant Physiol* **138**, 1809-1814.
Netherton, C., Moffat, K., Brooks, E. & Wileman, T. (2007). A guide to viral inclusions, membrane rearrangements, factories, and viroplasm produced during virus replication. *Adv Virus Res* **70**, 101-182.
Newman, M., Suk, F. M., Cajimat, M., Chua, P. K. & Shih, C. (2003). Stability and morphology comparisons of self-assembled virus-like particles from wild-type and mutant human hepatitis B virus capsid proteins. *J Virol* **77**, 12950-12960.
Neylon, C. (2008). Small angle neutron and X-ray scattering in structural biology: recent examples from the literature. *Eur Biophys J* **37**, 531-541.
Nguyen, D. H., Ludgate, L. & Hu, J. (2008). Hepatitis B virus-cell interactions and pathogenesis. *J Cell Physiol* **216**, 289-294.
Nie, Z., Tzeng, Y. K., Chang, H. C., Chiu, C. C., Chang, C. Y., Chang, C. M. & Tao, M. H. (2006). Microscopy-based mass measurement of a single whole virus in a cylindrical ion trap. *Angew Chem Int Ed Engl* **45**, 8131-8134.
Noe, F. & Fischer, S. (2008). Transition networks for modeling the kinetics of conformational change in macromolecules. *Curr Opin Struct Biol* **18**, 154-162.
O'Flaherty, S., Ross, R. P. & Coffey, A. (2009). Bacteriophage and their lysins for elimination of infectious bacteria. *FEMS Microbiol Rev* **33**, 801-819.
Oppenheim, A. B., Kobiler, O., Stavans, J., Court, D. L. & Adhya, S. (2005). Switches in bacteriophage lambda development. *Annu Rev Genet* **39**, 409-429.
Ouyang, Z., Takats, Z., Blake, T. A., Gologan, B., Guymon, A. J., Wiseman, J. M., Oliver, J. C., Davisson, V. J. & Cooks, R. G. (2003). Preparing protein microarrays by soft-landing of mass-selected ions. *Science* **301**, 1351-1354.
Packianathan, C., Katen, S. P., Dann, C. E., 3rd & Zlotnick, A. (2010). Conformational changes in the hepatitis B virus core protein are consistent with a role for allostery in virus assembly. *J Virol* **84**, 1607-1615.
Parent, K. N., Suhanovsky, M. M. & Teschke, C. M. (2007). Phage P22 procapsids equilibrate with free coat protein subunits. *J Mol Biol* **365**, 513-522.
Patient, R., Hourioux, C. & Roingeard, P. (2009). Morphogenesis of hepatitis B virus and its subviral envelope particles. *Cell Microbiol* **11**, 1561-1570.

Patton, J. T., Vasquez-Del Carpio, R., Tortorici, M. A. & Taraporewala, Z. F. (2007). Coupling of rotavirus genome replication and capsid assembly. *Adv Virus Res* **69**, 167-201.
Pe'ery, T. & Mathews, M. B. (2007). Viral Conquest of the Host Cell. In *Fields Virology*. Edited by D. M. Knipe, P. M. Howley, D. E. Griffin, R. A. Lamb, M. A. Martin, B. Roizman & S. E. Straus. Philadelphia: Lippincott Williams & Wilkins.
Pearson, M. N., Beever, R. E., Boine, B. & Arthur, K. (2009). Mycoviruses of filamentous fungi and their relevance to plant pathology. *Mol Plant Pathol* **10**, 115-128.
Peng, W. P., Yang, Y. C., Kang, M. W., Tzeng, Y. K., Nie, Z., Chang, H. C., Chang, W. & Chen, C. H. (2006). Laser-induced acoustic desorption mass spectrometry of single bioparticles. *Angew Chem Int Ed Engl* **45**, 1423-1426.
Pepinsky, R. B., Papayannopoulos, I. A., Campbell, S. & Vogt, V. M. (1996). Analysis of Rous sarcoma virus Gag protein by mass spectrometry indicates trimming by host exopeptidase. *J Virol* **70**, 3313-3318.
Perry, J. W. & Wobus, C. E. (2010). Endocytosis of murine norovirus 1 into murine macrophages is dependent on dynamin II and cholesterol. *J Virol* **84**, 6163-6176.
Pinkse, M. W., Maier, C. S., Kim, J. I., Oh, B. H. & Heck, A. J. (2003). Macromolecular assembly of *Helicobacter pylori* urease investigated by mass spectrometry. *J Mass Spectrom* **38**, 315-320.
Plasencia, M. D., Isailovic, D., Merenbloom, S. I., Mechref, Y. & Clemmer, D. E. (2008). Resolving and Assigning N-Linked Glycan Structural Isomers from Ovalbumin by IMS-MS. *J Am Soc Mass Spectrom* **19**, 1706-1715.
Plotkin, S. A. (2009). Vaccines: the fourth century. *Clin Vaccine Immunol* **16**, 1709-1719.
Prasad, B. V., Rothnagel, R., Jiang, X. & Estes, M. K. (1994). Three-dimensional structure of baculovirus-expressed Norwalk virus capsids. *J Virol* **68**, 5117-5125.
Prasad, B. V., Hardy, M. E., Dokland, T., Bella, J., Rossmann, M. G. & Estes, M. K. (1999). X-ray crystallographic structure of the Norwalk virus capsid. *Science* **286**, 287-290.
Pringle, S. D., Giles, K., Wildgoose, J. L., Williams, J. P., Slade, S. E., Thalassinos, K., Bateman, R. H., Bowers, M. T. & Scrivens, J. H. (2007). An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *Int J Mass Spectrom* **261**, 1-12.
Purves, R. W., Guevremont, R., Day, S., Pipich, C. W. & Matyjaszczyk, M. S. (1998). Mass spectrometric characterization of a high-field asymmetric waveform ion mobility spectrometer *Rev Sci Inst* **69**, 4094-4105.
Purves, R. W. & Guevremont, R. (1999). Electrospray Ionization High-Field Asymmetric Waveform Ion Mobility Spectrometry−Mass Spectrometry. *Anal Chem* **71**, 2346-2357.
Purves, R. W., Barnett, D. A., Ells, B. & Guevremont, R. (2000a). Investigation of bovine ubiquitin conformers separated by high-field asymmetric waveform ion mobility spectrometry: cross section measurements using energy-loss experiments with a triple quadrupole mass spectrometer. *J Am Soc Mass Spectrom* **11**, 738-745.
Purves, R. W., Barnett, D. A. & Guevremont, R. (2000b). Separation of protein conformers using electrospray-high field asymmetric waveform ion mobility spectrometry-mass spectrometry. *Int J Mass Spectrom* **197**, 163-177.
Purves, R. W., Barnett, D. A., Ells, B. & Guevremont, R. (2001). Elongated conformers of charge states +11 to +15 of bovine ubiquitin studied using ESI-FAIMS-MS. *J Am Soc Mass Spectrom* **12**, 894-901.
Rabe, B., Delaleau, M., Bischof, A., Foss, M., Sominskaya, I., Pumpens, P., Cazenave, C., Castroviejo, M. & Kann, M. (2009). Nuclear entry of hepatitis B virus capsids involves disintegration to protein dimers followed by nuclear reassociation to capsids. *PLoS Pathog* **5**, e1000563.
Reisdorph, N., Thomas, J. J., Katpally, U., Chase, E., Harris, K., Siuzdak, G. & Smith, T. J. (2003). Human rhinovirus capsid dynamics is controlled by canyon flexibility. *Virology* **314**, 34-44.
Robinson, E. W. & Williams, E. R. (2005). Multidimensional Separations of Ubiquitin Conformers in the Gas Phase: Relating Ion Cross Sections to H/D Exchange Measurements. *J Am Soc Mass Spectrom* **16**, 1427-1437.
Rohwer, F. & Edwards, R. (2002). The Phage Proteomic Tree: a genome-based taxonomy for phage. *J Bacteriol* **184**, 4529-4535.
Roos, W. H. & Wuite, G. L. (2009). Nanoindentation Studies Reveal Material Properties of Viruses. *Adv Mater* **21**, 1187-1192.
Roseman, A. M., Berriman, J. A., Wynne, S. A., Butler, P. J. & Crowther, R. A. (2005). A structural model for maturation of the hepatitis B virus core. *Proc Natl Acad Sci USA* **102**, 15821-15826.
Rosenthal, P. B. & Henderson, R. (2003). Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J Mol Biol* **333**, 721-745.
Rostom, A. A., Fucini, P., Benjamin, D. R., Juenemann, R., Nierhaus, K. H., Hartl, F. U., Dobson, C. M. & Robinson, C. V. (2000). Detection and selective dissociation of intact ribosomes in a mass spectrometer. *Proc Natl Acad Sci USA* **97**, 5185-5190.
Ruotolo, B. T., Gillig, K. J., Stone, E. G. & Russell, D. H. (2002). Peak capacity of ion mobility mass spectrometry: separation of peptides in helium buffer gas. *J Chromatogr B Analyt Technol Biomed Life Sci* **782**, 385-392.
Ruotolo, B. T., McLean, J. A., Gillig, K. J. & Russell, D. H. (2004a). Peak capacity of ion mobility mass spectrometry: the utility of varying drift gas polarizability for the separation of tryptic peptides. *J Mass Spectrom* **39**, 361-367.
Ruotolo, B. T., Tate, C. C. & Russell, D. H. (2004b). Ion mobility-mass spectrometry applied to cyclic peptide analysis: conformational preferences of gramicidin S and linear analogs in the gas phase. *J Am Soc Mass Spectrom* **15**, 870-878.
Ruotolo, B. T., Giles, K., Campuzano, I., Sandercock, A. M., Bateman, R. H. & Robinson, C. V. (2005). Evidence for macromolecular protein rings in the absence of bulk water. *Science* **310**, 1658-1661.
Ruotolo, B. T., Hyung, S. J., Robinson, P. M., Giles, K., Bateman, R. H. & Robinson, C. V. (2007). Ion mobility-mass spectrometry reveals long-lived, unfolded intermediates in the dissociation of protein complexes. *Angew Chem Int Ed Engl* **46**, 8001-8004.
Ruotolo, B. T., Benesch, J. L., Sandercock, A. M., Hyung, S. J. & Robinson, C. V. (2008). Ion mobility-mass spectrometry analysis of large protein complexes. *Nat Protoc* **3**, 1139-1152.
Scarff, C. A., Thalassinos, K., Hilton, G. R. & Scrivens, J. H. (2008). Travelling wave ion mobility mass spectrometry studies of protein structure: biological significance and comparison with X-ray crystallography and nuclear magnetic resonance spectroscopy measurements. *Rapid Commun Mass Spectrom* **22**, 3297-3304.
Scarff, C. A., Patel, V. J., Thalassinos, K. & Scrivens, J. H. (2009). Probing hemoglobin structure by means of traveling-wave ion mobility mass spectrometry. *J Am Soc Mass Spectrom* **20**, 625-631.
Schaefer, S. (2007). Hepatitis B virus taxonomy and hepatitis B virus genotypes. *World J Gastroenterol* **13**, 14-21.
Schinazi, R. F., Bassit, L. & Gavegnano, C. (2010). HCV drug discovery aimed at viral eradication. *J Viral Hepat* **17**, 77-90.
Scholten, A., Aye, T. T. & Heck, A. J. (2008). A multi-angular mass spectrometric view at cyclic nucleotide dependent protein kinases: *in vivo* characterization and structure/function relationships. *Mass Spectrom Rev* **27**, 331-353.
Seeger, C. & Mason, W. S. (2000). Hepatitis B virus biology. *Microbiol Mol Biol Rev* **64**, 51-68.
Seitz, S., Urban, S., Antoni, C. & Bottcher, B. (2007). Cryo-electron microscopy of hepatitis B virions reveals variability in envelope capsid interactions. *EMBO J* **26**, 4160-4167.
Sharon, M. & Robinson, C. V. (2007). The role of mass spectrometry in structure elucidation of dynamic protein complexes. *Annu Rev Biochem* **76**, 167-193.
Sheibani, A., Tabrizchi, M. & Ghaziaskar, H. S. (2008). Determination of aflatoxins B1 and B2 using ion mobility spectrometry. *Talanta* **75**, 233-238.
Shelimov, K. B. & Jarrold, M. F. (1996). "Denaturation" and Refolding of Cytochrome c *in Vacuo*. *J Am Chem Soc* **118**, 10313-10314.
Shelimov, K. B., Clemmer, D. E., Hudgins, R. R. & Jarrold, M. F. (1997). Protein structure in vacuo: Gas-phase confirmations of BPTI and cytochrome c. *J Am Chem Soc* **119**, 2240-2248.

Shelimov, K. B. & Jarrold, M. F. (1997). Conformations, unfolding, and refolding of apomyoglobin in vacuum: An activation barrier for gas-phase protein folding. *J Am Chem Soc* **119**, 2987-2994.
Shen Nan, L., Griffin, G. W., Horning, E. C. & Wentworth, W. E. (1974). Dependence of polyatomic ion mobilities on ionic size. *J Chem Phys* **60**, 4994-4999.
Shvartsburg, A. A. & Jarrold, M. F. (1996). An exact hard-spheres scattering model for the mobilities of polyatomic ions. *Chem Phys Lett* **261**, 86-91.
Shvartsburg, A. A., Hudgins, R. R., Dugourd, P. & Jarrold, M. F. (1997). Structural elucidation of fullerene dimers by high-resolution ion mobility measurements and trajectory calculation simulations. *J Phys Chem A* **101**, 1684-1688.
Shvartsburg, A. A., Hudgins, R. R., Dugourd, P. & Jarrold, M. F. (2001). Structural information from ion mobility measurements: applications to semiconductor clusters. *Chem Soc Rev* **30**, 26-35.
Shvartsburg, A. A., Bryskiewicz, T., Purves, R. W., Tang, K. Q., Guevremont, R. & Smith, R. D. (2006a). Field asymmetric waveform ion mobility spectrometry studies of proteins: Dipole alignment in ion mobility spectrometry? *J Phys Chem B* **110**, 21966-21980.
Shvartsburg, A. A., Li, F. M., Tang, K. Q. & Smith, R. D. (2006b). Characterizing the structures and folding of free proteins using 2-D gas-phase separations: Observation of multiple unfolded conformers. *Anal Chem* **78**, 3304-3315.
Shvartsburg, A. A. & Smith, R. D. (2008). Fundamentals of Traveling Wave Ion Mobility Spectrometry. *Anal Chem* **80**, 9689-9699.
Singh, P., Gonzalez, M. J. & Manchester, M. (2006). Viruses and their uses in nanotechnology. *Drug Dev Res* **67**, 23-41.
Singh, S. & Zlotnick, A. (2003). Observed hysteresis of virus capsid disassembly is implicit in kinetic models of assembly. *J Biol Chem* **278**, 18249-18255.
Sinz, A. (2007). Investigation of protein-ligand interactions by mass spectrometry. *Chem Med Chem* **2**, 425-431.
Siuzdak, G., Bothner, B., Yeager, M., Brugidou, C., Fauquet, C. M., Hoey, K. & Chang, C. M. (1996). Mass spectrometry and viral analysis. *Chem Biol* **3**, 45-48.
Siuzdak, G. (1998). Probing viruses with mass spectrometry. *J Mass Spectrom* **33**, 203-211.
Smith, D. P., Giles, K., Bateman, R. H., Radford, S. E. & Ashcroft, A. E. (2007). Monitoring copopulated conformational states during protein folding events using electrospray ionization-ion mobility spectrometry-mass spectrometry. *J Am Soc Mass Spectrom* **18**, 2180-2190.
Smith, D. P., Knapman, T. W., Campuzano, I., Malham, R. W., Berryman, J. T., Radford, S. E. & Ashcroft, A. E. (2009). Deciphering drift time measurements from travelling wave ion mobility spectrometry-mass spectrometry studies. *Eur J Mass Spectrom* **15**, 113-130.
Sobott, F., Benesch, J. L., Vierling, E. & Robinson, C. V. (2002a). 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.
Sobott, F., Hernandez, H., McCammon, M. G., Tito, M. A. & Robinson, C. V. (2002b). A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Anal Chem* **74**, 1402-1407.
St. Louis, R. H., Hill Jr, H. H. & Eiceman, G. A. (1990). Ion Mobility Spectrometry in Analytical Chemistry *Crit Rev Anal Chem* **21**, 321-355.
Steinmetz, N. F., Lin, T., Lomonossoff, G. P. & Johnson, J. E. (2009). Structure-based engineering of an icosahedral virus for nanomedicine and nanotechnology. *Curr Top Microbiol Immunol* **327**, 23-58.
Stephens, W. E. (1946). A Pulsed Mass Spectrometer with Time Dispersion. *Phys Rev* **69**, 691.
Steven, A. & Belnap, D. (2005). Electron microscopy and image processing: an essential tool for structural analysis of macromolecules. *Curr Protoc Protein Sci* **Chapter 17**, Unit 17 12.
Steven, A. C., Conway, J. F., Cheng, N., Watts, N. R., Belnap, D. M., Harris, A., Stahl, S. J. & Wingfield, P. T. (2005a). Structure, assembly, and antigenicity of hepatitis B virus capsid proteins. *Adv Virus Res* **64**, 125-164.
Steven, A. C., Heymann, J. B., Cheng, N., Trus, B. L. & Conway, J. F. (2005b). Virus maturation: dynamics and mechanism of a stabilizing structural transition that leads to infectivity. *Curr Opin Struct Biol* **15**, 227-236.
Stockley, P. G., Rolfsson, O., Thompson, G. S., Basnak, G., Francese, S., Stonehouse, N. J., Homans, S. W. & Ashcroft, A. E. (2007). A simple, RNA-mediated allosteric switch controls the pathway to formation of a *T*=3 viral capsid. *J Mol Biol* **369**, 541-552.
Stray, S. J., Ceres, P. & Zlotnick, A. (2004). Zinc ions trigger conformational change and oligomerization of hepatitis B virus capsid protein. *Biochemistry* **43**, 9989-9998.
Sun, J., Kitova, E. N. & Klassen, J. S. (2007). Method for stabilizing protein-ligand complexes in nanoelectrospray ionization mass spectrometry. *Anal Chem* **79**, 416-425.
Sun, S., Rao, V. B. & Rossmann, M. G. (2010). Genome packaging in viruses. *Curr Opin Struct Biol* **20**, 114-120.
Suttle, C. A. (2005). Viruses in the sea. *Nature* **437**, 356-361.
Synowsky, S. A., van den Heuvel, R. H., Mohammed, S., Pijnappel, P. W. & Heck, A. J. (2006). Probing genuine strong interactions and post-translational modifications in the heterogeneous yeast exosome protein complex. *Mol Cell Proteomics* **5**, 1581-1592.
Synowsky, S. A., van Wijk, M., Raijmakers, R. & Heck, A. J. (2009). Comparative multiplexed mass spectrometric analyses of endogenously expressed yeast nuclear and cytoplasmic exosomes. *J Mol Biol* **385**, 1300-1313.
Szymczyna, B. R., Taurog, R. E., Young, M. J., Snyder, J. C., Johnson, J. E. & Williamson, J. R. (2009). Synergy of NMR, computation, and x-ray crystallography for structural biology. *Structure* **17**, 499-507.
Tahallah, N., Pinkse, M., Maier, C. S. & Heck, A. J. (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 Commun Mass Spectrom* **15**, 596-601.
Tahallah, N., Van Den Heuvel, R. H., Van Den Berg, W. A., Maier, C. S., Van Berkel, W. J. & Heck, A. J. (2002). Cofactor-dependent assembly of the flavoenzyme vanillyl-alcohol oxidase. *J Biol Chem* **277**, 36425-36432.
Takats, Z., Wiseman, J. M., Gologan, B. & Cooks, R. G. (2004). Electrosonic spray ionization. A gentle technique for generating folded proteins and protein complexes in the gas phase and for studying ion-molecule reactions at atmospheric pressure. *Anal Chem* **76**, 4050-4058.
Tan, M., Meller, J. & Jiang, X. (2006). C-terminal arginine cluster is essential for receptor binding of norovirus capsid protein. *J Virol* **80**, 7322-7331.
Tao, L., McLean, J. R., McLean, J. A. & Russell, D. H. (2007). A collision cross-section database of singly-charged peptide ions. *J Am Soc Mass Spectrom* **18**, 1232-1238.
Taverner, T., Hernandez, H., Sharon, M., Ruotolo, B. T., Matak-Vinkovic, D., Devos, D., Russell, R. B. & Robinson, C. V. (2008). Subunit architecture of intact protein complexes from mass spectrometry and homology modeling. *Acc Chem Res* **41**, 617-627.
Thalassinos, K., Slade, S. E., Jennings, K. R., Scrivens, J. H., Giles, K., Wildgoose, J., Hoyes, J., Bateman, R. H. & Bowers, M. T. (2004). Ion mobility mass spectrometry of proteins in a modified commercial mass spectrometer. *Int J Mass Spectrom* **236**, 55-63.
Thomas, J. J., Bothner, B., Traina, J., Benner, W. H. & Siuzdak, G. (2004). Electrospray ion mobility spectrometry of intact viruses. *Spectrosc - Int J* **18**, 31-36.
Thomson, G. M., Schummers, J. H., James, D. R., Graham, E., Gatland, I. R., Flannery, M. R. & McDaniel, E. W. (1973). Mobility, diffusion, and clustering of K^+ ions in gases. *J Chem Phys* **58**, 2402-2411.
Thomson, J. J. (1913). On the Appearance of Helium and Neon in Vacuum Tubes. *Science* **37**, 360-364.
Tito, M. A., Tars, K., Valegard, K., Hajdu, J. & Robinson, C. V. (2000). Electrospray time-of-flight mass spectrometry of the intact MS2 virus capsid. *J Am Chem Soc* **122**, 3550-3551.
Toovey, S. (2007). Preventing rabies with the Verorab vaccine: 1985-2005 Twenty years of clinical experience. *Travel Med Infect Dis* **5**, 327-348.
Totrov, M. & Abagyan, R. (2008). Flexible ligand docking to multiple receptor conformations: a practical alternative. *Curr Opin Struct Biol* **18**, 178-184.
Trapane, T. L. & Lattman, E. E. (2007). Seventh Meeting on the Critical Assessment of Techniques for Protein Structure Prediction. *Proteins* **69 Suppl 8**, 1-2.

Trimpin, S., Plasencia, M., Isailovic, D. & Clemmer, D. E. (2007). Resolving oligomers from fully grown polymers with IMS-MS. *Anal Chem* **79**, 7965-7974.
Uetrecht, C., Versluis, C., Watts, N. R., Roos, W. H., Wuite, G. J., Wingfield, P. T., Steven, A. C. & Heck, A. J. (2008a). High-resolution mass spectrometry of viral assemblies: molecular composition and stability of dimorphic hepatitis B virus capsids. *Proc Natl Acad Sci USA* **105**, 9216-9220.
Uetrecht, C., Versluis, C., Watts, N. R., Wingfield, P. T., Steven, A. C. & Heck, A. J. (2008b). Stability and shape of hepatitis B virus capsids *in vacuo*. *Angew Chem Int Ed Engl* **47**, 6247-6251.
Vakhrushev, S. Y., Mormann, M. & Peter-Katalinic, J. (2006). Identification of glycoconjugates in the urine of a patient with congenital disorder of glycosylation by high-resolution mass spectrometry. *Proteomics* **6**, 983-992.
Vakhrushev, S. Y., Langridge, J., Campuzano, I., Hughes, C. & Peter-Katalinic, J. (2008). Ion mobility mass spectrometry analysis of human glycourinome. *Anal Chem* **80**, 2506-2513.
Valentine, S. J., Counterman, A. E. & Clemmer, D. E. (1997). Conformer-dependent proton-transfer reactions of ubiquitin ions. *J Am Soc Mass Spectrom* **8**, 954-961.
Valentine, S. J., Counterman, A. E. & Clemmer, D. E. (1999). A database of 660 peptide ion cross sections: use of intrinsic size parameters for bona fide predictions of cross sections. *J Am Soc Mass Spectrom* **10**, 1188-1211.
Valentine, S. J., Plasencia, M. D., Liu, X., Krishnan, M., Naylor, S., Udseth, H. R., Smith, R. D. & Clemmer, D. E. (2006). Toward plasma proteome profiling with ion mobility-mass spectrometry. *J Proteome Res* **5**, 2977-2984.
van Breukelen, B., Barendregt, A., Heck, A. J. & van den Heuvel, R. H. (2006). Resolving stoichiometries and oligomeric states of glutamate synthase protein complexes with curve fitting and simulation of electrospray mass spectra. *Rapid Commun Mass Spectrom* **20**, 2490-2496.
van de Graaf, R. J. (1928). A new method of determining the mobility of ions or electrons in gases. *Phil Mag* **6**, 210-217.
van den Heuvel, R. H. & Heck, A. J. (2004). Native protein mass spectrometry: from intact oligomers to functional machineries. *Curr Opin Chem Biol* **8**, 519-526.
van den Heuvel, R. H., van Duijn, E., Mazon, H. & other authors (2006). Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. *Anal Chem* **78**, 7473-7483.
van der Schoot, P. & Zandi, R. (2007). Kinetic theory of virus capsid assembly. *Phys Biol* **4**, 296-304.
van Duijn, E., Bakkes, P. J., Heeren, R. M., van den Heuvel, R. H., van Heerikhuizen, H., van der Vies, S. M. & Heck, A. J. (2005). Monitoring macromolecular complexes involved in the chaperonin-assisted protein folding cycle by mass spectrometry. *Nat Methods* **2**, 371-376.
van Duijn, E., Simmons, D. A., van den Heuvel, R. H., Bakkes, P. J., van Heerikhuizen, H., Heeren, R. M., Robinson, C. V., van der Vies, S. M. & Heck, A. J. (2006). Tandem mass spectrometry of intact GroEL-substrate complexes reveals substrate-specific conformational changes in the trans ring. *J Am Chem Soc* **128**, 4694-4702.
van Duijn, E., Barendregt, A., Synowsky, S., Versluis, C. & Heck, A. J. (2009). Chaperonin complexes monitored by ion mobility mass spectrometry. *J Am Chem Soc* **131**, 1452-1459.
van Duijn, E. (2010). Current limitations in native mass spectrometry based structural biology. *J Am Soc Mass Spectrom* **21**, 971-978.
van Gunsteren, W. F., Dolenc, J. & Mark, A. E. (2008). Molecular simulation as an aid to experimentalists. *Curr Opin Struct Biol* **18**, 149-153.
van Regenmortel, M. H. & Mahy, B. W. (2004). Emerging issues in virus taxonomy. *Emerg Infect Dis* **10**, 8-13.
Vellinga, J., Van der Heijdt, S. & Hoeben, R. C. (2005). The adenovirus capsid: major progress in minor proteins. *J Gen Virol* **86**, 1581-1588.
Vigerust, D. J. & Shepherd, V. L. (2007). Virus glycosylation: role in virulence and immune interactions. *Trends Microbiol* **15**, 211-218.
von Helden, G., Gotts, N. G. & Bowers, M. T. (1993a). Experimental evidence for the formation of fullerenes by collisional heating of carbon rings in the gas phase. *Nature* **363**, 60-63.
von Helden, G., Hsu, M. T., Gotts, N. & Bowers, M. T. (1993b). Carbon Cluster Cations with up to 84 Atoms - Structures, Formation Mechanism, and Reactivity. *J Phys Chem* **97**, 8182-8192.
Wang, I. N., Smith, D. L. & Young, R. (2000). Holins: the protein clocks of bacteriophage infections. *Annu Rev Microbiol* **54**, 799-825.
Weigand, K., Knaust, A. & Schaller, H. (2010). Assembly and export determine the intracellular distribution of hepatitis B virus core protein subunits. *J Gen Virol* **91**, 59-67.
Wetzel, R. (2006). Kinetics and thermodynamics of amyloid fibril assembly. *Acc Chem Res* **39**, 671-679.
White, J. M., Delos, S. E., Brecher, M. & Schornberg, K. (2008). Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme. *Crit Rev Biochem Mol Biol* **43**, 189-219.
White, L. J., Ball, J. M., Hardy, M. E., Tanaka, T. N., Kitamoto, N. & Estes, M. K. (1996). Attachment and entry of recombinant Norwalk virus capsids to cultured human and animal cell lines. *J Virol* **70**, 6589-6597.
White, L. J., Hardy, M. E. & Estes, M. K. (1997). Biochemical characterization of a smaller form of recombinant Norwalk virus capsids assembled in insect cells. *J Virol* **71**, 8066-8072.
Whitehouse, C. M., Dreyer, R. N., Yamashita, M. & Fenn, J. B. (1985). Electrospray Interface for Liquid Chromatographs and Mass Spectrometers. *Anal Chem* **57**, 675-679.
Whittaker, G. R., Kann, M. & Helenius, A. (2000). Viral entry into the nucleus. *Annu Rev Cell Dev Biol* **16**, 627-651.
Wild, B. J., Green, B. N., Cooper, E. K., Lalloz, M. R., Erten, S., Stephens, A. D. & Layton, D. M. (2001). Rapid identification of hemoglobin variants by electrospray ionization mass spectrometry. *Blood Cells Mol Dis* **27**, 691-704.
Wiley, W. C. & McLaren, I. H. (1955). Time-of-Flight Mass Spectrometer with Improved Resolution. *Rev Sci Inst* **26**, 1150-1157.
Williams, J. P., Giles, K., Green, B. N., Scrivens, J. H. & Bateman, R. H. (2008). Ion mobility augments the utility of mass spectrometry in the identification of human hemoglobin variants. *Rapid Commun Mass Spectrom* **22**, 3179-3186.
Williams, J. P. & Scrivens, J. H. (2008). Coupling desorption electrospray ionisation and neutral desorption/extractive electrospray ionisation with a travelling-wave based ion mobility mass spectrometer for the analysis of drugs. *Rapid Commun Mass Spectrom* **22**, 187-196.
Wingfield, P. T., Stahl, S. J., Williams, R. W. & Steven, A. C. (1995). Hepatitis core antigen produced in *Escherichia coli*: subunit composition, conformational analysis, and *in vitro* capsid assembly. *Biochemistry* **34**, 4919-4932.
Wolynes, P. G. (1995). Biomolecular folding *in vacuo*!!!(?). *Proc Natl Acad Sci USA* **92**, 2426-2427.
Wu, C., Hill, H. H. J. & Gamerdinger, A. P. (1998). Electrospray ionization-Ion mobility spectrometry as a field monitoring method for the detection of atrazine in natural water. *Field Analytical Chemistry & Technology* **2**, 155-161.
Wynne, S. A., Crowther, R. A. & Leslie, A. G. (1999). The crystal structure of the human hepatitis B virus capsid. *Mol Cell* **3**, 771-780.
Wyttenbach, T., vonHelden, G., Batka, J. J., Carlat, D. & Bowers, M. T. (1997). Effect of the long-range potential on ion mobility measurements. *J Am Soc Mass Spectrom* **8**, 275-282.
Wyttenbach, T., Witt, M. & Bowers, M. T. (2000). On the stability of amino acid zwitterions in the gas phase: The influence of derivatization, proton affinity, and alkali ion addition. *J Am Chem Soc* **122**, 3458-3464.
Wyttenbach, T., Kemper, P. R. & Bowers, M. T. (2001). Design of a new electrospray ion mobility mass spectrometer. *Int J Mass Spectrom* **212**, 13-23.
Xi, J. N., Graham, D. Y., Wang, K. N. & Estes, M. K. (1990). Norwalk virus genome cloning and characterization. *Science* **250**, 1580-1583.
Xie, Y., Zhai, J., Deng, Q., Tiollais, P., Wang, Y. & Zhao, M. (2010). Entry of hepatitis B virus: Mechanism and new therapeutic target. *Pathol Biol (Paris)*.
Yamada, T., Onimatsu, H. & Van Etten, J. L. (2006). Chlorella viruses. *Adv Virus Res* **66**, 293-336.
Yen, H. L. & Webster, R. G. (2009). Pandemic influenza as a current threat. *Curr Top Microbiol Immunol* **333**, 3-24.
Ying, W., Hao, Y., Zhang, Y. & other authors (2004). Proteomic analysis on structural proteins of Severe Acute Respiratory Syndrome coronavirus. *Proteomics* **4**, 492-504.

Zandi, R., van der Schoot, P., Reguera, D., Kegel, W. & Reiss, H. (2006). Classical nucleation theory of virus capsids. *Biophys J* **90**, 1939-1948.
Zanetti, A. R., Van Damme, P. & Shouval, D. (2008). The global impact of vaccination against hepatitis B: a historical overview. *Vaccine* **26**, 6266-6273.
Zeng, R., Ruan, H. Q., Jiang, X. S. & other authors (2004). Proteomic analysis of SARS associated coronavirus using two-dimensional liquid chromatography mass spectrometry and one-dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis followed by mass spectroemtric analysis. *J Proteome Res* **3**, 549-555.
Zhang, L., Zhang, Z. P., Zhang, X. E., Lin, F. S. & Ge, F. (2010a). Quantitative proteomics analysis reveals BAG3 as a potential target to suppress severe acute respiratory syndrome coronavirus replication. *J Virol* **84**, 6050-6059.
Zhang, X., Jin, L., Fang, Q., Hui, W. H. & Zhou, Z. H. (2010b). 3.3 A cryo-EM structure of a nonenveloped virus reveals a priming mechanism for cell entry. *Cell* **141**, 472-482.
Zhang, Y. (2008). Progress and challenges in protein structure prediction. *Curr Opin Struct Biol* **18**, 342-348.
Zhao, J., Trewhella, J., Corbin, J., Francis, S., Mitchell, R., Brushia, R. & Walsh, D. (1997). Progressive cyclic nucleotide-induced conformational changes in the cGMP-dependent protein kinase studied by small angle X-ray scattering in solution. *J Biol Chem* **272**, 31929-31936.
Zheng, H., Olia, A. S., Gonen, M., Andrews, S., Cingolani, G. & Gonen, T. (2008). A conformational switch in bacteriophage p22 portal protein primes genome injection. *Mol Cell* **29**, 376-383.
Zhu, X., Borchers, C., Bienstock, R. J. & Tomer, K. B. (2000). Mass spectrometric characterization of the glycosylation pattern of HIV-gp120 expressed in CHO cells. *Biochemistry* **39**, 11194-11204.
Zilch, L. W., Kaleta, D. T., Kohtani, M., Krishnan, R. & Jarrold, M. F. (2007). Folding and unfolding of helix-turn-helix motifs in the gas phase. *J Am Soc Mass Spectrom* **18**, 1239-1248.
Zlotnick, A., Cheng, N., Conway, J. F., Booy, F. P., Steven, A. C., Stahl, S. J. & Wingfield, P. T. (1996). Dimorphism of hepatitis B virus capsids is strongly influenced by the C-terminus of the capsid protein. *Biochemistry* **35**, 7412-7421.
Zlotnick, A., Johnson, J. M., Wingfield, P. W., Stahl, S. J. & Endres, D. (1999). A theoretical model successfully identifies features of hepatitis B virus capsid assembly. *Biochemistry* **38**, 14644-14652.
Zlotnick, A., Ceres, P., Singh, S. & Johnson, J. M. (2002). A small molecule inhibits and misdirects assembly of hepatitis B virus capsids. *J Virol* **76**, 4848-4854.
Zlotnick, A. (2005). Theoretical aspects of virus capsid assembly. *J Mol Recognit* **18**, 479-490.
Zlotnick, A. (2007). Distinguishing reversible from irreversible virus capsid assembly. *J Mol Biol* **366**, 14-18.

Chapter II

High-resolution mass spectrometry of viral assemblies: Molecular Composition and stability of dimorphic Hepatitis B virus capsids

Charlotte Uetrecht, Cees Versluis, Norman R. Watts, Wouter H. Roos, Gijs J.L. Wuite, Paul T. Wingfield, Alasdair C. Steven, Albert J.R. Heck

Proc Natl Acad Sci USA **105**, 9216-9220 (2008)

Abbreviations: $3^{C\rightarrow A}$ – a cysteine-free mutant of cp149; 61^{C} – a single cysteine-containing mutant of cp149; AFM – atomic force microscopy; CID – collision induced dissociation; cp – capsid protein; cp149 – cp from amino acid 1-149; EM – electron microscopy; ESI – electrospray ionisation; HBV – Hepatitis B virus; MS – mass spectrometry; *m/z* – mass to charge ratio; Q-ToF – quadrupole time of flight; SD – standard deviation; SEM – standard error of the mean; STEM – scanning transmission electron microscopy; ToF – time of flight

High-resolution mass spectrometry of viral assemblies: Molecular Composition and stability of dimorphic Hepatitis B virus capsids

Hepatitis B Virus (HBV) is a major human pathogen. In addition to its importance in human health, there is growing interest in adapting HBV and other viruses for drug delivery and other nanotechnological applications. In both contexts, precise biophysical characterisation of these large macromolecular particles is fundamental. HBV capsids are unusual in that they exhibit two distinct icosahedral geometries, nominally composed of 90 and 120 dimers with masses of ~ 3 and ~ 4 megadaltons, respectively. Here, a mass spectrometric approach was used to determine the masses of both capsids to within 0.1%. It follows that both lattices are complete, consisting of exactly 180 and 240 subunits. Nanoindentation experiments by atomic force microscopy (AFM) indicate that both capsids have similar stabilities. The data yielded a Young's modulus of ~ 0.4 GPa. This experimental approach, anchored on very precise and accurate mass measurements, appears to hold considerable potential for elucidating the assembly of viruses and other macromolecular particles.

1. Introduction

HBV is a major cause of liver disease in humans, [Blumberg, 1997] with more than 350 million people suffering from chronic infection. For the development of new antiviral drugs, further insight into the replication cycle and assembly pathway of the virus is needed. [Deres *et al.*, 2003] Moreover, there is a growing interest in HBV and other viral particles as vehicles for drug delivery and as platforms for nanoparticle technology. [Singh *et al.*, 2006] In this context, precise biophysical characterisation of these particles represents essential basic information.

HBV has an enveloped virion. Single-stranded viral RNA is packaged into the assembling capsid and, within this compartment, is reverse-transcribed into DNA. [Seeger & Mason, 2000; Steven *et al.*, 2005] The DNA-containing nucleocapsid then proceeds to envelopment. Both *in vivo* and *in vitro*, the capsid protein (cp) forms icosahedral capsids of two sizes, corresponding to triangulation numbers of $T = 3$ and $T = 4$, [Caspar & Klug, 1962] nominally consisting of 180 and 240 subunits, respectively. [Crowther *et al.*, 1994; Dryden *et al.*, 2006; Kenney *et al.*, 1995; Wingfield *et al.*, 1995] cp has a 140-residue N-terminal core domain connected to a 34-residue "protamine domain" by a 10-residue linker. [Watts *et al.*, 2002] The protamine domain binds RNA, while the core domain is necessary and sufficient for capsid assembly. The ratio of $T = 3$ to $T = 4$ capsids produced depends on the length of the linker and the conditions of assembly: the smaller $T = 3$ capsid becomes progressively more abundant as the linker is shortened. [Zlotnick *et al.*, 1996] The building block for capsid formation is a dimer stabilised *via* an intermolecular four-helix bundle [Böttcher *et al.*, 1997; Conway *et al.*, 1997; Wynne *et al.*, 1999] and a disulfide bond within the bundle (Cys^{61}). However, dimerisation and assembly also occur in the absence of the disulfide, *e.g.* when Cys^{61} is replaced with Ala. [Wingfield *et al.*, 1995; Zlotnick *et al.*, 1996] The capsid has protruding spikes at the dimer interfaces that display most of the antigenic epitopes and holes at the symmetry axes which allow infusion of nucleotides for reverse transcription. [Crowther *et al.*, 1994; Wynne *et al.*, 1999] While HBV capsid structure has been closely characterised by cryo-electron microscopy (EM) and X-ray crystallography, [Steven *et al.*, 2005] its biophysical properties and the assembly process are less well understood. [Ceres & Zlotnick, 2002; Kegel & van der Schoot, 2004] Here, we present biophysical data from macromolecular (tandem) mass spectrometry (MS) [Benesch *et al.*, 2007; Heck & van

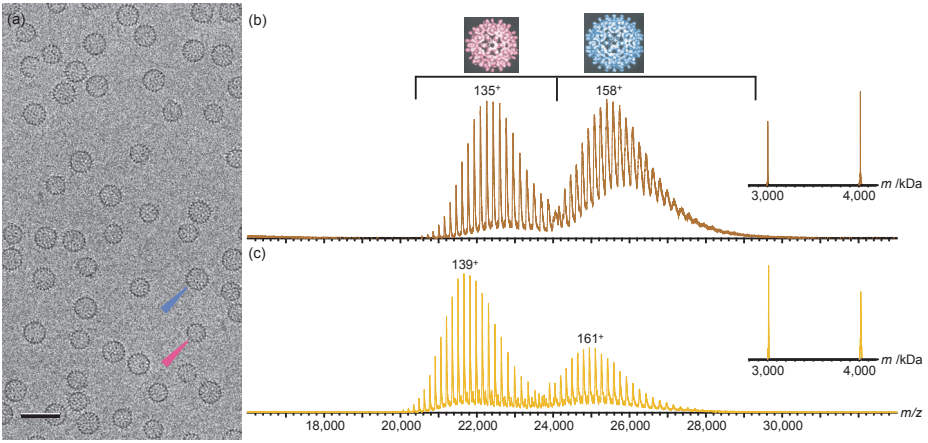


Fig. 1: Electron microscopy and mass spectrometry of HBV capsids
(a) Capsids visualised by cryo-electron microscopy at 38,000 x magnification. Two sizes of capsids are observed; large ones comprised of nominally 120 dimers with $T = 4$ geometry (blue arrowhead), and small ones comprised of nominally 90 dimers with $T = 3$ geometry (pink arrowhead). Bar = 50 nm. (b) Mass spectra of cp149 $3^{C\rightarrow A}$ capsids (composed of dimers without an intramolecular disulfide bond) and (c) cp149 61^{C} capsids (with an intramolecular disulfide bond) at 150 V accelerating voltage. The samples were at ~ 0.04 µM capsids (~ 8 µM monomer of capsid protein) in 200 mM ammonium acetate, pH 6.8. The distributions of peaks around *m/z* 22,000 and 25,000 represent the $T = 3$ and $T = 4$ capsids, respectively, as marked by the corresponding cryo-EM reconstructions in pink and blue. [Zlotnick *et al.*, 1996] For each distribution, the main charge state is indicated. The insets on the right show spectra convoluted to uncharged species. For a comparison of detected peaks with ones calculated for capsids composed of either 89, 90, 119, or 120 dimers, see appendix I Table S1.

den Heuvel, 2004] and AFM that illuminate the molecular composition and physical stability of HBV capsids. This analysis illustrates the potential of emerging biophysical methods, and particularly that of very precise and accurate mass measurements in the megadalton range, to cast new light on long-standing questions in macromolecular assembly as well as assembly products in nanotechnology.

2. Results

2.1. Mass spectrometry of the T = *3 and* T = *4 capsids*

Impressive progress in MS has enabled the measurement of accurate masses of intact macromolecules, contributing unique insights into their compositions, structures, topologies and stabilities. This approach has been applied to complexes as large as transcriptosomes [Lorenzen *et al.*, 2007a] and ribosomes. [Benjamin *et al.*, 1998] In macromolecular mass spectrometry, electrospray ionisation (ESI) is used to generate multiply charged ions of a protein or protein complex, typically yielding ion signals with an approximately Gaussian-shaped charge state distribution. With sufficient mass resolu-

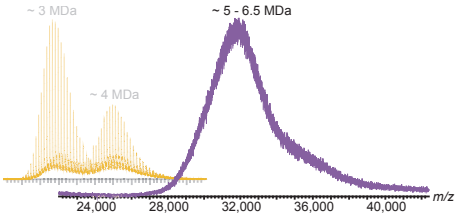


Fig. 2: Mass spectrum of HBV cp183 capsids recorded at 200 V accelerating voltage
The sample was 6 µM (cp monomer) in 200 mM ammonium acetate, pH 6.8. Cp183 incorporates small RNAs upon assembly in *E. coli.* [Wingfield *et al.*, 1995] The amount of RNA per particle is variable, giving rise to heterogeneity in masses so that the resulting mass spectra are not well resolved. They are, nevertheless, highly reproducible, exhibiting a broad maximum at ~ *m/z* 32,000 and a small shoulder on the high *m/z* side. The absence of resolved charge states precludes a mass assignment. Furthermore, no fragmentation products (that could hint at the capsid masses) were observed, even at the highest applicable accelerating voltage (200 V). However, we were able to estimate that the masses are between roughly 5,000 and 6,500 kDa (see 2.1.). The mass spectrum of cp149 61^{C} is given in the back for comparison.

tion, the charge and thus the mass of the analyte can be determined exactly from adjacent peaks in the spectra.

Viruses and capsids have been subjected to ESI-MS. [Fuerstenau *et al.*, 2001; Tito *et al.*, 2000] However, accurate masses could not be determined, due to insufficient mass resolution. Equipped with recent advances in technology, [Benesch *et al.*, 2007; Lorenzen *et al.*, 2007b; van den Heuvel *et al.*, 2006] we set out to analyse *in vitro* assembled HBV capsids, considered as otherwise well characterised test particles in the multi-megadalton range. We analysed the capsids of two cp149 variants (the cp149 construct consists of the core domain plus a nearly complete linker – see above); $3^{C\rightarrow A}$, a cysteine-free mutant of cp149; and 61^{C}, a single cysteine-containing mutant (cys^{61} is involved in the intradimer disulfide bridge). The preparations contained ~ 70% of the larger ($T = 4$) capsids and 30% of the smaller $T = 3$ capsids, and essentially no visible contaminants, as determined by EM (Fig. 1(a)). In our experiments to measure the capsids' molecular weights, spectra of unprecedentedly high mass resolution, with clearly separated charge states, were obtained (Figs. 1(b) and (c)). Both constructs showed a similar pattern with two well-resolved sets of ion signals around mass to charge ratios (*m/z*) of 22,000 and 25,000. These spectra yielded masses of 3,004 ± 3 kDa and 4,006 ± 3 kDa for the $3^{C\rightarrow A}$ $T = 3$ and $T = 4$ capsids, respectively; and 3,012 ± 3 kDa and 4,014 ± 5 kDa for the 61^{C} $T = 3$ and $T = 4$ capsids, respectively. These measurements are in very close agreement with the expected masses of capsids composed of 90 and 120 dimers (3,001 and 4,002 kDa for $3^{C\rightarrow A}$ and 3,007 and 4,009 kDa for 61^{C}). No particles with fewer or more than 90 or 120 dimers were detected (see also appendix I Table S1).

We also analysed cp183 capsids which have the full-length capsid protein. Upon expression and assembly in *Escherichia coli*, cp183 unspecifically packages small RNAs from the host. [Wingfield *et al.*, 1995] Although these particles were also highly purified, high resolution was not achieved in their spectra (*e.g.* Fig. 2), which we attribute to mass heterogeneity arising from the RNA contents. This outcome precluded exact mass measurements. However, from the appearance of the mass spectra of the cp149 capsids (Fig. 1) and assuming a somewhat simplified linear relationship between *m/z* and mass, we estimated masses between 5,000 and 6,500 kDa for the cp183 capsids. These numbers are in good agreement with mass measurements by scanning transmission electron microscopy (STEM) which yielded average values of 4,900 kDa for $T = 3$ cp183 capsids and 6,600 kDa for $T = 4$. [Wingfield *et al.*, 1995]

To further investigate the stoichiometry and stability of HBV capsids, we used macromolecular tandem MS. [Benesch *et al.*, 2007; Heck & van den Heuvel, 2004; Lorenzen *et al.*, 2007b] In this technique, a precursor ion is mass-selected in the first mass analyser where after, in the adjacent collision cell, an additional accelerating voltage collisionally activates the selected precursor ions, inducing their dissociation. In general, in tandem MS of protein complexes single highly charged subunits are ejected from the precursor ions, with the concomitant formation of relatively lower charged counter-complexes. The masses and charges of the fragments should add up to the mass and charge of the precursor. [Heck & van den Heuvel, 2004; van Duijn *et al.*, 2006] Figs. 3(a) and (b) show typical tandem mass spectra for the $T = 3$ $3^{C\rightarrow A}$ and 61^{C} capsids, respectively. Observing the expected conservation of overall mass and charge, the $T = 3$ and $T = 4$ $3^{C\rightarrow A}$ capsid ions were found to fragment *via* the ejection of cp monomers (Fig. 3(a)), even though the capsid proteins are known to be dimers in the particles and in solu-

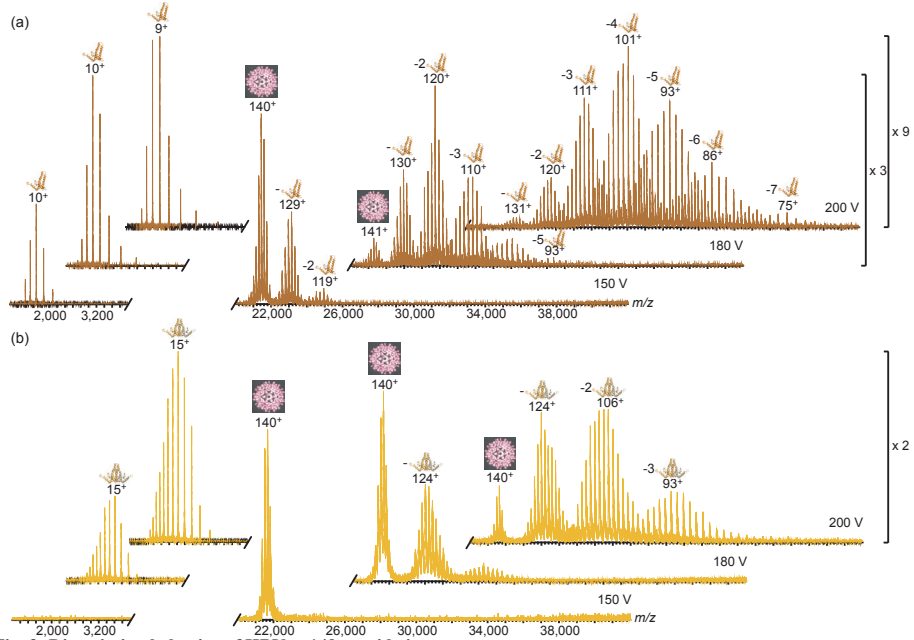


Fig. 3: Dissociation behavior of HBV cp149 capsids *in vacuo*
Tandem mass spectra of $T = 3$ $3^{C\rightarrow A}$ (a) and 61^{C} (B) capsids at *m/z* 21,700 show dissociated monomers ($3^{C\rightarrow A}$) and dimers (61^{C}) in the low mass range. As marker symbols, one or two subunits of the capsid crystal structure [Wynne *et al.*, 1999] are used, although we note that there is no evidence of a monomeric species that has this fold. Undissociated capsids (marked by EM reconstruction) [Zlotnick *et al.*, 1996] and dissociated counter-complexes (number of lost monomers or dimers, as indicated) are observed in the high mass range. Collision voltages were 150 V, 180 V and 200 V from front to back for both constructs. Charge states of the base peaks for each distribution are given. At the highest applicable collision voltage (200 V), $T = 3$ $3^{C\rightarrow A}$ loses up to seven monomers and 61^{C} up to three dimers. The spectra are normalised on their base peaks. Magnification levels in the high mass range are indicated when used.

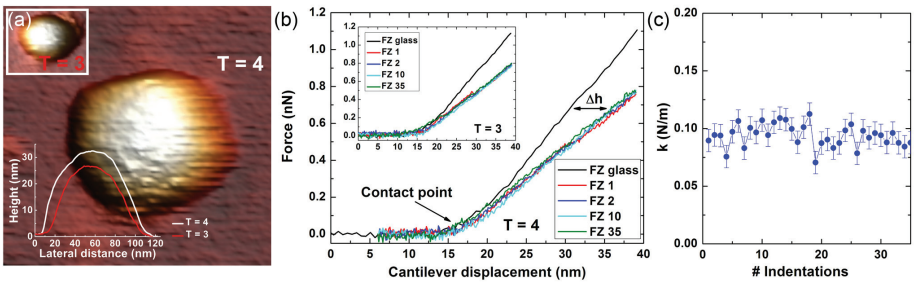


Fig. 4: Atomic force microscopy on HBV capsids
(a) AFM image of a $T = 4$ capsid, with a $T = 3$ particle inset at lower magnification (top left). The corresponding height profiles (bottom left inset) distinguish between the $T = 3$ and $T = 4$ capsids, detecting the 15% difference in diameter. The lateral apparent diameters are smeared by AFM tip convolution effects. (b) Example of force-distance (*FZ*) curves for a $T = 4$ and a $T = 3$ (inset) particle. The reference curve "*FZ* glass" is shown together with the 1st, 2nd, 10th and 35th *FZ* curves. Particle indentation is denoted by Δh. We measured average spring constants of 0.143 ± 0.007 N/m (± SEM, $N = 55$) and 0.093 ± 0.007 N/m (± SEM, $N = 46$) for the $T = 3$ and $T = 4$ capsids. (c) Spring constant k for one of the particles as a function of the indentation cycle. To test if subunits can be removed, the capsids were subjected to a sequence of small indentations. Even after 35 indentations, their stiffness was unchanged, hence it can be concluded that no subunits were removed. Furthermore, no significant decrease in height was observed after 35 indentations. Error bars show the standard deviation.

tion. In contrast, 61^{C} capsids fragment *via* the loss of dimers (Fig. 3(b)), presumably on account of the covalent disulfide bond. On measuring the energy-resolved fragmentation behavior of the capsids by tandem MS, we found that ejection of dimers from 61^{C} capsids required higher collision voltages, compared to the loss of cp monomers from $3^{C\rightarrow A}$ capsids. However, the RNA-containing cp183 capsids could not be dissociated in tandem MS (Fig. 2), probably on account of multiple interdimer disulfide bridges and RNA-protein interactions. [Wingfield *et al.*, 1995]

2.2. Physical stability of the T = *3 and* T = *4 capsids*

To date, little is known about the relative stability of the two capsids, in part because they have similar sizes and surface properties, making them hard to separate. However, the proposition that closer proximity of adjacent pentamers in polyhedral capsids is associated with reduced stability [Albertazzi *et al.*, 1999] raised the possibility that the $T = 3$ capsid may be less stable than $T = 4$. To investigate the mechanical properties of HBV capsids in solution, we used AFM to perform nanoindentation experiments on $3^{C\rightarrow A}$ capsids. [Ivanovska *et al.*, 2004; Roos *et al.*, 2007] The size difference between $T = 3$ and $T = 4$ (~ 15%) is readily detectable in height (above substrate) measurements made by AFM (Fig. 4(a)). After imaging, the capsids were indented with the AFM tip to determine the particle spring constant. Thin shell theory combined with finite element modelling has shown that, for particles similar in size to HBV capsids, the Young's modulus E may be calculated from the spring constant and the particle dimensions. [Gibbons & Klug, 2007; Landau & Lifshitz, 1986; Roos *et al.*, 2007] This analysis yielded $E = 0.37 \pm 0.02$ GPa (SEM, $N = 55$) and $E = 0.36 \pm 0.03$ GPa ($N = 46$) for the $T = 3$ and $T = 4$ particles, respectively. These values, essentially the same, are comparable to the E-moduli of soft plastics such as PTFE (Teflon). Additionally, these experiments showed that irreversible deformation occurs above an indentation force of 0.98 ± 0.06 nN (SEM, $N = 64$), with no significant difference between $T = 3$ and $T = 4$ capsids. To test for the dislodging of subunits, we performed series of small consecutive indentations (force < 0.8 nN) on both capsids (Figs. 4(b) and (c)). Such treatments have been shown to result in small fractures in and/or capsid protein removal from other capsids. [Ivanovska *et al.*, 2004] However, no irreversible damage could be induced in either capsid ($N = 11$). By all these criteria, therefore, the mechanical properties of $T = 3$ and $T = 4$ capsids are indistinguishable.

3. Discussion

The majority of fundamental biological processes are performed not by individual macromolecules but by large assemblies whose masses run into the multi-megadalton range. The present results demonstrate that very precise and accurate measurements may be performed by MS for particles of up to 4 MDa, at least. The high resolution achieved in the mass spectra reflects, in part, optimisation of experimental factors in instrumentation and sample preparation. In the modified Q-ToF instrument, [van den Heuvel *et al.*, 2006] the pressure conditions in the different chambers have been adjusted to improve the desolvation of the ions in the electrospray ionisation process, and to improve transmission of these large mass ions through the spectrometer. [Benesch *et al.*, 2007; Tahallah *et al.*, 2001; van Duijn *et al.*, 2005] However, even with these modifications, the peak width and thus the mass resolution are largely determined by the homogeneity and purity of the sample. In this study, sample homogeneity was checked by collision induced dissociation (CID) which revealed no detectable heterogeneity other than the expected dimorphism. Issues of homogeneity become more important as the stoichiometry of

a macromolecular complex increases. As pointed out by Benesch *et al.,* [Benesch *et al.*, 2006] the probability of an oligomer being composed entirely of pure monomer, P_O, depends on the monomer purity, P_M, according to: $P_O\,[\%] = P_M\,[\%]^{\,n} / 100^{\,n-1}$. For a capsid of 240 subunits with a monomeric purity of 99% the intact capsid purity would drop to less than 9% (with a monomer purity of 99.9% this would be 79%). The tendency of many biological assemblies to be naturally heterogeneous due to posttranslational modifications or binding diverse cofactors (*e.g.* regulatory proteins, RNA, lipids) may broaden the ion signals in their mass spectra as was observed for the cp183 capsids. With the HBV cp149 capsids, the homogeneity of the cp monomers combined with the bidispersity of their assembly products facilitated analysis at high mass resolution.

3.1. HBV capsids are complete

In the assembly of a polymeric particle of predefined size, one may distinguish three stages: initiation, elongation, and completion. Of these, completion has been little studied because, experimentally, it is relatively inaccessible. Nevertheless, it has been discussed in the context of kinetic trapping in reversible and irreversible assembly pathways based on multiple weak interactions. [Zlotnick, 2007] Moreover, an extreme case of non-completion has been documented in the *in vitro* reassembly of bacteriophage HK97 procapsids (which are $T = 7$ icosahedral particles) with a specific point mutation in the capsid protein: in this reaction, completion never takes place and the capsids' vertices remain vacant. [Li *et al.*, 2005; Ross *et al.*, 2006]

For *in vitro* assembly of capsids from HBV core domain dimers, kinetic and other data indicate that the initiation complex is a trimer of dimers. [Zlotnick *et al.*, 1999] (Initiation *in vivo* is a more complicated process. [Steven *et al.*, 2005]) Elongation involves the addition of dimers, but completion has not been previously addressed. In principle, insertion of the last dimer into an assembling lattice faces a steric challenge with low tolerances, raising the question of whether such insertion actually takes place or is kinetically prohibited. If HBV capsids were to have vacancies at one or a few sites, they would not be detected in either X-ray crystallography or cryo-EM reconstructions on account of being implicitly averaged with occupied sites. Mass measurements of capsids by STEM yielded values that were consistent with the 90-dimer and 120-dimer scenario [Wingfield *et al.*, 1995] but unable to rule out the possibility of one or a few missing dimers. The present analysis settles this question.

3.2. Future prospects

The capability to perform very precise mass measurements makes it possible to determine the exact molecular stoichiometries of complexes in this size range (0.1% uncertainty in the mass of a 4 MDa particle is generally less than a single protein subunit). The prospect of characterising other viruses and macromolecular assemblies – both native and synthetic - with this exactitude opens up a new horizon in biomolecular analysis. For example, in the context of HBV assembly, mass analyses of capsids assembled from subunits of different sizes (*e.g.* with C-terminal appendages) should yield information about assembly intermediates; incubation of homomeric capsids with dimer pools of subunits with different masses should allow rates of exchange ("breathing") to be measured; and time-resolved monitoring of the assembly process may also reveal stable intermediates. Similarly, capsids' contents of minor proteins which may be present in only a few copies per virion but are nevertheless of functional importance may now be quantitated with confidence.

4. Materials and methods

4.1. Proteins

The cp149 $3^{C\rightarrow A}$ and cp149 61^{C} capsids used for MS were prepared as described previously. [Zlotnick *et al.*, 1996] Cp183 capsids were purified without dissociation, [Wingfield *et al.*, 1995] as were the cp149 $3^{C\rightarrow A}$ capsids used for AFM.

4.2. Electron microscopy

Negative staining EM was performed as described. [Watts *et al.*, 2002] These micrographs were used to determine the $T = 4$ to $T = 3$ ratio. Cryo-EM was performed as described by Dr. N. Cheng. [Wingfield *et al.*, 1995]

4.3. Mass spectrometry

High resolution and tandem mass spectra were recorded on a modified Q-TOF 1 instrument (Waters, UK) [van den Heuvel *et al.*, 2006] in positive ion mode. Xenon was used as collision gas to increase transmission of HBV capsids. [Lorenzen *et al.*, 2007b] The voltages and pressures were optimised for large noncovalent protein complexes. [Sobott *et al.*, 2002] Briefly, the capillary and cone voltages were kept at 1350 and 150 V, respectively. The collision or accelerating voltage was varied from 10 to 200 V, while the xenon pressure was 1.5×10^{-2} mbar. The pressure in the source region was increased to approximately 10 mbar. Capsids assembled from cp149 $3^{C\rightarrow A}$, cp149 61^{C} and cp183 were introduced to the mass spectrometer at 6 - 8 μM monomer concentrations in 200 mM ammonium acetate, pH 6.8. Masses were determined and averaged from all detected peaks in a charge state distribution. Mass assignments of at least three independent measurements were averaged again and their standard deviation is given. The determination of the mass error from the peak width leads to similar uncertainties whereas the error within a mass assignment is lower. The performance of the instrument was tested with CsI and a calibration was applied where necessary. Capillaries for ESI were prepared in-house from borosilicate glass tubes of 1.2 mm OD and 0.68 mm ID with filament (World Precision Instruments, USA), using a P-97 micropipette puller (Sutter Instruments, USA), and gold-coated using a Scancoat six Pirani 501 sputter coater (Edwards Laboratories, USA). Capillary tips were opened on the sample cone of the instrument.

4.4. Atomic Force Microscopy

AFM experiments were conducted in liquid with a Nanotec Electronica (Madrid, Spain) instrument in jumping mode, [de Pablo *et al.*, 1998] using rectangular gold-coated cantilevers (Olympus, Japan) with a tip apex nominal value of < 20 nm. The cantilevers, calibrated by the method of Sader *et al.*, [Sader *et al.*, 1999] had a spring constant of 0.52 ± 0.002 (SD) N/m. Cover slips were cleaned by immersion in an 86% (v/v) ethanol solution saturated with KOH for 14 h and subsequently rinsed thoroughly with deionised water. After drying, they were rendered hydrophobic by placing them in a sealed container under a saturated hexamethyldisilazane (Fluka, Germany) vapor pressure for 14 h. [Ivanovska *et al.*, 2004] Imaging was performed at room temperature. A droplet of capsid suspension at 4.2 μM monomer concentration in 50 mM Tris at pH 7.5 was incubated for 20 min on a hydrophobic glass cover slip before starting the measurements.

The Young's modulus calculation was performed with values for the shell thickness $t = 2.4$ nm (2.1 nm) and average radius $R_{average} = 11.9$ nm (13.55 nm) for the $T = 3$ ($T = 4$) particles. [Zlotnick *et al.*, 1996] The radii are taken without the spikes as these are assumed not to contribute significantly to the overall mechanical properties of the contiguous shells. Similarly, although the holes make up ~ 16% of the capsid surface, [Zlotnick *et al.*, 1996] we approximated the capsids as closed hollow spheres.

References

Albertazzi, E., Domene, C., Fowler, P. W., Heine, T., Seifert, G., Van Alsenoy, C. & Zerbetto, F. (1999). Pentagon adjacency as a determinant of fullerene stability. *Phys Chem Chem Phys* **1**, 2913-2918.

Benesch, J. L., Aquilina, J. A., Ruotolo, B. T., Sobott, F. & Robinson, C. V. (2006). Tandem mass spectrometry reveals the quaternary organiza-

tion of macromolecular assemblies. *Chem Biol* **13**, 597-605.
Benesch, J. L., Ruotolo, B. T., Simmons, D. A. & Robinson, C. V. (2007). Protein complexes in the gas phase: technology for structural genomics and proteomics. *Chem Rev* **107**, 3544-3567.
Benjamin, D. R., Robinson, C. V., Hendrick, J. P., Hartl, F. U. & Dobson, C. M. (1998). Mass spectrometry of ribosomes and ribosomal subunits. *Proc Natl Acad Sci USA* **95**, 7391-7395.
Blumberg, B. S. (1997). Hepatitis B virus, the vaccine, and the control of primary cancer of the liver. *Proc Natl Acad Sci USA* **94**, 7121-7125.
Böttcher, B., Wynne, S. A. & Crowther, R. A. (1997). Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. *Nature* **386**, 88-91.
Caspar, D. L. & Klug, A. (1962). Physical principles in the construction of regular viruses. *Cold Spring Harb Symp Quant Biol* **27**, 1-24.
Ceres, P. & Zlotnick, A. (2002). Weak protein-protein interactions are sufficient to drive assembly of hepatitis B virus capsids. *Biochemistry* **41**, 11525-11531.
Conway, J. F., Cheng, N., Zlotnick, A., Wingfield, P. T., Stahl, S. J. & Steven, A. C. (1997). Visualization of a 4-helix bundle in the hepatitis B virus capsid by cryo-electron microscopy. *Nature* **386**, 91-94.
Crowther, R. A., Kiselev, N. A., Bottcher, B., Berriman, J. A., Borisova, G. P., Ose, V. & Pumpens, P. (1994). Three-dimensional structure of hepatitis B virus core particles determined by electron cryomicroscopy. *Cell* **77**, 943-950.
de Pablo, P. J., Colchero, J., Gomez-Herrero, J. & Baro, A. M. (1998). Jumping mode scanning force microscopy. *Appl Phys Lett* **73**, 3300-3302.
Deres, K., Schroder, C. H., Paessens, A. & other authors (2003). Inhibition of hepatitis B virus replication by drug-induced depletion of nucleocapsids. *Science* **299**, 893-896.
Dryden, K. A., Wieland, S. F., Whitten-Bauer, C., Gerin, J. L., Chisari, F. V. & Yeager, M. (2006). Native hepatitis B virions and capsids visualized by electron cryomicroscopy. *Mol Cell* **22**, 843-850.
Fuerstenau, S. D., Benner, W. H., Thomas, J. J., Brugidou, C., Bothner, B. & Siuzdak, G. (2001). Mass spectrometry of an intact virus. *Angew Chem Int Ed Engl* **40**, 542-544.
Gibbons, M. M. & Klug, W. S. (2007). Nonlinear finite-element analysis of nanoindentation of viral capsids. *Phys Rev E Stat Nonlin Soft Matter Phys* **75**, 031901.
Heck, A. J. & van den Heuvel, R. H. (2004). Investigation of intact protein complexes by mass spectrometry. *Mass Spectrom Rev* **23**, 368-389.
Ivanovska, I. L., de Pablo, P. J., Ibarra, B., Sgalari, G., MacKintosh, F. C., Carrascosa, J. L., Schmidt, C. F. & Wuite, G. J. (2004). Bacteriophage capsids: tough nanoshells with complex elastic properties. *Proc Natl Acad Sci U S A* **101**, 7600-7605.
Kegel, W. K. & van der Schoot, P. (2004). Competing hydrophobic and screened-coulomb interactions in hepatitis B virus capsid assembly. *Biophys J* **86**, 3905-3913.
Kenney, J. M., von Bonsdorff, C. H., Nassal, M. & Fuller, S. D. (1995). Evolutionary conservation in the hepatitis B virus core structure: comparison of human and duck cores. *Structure* **3**, 1009-1019.
Landau, L. D. & Lifshitz, E. M. (1986). *Theory of Elasticity*, 3rd edn. Oxford: Elsevier.
Li, Y., Conway, J. F., Cheng, N., Steven, A. C., Hendrix, R. W. & Duda, R. L. (2005). Control of virus assembly: HK97 "Whiffleball" mutant capsids without pentons. *J Mol Biol* **348**, 167-182.
Lorenzen, K., Vannini, A., Cramer, P. & Heck, A. J. (2007a). Structural biology of RNA polymerase III: mass spectrometry elucidates subcomplex architecture. *Structure* **15**, 1237-1245.
Lorenzen, K., Versluis, C., van Duijn, E., van den Heuvel, R. H. H. & Heck, A. J. R. (2007b). Optimizing macromolecular tandem mass spectrometry of large non-covalent complexes using heavy collision gases. *Int J Mass Spectrom* **268**, 198-206.
Roos, W. H., Ivanovska, I. L., Evilevitch, A. & Wuite, G. J. (2007). Viral capsids: mechanical characteristics, genome packaging and delivery mechanisms. *Cell Mol Life Sci* **64**, 1484-1497.
Ross, P. D., Conway, J. F., Cheng, N., Dierkes, L., Firek, B. A., Hendrix, R. W., Steven, A. C. & Duda, R. L. (2006). A free energy cascade with locks drives assembly and maturation of bacteriophage HK97 capsid. *J Mol Biol* **364**, 512-525.
Sader, J. E., Chon, J. W. M. & Mulvaney, P. (1999). Calibration of rectangular atomic force microscope cantilevers. *Rev Sci Instrum* **70**, 3967-3969.
Seeger, C. & Mason, W. S. (2000). Hepatitis B virus biology. *Microbiol Mol Biol Rev* **64**, 51-68.
Singh, P., Gonzalez, M. J. & Manchester, M. (2006). Viruses and their uses in nanotechnology. *Drug Dev Res* **67**, 23-41.
Sobott, F., Hernandez, H., McCammon, M. G., Tito, M. A. & Robinson, C. V. (2002). A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Anal Chem* **74**, 1402-1407.
Steven, A. C., Conway, J. F., Cheng, N., Watts, N. R., Belnap, D. M., Harris, A., Stahl, S. J. & Wingfield, P. T. (2005). Structure, assembly, and antigenicity of hepatitis B virus capsid proteins. *Adv Virus Res* **64**, 125-164.
Tahallah, N., Pinkse, M., Maier, C. S. & Heck, A. J. (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 Commun Mass Spectrom* **15**, 596-601.
Tito, M. A., Tars, K., Valegard, K., Hajdu, J. & Robinson, C. V. (2000). Electrospray time-of-flight mass spectrometry of the intact MS2 virus capsid. *J Am Chem Soc* **122**, 3550-3551.
van den Heuvel, R. H., van Duijn, E., Mazon, H. & other authors (2006). Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. *Anal Chem* **78**, 7473-7483.
van Duijn, E., Bakkes, P. J., Heeren, R. M., van den Heuvel, R. H., van Heerikhuizen, H., van der Vies, S. M. & Heck, A. J. (2005). Monitoring macromolecular complexes involved in the chaperonin-assisted protein folding cycle by mass spectrometry. *Nat Methods* **2**, 371-376.
van Duijn, E., Simmons, D. A., van den Heuvel, R. H., Bakkes, P. J., van Heerikhuizen, H., Heeren, R. M., Robinson, C. V., van der Vies, S. M. & Heck, A. J. (2006). Tandem mass spectrometry of intact GroEL-substrate complexes reveals substrate-specific conformational changes in the trans ring. *J Am Chem Soc* **128**, 4694-4702.
Watts, N. R., Conway, J. F., Cheng, N., Stahl, S. J., Belnap, D. M., Steven, A. C. & Wingfield, P. T. (2002). The morphogenic linker peptide of HBV capsid protein forms a mobile array on the interior surface. *Embo J* **21**, 876-884.
Wingfield, P. T., Stahl, S. J., Williams, R. W. & Steven, A. C. (1995). Hepatitis core antigen produced in *Escherichia coli*: subunit composition, conformational analysis, and *in vitro* capsid assembly. *Biochemistry* **34**, 4919-4932.
Wynne, S. A., Crowther, R. A. & Leslie, A. G. (1999). The crystal structure of the human hepatitis B virus capsid. *Mol Cell* **3**, 771-780.
Zlotnick, A., Cheng, N., Conway, J. F., Booy, F. P., Steven, A. C., Stahl, S. J. & Wingfield, P. T. (1996). Dimorphism of hepatitis B virus capsids is strongly influenced by the C-terminus of the capsid protein. *Biochemistry* **35**, 7412-7421.
Zlotnick, A., Johnson, J. M., Wingfield, P. W., Stahl, S. J. & Endres, D. (1999). A theoretical model successfully identifies features of hepatitis B virus capsid assembly. *Biochemistry* **38**, 14644-14652.
Zlotnick, A. (2007). Distinguishing reversible from irreversible virus capsid assembly. *J Mol Biol* **366**, 14-18.

Chapter III

Stability and shape of Hepatitis B virus capsids *in vacuo*

Charlotte Uetrecht, Cees Versluis, Norman R. Watts, Paul T. Wingfield, Alasdair C. Steven, Albert J.R. Heck

Angew Chem Int Ed Engl **47**, 6247-6251 (2008)

Chapter IV

Norwalk virus assembly and stability monitored by mass spectrometry

Glen K. Shoemaker, Esther van Duijn, Sue E. Crawford, Charlotte Uetrecht, Marian Baclayon, Wouter H. Roos, Gijs J.L. Wuite, Mary K. Estes, B.V. Venkataram Prasad and Albert J.R. Heck

This research was originally published in *Mol Cell Proteomics* **9**, 1742-1751 (2010)

Abbreviations: AFM – atomic force microscopy; CID – collision induced dissociation; EM – electron microscopy; ESI – electrospray ionisation; HBV – Hepatitis B virus; IMMS – ion mobility mass spectrometry; MS – mass spectrometry; NVLP – recombinant norovirus-like particles; *m/z* – mass to charge ratio; *Ω* – collision cross section; Q-ToF – quadrupole time of flight; Sf9 cells – *Spodopter frugiperda* cells; VP – viral protein

Norwalk virus assembly and stability monitored by mass spectrometry

Viral capsid assembly, in which viral proteins self-assemble into complexes of well-defined architecture, is a fascinating biological process. Although viral structure and assembly processes have been the subject of many excellent structural biology studies in the past, questions still remain regarding the intricate mechanisms that underlie viral structure, stability, and assembly. Here we use native mass spectrometry (MS) based techniques to study the structure, stability and assembly of recombinant norovirus-like particles (NVLP). While there exists detailed structural information on the fully assembled capsid, less information is available on potential capsid (dis)assembly intermediates, largely due to the inherent heterogeneity and complexity of the (dis)assembly pathways. We use native MS and atomic force microscopy (AFM) to investigate the (dis)assembly of the NVLPs as a function of solution pH, ionic strength, and VP1 protein concentration. Native MS analysis at physiological pH revealed the presence of the complete capsid ($T = 3$) consisting of 180 copies of VP1. The mass of these capsid particles extends over 10 MegaDa, ranking them amongst the largest protein complexes ever analysed by native MS. While very stable under acidic conditions, the capsid was found to be sensitive to alkaline treatment. At elevated pH, intermediate structures consisting of 2, 4, 6, 18, 40, 60 and 80 copies of VP1 were observed, with the $VP1_{60}$ (3.36 MDa) and $VP1_{80}$ (4.48 MDa) species being most abundant. AFM imaging and ion mobility mass spectrometry (IMMS) confirmed the formation of these latter midsize spherical particles at elevated pH. All these VP1 oligomers could be reversely assembled into the original capsid ($VP1_{180}$). From the MS data collected over a range of experimental conditions, we suggest a disassembly model in which the $T = 3$ $VP1_{180}$ particles dissociate into smaller oligomers upon alkaline treatment, predominantly dimers, prior to reassembly into $VP1_{60}$ and $VP1_{80}$ species.

1. Introduction

Accounting for most cases of non-bacterial gastroenteritis, the norovirus represents an important human pathogen. [Atmar & Estes, 2006; Estes *et al.*, 2006] It is the most predominant pathogen within the family, *Caliciviridae*, which also includes *Sapovirus*, *Vesivirus*, and *Lagovirus*. [Green *et al.*, 2000] The prototypical strain of the human noroviruses is the Norwalk virus. It is a small (7.7 kb genome) non-enveloped, single stranded RNA virus. Its genome contains three open reading frames, encoding for the major capsid protein (VP1), the minor capsid protein (VP2) and a non-structural polyprotein. [Jiang *et al.*, 1993; Xi *et al.*, 1990] VP1 forms homodimers and the mature norovirus capsids ($T = 3$) are composed of 90 VP1 dimers, [Ausar *et al.*, 2006; Prasad *et al.*, 1994] and possibly a few copies of VP2 that are thought to stabilise the icosahedral structure as well as effecting the expression of VP1. [Ausar *et al.*, 2006; Bertolotti-Ciarlet *et al.*, 2003] Due to a lack of suitable animal models or *in vitro* cell culture systems, structural studies so far have been largely focused on NVLP, which are spontaneously assembled during the expression of recombinant VP1 and VP2 in insect cells. [Xi *et al.*, 1990] Importantly, these empty noninfectious particles have been demonstrated to be morphologically and antigenically similar to the genuine virion. [Xi *et al.*, 1992]

The NVLPs have been studied extensively using X-ray crystallography and electron microscopy (EM), which have provided a detailed image of the intact capsid, revealing the $T = 3$ icosahedral organisation. [Bertolotti-Ciarlet *et al.*, 2002; Prasad *et al.*, 1994; Prasad *et al.*, 1999; Xi *et al.*, 1992] The VP1 monomer structure is principally composed of two domains; an S domain

consisting of the 225 N-terminal residues and a C-terminal P domain. In the intact capsid, the S domain forms a contiguous protein shell with a diameter of ~30 nm, while the P domain forms prominent protrusions, which give the NVLPs a diameter of ~40 nm. A remarkable feature of the NVLPs is that a single protein is responsible for directing capsid assembly and host interactions. The NVLPs thus represent a simple model to study the assembly of iscosahedral viruses. While the requirements for capsid assembly have been investigated previously, [Ausar *et al.*, 2006; Bertolotti-Ciarlet *et al.*, 2002] there is little information regarding intermediates along the (dis)assembly pathway. Obtaining such information can be quite difficult, owing to the inherent heterogeneity of capsid assembly. An emerging technique for interrogating such heterogeneous protein assemblies is native electrospray ionisation (ESI) MS.

Long regarded as a tool for small molecule analysis and more recently proteomics investigations, the utility of MS in structural biology is increasingly applied and accepted. [Heck, 2008; Ruotolo *et al.*, 2005; Sharon & Robinson, 2007; Uetrecht *et al.*, 2008a] Native MS exploits the gentle ionisation conditions afforded by ESI to transfer intact noncovalently bound protein assemblies into the gas phase. Determining the mass of these complexes with high accuracy allows the oligomeric stoichiometry to be unambiguously deduced. Traditionally challenging targets for structural biology, including complexes in the megadalton range, [Kaddis *et al.*, 2007; Uetrecht *et al.*, 2008a; Uetrecht *et al.*, 2008b] heterogeneous or polydisperse, [Synowsky *et al.*, 2006; van Duijn *et al.*, 2005] and membrane-bound protein assemblies [Barrera *et al.*, 2009] can now be interrogated in this manner. Furthermore, selective dissociation of these assemblies in both the gas and solution phases allows the designation of sub-complexes, noncovalently bound species smaller than the original protein complex. Combining the knowledge obtained from such data provides information regarding subunit organisation at both the architecture and sub-architecture level, allowing the generation of low-resolution maps of the overall three-dimensional structure of protein complexes. [Lorenzen *et al.*, 2007a; Taverner *et al.*, 2008; Zhou *et al.*, 2008] Additional information can also be obtained through the combination of tandem MS techniques. Collision induced dissociation (CID), for example, can be used to selectively dissociate specific protein assemblies and thus provide information regarding the stability and aid in the assignment of stoichiometry for a given complex. Another tandem MS approach, IMMS, provides additional information regarding the shape of gaseous protein complexes. In IMMS, in addition to separation based on their mass to charge ratio (*m/z*), ions are also passed through an ion mobility cell with a counter flow of neutral background gas where they are separated based on their size and shape. [Hoaglund *et al.*, 1998; Ruotolo *et al.*, 2005]

The ability to perform mass measurements of intact viruses has been exploited by several groups but is often limited by mass resolution, which is impeded by the incomplete desolvation of the large protein assemblies during the ionisation process. Siuzdak [Siuzdak *et al.*, 1996] and Robinson [Tito *et al.*, 2000] pioneered the analyses of viruses using MS. More recently Uetrecht *et al.* reported ESI-MS data on the Hepatitis B virus (HBV) capsid. [Uetrecht *et al.*, 2008a; Uetrecht *et al.*, 2008b] In these studies sufficient mass resolution was obtained to determine the accurate mass and stoichiometry of the $T = 3$ and $T = 4$ HBV capsids, despite their large mass of 3 and 4 million Da, respectively. In addition to being able to measure the mass and stoichiometry of protein assemblies, the capacity of native MS to analyse simultaneously a heterogeneous population of assembly intermediates makes it a powerful technique to study virus assembly. [Morton *et al.*, 2008]

In the work described here, the disassembly of NVLPs is monitored over a range of solution conditions using native ESI-MS, providing insights into their stability and factors which govern icosahedral assembly for this model calicivirus. Unravelling the details of these complex structures and the associated self-assembly pathways that lead to their efficient and precise construction may play an important role in the development of anti-viral therapeutics and in the field of nanotechnology where there is much interest in the fundamentals of particle self-assembly.

2. Materials and methods

2.1. Preparation of the NVLPs for ESI-MS analysis

The native ESI-MS measurements were carried out using recombinant VP1 which was expressed in *Spodopter frugiperda* (Sf9) cells using a baculovirus expression system and purified as described previously. [Xi *et al.*, 1992] Briefly, Sf9 cells were harvested 5-7 days after baculovirus infection and then purified using centrifugation with a CsCl gradient of 1.362 g/cm^3. The purified NVLPs (400 μM VP1) were stored in water at 4 ºC. For ESI-MS analysis, the NVLPs were exchanged into an aqueous ammonium acetate buffer (50-500 mM) at various pH values using an Amicon Ultra 0.5 ml centrifugal filter (Millipore, USA) with a molecular weight cutoff of 10 kDa. The pH of the aqueous ammonium acetate solution was adjusted by using either an aqueous solution of ammonia or acetic acid. All measurements at the different experimental conditions were repeated at least three times on different days. Additionally, results from two different preparations of reconstituted NVLPs yielded consistent results.

2.2. Mass spectrometry

High-resolution and tandem mass spectra were recorded on a modified Q-ToF 1 instrument (Waters, UK) in positive ion mode. [van den Heuvel *et al.*, 2006] To enhance the transmission of the large ions corresponding to NVLPs, xenon, at a pressure of 2 x 10^{-2} mbar, was used in the collision cell. [Lorenzen *et al.*, 2007b] The voltages and pressures were also optimised for large noncovalent protein complexes. [Sobott *et al.*, 2002] Briefly, the capillary and cone voltages were kept constant at 1450 V and 165 V, respectively. The voltage before the collision cell was varied from 10 to 400 V, but generally left at 50 V for the accumulation of native ESI mass spectra. Ions were introduced into the source under an elevated pressure of 10 mbar. Ion mobility measurements were performed on a Synapt HDMS (Waters, UK). [Pringle *et al.*, 2007] To generate intact gas-phase ions from large protein complexes in solution, the source was maintained at 6.1 mbar and voltages of 1400 V and 164 V were applied to the capillary and sample cone, respectively. Xenon was used as the background gas in the trap and transfer ion guides at a flow rate of 4 ml/min. The voltages in the trap and transfer were 20 V and 25 V, respectively. The gas in the ion mobility cell was nitrogen at a flow rate of 25 ml/min and a ramped wave height of 10 – 30 V with a velocity of 250 m/s. Ramped wave heights were previously shown to provide better results for protein complexes in the megadalton range. [Uetrecht *et al.*, 2008b] Collision cross sections (Ω) were determined from the measured drift times through calibration using proteins of known cross sections as described previously. [Ruotolo *et al.*, 2008; Uetrecht *et al.*, 2008b; van Duijn *et al.*, 2009] Denatured ESI-MS analysis of the VP1 monomers was performed on an LCT instrument (Waters, UK). The performance of the instruments was tested with an aqueous CsI solution and a calibration was applied where necessary. ESI tips were prepared in-house from borosilicate glass tubes of 1.2 mm OD and 0.68 mm (World Precision Instruments, USA) by using a P-97 micropipette puller (Sutter Instruments, USA). The ESI tips were gold-coated using a Scancoat six Pirani 501 sputter coater (Edwards Laboratories, USA).

2.3. Atomic force microscopy

AFM experiments were conducted in aqueous ammonium acetate solutions with a Nanotec Electronica (Spain) instrument operating in jumping mode [de Pablo *et al.*, 1998] using rectangular gold-coated cantilevers (Olympus, the Netherlands) with a tip apex nominal value of < 20 nm. The cantilevers were calibrated as described previously [Sader *et al.*, 1999] and were found to have a spring constant of

0.052 ± 0.004 N/m. Glass coverslips were cleaned by immersion in an 86% (v/v) ethanol solution saturated with KOH for 14 h. After being rinsed thoroughly with deionised water and left to dry, the coverslips were rendered hydrophobic by incubating them under a saturated hexamethyldisilazane (Fluka, the Netherlands) vapour for 14 h. [Ivanovska *et al.*, 2004] For imaging, a droplet containing VP1 (0.2 - 10 µM) in an aqueous ammonium acetate buffer (500 mM) at various pH values was placed on the hydrophobic glass coverslip and analysed after a short (5 min) incubation.

3. Results

3.1. ESI-MS analysis of NVLPs

To first investigate whether the intact capsid can be detected using ESI-MS, the NVLPs were analysed at neutral pH, conditions that are known from previous EM studies to be favorable for the formation of intact capsid. [Ausar *et al.*, 2006; Xi *et al.*, 1992] A mass spectrum at pH 7 in a 250 mM aqueous ammonium acetate buffer is presented in Fig. 1(a). A broad distribution of ions is present in the high *m/z* region of the mass spectrum, corresponding to intact NVLPs, as well as a small amount of VP1 dimers (around *m/z* 5000). Since the individual charge states of the capsid ions are unresolved, the mass cannot be precisely assigned. Importantly, however, the ions corresponding to NVLPs are detected in the expected region of the mass spectrum. To demonstrate this, an inset of two other virus capsids analysed by ESI-MS is included in Fig. 1(a). This spectrum corresponds to the two HBV capsids which exhibit $T = 3$ and $T = 4$ symmetry that possess masses of 3 and 4 MDa, respectively. [Uetrecht *et al.*, 2008a; Uetrecht *et al.*, 2008b] This data is also represented as a plot of charge state *vs.* the square root of the mass for several proteins and protein assemblies analysed by ESI-MS, which is expected to be linear according to the charge residue model for ESI. [Felitsyn *et al.*, 2002; Heck & van den Heuvel, 2004] As can be seen in Fig. 1(b), the plot demonstrates excellent linearity and the NVLPs are detected around the expected charge state. With a mass of ~10.1 MDa, the NVLPs are among the largest protein assemblies analysed by native ESI-MS. We attempted to subject these capsid ions to high energy CID to enhance desolvation, but unfortunately the individual charge states remained unresolved. At even

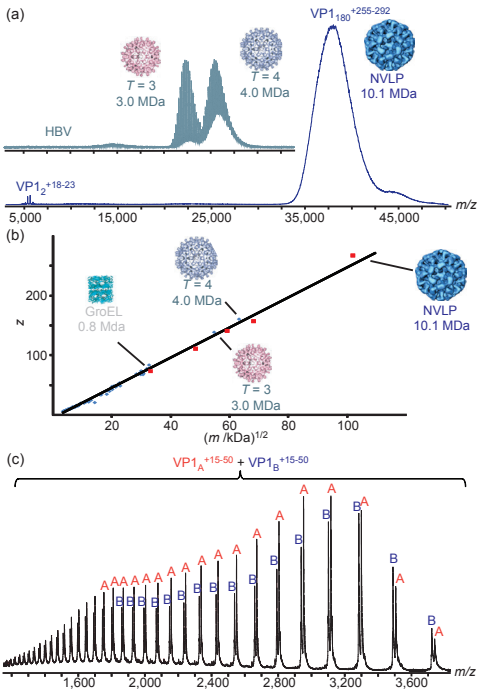


Fig. 1: Native ESI-MS analysis of the NVLPs
(a) Representative native ESI mass spectrum of an aqueous solution containing NVLPs (30 µM VP1) in a 250 mM aqueous ammonium acetate buffer at pH 7. For comparison, a native ESI mass spectrum of the HBV virus capsid is also shown. [Uetrecht *et al.*, 2008a] (b) A plot of charge state *vs* the square root of the mass for several macromolecular assemblies analysed by ESI-MS (grey diamonds), [Lorenzen *et al.*, 2008] including the NVLPs, VP1 oligomers ($VP1_{80}$, $VP1_{60}$, $VP1_{40}$, and $VP1_{18}$) (red cubes) as well as the $T = 3$ and $T = 4$ HBV capsids (grey diamonds). (c) Denatured ESI mass spectrum obtained from a solution containing NVLPs (16 µM VP1) in 20% (v/v) acetonitrile and 1% (v/v) formic acid, demonstrating the mass heterogeneity of the VP1 monomers.

higher voltages in the collision cell, dissociation of the capsid was observed which proceeded with the loss of VP1 monomers and concomitant high mass fragments, which could not be well resolved (data not shown).

To confirm the identity and homogeneity of the major capsid protein (VP1), the NVLPs were also analysed with ESI-MS under denaturing conditions. The NVLPs were denatured in an acetonitrile/water solution that was acidified using formic acid (Fig. 1(c)). The denatured mass spectrum revealed the existence of two nearly equally abundant VP1 monomer species with masses of 56,077 ± 2 Da and 55,813 ± 2 Da, respectively. The relative abundance of each monomer was determined from Fig. 1(c) and an average weighted monomer mass of 55,960 Da was used in all subsequent stoichiometry assignments for the VP1 oligomers. The two monomer species differ in mass by 264 Da, which is consistent with two methionine residues. Thus the two co-existing VP1 monomer species likely arise from differences in posttranslational methionine processing as there are three sequential methionines located at the VP1 N-terminus. This mass heterogeneity in the VP1 monomers is an important contributing factor for the lack of charge state resolution for the NVLP ions (Fig. 1(a)). Notably, similar heterogeneity in VP1 monomer species has been previously reported using SDS-PAGE. [White *et al.*, 1997] Absent from the mass spectra was the minor structural protein, VP2. It was expected from previous experiments that a few copies of VP2 could be present per capsid, but no VP2 was detected in the denatured MS analysis or on the SDS-PAGE gels. Its precise stoichiometry and role in capsid assembly thus remain unclear.

3.2. Analysis of NVLPs stability and (dis)assembly intermediates

The stability of NVLPs over a range of solution conditions has been probed using EM and various spectroscopic techniques. [Ausar *et al.*, 2006; Bertolotti-Ciarlet *et al.*, 2002] From this work it was discovered that while the NVLPs are stable under acidic conditions, they were quite sensitive to even slightly alkaline treatment. Consistent with these previous studies, under acidic pH, ions corresponding exclusively to the intact capsid were observed in the mass spectrum (data not shown). Even at pH 2.5, intact NVLPs dominated the spectrum, although some small VP1 oligomers, predominantly VP1 monomers and dimers, were also detected. Conversely, upon alkaline treatment, extensive capsid dissociation was observed. Representative mass spectra of the NVLPs in a 250 mM ammonium acetate buffer over a pH range of 6 - 9 are given in Fig. 2. Additionally, the ion intensities for all species were summed for each solution pH and are represented in the bar graph included in Fig. 2. It should be noted that for all the spectra collected in Fig. 2, all other important variables such as VP1 concentration and ionic strength were kept constant. As can be seen in Fig. 2, at pH 8 extensive capsid dissociation was observed as VP1 dimers were quite abundant in the spectrum. Several (dis)assembly intermediates were also detected, consisting of 4, 6, 18, and 60 copies of VP1, with the $VP1_{60}$ species (3.4 MDa) being the most populated of the VP1 oligomers at this pH. Interestingly, smaller capsids were also reported in a previous EM study. [White *et al.*, 1997] These smaller particles possessed a diameter of around 25 nm, and were speculated to be $T = 1$ particles made up of 60 copies of VP1. Thus it appears that at this pH, the intact capsid can disassemble, initially forming smaller VP1 oligomers, predominantly dimers. These small VP1 oligomers can then interact to reform higher order oligomers, predominantly $VP1_{60}$ at pH 8. For the norovirus, it has been suggested that in solution the VP1 dimers may exist in two distinct conformations and that the N-terminus of VP1

acts as a switch to distinguish between these two dimer conformations guiding $T = 3$ icosahedral assembly. [Harrison, 2001; Prasad *et al.*, 1999] Thus there are perhaps subtle pH-induced conformational changes within the VP1 dimers, affecting the switching mechanism and leading to the formation of alternate capsids with different symmetry and stoichiometry.

A further increase of the solution pH to 9 led to the complete loss of the full $VP1_{180}$ capsid and the formation of more mid-size VP1 oligomers. At the low *m/z* region, a series of VP1 oligomers is observed; $VP1_2$, $VP1_4$, $VP1_6$ and a small amount of $VP1_8$ and $VP1_{10}$. In the higher *m/z* region, again the $VP1_{60}$ species was detected in high abundance. Additionally, VP1 oligomers consisting of 18, 40, and 80 copies of VP1 were observed. Of all the VP1 oligomers detected, the $VP1_{60}$, $VP1_{80}$, and $VP1_{18}$ are the most populated of the higher order species. Interestingly, Middaugh *et al.* postulated that three larger oligomeric forms of VP1 may exist at alkaline pH based on fitting their dynamic light scattering data to a multi-modal distribution. [Ausar *et al.*, 2006] Using ESI-MS we can directly observe these species and assign their stoichiometry. This demonstrates one of the strengths of an MS based approach to study protein assemblies as it is possible to simultaneously detect all oligomers within a heterogeneous population.

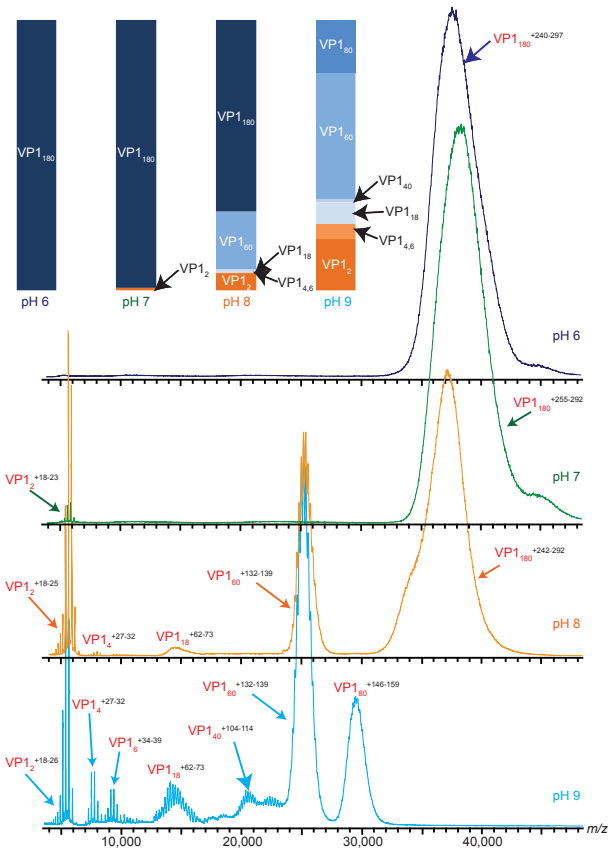


Fig. 2: Stability of the NVLPs over a pH range of 6 - 9 monitored by native ESI-MS
Mass spectra were obtained from aqueous solutions of the NVLPs (30 µM VP1) in a 250 mM ammonium acetate buffer at the pH values indicated. The summed intensities for all observed VP1 oligomers at each pH are summarised in the bar graphs in the upper left corner.

To confirm the identity of these larger VP1 oligomers, ions corresponding to each species were selected in the quadrupole and subjected to high energy CID (Fig. 3). Representative CID spectra of the $VP1_{60}$ species are given in Fig. 3(a). Dissociation proceeded by the sequential loss of highly charged VP1 monomers, with at most 4 monomers being ejected at the highest collision voltage investigated (400 V). The series of high mass dissociation products differing by one VP1 monomer confirms the oligomeric state of the parent 60-mer. Unfortunately $VP1_{80}$ ions could not be selected using the quadrupole mass filter

for tandem MS, as their *m/z* value is above the limit for precursor ion selection. Instead, the ions were analysed with a sufficiently high energy in the collision cell to improve desolvation of the high mass ions but not cause any fragmentation (*i.e.*, 200 V in the collision cell). As can be seen in Fig. 3(b), this greatly enhanced the resolution of the individual charge states, making the determination of the stoichiometry unambiguous. To further support the assignment of these stoichiometries, simulated mass spectra of the two VP1 oligomers were superimposed over the experimental mass spectrum of the $VP1_{60}$ and $VP1_{80}$ at high desolvation energy (appendix I Fig. S2), and exhibited excellent agreement. Simulations of these spectra using other stoichiometries, for instance $VP1_{72}$ and $VP1_{84}$ which represent closely related and allowed icosahedral symmetries, clearly did not match the experimental data, thus confirming the stoichiometry of the $VP1_{80}$ species. The charge state distributions for the $VP1_{80}$ and $VP1_{60}$ oligomers were also assigned assuming a stoichiometry of ± 1 VP1 dimer (appendix I Table S4 and S5). From these tables it can be seen that a stoichiometry of $VP1_{80}$ and $VP1_{60}$ represents the best fit to the experimental data. Additionally, the VP1 oligomers were included in the plot of the charge state *vs.* the square root of mass. As can be seen in Fig. 1(b), the VP1 oligomers of the different stoichiometries are all detected at the expected charge state.

It has been demonstrated for other icosahedral viruses, such as cowpea chlorotic mottle virus, that capsid assembly can also be influenced by ionic strength. [Zlotnick *et al.*, 2000] To

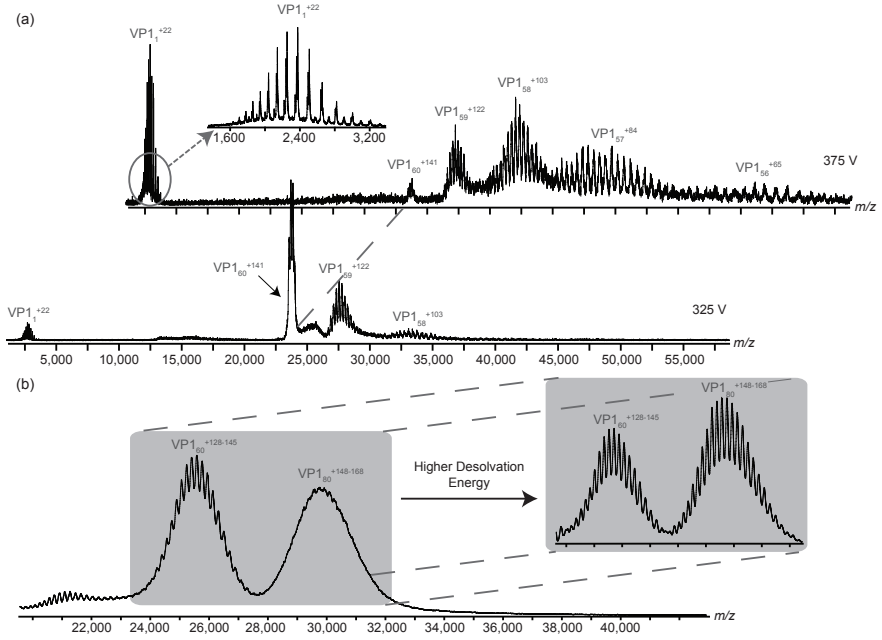


Fig. 3: Illustrative ESI mass spectra demonstrating the identification of the higher order $VP1_{60}$ and $VP1_{80}$ oligomers. (a) Representative CID mass spectra of isolated ions corresponding to $VP1_{60}$, excited with a voltage of 325 V and 375 V before the collision cell. All CID spectra were obtained from aqueous solutions of the NVLPs (30 μM VP1) in a 250 mM ammonium acetate buffer (pH 9). (b) When selection of specific ions was not possible, as was the case for the high *m/z* $VP1_{80}$ ions, spectra were collected at a higher desolvation energy in the collision cell (200 V) to improve resolution and mass assignment.

investigate the effect of ionic strength on the assembly of these VP1 oligomers, ESI-MS analysis was performed after dilution of the intact NVLPs into various ammonium acetate buffers with a concentration range of 50-500 mM and a pH range of 7 - 9. The ion intensities for each VP1 oligomer were summed over all charge states at each solution composition and are summarised as bar graphs in Fig. 4. Over the entire range of ionic strength, ions corresponding to intact NVLPs dominated the mass spectra at physiological pH (Fig. 4(a)), with only a small amount of VP1 dimer being detected. Ionic strength had a much more significant effect under alkaline conditions. Here, extensive capsid dissociation was observed at each ionic strength value as evidenced by the formation of $VP1_2$ and other VP1 oligomers. Interestingly, at pH 8, the intact capsid was most abundant at an intermediate ionic strength of 100 mM (Fig. 4(b)). At this pH, the ionic strength also seems to govern which higher order VP1 oligomers are preferentially formed. Below 100 mM, a small amount of $VP1_{80}$ is detected and no $VP1_{60}$. This is in sharp contrast to the results at higher ionic strengths, in which significant $VP1_{60}$ is detected and the $VP1_{80}$ is completely absent from the mass spectrum. At even lower ionic strengths (25 mM, data not shown) only small VP1 oligomers were detected. Another interesting trend at this pH is that the abundance of the $VP1_2$ decreases with increasing salt concentration, with a concomitant increase in $VP1_{60}$ abundance. Thus it appears that only above a certain threshold ionic strength VP1 dimers can reassemble into the $VP1_{60}$ species.

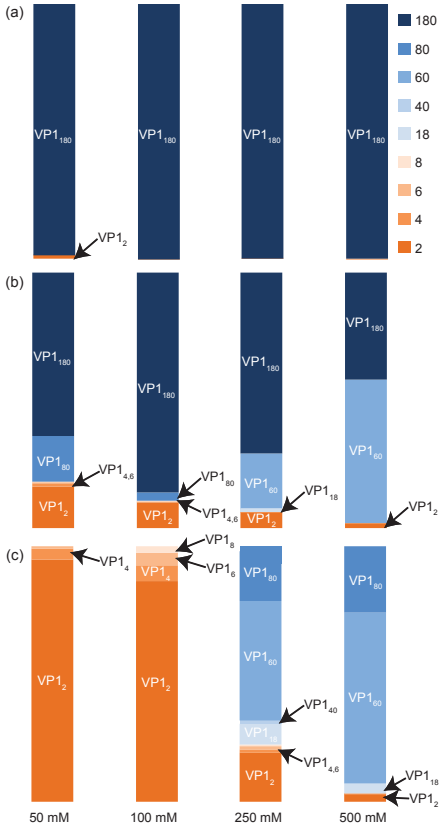


Fig. 4: The effect of ionic strength on the assembly of VP1 oligomers
Assembly was monitored by ESI-MS at a solution pH of (a) 7, (b) 8, and (c) 9 under a range of ammonium acetate concentrations (50-500 mM).

These trends are even more evident at pH 9 (Fig. 4(c)). At this pH there is no intact $VP1_{180}$ at any of the ionic strengths investigated. As was observed at pH 8, at higher ionic strengths there is an increasing propensity to form the high mass VP1 oligomers, such as $VP1_{60}$ and $VP1_{80}$, with a concomitant decrease in $VP1_2$ abundance. In fact, below a certain ionic strength threshold (100 mM), there are no higher order oligomers present, as almost exclusively VP1 dimers were detected in the mass spectrum.

Another important factor in the formation of protein assemblies is the concentration of their protein building blocks. To investigate the effect of VP1 concentration on the assembly of these novel VP1 oligomers, serial dilutions of a VP1 stock solution (160 µM) in a 250 mM ammonium acetate buffer (pH 9) were made and analysed by ESI-MS. Not surprisingly it was found that at low concentrations of VP1, the large oligomers were very low in abundance and $VP1_2$ was the dominant species in the spectrum (appendix I Fig. S3).

With increasing VP1 concentration, $VP1_{60}$ and $VP1_{80}$ became more prevalent. Interestingly, the ratio of $VP1_{60}/VP1_{80}$ was also highly dependent on VP1 concentration. At a VP1 concentration of 23 μM (Fig. S3(b)) both oligomers were detected with similar intensities by ESI-MS. However, at the highest concentration investigated, 80 μM VP1, the $VP1_{80}$ species was approximately eight times more abundant than the $VP1_{60}$ oligomer.

The assembly of these higher order VP1 oligomers was also found to be a reversible process (Fig. 5). In this experiment, the intact NVLPs were first exchanged into conditions favorable for the formation of $VP1_{60}$ and $VP1_{80}$, namely 250 mM ammonium acetate, pH 9. An aliquot was removed from this stock solution and diluted into the same buffer and analysed by ESI-MS (Fig. 5(a)). From this stock solution, two additional aliquots were removed and diluted ten-fold into buffer with a different ionic strength or pH. One aliquot was diluted ten-fold into a lower ionic strength buffer (50 mM ammonium acetate) at the same pH prior to ESI-MS analysis (Fig. 5(b)). Consistent with Fig. 4(c), the $VP1_{60}$ and $VP1_{80}$ species were not present, owing to the low ionic strength. Given that Fig. 5(b) was generated minutes after dilution into the low ionic strength buffer, the VP1 oligomers appear to be rapidly interconverting with the range of possible species being determined by the solution conditions (*i.e.* pH, ionic strength). The mass spectrum in Fig. 5(c) was generated by diluting the VP1 stock solution into a buffer of the same ionic strength but at pH 6. Under these conditions intact NVLPs are reassembled and dominate the mass spectrum. Thus these experiments demonstrate the amazing ability of the VP1 to reversibly and rapidly form higher order structures depending upon the solution conditions. It has been recently suggested from molecular dynamics simulations of capsid assembly that strong reversible steps are essential to capsid formation as they diminish the effect of kinetic traps. [Rapaport, 2008] This definitely appears to be the case for the NVLPs and the VP1 oligomers described here and may explain why norovirus is such a robust and environmentally persistent virion.

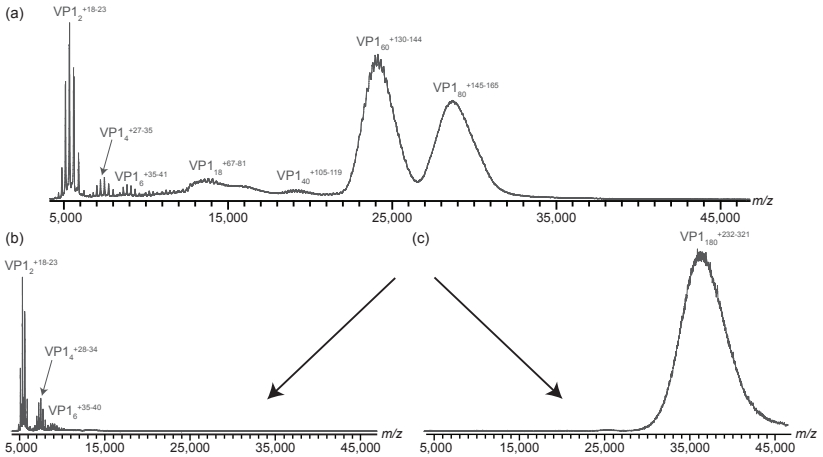


Fig. 5: Representative ESI mass spectra demonstrating that the assembly of the various VP1 oligomers is reversible The solutions were prepared by diluting a stock solution of VP1 (160 μM) in a 250 mM ammonium acetate buffer (pH 9) ten-fold into (a) 250 mM ammonium acetate buffer (pH 9), (b) 50 mM ammonium acetate buffer (pH 9), and (c) 250 mM ammonium acetate (pH 6).

3.3. Structural analysis of VP1 (dis)assembly intermediates

To confirm the presence of higher order VP1 oligomers (*i.e.* $VP1_{60}$ and $VP1_{80}$) and to gain insights into the shape of the various VP1 assemblies, AFM images were collected under the same solution conditions as used in the MS analysis. From the MS data, it was found that at a pH < 7, exclusively $VP1_{180}$ is present at all ionic strengths investigated (50 – 500 mM). Therefore, to confirm these results, AFM experiments were performed in buffers containing VP1 in a 500 mM ammonium acetate buffer at pH 6. Fig. 6 summarises the height distribution of the imaged particles. A representative AFM image and height profile for one of the imaged $T = 3$ particles is included as an inset. Importantly, the particles were found to be spherical and to possess an average diameter of 40 ± 3 nm which is in excellent agreement with the expected value of 38 nm for $VP1_{180}$ obtained from EM and X-ray crystallographic studies. [Prasad *et al.*, 1994]

Next, we sought to obtain images of the intermediate VP1 assemblies, namely $VP1_{60}$ and $VP1_{80}$. From the MS data, it is clear that the formation of both of these assemblies requires a high pH and a high ionic strength (Fig. 4(c)). Thus VP1 was analysed by AFM in a 500 mM ammonium acetate buffer at pH 9 (Fig. 6(b)). Consistent with the MS data, no $VP1_{180}$ particles were detected and a broad distribution of smaller particles was present. Interestingly the particles were spherical, suggesting that these smaller VP1 oligomers are indeed capsids of various stoichiometries and symmetries. The most abundant of these particles was a distribution of species around 22 nm. Under these conditions it was found from MS that the most abundant VP1 oligomer was $VP1_{60}$. Combining information regarding stoichiometry obtained from MS with the structural data afforded by AFM provides strong evidence that the $VP1_{60}$ oligomer is a spherical capsid with $T = 1$ icosahedral symmetry. Additional support for this claim is provided by previous EM measurements, which, as mentioned earlier, also observed a $T = 1$ capsid with a diameter of 23 nm. [White *et al.*, 1997] Even more intriguing is the $VP1_{80}$ oligomer. Although most of the particles possess diameters of around 23 nm, there were some larger species detected in the AFM analysis likely corresponding to $VP1_{80}$. However, due to their low abundance it is difficult to make any definitive conclusions regarding the morphology of the $VP1_{80}$ species. It is interesting to note that at this concentration of VP1 (10 μM) and at this solution composition, high ionic strength and high pH, very little $VP1_{80}$ was detected by ESI-MS as well (appendix I Fig. S3). From the MS data, a higher VP1 concentration should lead to the formation of more $VP1_{80}$. It is difficult to perform the AFM analysis at such high VP1 concentrations since the glass plate becomes overcrowded with particles, precluding the height measurements of individual particles. Importantly, AFM experiments at higher VP1 concentrations yielded a distribution of particles with higher diameters which is consistent with the MS results.

To gain further structural insights into these VP1 oligomers, particularly the $VP1_{80}$ species, all oligomers were interrogated using IMMS. In IMMS the mobility of ions through a pressurized tube is measured, from which information may be gathered about the Ω of the ions, which may be used to obtain low resolution structural information about the species, like volume/shape/conformation, as reviewed recently. [Uetrecht *et al.*, 2010] The experimentally determined Ω for the $VP1_{180}$, $VP1_{80}$, and $VP1_{60}$ oligomers are plotted in Fig. 7 *versus* their mass (in MDa). Interestingly, the measured Ω of NLVP oligomers follow a linear trend and increase with mass, suggesting that these assemblies possess a similar (viral shell like) geometry. Also included in the plot are Ω for the two morphologies observed for the HBV capsids, reported pre-

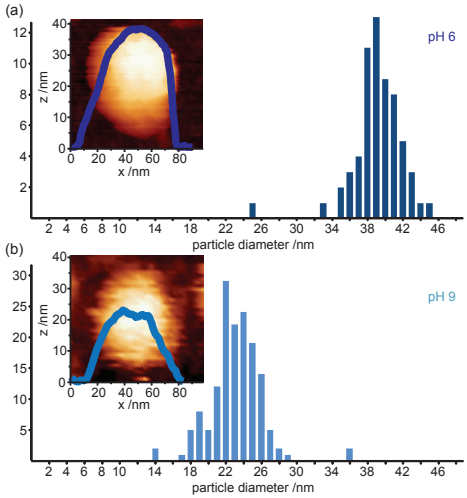


Fig. 6: AFM imaging of the NVLPs and VP1 oligomers
Particle size distribution histograms constructed from AFM data obtained for NVLPs incubated in a 500 mM ammonium acetate buffer at (a) pH 6 and (b) pH 9. Representative AFM images and height profiles of the (a) $VP1_{180}$ and (b) $VP1_{60}$ are also included as insets. Imaging of the particles at pH 9 was difficult, possibly as a result of improper attachment. However, to allow comparison of the results at pH 6 and pH 9, similar hydrophobic surfaces were used in both experiments.

viously. [Uetrecht *et al.*, 2008b] In this work it was demonstrated by IMMS that the HBV capsids retain most likely a hollow spherical geometry upon transfer to the gas phase and throughout the IMMS analysis. As can be seen in Fig. 7, the HBV capsids follow the same linear increase in Ω with mass observed for the VP1 oligomers, suggesting that also all observed NLVP stoichiometries (*i.e.* $VP1_{60}$, $VP1_{80}$ and $VP1_{180}$) have also adopted a hollow, closed geometry. This result is consistent with the AFM images and previous EM measurements that indicate the $VP1_{60}$ and $VP1_{180}$ adopt a spherical icosahedral geometry. Most interestingly, the $VP1_{80}$ species fits on the same trend line, thus it appears that the $VP1_{80}$ also adopts a closed, hollow geometry as opposed to an extended intermediate along the pathway to the full $T = 3$ $VP1_{180}$ capsid or a completely collapsed globular assembly. Although a structure consisting of 80 copies of VP1 does not fit with any structures possessing icosahedral symmetry, there is precedence in the literature about the P domain of VP1 forming assemblies with different symmetry. [Tan *et al.*, 2008]

4. Discussion

The stability of the NVLPs was investigated over a range of solution conditions using native ESI-MS. From this data it was shown that upon alkaline treatment the NVLPs disassemble and the corresponding VP1 dimers that are released can reassemble into various VP1 oligomers based on the solution pH and ionic strength. Higher ionic strength and solution pH drives the assembly towards the formation of species consisting of 60 and 80 copies of VP1. It was also demonstrated that the capsid formation of the various VP1 oligomers was completely reversible. Structural analysis using AFM and IMMS of the fully assembled and large VP1 oligomers revealed that these particles possess a spherical geometry. This was to be expected for the fully assembled NVLPs and for the $VP1_{60}$ species, but the precise

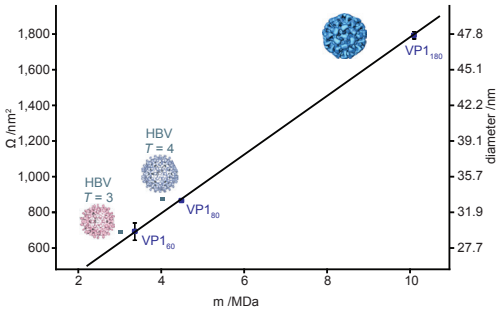


Fig. 7: Shape of NVLPs and VP1 oilgomers determined by IMMS.
A plot of the calibrated Ω (nm^2) for the fully assembled NVLPs and VP1 (dis)assembly intermediates (blue) *vs.* mass. For each species, the Ω increased linearily with mass, suggesting that these species adopt a similar hollow geometry. Included on the plot are values obtained for the HBV (grey) capsid obtained previously. [Uetrecht *et al.*, 2008b] The diameters for each species, assuming a spherical geometry, are also included.

structure of the $VP1_{80}$ remains intriguing. Overall, capsid (dis)assembly was found to be a very heterogeneous process and very sensitive to solution pH, ionic strength, and VP1 concentration. This work highlights the application of native ESI-MS to such analyses, where the simultaneous detection of all oligomeric forms of a protein assembly can provide insights into the factors governing a process such as viral assembly.

References

Atmar, R. L. & Estes, M. K. (2006). The epidemiologic and clinical impartance of norovirus infection. *Gastroenterol Clin North Am* **35**, 275.

Ausar, S. F., Foubert, T. R., Hudson, M. H., Vedvick, T. S. & Middaugh, C. R. (2006). Conformational stability and disassembly of Norwalk virus-like particles - Effect of pH and temperature. *J Biol Chem* **281**, 19478-19488.

Barrera, N. P., Isaacson, S. C., Zhou, M. & other authors (2009). Mass spectrometry of membrane transporters reveals subunit stoichiometry and interactions. *Nat Methods* **6**, 585-U549.

Bertolotti-Ciarlet, A., White, L. J., Chen, R., Prasad, B. V. V. & Estes, M. K. (2002). Structural requirements for the assembly of Norwalk virus-like particles. *J Virol* **76**, 4044-4055.

Bertolotti-Ciarlet, A., Crawford, S. E., Hutson, A. M. & Estes, M. K. (2003). The 3 ‘ end of Norwalk virus mRNA contains determinants that regulate the expression and stability of the viral capsid protein VP1: a novel function for the VP2 protein. *J Virol* **77**, 11603-11615.

de Pablo, P. J., Colchero, J., Gomez-Herrero, J. & Baro, A. M. (1998). Jumping mode scanning force microscopy. *Appl Phys Lett* **73**, 3300-3302.

Estes, M. K., Prasad, B. V. V. & Atmar, R. L. (2006). Noroviruses everywhere: has something changed? *Curr Opin Infect Dis* **19**, 467-474.

Felitsyn, N., Peschke, M. & Kebarle, P. (2002). Origin and number of charges observed on multiply-protonated native proteins produced by ESI. *Int J Mass Spectrom* **219**, 39-62.

Green, K. Y., Ando, T., Balayan, M. S. & other authors (2000). Taxonomy of the caliciviruses. *J Infect Dis* **181 Suppl 2**, S322-330.

Harrison, S. C. (2001). The familiar and the unexpected in structures of icosahedral viruses. *Curr Opin Struct Biol* **11**, 195-199.

Heck, A. J. R. & van den Heuvel, R. H. H. (2004). Investigation of intact protein complexes by mass spectrometry. *Mass Spectrom Rev* **23**, 368-389.

Heck, A. J. R. (2008). Native mass spectrometry: a bridge between interactomics and structural biology. *Nat Methods* **5**, 927-933.

Hoaglund, C. S., Valentine, S. J., Sporleder, C. R., Reilly, J. P. & Clemmer, D. E. (1998). Three-dimensional ion mobility TOFMS analysis of electrosprayed biomolecules. *Anal Chem* **70**, 2236-2242.

Ivanovska, I. L., de Pablo, P. J., Ibarra, B., Sgalari, G., MacKintosh, F. C., Carrascosa, J. L., Schmidt, C. F. & Wuite, G. J. L. (2004). Bacteriophage capsids: Tough nanoshells with complex elastic properties. *Proc Natl Acad Sci USA* **101**, 7600-7605.

Jiang, X., Wang, M., Wang, K. N. & Estes, M. K. (1993). Sequence and genomic organization of Norwalk virus. *Virology* **195**, 51-61.

Kaddis, C. S., Lomeli, S. H., Yin, S., Berhane, B., Apostol, M. I., Kickhoefer, V. A., Rome, L. H. & Loo, J. A. (2007). Sizing large proteins and protein complexes by electrospray ionization mass spectrometry and ion mobility. *J Am Soc Mass Spectrom* **18**, 1206-1216.

Lorenzen, K., Vannini, A., Crarner, P. & Heck, A. J. R. (2007a). Structural biology of RNA polymerase III: Mass spectrometry elucidates subcomplex architecture. *Structure* **15**, 1237-1245.

Lorenzen, K., Versluis, C., van Duijn, E., van den Heuvel, R. H. H. & Heck, A. J. R. (2007b). Optimizing macromolecular tandem mass spectrometry of large non-covalent complexes using heavy collision gases. *Int J Mass Spectrom* **268**, 198-206.

Lorenzen, K., Olia, A. S., Uetrecht, C., Cingolani, G. & Heck, A. J. R. (2008). Determination of stoichiometry and conformational changes in the first step of the P22 tail assembly. *J Mol Biol* **379**, 385-396.

Morton, V. L., Stockley, P. G., Stonehouse, N. J. & Ashcroft, A. E. (2008). Insights into virus capsid assembly from non-covalent mass spectrometry. *Mass Spectrom Rev* **27**, 575-595.

Prasad, B. V. V., Rothnagel, R., Jiang, X. & Estes, M. K. (1994). 3-Dimensional strucutre of baculovirus-expressed Norwalk virus capsids. *J Virol* **68**, 5117-5125.

Prasad, B. V. V., Hardy, M. E., Dokland, T., Bella, J., Rossmann, M. G. & Estes, M. K. (1999). X-ray crystallographic structure of the Norwalk virus capsid. *Science* **286**, 287-290.

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

Rapaport, D. C. (2008). Role of Reversibility in Viral Capsid Growth: A Paradigm for Self-Assembly. *Phys Rev Lett* **101**.

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

Ruotolo, B. T., Benesch, J. L. P., Sandercock, A. M., Hyung, S. J. & Robinson, C. V. (2008). Ion mobility-mass spectrometry analysis of large protein complexes. *Nat Protoc* **3**, 1139-1152.

Sader, J. E., Chon, J. W. M. & Mulvaney, P. (1999). Calibration of rectangular atomic force microscope cantilevers. *Rev Sci Instrum* **70**, 3967-3969.

Sharon, M. & Robinson, C. V. (2007). The role of mass Spectrometry in structure elucidation of dynamic protein complexes. *Annu Rev Biochem* **76**, 167-193.

Siuzdak, G., Bothner, B., Yeager, M., Brugidou, C., Fauquet, C. M., Hoey, K. & Chang, C. M. (1996). Mass spectrometry and viral analysis. *Chem Biol* **3**, 45-48.

Sobott, F., Hernandez, H., McCammon, M. G., Tito, M. A. & Robinson, C. V. (2002). A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Anal Chem* **74**, 1402-1407.

Synowsky, S. A., van den Heuvel, R. H. H., Mohammed, S., Pijnappel, W. & Heck, A. J. R. (2006). Probing genuine strong interactions and post-translational modifications in the heterogeneous yeast exosome protein complex. *Mol Cell Proteomics* **5**, 1581-1592.

Tan, M., Fang, P. G., Chachiyo, T., Xia, M., Huang, P. W., Fang, Z. Y., Jiang, W. & Jiang, X. (2008). Noroviral P particle: Structure, function and applications in virus-host interaction. *Virology* **382**, 115-123.

Taverner, T., Hernandez, H., Sharon, M., Ruotolo, B. T., Matak-Vinkovic, D., Devos, D., Russell, R. B. & Robinson, C. V. (2008). Subunit architecture of intact protein complexes from mass spectrometry and homology modeling. *Acc Chem Res* **41**, 617-627.

Tito, M. A., Tars, K., Valegard, K., Hajdu, J. & Robinson, C. V. (2000). Electrospray time-of-flight mass spectrometry of the intact MS2 virus capsid. *J Am Chem Soc* **122**, 3550-3551.

Uetrecht, C., Versluis, C., Watts, N. R., Roos, W. H., Wuite, G. J. L., Wingfield, P. T., Steven, A. C. & Heck, A. J. R. (2008a). High-resolution mass spectrometry of viral assemblies: Molecular composition and stability of dimorphic hepatitis B virus capsids. *Proc Natl Acad Sci USA* **105**, 9216-9220.

Uetrecht, C., Versluis, C., Watts, N. R., Wingfield, P. T., Steven, A. C. & Heck, A. J. R. (2008b). Stability and shape of hepatitis B virus capsids *in vacuo*. *Angew Chem Int Ed Engl* **47**, 6247-6251.

Uetrecht, C., Rose, R. J., Duijn, E. v., Lorenzen, K. & Heck, A. J. R. (2010). Ion mobility mass spectrometry of proteins and protein assemblies. *Chem Soc Rev* **39**, 1633 - 1655.
van den Heuvel, R. H. H., van Duijn, E., Mazon, H. & other authors (2006). Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. *Anal Chem* **78**, 7473-7483.
van Duijn, E., Bakkes, P. J., Heeren, R. M. A., van den Heuvell, R. H. H., van Heerikhuizen, H., van der Vies, S. M. & Heck, A. J. R. (2005). Monitoring macromolecular complexes involved in the chaperonin-assisted protein folding cycle by mass spectrometry. *Nat Methods* **2**, 371-376.
van Duijn, E., Barendregt, A., Synowsky, S., Versluis, C. & Heck, A. J. R. (2009). Chaperonin Complexes Monitored by Ion Mobility Mass Spectrometry. *J Am Chem Soc* **131**, 1452-1459.
White, L. J., Hardy, M. E. & Estes, H. K. (1997). Biochemical characterization of a smaller form of recombinant Norwalk virus capsids assembled in insect cells. *J Virol* **71**, 8066-8072.
Xi, J. A., Graham, D. Y., Wang, K. N. & Estes, M. K. (1990). Norwalk virus genome cloning and characterization. *Science* **250**, 1580-1583.
Xi, J. A., Min, W., Graham, D. Y. & Estes, M. K. (1992). Expression, self-assembly, and antigenicity of the Norwalk virus capsid protein. *J Virol* **66**, 6527-6532.
Zhou, M., Sandercock, A. M., Fraser, C. S. & other authors (2008). Mass spectrometry reveals modularity and a complete subunit interaction map of the eukaryotic translation factor eIF3. *Proc Natl Acad Sci USA* **105**, 18139-18144.
Zlotnick, A., Aldrich, R., Johnson, J. M., Ceres, P. & Young, M. J. (2000). Mechanism of capsid assembly for an icosahedral plant virus. *Virology* **277**, 450-456.

Chapter V

Interrogating viral capsid assembly with ion mobility mass spectrometry: from sheet- to sheath-like structures

Charlotte Uetrecht, Ioana M. Barbu, Glen K. Shoemaker, Esther van Duijn, Albert J.R. Heck

submitted

Abbreviations: cp – capsid protein; EM – electron microscopy; ESI – electrospray ionisation; HBV – Hepatitis B virus; IMMS – ion mobility mass spectrometry; MS – mass spectrometry; NVLP – recombinant norovirus-like particle; *m/z* – mass to charge ratio; *Ω* – collision cross section; PA – projection approximation; PDB – Protein Data Bank; Q-ToF – quadrupole time of flight; *T* – triangulation number; VAO – vanillyl-alcohol oxidase

Interrogating viral capsid assembly with ion mobility mass spectrometry: from sheet- to sheath-like structures

Viral coat proteins of Hepatitis B virus (HBV) and norovirus can spontaneously produce capsids *in vitro* with morphologies identical to the native virion. The mechanisms for self-assembly are still poorly understood, although it has been suggested that certain smaller oligomers act as nucleation intermediates for assembly. After formation of the nucleus, the reaction progresses fast to completion. Therefore, the nucleus and larger viral assembly intermediates are generally low abundant and difficult to monitor. Here, we characterised small oligomers of HBV and norovirus under equilibrium conditions using native ion mobility mass spectrometry (IMMS). The combination of this technique and computational modelling enabled us, for the first time, to identify structural features of these potential intermediates. Instead of closed icosahedral capsids or more general globular shapes, these oligomers exhibit sheet-like structures suggesting that they are assembly competent. Based on the data we propose pathways for the formation of both HBV and norovirus capsids.

1. Introduction

Although, it is now well established that a huge variety of protein complexes form the basis for cellular regulation, relatively little information is yet available on the general mechanisms through which proteins assemble into complexes. [Alberts, 1998; Robinson *et al.*, 2007] This is already true for proteins that do associate into homo-oligomers (*e.g.* GroEL chaperone machineries), [van Duijn *et al.*, 2009] but even more for heterogeneous multi-component functional machineries like the ribosome [Sykes & Williamson, 2009] or the proteasome. [Besche *et al.*, 2009; Sharon *et al.*, 2007] A specific class of self-associating protein complexes are viruses, which often have a characteristic capsid shell built up from hundreds of copies of the viral structural/capsid proteins (cp). [Caspar & Klug, 1962] Many viral capsid proteins can *in vitro* self-assemble into capsids of specific geometries mimicking the morphology and structure of the virions generated *in vivo* thus providing ideal model systems to investigate self-assembly. It is generally established that viral capsid association proceeds rapidly to completion after formation of an assembly nucleus, which can be different depending on the virus. [Zlotnick, 2005] As a consequence the nucleus and small intermediates are often very low abundant, hampering elucidation of their exact identity and structure. For some viruses one or more of the assembly intermediates have been trapped in over-nucleated reactions, [Stockley *et al.*, 2007] but in most cases theoretical models and the entire capsid structure provide the only hints towards the structure of these intermediates. [Prasad *et al.*, 1999; Zlotnick *et al.*, 1999] Here, we identify and structurally investigate self-assembly intermediates of two well studied model systems originating from important human pathogens: norovirus and HBV. Noroviruses are the main cause of viral gastroenteritis and HBV infections can eventually lead to liver cancer or cirrhosis. [Glass *et al.*, 2009; Sarin *et al.*, 2006] Both viruses have in common that they form capsids of icosahedral symmetry using dimers as the primary building blocks. [Crowther *et al.*, 1994; Prasad *et al.*, 1996; Prasad *et al.*, 1999; Wynne *et al.*, 1999] Icosahedral capsids are defined by the triangulation (T) number, which identifies the amount of monomers (T x 60) in the structure. [Caspar & Klug, 1962] In the native virus, the capsid contains the genome. Subsequently, the capsid can be enveloped by a lipid bilayer. For the norovirus, the capsid is non-enveloped and possesses $T = 3$ symmetry (number of monomers, $N = 180$) with a mass of approximately 10 million Da. The

norovirus cp, VP1, mediates host interaction and infection as well as RNA encapsulation. [Hutson *et al.*, 2003; Prasad *et al.*, 1999] In contrast, the cp in HBV only encloses the genome and assures the interaction with the proteins located in the envelope. [Steven *et al.*, 2005] A quite unique feature of HBV cp is the formation of two different capsid morphologies both *in vivo* and *in vitro*, namely the $T = 3$ ($N = 180$) and $T = 4$ ($N = 240$) with a molecular mass of 3 and 4 million Da, respectively. [Dryden *et al.*, 2006; Uetrecht *et al.*, 2008a; Zlotnick *et al.*, 1996] The proposed assembly pathway of both viruses proceeds via nucleation followed by sequential addition of dimers, [Prasad *et al.*, 1999; Zlotnick *et al.*, 1999] as opposed to preassembly of larger oligomers, which has been observed in for example HK97. [Duda *et al.*, 1995] However, the hypothetical nuclei vary for HBV and norovirus capsid formation. Theoretical models have proposed a hexameric nucleus for HBV [Ceres & Zlotnick, 2002; Zlotnick *et al.*, 1999] and crystallography has indicated a decameric nucleus for norovirus. [Prasad *et al.*, 1999]

To investigate viral capsid assembly, the free building blocks, the intact capsid as well as any possible oligomeric intermediates would ideally be monitored simultaneously with a resolution sufficient to distinguish all species present. A technique distinctively able to do this is native mass spectrometry (MS). [Heck, 2008; Morton *et al.*, 2008] For instance, we have recently demonstrated the strength of this technique in evaluating the exact mass, stoichiometry and gas-phase structure of HBV capsids, as well as the morphological variation present within *in vitro* reconstituted norovirus-like particles (NVLP). [Shoemaker *et al.*, 2010; Uetrecht *et al.*, 2008a; Uetrecht *et al.*, 2008b] When native MS is hyphenated with ion mobility (IMMS), insights into the average shape and structure of the ionised gaseous protein complexes can be obtained by determining the collision cross section (Ω). [Lorenzen *et al.*, 2008; Ruotolo *et al.*, 2005; Uetrecht *et al.*, 2010] Using IMMS, we could establish that the hollow icosahedral structure of the viral capsids of HBV and the norovirus is largely retained in the gas phase following electrospray ionisation (ESI) under native conditions. [Shoemaker *et al.*, 2010; Uetrecht *et al.*, 2008b]

Native mass spectra of viral capsid proteins are typically dominated by the intact viral shells ($N = 180$ for NVLP, $N = 180$ and 240 for HBV) with sometimes low abundant residual building blocks present at the low mass range ($N = 2$). [Shoemaker *et al.*, 2010; Uetrecht *et al.*, 2008a] Due to its sensitivity and inherent strong mass resolving power, however, we have also been able to detect by MS other low abundant oligomers ($2 < N < 40$), [Shoemaker *et al.*, 2010] enabling us to optimise experimental conditions to enhance their relative abundance. These smaller oligomeric species possibly represent intermediates in viral capsid assembly. Here, we applied native IMMS to gather structural information on these oligomers of HBV and norovirus under equilibrium conditions. We obtained experimental Ω values for all HBV and norovirus species present. For extensive comparison to a protein complex standard, we also determined Ω values of a large number of globular protein assemblies with known structural features. The data on the viral oligomers were additionally compared with theoretical modelled Ω values [Mesleh *et al.*, 1996; Shvartsburg & Jarrold, 1996] derived from partial crystal structures of the intact capsids, corresponding to assemblies formed by a specific number of capsid proteins ($2 < N < 40$). We hypothesise that oligomers that are on-pathway to capsid assembly should adopt a relatively planar sheet-like structure, similar to that exhibited in the intact capsid shell. Alternatively, non-assembly directed aggregates would adopt a more globular conformation as the protein complex standard. In addition, the complete intact capsids and capsid-like norovirus assemblies ($N = 60$ and 80) have a degree of compactness in agreement with a hollow configuration as previously

proposed. [Shoemaker *et al.*, 2010; Uetrecht *et al.*, 2008b] From these data we conclude that the observed viral oligomeric species are likely to be intermediates in an assembly competent conformation even under equilibrium conditions. Comparative analysis of the oligomers under conditions that either do or do not facilitate capsid formation revealed significant conformational differences, and enabled us to propose assembly pathways for HBV and norovirus including the identities of the putative assembly nuclei.

2. Results

Previously, we demonstrated that in addition to intact NVLPs, consisting of 180 cp subunits, also oligomers with different stoichiometries, ranging from dimers to 18-mers and even larger oligomers consisting of 40, 60 and 80 cp, can exist at alkaline pH. [Shoemaker *et al.*, 2010] These results prompted us to further optimise the conditions to observe these lower abundant potential intermediates of the NVLPs so that they could be investigated by IMMS. Best results were obtained at a pH of 9.0, and ionic strength of 250 mM ammonium acetate, as illustrated in Fig. 1(a).

For HBV, we used a C-terminally truncated cp mutant ending after amino acid 140, which is the smallest construct that is still capable of capsid formation. [Zlotnick *et al.*, 1996] For this mutant the assembly is, however, less efficient resulting in a relatively higher abundance of the potential intermediates. Nevertheless, *Ω* results on longer HBV cp constructs were very similar (data not shown). As shown in Fig. 1(b), HBV cp oligomers from dimers up to 28-mers could be observed at a pH of 6.8 in 200 mM ammonium acetate. Notably, in these spectra, the intact $T = 3$ and $T = 4$ capsid charge state distributions are less well resolved compared to the constructs and conditions used in previous studies. [Uetrecht *et al.*, 2008a; Uetrecht *et al.*, 2008b] This is likely due to the low activation energies used in the instrument to prevent possible fragmentation and unfolding of smaller assemblies.

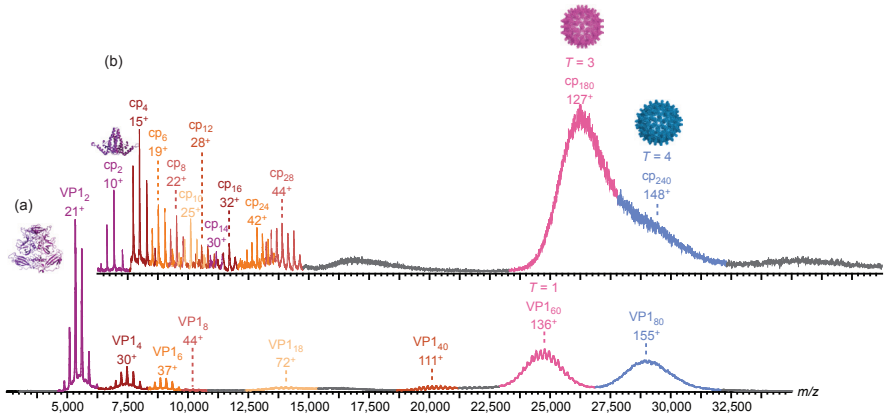


Fig. 1: Mass spectra of capsid intermediates
(a) Mass spectrum of the norovirus cp VP1 electrosprayed at 10 µM based on the monomer from 250 mM ammonium acetate, pH 9.0. Under these conditions oligomers up to 18-mers could be detected as well as 40-, 60- and 80-mers. (b) Mass spectrum of HBV cp140 (15 µM) in 200 mM ammonium acetate, pH 6.8, whereby oligomers of up to 28-mers and intact capsids are detected. The oligomer size increases by additional dimeric building blocks. The number of monomers in the observed oligomers is indicated above the peaks together with the charge of the most abundant charge state. The different charge state distributions are colour coded. The HBV capsids of cp140 do not show charge state resolution due to low activation energies and a mix of complete assembled capsids with species lacking one or two dimers.

2.1. Viral oligomers have extended sheet-like structures

Following optimisation of the conditions to detect especially the oligomers up to 40 subunits, we subjected the cp samples to IMMS to gain further structural insights. After appropriate calibration, individual experimental Ω values could be obtained for all detected norovirus and HBV cp oligomers as presented in Fig. 2(a) and in appendix I Table S6. Plotting the experimental Ω results *versus* the mass of the species revealed that both the intermediates of NVLPs (up to $N = 20$) and HBV capsids show a linear correlation between mass and Ω. Interestingly, such a linear correlation between mass and Ω is consistent with a sheet-like structure for these oligomers (as would be expected for on-pathway intermediates), but in conflict with the generally observed trend for globular protein complexes. Also, the capsids and capsid-like assemblies are linearly correlated in mass and Ω albeit with a shallower slope. This direct dependence of the ions shape on mass suggests a similar architecture in case of the potential intermediates as well as in the case of the larger capsid-like particles. The steeper slope indicates a more extended sheet-like structure for the smaller intermediates relative to the closed or spherical structures of the capsids and capsid-like 60- and 80-mers. [Shoemaker *et al.*, 2010; Uetrecht *et al.*, 2008b]

To further probe our hypothesis, computational modelling of Ω values was performed using structures deposited in the Protein Data Bank (PDB). With mobcal [Mesleh *et al.*, 1996; Shvartsburg & Jarrold, 1996] (see material and methods), theoretical Ω data were calculated from the available capsid crystal structures (1IHM and 1QGT) [Prasad *et al.*, 1999; Wynne *et al.*, 1999] for the capsids and the observed oligomers assuming they have sheet-like structures representing partial capsids. We additionally analysed globular protein complexes covering a mass range similar to that of the viral oligomers (Fig. 2(b)). To ensure comparable results, great effort was taken to use consistent experimental conditions for all samples. For an explanation of the deviation between modelled and experimental data for capsids and capsid-like assemblies see material and methods. Importantly, the agreement between calculated and experimentally derived Ω values is very satisfactory for both the small cp oligomers and the globular protein complexes (Fig. 2 and Table S6) with the deviations being impressively low, especially for the sheet-like viral intermediates. For globular proteins indeed the expected non-linear correlation between mass and the Ω was observed (Fig. 2(b)). The higher deviation between modelled and experimental values for globular protein complexes could be due to the fact that these complexes often contain more flexible regions which tend to extend in the gas phase, but are unresolved in the crystal structures. [Derewenda, 2004] Nevertheless, additional technical improvements are required to assess the correctness of Ω values for assemblies above several hundred kDa in mass (see material and methods). The impact of unfolding to the discrepancy between calculation and experiment is, without further investigation, difficult to estimate. However, to minimise this effect, all applied voltages in the ion mobility analysis were kept as low as possible to avoid extensive excitation of the ions.

Unfolding increases the conformational diversity, which is reflected by the peak width in both Ω and original drift time traces for a particular ion. For individual charge states of the HBV oligomeric species, these drift time traces are narrow and unchanged over a broad range of activation energies indicating a high gas phase stability of their structures. This even enabled the discrimination between oligomers of different stoichiometry overlapping in *m/z* (Fig. 3(a)). Moreover, under the experimental conditions chosen the unfolding of most large globular

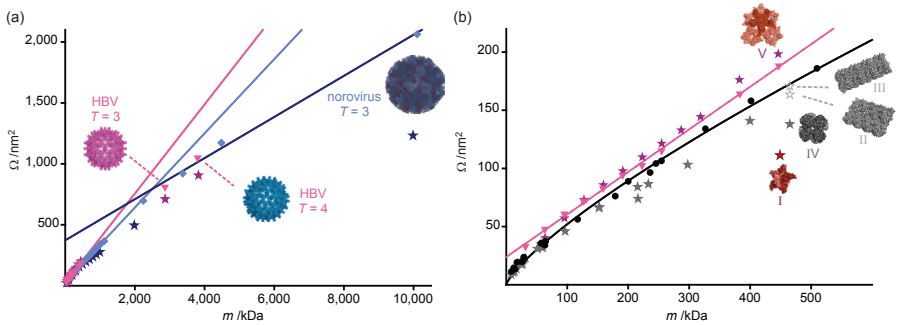


Fig. 2: Experimental and theoretical *Ω* values of capsid protein oligomers
(a) Experimental *Ω* values of the norovirus intermediates (blue diamonds) and those for the HBV intermediates (pink triangles) are plotted *versus* the mass of each species. Theoretical *Ω* values assuming a sheet-like structure were calculated from partial crystal structures of the capsid (norovirus: blue stars and HBV: purple stars). Both experimental and modelled *Ω* values reveal a linear correlation between *Ω* and mass of the oligomers (see also appendix I Table S6), for both norovirus (blue line) and HBV (pink line). Significant deviations occur when the oligomers are transformed into capsid-like structures as indicated by the dark blue straight line. (b) A zoom in on species below 600 kDa again showing the *Ω* data of the HBV intermediates (experimental: pink triangles, theoretical: purple stars). For comparison, experimental (black dots) and theoretical *Ω* results (grey stars) are plotted for several globular protein complexes. These complexes exhibit a non-linear relation between *Ω* and mass as expected for globular assemblies, and revealing a decreasing degree of compactness towards higher masses for the viral oligomers. The globular complexes are significantly smaller than the viral intermediates of similar mass. A power law function to guide the eye was fitted to the data. Moreover, the reduction in modelled *Ω* upon artificial collapse of the HBV cp140 28-mers (I, dark red star) and the increase in modelled *Ω* after planar extension of the VAO 8-mer (II and III, empty grey stars) reflect the strikingly different geometry of viral oligomers relative to globular assemblies. Additionally, the structures of the native VAO 8-mer (IV, grey star) and the sheet-like HBV 28-mer (V, purple star) are depicted next to the corresponding modelled *Ω* values.

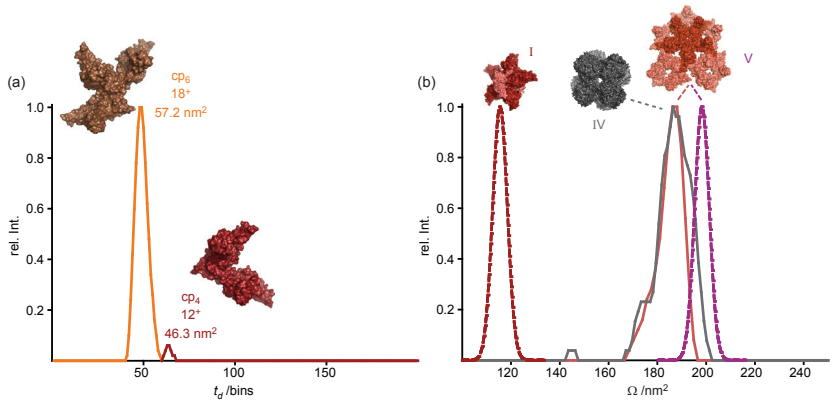


Fig. 3: Ion mobility data of selected HBV oligomers
(a) Drift time trace (in bins) of HBV cp140 at *m/z* 5,291 shows two different species that are well separated; a hexamer (charge state 18^+, orange) at ~50 bins and a tetramer (12^+, red) at ~60 bins. The corresponding *Ω* values and structures are indicated. (b) Overlay of experimental drift time traces converted to *Ω* of the cp140 28-mer (structure V, 447 kDa, solid copper line) and the VAO 8-mer (IV, 510 kDa, solid grey line) both appearing at *m/z* 10,400. Their *Ω* as well as the peak width is similar, even though the VAO 8-mer has a substantially higher mass (~14%). This illustrates that globular proteins of the same mass are generally more compact than the viral intermediates. Apparently, unfolding is limited in the performed experiments as the VAO trace only shows a small shoulder around 200 nm² compared to the cp140 28-mer trace. Additionally, hypothetical Gaussian traces based on the *Ω* of the HBV 28-mer modelled from the crystal structure (V) of the virus capsid (dashed purple line) and a collapsed structure (I, dashed dark red line) are shown.

protein complexes seems to be very restricted. The conformational diversity of the HBV 28-mer and the vanillyl-alcohol oxidase (VAO) 8-mer is comparable at similar *m/z*. Only a minor shoulder indicates a fraction of extended VAO. This result strongly supports our hypothesis, as these two ions have an almost identical Ω despite the ~14% higher mass of the VAO 8-mer, revealing that the native VAO is significantly more compact than the sheet-like HBV 28-mer. Interestingly, if the PDB structure of this globular VAO oligomer (IV in Fig. 2(b)) is artificially rearranged to a planar (II) or rod-like structure (III) the modelled Ω falls closer to a region as observed for the viral intermediates (Fig. 2(b) and Table S6). Likewise, when a sheet-like partial capsid structure (V) is virtually collapsed (I) it leads to a Ω better resembling that of a globular protein complex (see also Fig. 3(b)).

2.2. Implications for the assembly pathway

Recently, Packianathan *et al.* produced an HBV cp149 mutant that is assembly incompetent. [Packianathan *et al.*, 2009] Interestingly, this mutant crystallised in a hexameric state (3KXS) displaying only marginal differences in structure to the triangular hexamers from the 3-fold symmetry axis in the capsid (1QGT, Fig. 4 and 5). However, the mutant did not exhibit all the necessary interfaces to form larger sheets or even full capsids. This inspired us to analyse HBV cp hexamers under conditions facilitating or preventing assembly to determine whether or not a distinction between those closely related structures could be made using IMMS. We therefore used the fact that HBV cp does not significantly assemble at low ionic strength and alkaline pH upon exchange to the appropriate buffer (*i.e.* to 10 mM ammonium acetate, pH 9.6). Subsequently, the protein can be diluted to an appropriate concentration in a pH 6.8 buffer. As illustrated in Fig 5(a), the mass spectra under non-assembly conditions (low ionic strength) demonstrated that indeed no species larger than the hexamer were present in solution. Furthermore, the hexamer was relatively higher abundant than under assembly conditions (see Fig. 1), indicating that it accumulates rather than proceeds towards larger oligomers. Intriguingly, the experimentally determined Ω for the HBV hexamer was found to be very much dependent on the ionic strength, decreasing about 17% (from 69.8 ± 3.1 nm^2 to 50.3 ± 1.6 nm^2 as determined in several independent experiments for the 19$^+$ ion) when going to low ionic strength (Fig 5(a) and appendix I Table S6). This is at first glance unexpected as the modelled Ω is similar for the triangular hexamer in the capsid (53.4 nm^2) and the assembly incompetent mutant (54.0 nm^2, Table S6 and S7). At low ionic strength, the experimental Ω was smaller most likely indicating a structural collapse in the gas phase (Fig. 5(a)). This would be due to the decreased number of interactions between individual subunits relative to the triangular hexamer in a capsid. The hexamer capable of capsid assembly exhibited a larger Ω more in agreement with a Ω calculated for a partial vertex (5-fold axis) or a quasi-6-fold hexamer symmetry (Table S7). The hexamers extracted from the different symmetry axes are superimposed onto the capsid structure in Fig. 4. The difference in abundance and structure for the hexamer under the two conditions indicates that

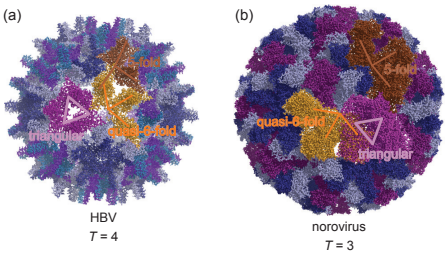


Fig. 4: Capsid symmetry of HBV and norovirus
The crystal structures of the HBV $T = 4$ capsid ((a) 1QGT) and the norovirus $T = 3$ capsid ((b) 1IHM) are depicted. For calculation of theoretical Ω values (Table S7), the highlighted hexamers ($N = 6$) with a triangular (purple), 5-fold (brown) and quasi-6-fold (orange) arrangement were extracted from the capsid structures.

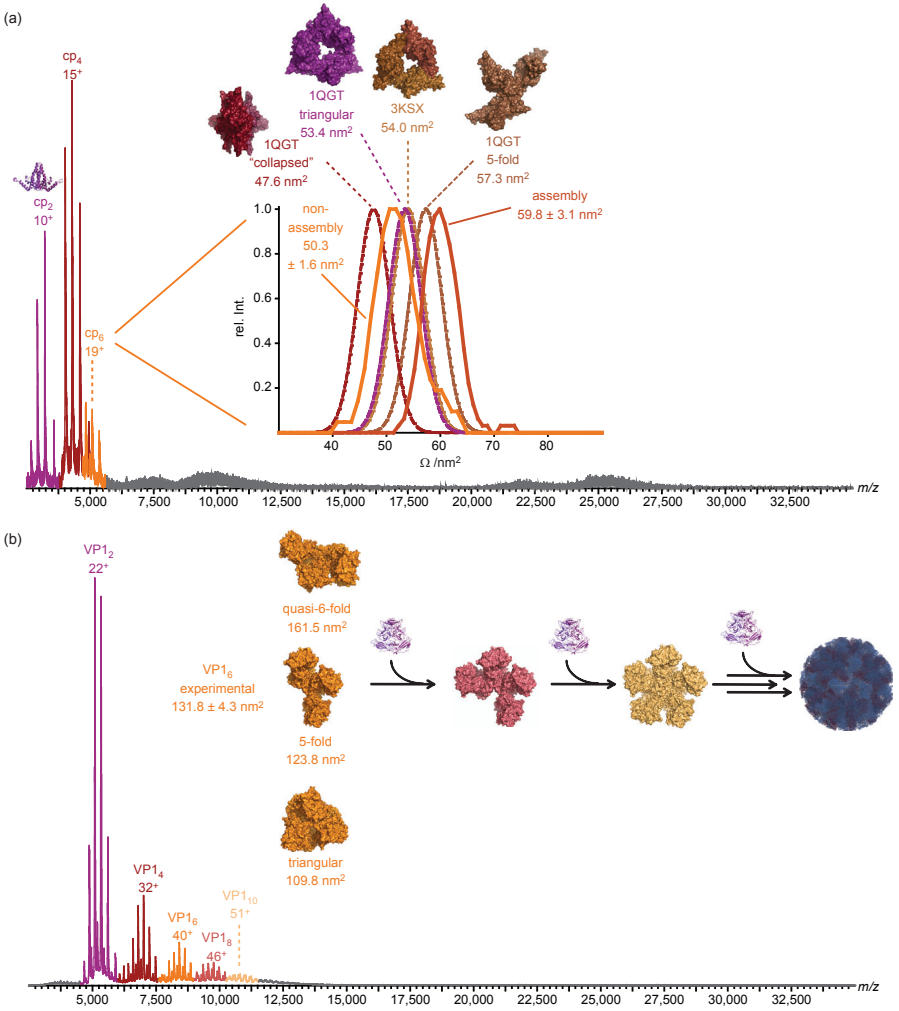


Fig. 5: HBV and norovirus cp under non-assembly conditions
(a) Mass spectrum of 8 µM cp142 under non-assembly conditions (in 10 mM ammonium acetate, pH 6.8) whereby primarily dimers, tetramers and hexamers are detected. Some residual capsid is always observed, but no other oligomers larger than the hexamer could be identified with any cp construct under these conditions. The experimental Ω traces of the hexamer (charge state 19^+) under non-assembly (solid orange line) and assembly conditions (solid copper line) corresponding to low and high ionic strength (inset), respectively, show that under non-assembly conditions the hexamer is substantially more compact. For comparison, the modelled Ω traces and values for the triangular hexamer of 3KXS (non-assembling mutant, dashed light brown line) or 1QGT (intact virus, dashed purple line) and the 5-fold hexamer of 1QGT (dashed brown line) are shown together with an artificially collapsed hexamer based on 1QGT (dashed dark red line). (b) Mass spectrum of norovirus cp at non-assembly conditions; 10 µM in 100 mM ammonium acetate, pH 9.0. Only smaller oligomers up to the decamer are detected in agreement with the proposed assembly nucleus. The inset depicts a model for the proposed assembly pathway, indicating that the observed hexamer is on pathway to the 5-fold vertex. Averaged Ω values of the two HBV hexamers (19^+ ion) and the norovirus hexamer were determined from several independent datasets. Labelling and colour coding in the mass spectra is consistent with Fig. 1.

the nucleation in case of HBV is probably related to a conformational change in the hexamer as previously suggested. [Ceres & Zlotnick, 2002; Packianathan *et al.*, 2009; Zlotnick *et al.*, 1999] Interestingly, a nucleation structure resembling a partial 5-fold vertex is seemingly in better agreement with our results than the triangular arrangement. Such a conformation may only be populated under certain conditions and therefore inhibit HBV capsid assembly under non-assembly conditions.

We previously observed that norovirus cp is also not capable of forming capsids or capsid-like structures at alkaline pH and low ionic strength (Fig. 5(b)). The largest oligomer detected under these experimental conditions is a decamer, the proposed assembly nucleus for NVLPs. As for HBV, modelled Ω for these norovirus oligomers are dependent on the symmetry axis selected (110, 124 and 162 nm^2 for the triangular, 5- and quasi-6-fold symmetric hexamer, respectively (Table S7)). Strikingly, the hexamer structure extracted from the 5-fold axis shows best agreement with the experimental data (131.8 ± 4.3 nm^2 from four independent datasets). In general, the Ω values calculated for capsid fragments of norovirus and HBV based on a pentameric vertex were closer to the experimental results and were therefore used in Table S6. This leads us to a model wherein assembly proceeds for both norovirus and HBV *via* the addition of dimers to a 5-fold symmetric nucleation structure (Fig. 5(b)), although the nuclei of HBV (hexamer) and norovirus (decamer) are different in stoichiometry.

3. Discussion

Here, we combined native MS and ion mobility not only to detect small oligomeric viral capsid assembly intermediates of norovirus and HBV, but also to structurally characterise these oligomers. The transient nature of intermediates leading to their low-abundance combined with the heterogeneous population omit structural analyses using other techniques. Even though absolute Ω cannot be directly obtained in our instrument, our results are still of high importance to viral assembly. Additionally, experiments were carefully constructed to draw clear albeit primarily qualitative conclusions. The Ω values of the intermediates were compared with modelling data, indicating that the observed small oligomeric species exhibit an extended sheet-like conformation. Further evidence was provided by their Ω values being significantly larger than observed for globular protein complexes similar in mass. The modelling also enabled us to determine the likely symmetry arrangement in these smaller oligomers, identifying a conformational change in the HBV hexamer as a potential nucleation step. Both norovirus and HBV appear to share a common assembly pathway starting with a nucleus exhibiting a 5-fold symmetry axis, whereupon assembly proceeds by the addition of dimers, even though their biology and assembly nucleus are different. [Hutson *et al.*, 2003; Prasad *et al.*, 1999; Steven *et al.*, 2005; Zlotnick *et al.*, 1999] The pentameric vertex is common to all icosahedral capsids, therefore the decision about the final size is apparently only made after nucleation in case of HBV and norovirus. This may explain the inherent ability of certain viruses, especially of those studied here, [Dryden *et al.*, 2006; Shoemaker *et al.*, 2010; Zlotnick *et al.*, 1996] to assemble into various stoichiometries under appropriate conditions. [Dokland, 2000; Hsu *et al.*, 2006] Further structural investigations are necessary to prove the postulated symmetry arrangement in the viral oligomers. Nevertheless, the observed intermediates appear to be assembly competent on-pathway products rather than dead-end aggregates of capsid proteins in agreement with the experimentally observed reversible formation and re-equilibration of such small intermediates

with the intact capsids. [Hilmer *et al.*, 2008; Parent *et al.*, 2007; Shoemaker *et al.*, 2010] This may be a prerequisite for disassembly and genome release in the cellular environment. [Rabe *et al.*, 2009] Intriguingly, this ability is even present under equilibrium conditions supporting the theoretical models describing viral capsids as dynamic entities able to exchange individual building blocks with free dimers in solution. [Zlotnick, 2005; Zlotnick, 2007]

4. Material and methods

4.1. Native ion mobility mass spectrometry

Ion mobility mass spectrometry experiments were performed on a Synapt HDMS G1 (Waters, UK) using nano-ESI in positive ion mode. [Pringle *et al.*, 2007] To obtain high resolution spectra of viral samples and confirm the identity of oligomeric species in tandem MS a modified QToF1 (Waters, UK) [van den Heuvel *et al.*, 2006] was used. Gold-coated glass capillaries were prepared in-house as previously reported. [van Duijn *et al.*, 2006] Most proteins were analysed with argon as collision and nitrogen as ion mobility gas, only for the viral proteins xenon was the preferred trap gas to increase the transmission of the oligomers across the entire mass range. [Lorenzen *et al.*, 2007] Some intermediates could be detected with argon as trap gas providing identical results (data not shown). Proteins were either buffer-exchanged to ammonium acetate at appropriate pH *via* centrifugal filter units (GE Healthcare, USA or Millipore, USA) with a molecular weight cutoff of 5 kDa or directly dissolved in this volatile buffer and sprayed at monomer concentrations between 5-10 μM. Instrumental settings were adjusted to preserve noncovalent complexes and minimise unfolding. [Sobott *et al.*, 2002; Tahallah *et al.*, 2001]

The NVLPs used in this study were formed using the recombinant cp VP1 which was expressed in *Spodopter frugiperda* (Sf9) cells using a baculovirus expression system and purified as described previously. [Jiang *et al.*, 1992] The purified recombinant NVLPs (400 μM cp) were stored in water at 4 ºC. For ESI-MS analysis, the NVLPs were exchanged into an aqueous ammonium acetate buffer (50-500 mM) at various pH values using centrifugal filter units with a molecular weight cutoff of 10 kDa. The pH of the aqueous ammonium acetate solution was adjusted by using either an aqueous solution of ammonia or acetic acid.

HBV cp140 was over-expressed in *Escherichia coli*. The obtained dimers were assembled according to Wingfield *et al.* [Wingfield *et al.*, 1995] The capsid fraction was purified using size exclusion chromatography and then dialysed to 200 mM ammonium acetate, pH 6.8, at room temperature. Samples were analysed at 15 μM monomer concentration. Results obtained with different instrumental settings and activation energies were in close agreement and within the general standard deviation (~5%) of the experiment further supporting the reliability of the data (not shown). HBV cp142 dimers were directly exchanged to 10 mM ammonium acetate, pH 9.6, via centrifugal filter units. To remove residual capsids, the sample was first passed through a 100 kDa cutoff filter and the flow-through was concentrated with a 5 kDa cutoff filter. Mass spectrometric analysis was performed in 10 mM ammonium acetate at pH 6.8. For comparison, HBV cp142 capsids were assembled and purified as cp140 and analysed at 8 μM monomer concentration.

All complexes below ~1 MDa were analysed with both a ramped and fixed travelling voltage wave in the ion mobility device and Ω values were in close agreement (Table S6). For larger complexes, ramped waves provide better transmission, lower standard deviations as well as limit unfolding and were therefore preferred. [Uetrecht *et al.*, 2008b] All results presented in the figures were obtained with a ramped wave height. Drift times were calibrated with denatured myoglobin, cytochrome C and ubiquitin according to Ruotolo *et al.* [Ruotolo *et al.*, 2005; Ruotolo *et al.*, 2008] We have previously demonstrated that calibration can also be performed with ramped settings and extrapolated for larger assemblies. [Uetrecht *et al.*, 2008b] We did not apply a step-wise calibration here, since this procedure is highly dependent on the protein complex used for the intermediate step and the experimental settings, which may cause unfolding and therefore aligning dissimilar structures. Subtle changes in the calibration values did not significantly affect the Ω results for the species of primary interest, but massively increased the Ω values of the capsids. Therefore,

the theoretical Ω results for the intact capsids are substantially lower than the corresponding experimental values (Fig. 2(a) and Table S6). This points out the need for appropriate calibrants at the high mass end to reduce the effect of extrapolation. [Uetrecht *et al.*, 2010] Here, a consistent calibration procedure provided higher confidence for the comparison of trends observed in Fig. 2. Most presented data was taken from a representative set of experiments to exclude that effects of the calibration procedure mask the general trend. Ω values for the hexamers of norovirus and HBV given in Fig. 5 were obtained from multiple independent sets to confirm the symmetry arrangement.

4.2. Molecular modelling

The structure of the viral oligomers, empty capsids and globular proteins were investigated using molecular modelling in order to validate the IMMS data. For this reason, the Ω values of the potential gas phase structures are calculated with the projection approximation (PA) model implemented in mobcal, described in detail elsewhere. [Mesleh *et al.*, 1996; Shvartsburg & Jarrold, 1996] Even though other programs are available using a similar algorithm, the well established mobcal was chosen. [Ruotolo *et al.*, 2005; Smith *et al.*, 2009; Uetrecht *et al.*, 2010] No adaptation or calibration of the calculated Ω values to fit the experimental data was performed. Since the PA neglects effects of structural details, it should be less suitable for non-spherical assemblies. Therefore, we also tested the exact hardsphere scattering and trajectory method implemented in mobcal for the sheet-like viral oligomers and globular protein complexes. Surprisingly, both methods gave similar Ω values within 1%. A relative offset of ~35% towards higher values than with the PA was assessed consistently, resulting in higher Ω values than observed experimentally. Importantly, all methods resulted in the same trend lines for globular and sheet-like assemblies. The PA was previously shown to describe experimental data of proteins and their complexes well. [Ruotolo *et al.*, 2005; Smith *et al.*, 2009; van Duijn *et al.*, 2009] Furthermore, discrepancies observed for large globular protein complexes can be explained by the input of incomplete crystal structures accompanied by higher flexibility and therefore the PA algorithm was applied.

The geometries of the intermediate oligomers are extracted from the structures of the empty viral capsids (1QGT – $T = 4$ and 1IHM – $T = 3$) [Prasad *et al.*, 1999; Wynne *et al.*, 1999] using pymol (Schrödinger, USA). The structures were based on the 5-fold axis as starting point unless stated otherwise. A pseudo-model of the HBV $T = 3$ capsid [Roos *et al.*, 2010] was kindly provided by A.C. Steven for the calculation of its Ω showing good agreement with previous results as was also observed for $T = 4$. [Uetrecht *et al.*, 2008b; Uetrecht *et al.*, 2010] For the calculation of Ω values of the non-assembly hexamer structure, 3KXS was used. [Packianathan *et al.*, 2009] The PDB identifiers of the globular proteins can be found in Table S6. The collapsed, globular forms of the HBV hexamer, 28-mer and extended forms of VAO were constructed using pymol.

References

Alberts, B. (1998). The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* **92**, 291-294.
Besche, H. C., Peth, A. & Goldberg, A. L. (2009). Getting to first base in proteasome assembly. *Cell* **138**, 25-28.
Caspar, D. L. & Klug, A. (1962). Physical principles in the construction of regular viruses. *Cold Spring Harb Symp Quant Biol* **27**, 1-24.
Ceres, P. & Zlotnick, A. (2002). Weak protein-protein interactions are sufficient to drive assembly of hepatitis B virus capsids. *Biochemistry* **41**, 11525-11531.
Crowther, R. A., Kiselev, N. A., Bottcher, B., Berriman, J. A., Borisova, G. P., Ose, V. & Pumpens, P. (1994). 3-Dimensional Structure of Hepatitis-B Virus Core Particles Determined by Electron Cryomicroscopy. *Cell* **77**, 943-950.
Derewenda, Z. S. (2004). The use of recombinant methods and molecular engineering in protein crystallization. *Methods* **34**, 354-363.
Dokland, T. (2000). Freedom and restraint: themes in virus capsid assembly. *Structure* **8**, R157-R162.
Dryden, K. A., Wieland, S. F., Whitten-Bauer, C., Gerin, J. L., Chisari, F. V. & Yeager, M. (2006). Native hepatitis B virions and capsids visualized by electron cryomicroscopy. *Mol Cell* **22**, 843-850.
Duda, R. L., Martincic, K., Xie, Z. & Hendrix, R. W. (1995). Bacteriophage HK97 head assembly. *FEMS Microbiol Rev* **17**, 41-46.
Glass, R. I., Parashar, U. D. & Estes, M. K. (2009). Norovirus gastroenteritis. *N Engl J Med* **361**, 1776-1785.
Heck, A. J. (2008). Native mass spectrometry: a bridge between interactomics and structural biology. *Nat Methods* **5**, 927-933.
Hilmer, J. K., Zlotnick, A. & Bothner, B. (2008). Conformational equilibria and rates of localized motion within hepatitis B virus capsids. *J Mol Biol* **375**, 581-594.
Hsu, C., Singh, P., Ochoa, W., Manayani, D. J., Manchester, M., Schneemann, A. & Reddy, V. S. (2006). Characterization of polymorphism displayed by the coat protein mutants of tomato bushy stunt virus. *Virology* **349**, 222-229.
Hutson, A. M., Atmar, R. L., Marcus, D. M. & Estes, M. K. (2003). Norwalk Virus-Like Particle Hemagglutination by Binding to H Histo-Blood Group Antigens. *J Virol* **77**, 405-415.

Jiang, X., Wang, M., Graham, D. Y. & Estes, M. K. (1992). Expression, self-assembly, and antigenicity of the Norwalk virus capsid protein. *J Virol* **66**, 6527-6532.
Lorenzen, K., Versluis, C., van Duijn, E., van den Heuvel, R. H. H. & Heck, A. J. R. (2007). Optimizing macromolecular tandem mass spectrometry of large non-covalent complexes using heavy collision gases. *Int J Mass Spectrom* **268**, 198-206.
Lorenzen, K., Olia, A. S., Uetrecht, C., Cingolani, G. & Heck, A. J. (2008). Determination of stoichiometry and conformational changes in the first step of the P22 tail assembly. *J Mol Biol* **379**, 385-396.
Mesleh, M. F., Hunter, J. M., Shvartsburg, A. A., Schatz, G. C. & Jarrold, M. F. (1996). Structural information from ion mobility measurements: Effects of the long-range potential. *J Phys Chem* **100**, 16082-16086.
Morton, V. L., Stockley, P. G., Stonehouse, N. J. & Ashcroft, A. E. (2008). Insights into virus capsid assembly from non-covalent mass spectrometry. *Mass Spectrom Rev* **27**, 575-595.
Packianathan, C., Katen, S. P. & Zlotnick, A. (2010). Conformational changes in the Hepatitis B virus core protein are consistent with a role for allostery in virus assembly. *J Virol* **84**, 1607-1615.
Parent, K. N., Suhanovsky, M. M. & Teschke, C. M. (2007). Phage P22 procapsids equilibrate with free coat protein subunits. *J Mol Biol* **365**, 513-522.
Prasad, B. V., Hardy, M. E., Jiang, X. & Estes, M. K. (1996). Structure of Norwalk virus. *Arch Virol Suppl* **12**, 237-242.
Prasad, B. V., Hardy, M. E., Dokland, T., Bella, J., Rossmann, M. G. & Estes, M. K. (1999). X-ray crystallographic structure of the Norwalk virus capsid. *Science* **286**, 287-290.
Pringle, S. D., Giles, K., Wildgoose, J. L., Williams, J. P., Slade, S. E., Thalassinos, K., Bateman, R. H., Bowers, M. T. & Scrivens, J. H. (2007). An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *Int J Mass Spectrom* **261**, 1-12.
Rabe, B., Delaleau, M., Bischof, A., Foss, M., Sominskaya, I., Pumpens, P., Cazenave, C., Castroviejo, M. & Kann, M. (2009). Nuclear entry of hepatitis B virus capsids involves disintegration to protein dimers followed by nuclear reassociation to capsids. *PLoS Pathog* **5**, e1000563.
Robinson, C. V., Sali, A. & Baumeister, W. (2007). The molecular sociology of the cell. *Nature* **450**, 973-982.
Roos, W. H., Gibbons, M. M., Arkhipov, A. & other authors (2010). Squeezing protein shells: how continuum elastic models, molecular dynamics simulations and experiments coalesce at the nanoscale. *Biophys J* **99**, pp. 1175 - 1181**.**
Ruotolo, B. T., Giles, K., Campuzano, I., Sandercock, A. M., Bateman, R. H. & Robinson, C. V. (2005). Evidence for macromolecular protein rings in the absence of bulk water. *Science* **310**, 1658-1661.
Ruotolo, B. T., Benesch, J. L., Sandercock, A. M., Hyung, S. J. & Robinson, C. V. (2008). Ion mobility-mass spectrometry analysis of large protein complexes. *Nat Protoc* **3**, 1139-1152.
Sarin, S. K., Sollano, J. D., Chawla, Y. K. & other authors (2006). Consensus on extra-hepatic portal vein obstruction. *Liver Int* **26**, 512-519.
Sharon, M., Witt, S., Glasmacher, E., Baumeister, W. & Robinson, C. V. (2007). Mass spectrometry reveals the missing links in the assembly pathway of the bacterial 20 S proteasome. *J Biol Chem* **282**, 18448-18457.
Shoemaker, G. K., van Duijn, E., Crawford, S. E. & other authors (2010). Norwalk virus assembly and stability monitored by mass spectrometry. *Mol Cell Proteomics* **9**, 1742-1751.
Shvartsburg, A. A. & Jarrold, M. F. (1996). An exact hard-spheres scattering model for the mobilities of polyatomic ions. *Chem Phys Lett* **261**, 86-91.
Smith, D. P., Knapman, T. W., Campuzano, I., Malham, R. W., Berryman, J. T., Radford, S. E. & Ashcroft, A. E. (2009). Deciphering drift time measurements from travelling wave ion mobility spectrometry-mass spectrometry studies. *Eur J Mass Spectrom* **15**, 113-130.
Sobott, F., Hernandez, H., McCammon, M. G., Tito, M. A. & Robinson, C. V. (2002). A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Anal Chem* **74**, 1402-1407.
Steven, A. C., Conway, J. F., Cheng, N., Watts, N. R., Belnap, D. M., Harris, A., Stahl, S. J. & Wingfield, P. T. (2005). Structure, assembly, and antigenicity of hepatitis B virus capsid proteins. *Adv Virus Res* **64**, 125-164.
Stockley, P. G., Rolfsson, O., Thompson, G. S., Basnak, G., Francese, S., Stonehouse, N. J., Homans, S. W. & Ashcroft, A. E. (2007). A simple, RNA-mediated allosteric switch controls the pathway to formation of a *T*=3 viral capsid. *J Mol Biol* **369**, 541-552.
Sykes, M. T. & Williamson, J. R. (2009). A complex assembly landscape for the 30S ribosomal subunit. *Annu Rev Biophys* **38**, 197-215.
Tahallah, N., Pinkse, M., Maier, C. S. & Heck, A. J. (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 Commun Mass Spectrom* **15**, 596-601.
Uetrecht, C., Versluis, C., Watts, N. R., Roos, W. H., Wuite, G. J., Wingfield, P. T., Steven, A. C. & Heck, A. J. (2008a). High-resolution mass spectrometry of viral assemblies: molecular composition and stability of dimorphic hepatitis B virus capsids. *Proc Natl Acad Sci USA* **105**, 9216-9220.
Uetrecht, C., Versluis, C., Watts, N. R., Wingfield, P. T., Steven, A. C. & Heck, A. J. (2008b). Stability and shape of hepatitis B virus capsids *in vacuo*. *Angew Chem Int Ed Engl* **47**, 6247-6251.
Uetrecht, C., Rose, R. J., Duijn, E. v., Lorenzen, K. & Heck, A. J. R. (2010). Ion mobility mass spectrometry of proteins and protein assemblies. *Chem Soc Rev* **39**, 1633 - 1655.
van den Heuvel, R. H., van Duijn, E., Mazon, H. & other authors (2006). Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. *Anal Chem* **78**, 7473-7483.
van Duijn, E., Simmons, D. A., van den Heuvel, R. H., Bakkes, P. J., van Heerikhuizen, H., Heeren, R. M., Robinson, C. V., van der Vies, S. M. & Heck, A. J. (2006). Tandem mass spectrometry of intact GroEL-substrate complexes reveals substrate-specific conformational changes in the trans ring. *J Am Chem Soc* **128**, 4694-4702.
van Duijn, E., Barendregt, A., Synowsky, S., Versluis, C. & Heck, A. J. (2009). Chaperonin complexes monitored by ion mobility mass spectrometry. *J Am Chem Soc* **131**, 1452-1459.
Wingfield, P. T., Stahl, S. J., Williams, R. W. & Steven, A. C. (1995). Hepatitis core antigen produced in *Escherichia coli*: subunit composition, conformational analysis, and *in vitro* capsid assembly. *Biochemistry* **34**, 4919-4932.
Wynne, S. A., Crowther, R. A. & Leslie, A. G. (1999). The crystal structure of the human hepatitis B virus capsid. *Mol Cell* **3**, 771-780.
Zlotnick, A., Cheng, N., Conway, J. F., Booy, F. P., Steven, A. C., Stahl, S. J. & Wingfield, P. T. (1996). Dimorphism of hepatitis B virus capsids is strongly influenced by the C-terminus of the capsid protein. *Biochemistry* **35**, 7412-7421.
Zlotnick, A., Johnson, J. M., Wingfield, P. W., Stahl, S. J. & Endres, D. (1999). A theoretical model successfully identifies features of hepatitis B virus capsid assembly. *Biochemistry* **38**, 14644-14652.
Zlotnick, A. (2005). Theoretical aspects of virus capsid assembly. *J Mol Recognit* **18**, 479-490.
Zlotnick, A. (2007). Distinguishing reversible from irreversible virus capsid assembly. *J Mol Biol* **366**, 14-18.

Chapter VI

Subunit exchange rates in Hepatitis B virus capsids are geometry- and temperature-dependent

Charlotte Uetrecht, Norman R. Watts, Stephen J. Stahl, Paul T. Wingfield, Alasdair C. Steven and Albert J. R. Heck

Phys Chem Chem Phys **12,** 13368 - 13371 (2010)

Abbreviations: CID – collision induced dissociation; cp – capsid protein; cp149 – cp from amino acid 1-149; $cp149_d$ – dimeric capsid protein; $cp149_c$ – capsid; EM – electron microscopy; HBV – Hepatitis B virus; MS – mass spectrometry; *m/z* – mass to charge ratio; Q-ToF – quadrupole time of flight; ToF – time of flight

Subunit exchange rates in Hepatitis B virus capsids are geometry- and temperature-dependent

1. Introduction

Hepatitis B virus (HBV) is a major cause of liver disease in humans, and the leading cause of hepatic cancer. [Lesmana *et al.*, 2006] The development of new antiviral therapies targeting the viral capsid, and a growing interest in utilising engineered adaptations thereof as drug delivery vehicles, is driving the need for precise biophysical characterisation of these particles, including their dynamic properties. [Singh *et al.*, 1996] The structures of HBV capsids and their building-blocks have been determined by cryo-electron microscopy (EM), [Steven *et al.*, 2005] X-ray crystallography, [Packianathan *et al.*, 2010; Wynne *et al.*, 1999] and NMR spectroscopy. [Freund *et al.*, 2008] Native (intra-viral) capsids are polymeric shells of a 183-residue polypeptide, whose first 149 residues suffice for capsid assembly. This assembly domain forms dimers wherein the two chains pair *via* formation of a central four-helix bundle. These dimers serve as building blocks that naturally assemble into capsids of two sizes, corresponding to triangulation numbers of $T = 3$ (90 dimers and 32.0 nm diameter) and $T = 4$ (120 dimers and 35.5 nm diameter), wherein the four-helix bundles form the capsids' characteristic protrusions. [Steven *et al.*, 2005] Recently, we have employed tandem mass spectrometry (tandem MS) to investigate the molecular composition of both HBV capsids and their stability *in vacuo*. [Uetrecht *et al.*, 2008a; Uetrecht *et al.*, 2008b] These particles proved to be particularly suitable for analysis by MS, yielding mass measurements with a precision of < 0.1% and thereby demonstrating for the first time that the lattices of both capsids are complete, consisting of exactly 90 and 120 dimers, respectively.

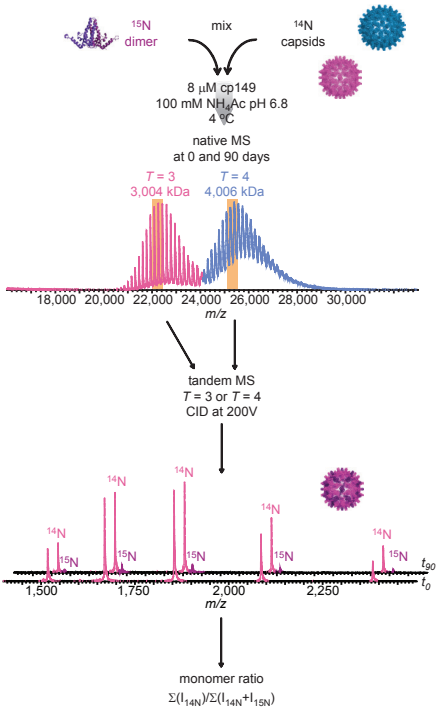


Fig. 1: Monitoring HBV capsid breathing by tandem MS
^{15}N-cp149$_d$ was mixed with cp149$_c$ and then analysed at intervals for up to 90 days by tandem MS. Precursor ions of $T = 3$ and $T = 4$ cp149$_c$ were selected (represented by vertical tan-colored bands) and individually subjected to CID. The relative amount of ^{15}N-cp149$_d$ incorporated into the intact cp149$_c$ was determined from the ion intensities of the ejected monomers in the tandem MS spectrum, as shown for the $T = 3$ cp149$_c$. Monomer ratios were obtained based on the summed intensities of the different charge states.

Any viral capsid must be very stable to afford protection to the enclosed genome during transit between hosts; however, it must also be capable of releasing the genome - and in some cases, completely disassembling - when the appropriate compartment of an infected cell is reached. Similarly, effective drug delivery requires cargo release and/or vehicle disassembly at the targeted site. Whereas capsid structures determined by X-ray crystallography and cryo-EM give the impression of static stable structures, there is evidence

- particularly for picornaviruses, which have icosahedral capsids of similar size to those of HBV - that capsids can be quite dynamic. [Katpally *et al.*, 2009; Lewis *et al.*, 1998; Li *et al.*, 1994] Thus polypeptides visualised to be internal to the closed capsid shell were found to be accessible to proteases and antibodies in the external milieu, implying that they are transiently exposed on the outer surface. This phenomenon has been termed "breathing".

Whereas the assembly of HBV capsids has been studied extensively (see above), little is yet known about their dynamic properties and disassembly. Singh and Zlotnick observed hysteresis in HBV capsid assembly/disassembly and proposed that weak protein-protein interactions can account for both capsid stability and breathing. [Singh & Zlotnick, 2003] Here we have exploited the high mass discrimination achievable by tandem MS, in conjunction with isotopic labelling, to examine dimer exchange under dilute conditions in the absence of denaturants, for both capsid forms.

2. Materials and methods

The proteins analysed were prepared as follows. The 149-residue assembly domain (hereafter cp149) was expressed in *E. coli*, purified as dimers ($cp149_d$), and stored in 100 mM sodium carbonate (pH 9.6), as described previously. [Wingfield *et al.*, 1995; Zlotnick *et al.*, 1996] The three cysteine residues in the native protein were substituted with alanine in the construct used. The corresponding ^{15}N-labelled protein (^{15}N-$cp149_d$) was prepared in a similar manner. $cp149_d$ was induced to assemble into capsids ($cp149_c$) by adjusting solution conditions to 150 mM sodium chloride and 100 mM HEPES (pH 7.0) by gel filtration. [Wingfield *et al.*, 1995; Zlotnick *et al.*, 1996] The ratio of $T = 3 : T = 4$ was 43 : 57, as determined by counting 1261 capsids in cryo-EM (the wild-type cp149 assembly domain gives a majority of $T = 4$ capsids [Zlotnick *et al.*, 1996]). To observe exchange, it was necessary to co-incubate the proteins under conditions where $cp149_c$ and $cp149_d$ co-exist under equilibrium conditions. cp149 remains dimeric at low ionic strength and high pH (ca. 9.6) at 4 ºC, but assembles into $cp149_c$ when the ionic strength is raised and the pH is adjusted to neutral, as described above. Raising the temperature also drives the assembly reaction. Once assembled, however, $cp149_c$ remain fully assembled at low ionic strength and high pH. [Singh & Zlotnick, 2003] In view of these considerations, the incubations described below were conducted under conditions of low ionic strength, near-neutral pH, and low temperature, unless otherwise indicated. Ammonium acetate was employed because previous experiments indicated that the proteins were properly folded, assembled, and soluble in this buffer and also because of its compatibility with the MS. [Uetrecht *et al.*, 2008a; Uetrecht *et al.*, 2008b] To test for dimer exchange, $cp149_c$ were mixed with ^{15}N-$cp149_d$ in equimolar ratio (in terms of dimers) at a total protein concentration of 8 μM in 100 mM ammonium acetate (pH 6.8) at 4 °C. Controls to assess the long term integrity of the $cp149_c$ and ^{15}N-$cp149_d$ were prepared under identical conditions. Immediately after mixing, and after 90 days of incubation, samples were withdrawn and analysed by tandem MS using a modified Q-ToF1 instrument (Waters, UK) as described previously. [Heck, 2008; van den Heuvel *et al.*, 2006] Precursor ions at *m/z* values of 21,700 and 24,800, corresponding to $T = 3$ and $T = 4$ $cp149_c$ respectively, were selected and subjected to collision induced dissociation (CID) at an accelerating voltage of 200 V, whereby, as shown previously, cp149 monomers were ejected from $cp149_c$. [Uetrecht *et al.*, 2008a] The relative amount of ^{15}N-cp149 monomer, compared to the amount of ^{14}N-cp149 monomer, is therefore a measure of the amount of $cp149_d$ exchanged between the capsids and free ones in solution. In other words, monomers constitute the readout but dimers are exchanged in solution. Denaturation studies of dimers show that they are very stable, with no evidence of separation of monomers prior to global unfolding. [Steven *et al.*, 2005]

3. Results

To monitor for dimer exchange, ^{14}N-containing $cp149_c$ were incubated with ^{15}N-$cp149_d$ and

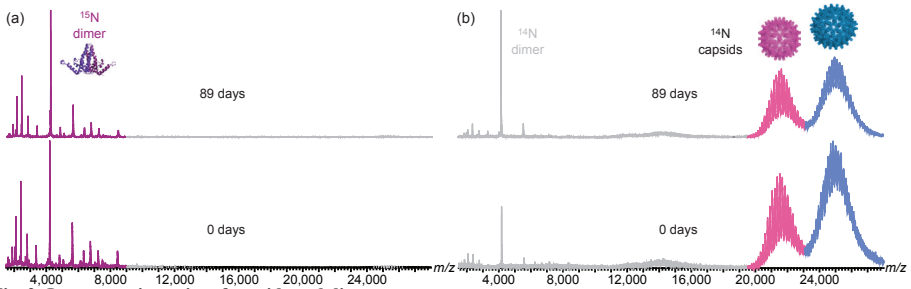


Fig. 2: Long-term integrity of capsids and dimers
MS of the ^{15}N-cp149$_d$ (a) and cp149$_c$ (b) controls at 0 and 89 days of incubation at 4 °C shows that the proteins were stable. The ^{15}N-cp149$_d$ did not assemble into capsids, and cp149$_c$ did not dissociate to dimers.

analysed by tandem MS as described above. As depicted in Fig. 1, after 90 days a significant number of ^{15}N-cp149 monomer ions were detected, indicating that cp149$_d$ exchange had occurred between the $T = 3$ cp149$_c$ and the soluble pool. In contrast, essentially no incorporation was detected in the case of $T = 4$ cp149$_c$. The long-term integrity of the cp149$_c$ and ^{15}N-cp149$_d$ controls incubated separately was assessed and confirmed as shown in Fig. 2. Both proteins were found to be completely stable under these conditions for at least 89 days; neither did cp149$_d$ assemble to cp149$_c$ nor did cp149$_c$ dissociate to cp149$_d$. Dynamic light scattering and circular dichroism measurements of the proteins did not change over this time period (not shown).

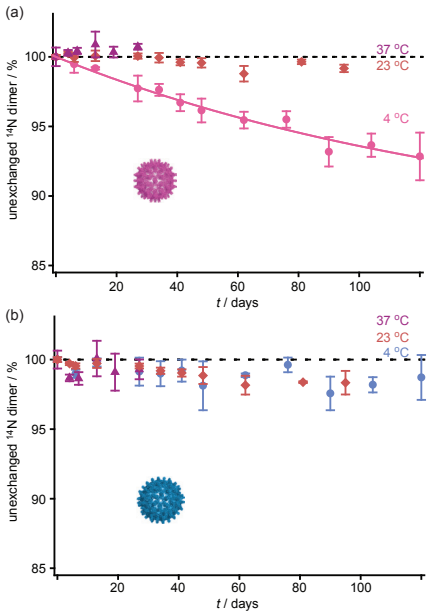


Fig. 3: Subunit exchange depends on capsid geometry and temperature
Dimer exchange was monitored by tandem MS at 4 °C, 23 °C, and 37 °C for both $T = 3$ (a) and $T = 4$ (b) cp149$_c$ for up to 120 days. No significant exchange was observed at 23 °C (orange) and 37 °C (purple) for either geometry. At 4 °C, the $T = 3$ cp149$_c$ exchanged cp149$_d$ with free ones in solution (pink), an exponential fit is shown indicating an endpoint of approximately 13%. $T = 4$ cp149$_c$ do not show significant levels of exchange at the same temperature (blue). Error bars correspond to the standard deviation of triplicate analysis. The dashed line represents the 100% level.

3.1. T = *3 capsids exchange at low temperatures*

To investigate the temperature dependence of cp149$_d$ exchange, we incubated cp149$_c$ with ^{15}N-cp149$_d$ at different temperatures. Samples were withdrawn at intervals for up to 120 days and analysed by tandem MS. The results are shown in Fig. 3. No significant exchange of cp149$_d$ was observed in either capsid at the elevated temperatures (23 °C and 37 °C) but a significant portion of the cp149$_d$ in $T = 3$ cp149$_c$ was exchanged at 4 °C (Fig. 3(a)). An exponential fit to the data suggests that only a subfraction of the dimers are subject to exchange. No significant exchange of cp149$_d$ was observed in the case of $T = 4$ cp149$_c$ at any of the tested temperatures (Fig. 3(b)).

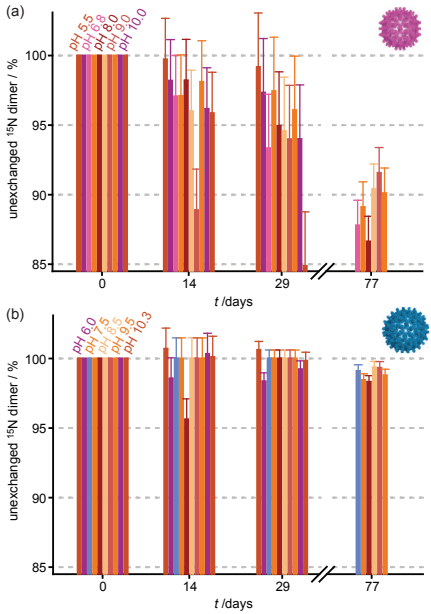


Fig. 4: Exchange rates are largely unaffected by change in pH
Dimer exchange was monitored in samples taken from incubations with pH values ranging from 5.5 to 10.3, for both $T = 3$ (a) and $T = 4$ (b) $cp149_c$. Significant exchange occurred only in the $T = 3$ form, but there was no clear pH dependency for either geometry.

3.2. Dimer exchange is pH independent

To detect a potential pH-dependence for $cp149_d$ exchange, incubations were done at pH values ranging from 5.5 to 10.3 in 100 mM ammonium acetate at 4 °C. Samples were withdrawn at intervals for up to 80 days and analysed by MS. The results are shown in Fig. 4. For both $T = 3$ and $T = 4$ $cp149_c$ there was no significant pH dependence. Both capsid geometries appeared stable across this pH range as assessed by native gel electrophoresis (not shown), however, the protein was observed to aggregate at more acidic pH. Attempts to assess the ionic strength dependence of this system by additions of sodium chloride were precluded by the incompatibility of this salt with the MS.

4. Discussion

Both forms of $cp149_c$ are stable, typically requiring >3 M urea and pH 9.6 for dissociation. As to their relative stabilities, simple curvature or vertex-proximity considerations predict that the $T = 3$ structure should be less stable, [Albertazzi *et al.*, 1999] *i.e.* less resistant to mechanical stress. However, atomic force microscopy measurements of deformability have detected no difference between the two forms. [Uetrecht *et al.*, 2008a] On the other hand, results from CID during tandem MS - in which the capsids are subjected to a different kind of stress (being transferred *in vacuo*) - showed the $T = 3$ form to better maintain its integrity under these conditions. [Uetrecht *et al.*, 2008b] The results presented here demonstrate that the $T = 3$ capsid exchanges dimers with the soluble pool slowly but definitely over a period of months, whereas the $T = 4$ capsid exchanges much more slowly, if at all. This raises the questions of where and how exchange might occur.

Capsid assembly is increased *in vitro* by raising the temperature (our unpublished observations), and, as shown here, dimer exchange is greater in the cold. Assembly of capsids is driven by the burial of hydrophobic surface area. [Ceres & Zlotnick, 2002] As it is well known that hydrophobic interactions decrease in strength with de-

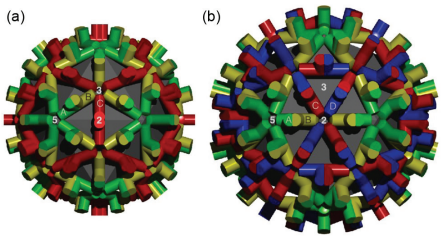


Fig. 5: Surface lattice of $T = 3$ and $T = 4$ HBV capsids
Shown is the arrangement of dimers in the $T = 3$ (a) and $T = 4$ (b) icosahedral geometries. The quasi-equivalent sites are labelled in both cases ($T = 3$; A, B, C and $T = 4$; A, B, C, D). Note that sites with the same symbol do not have identical structures. The 5, 3, and 2-fold axes of symmetry are also indicated. In the $T = 3$ lattice, the red C dimer lies on the two-fold axis of symmetry, and across two facets of the icosahedron. No such environment exists in the $T = 4$ geometry. Figure reproduced from [Steven *et al.*, 2005].

creasing temperature [Scheraga *et al.*, 1962] this suggests that exchange occurs by the selective extraction of dimers with weaker hydrophobic contacts. In the two HBV capsids, there is a total of seven quasi-equivalent subunits with their associated environments; three (A, B, C) in the case of $T = 3$ and four (A, B, C, D) in the case of $T = 4$ (Fig. 5). [Steven *et al.*, 2005] Similarly, two kinds of dimers are distinguished by quasi-equivalence in each capsid: 60 A-B dimers and 30 C-C dimers in the $T = 3$ capsid, and 60 A-B dimers and 60 C-D dimers in the $T = 4$ capsid. The data suggest that only a small subset of the dimers in the $T = 3$ capsid are exchangeable. In this regard we note that the $T = 3$ C-C dimer is the only one within the two geometries that lies across two facets of the icosahedron (Fig. 5(a)). It is plausible that this extra curvature results in less stable contacts between the dimer and its neighbors.

In conclusion, we report the first measurements of the rates at which the two polymorphic variants of the HBV capsid exchange dimeric building-blocks with the soluble pool. Exchange is slow but real for the $T = 3$ capsid (~7% in 3 months) but unmeasurably slow for the $T = 4$ capsid. The results have implications for drug delivery vehicle stability and formulation, and for the process of uncoating during the course of natural HBV infection (also termed "breathing"), [Rabe *et al.*, 2009] for instance, that the dissociation probably does not proceed spontaneously but instead involves an active mechanism.

References

Albertazzi, E., Domene, C., Fowler, P. W., Heine, T., Seifert, G., Van Alsenoy, C. & Zerbetto, F. (1999). Pentagon adjacency as a determinant of fullerene stability. *Phys Chem Chem Phys* **1**, 2913-2918.

Ceres, P. & Zlotnick, A. (2002). Weak protein-protein interactions are sufficient to drive assembly of hepatitis B virus capsids. *Biochemistry* **41**, 11525-11531.

Freund, S. M., Johnson, C. M., Jaulent, A. M. & Ferguson, N. (2008). Moving towards high-resolution descriptions of the molecular interactions and structural rearrangements of the human hepatitis B core protein. *J Mol Biol* **384**, 1301-1313.

Heck, A. J. (2008). Native mass spectrometry: a bridge between interactomics and structural biology. *Nat Methods* **5**, 927-933.

Katpally, U., Fu, T. M., Freed, D. C., Casimiro, D. R. & Smith, T. J. (2009). Antibodies to the buried N terminus of rhinovirus VP4 exhibit cross-serotypic neutralization. *J Virol* **83**, 7040-7048.

Lesmana, L. A., Leung, N. W., Mahachai, V., Phiet, P. H., Suh, D. J., Yao, G. & Zhuang, H. (2006). Hepatitis B: overview of the burden of disease in the Asia-Pacific region. *Liver Int* **26 Suppl 2**, 3-10.

Lewis, J. K., Bothner, B., Smith, T. J. & Siuzdak, G. (1998). Antiviral agent blocks breathing of the common cold virus. *Proc Natl Acad Sci USA* **95**, 6774-6778.

Li, Q., Yafal, A. G., Lee, Y. M., Hogle, J. & Chow, M. (1994). Poliovirus neutralization by antibodies to internal epitopes of VP4 and VP1 results from reversible exposure of these sequences at physiological temperature. *J Virol* **68**, 3965-3970.

Packianathan, C., Katen, S. P., Dann, C. E., 3rd & Zlotnick, A. (2010). Conformational changes in the hepatitis B virus core protein are consistent with a role for allostery in virus assembly. *J Virol* **84**, 1607-1615.

Rabe, B., Delaleau, M., Bischof, A., Foss, M., Sominskaya, I., Pumpens, P., Cazenave, C., Castroviejo, M. & Kann, M. (2009). Nuclear entry of hepatitis B virus capsids involves disintegration to protein dimers followed by nuclear reassociation to capsids. *PLoS Pathog* **5**, e1000563.

Scheraga, H. A., Nemethy, G. & Steinberg, I. Z. (1962). The contribution of hydrophobic bonds to the thermal stability of protein conformations. *J Biol Chem* **237**, 2506-2508.

Singh, P., Gonzales, M. J. & Manchester, M. (1996). Viruses and their uses in nanotechnology. *Drug Dev Res* **67**, 23-41.

Singh, S. & Zlotnick, A. (2003). Observed hysteresis of virus capsid disassembly is implicit in kinetic models of assembly. *J Biol Chem* **278**, 18249-18255.

Steven, A. C., Conway, J. F., Cheng, N., Watts, N. R., Belnap, D. M., Harris, A., Stahl, S. J. & Wingfield, P. T. (2005). Structure, assembly, and antigenicity of hepatitis B virus capsid proteins. *Adv Virus Res* **64**, 127-165.

Uetrecht, C., Versluis, C., Watts, N. R., Roos, W. H., Wuite, G. J., Wingfield, P. T., Steven, A. C. & Heck, A. J. (2008a). High-resolution mass spectrometry of viral assemblies: molecular composition and stability of dimorphic hepatitis B virus capsids. *Proc Natl Acad Sci USA* **105**, 9216-9220.

Uetrecht, C., Versluis, C., Watts, N. R., Wingfield, P. T., Steven, A. C. & Heck, A. J. (2008b). Stability and shape of hepatitis B virus capsids *in vacuo*. *Angew Chem Int Ed Engl* **47**, 6247-6251.

van den Heuvel, R. H., van Duijn, E., Mazon, H. & other authors (2006). Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. *Anal Chem* **78**, 7473-7483.

Wingfield, P. T., Stahl, S. J., Williams, R. W. & Steven, A. C. (1995). Hepatitis core antigen produced in *Escherichia coli*: subunit composition, conformational analysis, and *in vitro* capsid assembly. *Biochemistry* **34**, 4919-4932.

Wynne, S. A., Crowther, R. A. & Leslie, A. G. (1999). The crystal structure of the human hepatitis B virus capsid. *Mol Cell* **3**, 771-780.

Zlotnick, A., Cheng, N., Conway, J. F., Booy, F. P., Steven, A. C., Stahl, S. J. & Wingfield, P. T. (1996). Dimorphism of hepatitis B virus capsids is strongly influenced by the C-terminus of the capsid protein. *Biochemistry* **35**, 7412-7421.

Chapter VII

Summary

Abbreviations: CID – collision induced dissociation; cp – capsid protein; EM – electron microscopy; ESI – electrospray ionisation; HBV – Hepatitis B virus; IM – ion mobility; IMMS – ion mobility mass spectrometry; MS – mass spectrometry; m/z – mass to charge ratio; NVLP – norovirus-like particle

1. Summarising discussion

Nowadays, it has been accepted that most proteins do not act as individuals in the cellular context. Instead, they form large and complex interaction networks involving also transient assemblies. [Alberts, 1998] Proteomics studies reveal glimpses at such networks, but lack the information how the assembly of a protein complex is determined. [Heck, 2008] Often, additional proteins like chaperones are required for successful complex formation increasing the complexity of the system and hampering more detailed *in vitro* analyses. [van Duijn *et al.*, 2005] In light of this, viruses and more importantly the protein shells or capsids, they contain, represent ideal model systems to gain deeper insights to protein complex (self)-assembly. [Steinmetz *et al.*, 2009] Besides this general mechanistic aspect, understanding of the viral assembly processes is crucial for various applications in healthcare and technology. [Mitragotri & Lahann, 2009; Singh *et al.*, 2006] Consequently, viruses and their assembly have been extensively characterised in the past decades, even though many open questions still remain including the initiation and completion of capsid formation.

1.1. Thesis overview

The completion problem has been addressed in chapter II, where we analysed the $T = 3$ and $T = 4$ icosahedral capsids of Hepatitis B virus (HBV) by native mass spectrometry (MS). From atomic structures, a lot of information regarding the subunit interactions in the capsids can be deduced. Due to the averaging procedures applied in X-ray crystallography and electron microscopy (EM), it remains unclear whether the last capsid protein (cp) dimer becomes incorporated. Previous instrumental optimisations in native MS enabled an unprecedented mass resolution for the ~3 and 4 MDa species. The obtained mass accuracy of 0.1% allowed the precise determination of the capsid stoichiometries. Tandem MS on the selected capsid ions further confirmed the results, showing that the two capsid geometries of HBV assemble indeed completely. However, mass resolution is highly dependent on the sample heterogeneity. The full length cp packages varying amounts of small RNAs, which obscured the exact mass measurement. Since the $T = 3$ and $T = 4$ geometries only differ in the number of cp (180 and 240 for $T = 3$ and $T = 4$, respectively), it is tempting to speculate about the influence of capsid curvature on the structural integrity. Interestingly, the mechanical properties of the two capsid sizes were indistinguishable when probed by atomic force microscopy (AFM) in solution.

The capsids are surprisingly stable during transfer into the gas phase and through the mass spectrometer. This was shown by ion mobility MS (IMMS) in chapter III. The capsid radii estimated from the IM data were in good agreement with EM results. Surprisingly, the hollow spherical shape of the capsids was even preserved upon collision induced dissociation (CID) of an individual building block from the capsids. For each geometry at least two conformers were observed, which displayed similar electrospray ionisation (ESI) properties. In conclusion, fast interconverting capsid conformers apparently exist in solution that can be locked in either state during the desolvation process prior to MS analysis. As the capsid structure can be retained in the gas phase, the differential dissociation behaviour in tandem MS could report stability modulations in the two geometries. The studied cp construct forms an excess of the $T = 4$ particles, which may relate to a higher stability of these. However, in the gas phase $T = 3$ capsids showed a higher resistance towards CID. So far, one other investigation compared the two morphologies indicating an increased tolerance for amino acid insertion at the spike tips of the cp in $T = 3$.

[Bottcher *et al.*, 2006]

In chapter IV, we investigated the disassembly pathways of norovirus-like particles (NVLP). Due to a length variance in the cp, charge state resolution was not achieved for the $T = 3$ capsid. Nevertheless, a peak was observed at the mass to charge ratio (m/z) expected for a 10 MDa particle. The detected disassembly products were highly dependent on the pH and ionic strength of the ammonium acetate solution. At acidic pH chiefly cp dimers were present, whereas also higher order oligomers formed at basic pH and high ionic strength. Those larger oligomers most likely arose from initial dissociation of the $T = 3$ capsids into dimers and subsequent reassembly. The main products of this pathway were identified as 60- and 80-mers. The transition between $T = 3$ particles, dimers and the other oligomers was shown to be reversible. Previously, EM studies had already indicated that $T = 1$ capsids consisting of 60 cp subunits formed at basic pH. Using IMMS and AFM we could confirm that the 60-mers indeed were in agreement with a spherical assembly of similar size as the $T = 1$ particles in EM. However, an 80-mer had not been reported before. Despite of a subunit number which is at odds with an icosahedral symmetry, the IMMS analysis suggested that the 80-mer also forms a spherical capsid-like structure. The symmetry of the 80-mer is still unknown and further investigations hopefully elucidate its nature.

A major question is the pathway of virus or capsid assembly. The oligomers being intermediate between building blocks and capsids are low abundant and therefore basically inaccessible to classical structural analyses. In chapter V, we used the high sensitivity in IMMS to structurally characterise the observed oligomers of norovirus at basic pH and oligomers of HBV cp coexisting with the capsids. The reversible assembly of NVLP may indicate on-pathway products even under equilibrium conditions. We compared the shape of the viral oligomers to more globular protein complexes and modelled collision cross sections. For both viruses, the potential intermediates exhibited extended sheet-like structures as would be expected for on-pathway products, whereas aggregates would more resemble the dimensions of globular protein assemblies. Additional investigations under solution conditions that suppressed formation of capsids or capsid-like particles suggested a decameric assembly nucleus for norovirus and a hexameric for HBV capsids. The nucleus formation in HBV is likely accompanied by a conformational change. Based on the IMMS data, we proposed a common assembly pathway for the two viruses. Initially, a 5-fold vertex forms and subsequently grows by addition of dimers.

Chapter VI highlights that native MS cannot only be applied to monitor capsid stability but also dynamics. Theoretical assembly models predicted that dimers incorporated in capsids should exchange with free dimers in solution. A process termed “capsid breathing”. [Ceres & Zlotnick, 2002] We incubated preassembled HBV $T = 3$ and $T = 4$ capsids together with isotopically labelled cp dimers for prolonged times. The conditions were chosen to ensure the integrity of both dimers and capsids, in other words in which neither the capsids dissociated nor the dimers assembled. At multiple time points tandem MS was selectively performed on $T = 3$ or $T = 4$ capsids to determine the amount of isotopically labelled dimer incorporated in the otherwise unlabelled capsids. In line with the hydrophobic nature of the cp interactions in the capsids, exchange was only observed at low temperatures. Overall, the process was very slow and only the $T = 3$ capsids significantly incorporated dimers over the investigated timeframe of several month. The exchange rate was consistently slow across a broad pH range, which could explain

the strong hysteresis observed in HBV capsid disassembly.

1.2. Challenges

Generally, native MS is restricted to solution conditions, which avoid involatile salts or large amounts of denaturants. This complicates the interrogation of disassembly pathways for extremely stable capsids. For example, HBV capsids have been shown to efficiently dissociate into dimers at pH 9.6 in presence of molar urea concentrations. Even though urea is slightly volatile, such large amounts are incompatible with native MS analysis. Possible replacements for the classical denaturants are MS friendly organic modifiers. However, these can lead to alternate disassembly pathways.

Ideally, dissociation and assembly processes would be followed in real time to directly identify on-pathway intermediates. There are only few examples in the literature in which this was successfully conceived. [Stockley *et al.*, 2007] For most viruses, capsid assembly occurs very fast or not at all on detectable levels. The same holds true for the disassembly of NVLPs. ESI sources equipped with stopped-flow devices as used in protein folding studies could enable the time resolved monitoring of product formation in either direction. [Konermann *et al.*, 1997]

To correctly assign masses in native MS, the homogeneity of the sample is crucial. This effect becomes more important with increasing assembly size, especially in terms of the component number. [Benesch *et al.*, 2006] As we have seen, viral capsids are constituted by hundreds of copies of the cp subunits. The high multiplicity leads to a loss of mass resolution for the capsid signal, if only a low percentage of cp possesses a different mass, for example due to a lacking amino acid. Also heterogeneity in the genome content can easily hamper the exact mass determination asking for well controlled sample production, purification and preparation.

1.3. Outlook

Encapsulated material, *e.g.* the viral genome, is likely to affect the capsid structure and assembly process. Therefore, future projects aim at analysing the ionisation behaviour, the gas phase structure and stability of filled particles. Using organic modifiers or extreme pH, the solution stability and dissociation products of packed and empty particles could be compared in native MS. Additionally, influences of the nucleic acid on the mechanical properties may be monitored by AFM. Further development of soft-landing techniques could improve the sample quality and recovery to also collect minor species. Together with the *m/z* selection possible in quadrupole MS, this could enable structural investigations with EM or AFM on low abundant viral intermediates. [Benesch *et al.*, 2010] Another technique that may become available in the near future intends to utilise ultra-short and intense X-ray laser pulses to obtain high resolution structures of individual protein complexes in the gas phase. In this case, neither soft-landing of ions nor protein crystallisation is required. [Neutze *et al.*, 2000] Such methods are complementary to the here applied native MS and IMMS approaches and could provide higher resolution structural information invaluable to structural biologists.

2. Nederlandse Samenvatting

De meeste eiwitten spelen geen individuele rol in de cel maar zijn onderdeel van grote en complexe interactieve netwerken, waaronder ook kortstondige netwerken. Proteomics studies hebben deze netwerken grotendeels in kaart gebracht maar hierbij ontbreekt veelal informatie over de samenstelling van een eiwitcomplex. Vaak zijn andere eiwitten, zoals chaperone eiwitten, noodzakelijk voor een succesvolle complexvorming; hierdoor neemt de complexiteit van het syteem toe en wordt een meer gedetailleerde in vitro analyse verhinderd.

Virussen, of simpeler de eiwitomhulsels van virussen (ook wel capsiden genoemd), zijn uitermate geschikt als modelsysteem om meer inzicht te verkrijgen in de mechanismen van eiwitten om complexen te vormen. Kennis over deze virale eiwitcomplexvorming is ook van belang voor verscheidene technologische toepassingen en voor toepassingen in de gezondheidszorg. Derhalve is er de laatste decennia hard gewerkt aan de karakterisering van virussen en hun complexvorming, maar er zijn nog veel vragen onbeantwoord waaronder de initiatie en de voltooiing van capsidevorming.

In Hoofdstuk II wordt onderzoek naar de voltooing van capsiden beschreven. $T = 3$ en $T = 4$ icosahedrale capsiden van het Hepatitis B virus (HBV) worden met behulp van natieve massaspectrometrie (MS) geanalyseerd. Veel informatie over de interacties tussen subunits in capsiden kan worden afgeleid uit de atomaire structuren. Zowel kristallografie als elektronenmicroscopie (EM) kunnen echter geen uitsluitsel geven over de incorporatie van het laatste capside eiwit dimeer omdat bij deze technieken het gemiddelde van meerdere metingen wordt gebruikt. Tegenwoordig kan, door instrumentele verbeteringen, met natieve masssaspectrometrie een ongekend hoge massaresolutie bereikt worden bij metingen aan ~3 en 4 MDa complexen. Met de verkregen massanauwkeurigheid van 0,1% was het mogelijk om de capside stoichiometrie precies vast te stellen. Met tandem MS werd bevestigd dat beide HBV capside vormen volledig zijn. De massa van HBV gevuld met heterogeen endogeen RNA kon niet worden bepaald veroorzaakt door de complexe samenstelling van het RNA. De massaresolutie is dus sterk afhankelijk van de heterogeniteit van het monster. Aangezien de $T = 3$ en $T = 4$ capsiden alleen verschillen in het aantal capside eiwitten (180 en 240 voor $T = 3$ en $T = 4$, respectievelijk), is het verleidelijk te speculeren dat de capside kromming invloed heeft op de structurele integriteit. Er werd echter geen verschil in mechanische eigenschappen gevonden tussen de twee capside vormen in oplossing met ‘atomic force’ microscopie (AFM).

Met behulp van ion mobility MS (IMMS) wordt in Hoofdstuk III aangetoond dat capsiden verrassend stabiel zijn tijdens de overgang naar de gasfase en ook verder in het vacuum van de massaspectrometer. De capside afmetingen die zijn afgeleid uit de IMMS metingen zijn in goede overeenstemming met de EM resultaten. De holle, bolvormige vorm van de capsiden bleef zelfs intact na botsings-geïnduceerde dissociatie (collision-induced dissociation, CID). Van elke capside vorm werden tenminste twee conformaties gevonden, die vergelijkbare elektrospray ionisatie (ESI) eigenschappen vertoonden. Samenvattend, lijken er in oplossing verschillende capsideconformaties te bestaan, die dynamisch naar een andere conformatie kunnen overgaan. Tijdens het desolvatieproces, dat voor de MS analyse optreedt, wordt deze overgang stop gezet en zit het capside ‘gevangen’ in een bepaalde conformatie. Doordat de capsidestructuur in de gasfase behouden blijft, kan een verschil in dissociatiegedrag in tandem MS experimenten wijzen op stabiliteitsveranderingen tussen de twee vormen. Het capside eiwitconstruct

dat in Hoofdstuk III bestudeerd wordt vormt een overmaat aan $T = 4$ deeltjes, wat zou kunnen samenhangen met een verhoogde stabiliteit van deze deeltjes. In de gasfase vertonen de $T = 3$ deeltjes echter een hogere weerstand tegen CID. In de literatuur is slechts één ander onderzoek beschreven waarin de stabiliteit van twee vormen vergeleken werd; hierin wordt een verhoogde tolerantie voor aminozuurinsertie bij de uitstekende punten van het capside eiwit van $T = 3$ gevonden.

In hoofdstuk IV wordt de assemblage en dissociatie van norovirus-achtige deeltjes (NVLP) onderzocht. Door variatie in lengte van capside monomeer eiwitten, was er onvoldoende resolutie voor het exact vaststellen van de lading en massa. Desalniettemin werd er wel een brede piek gedetecteerd op de verwachte massa/lading verhouding voor een 10 MDa deeltje, de massas van het $T = 3$ NLVP deeltje. De NLVP grootte bleek sterk afhankelijk van de pH en de ionsterkte van de ammoniumacetaatbuffer: bij lage pH werden voornamelijk capside eiwit dimeren gemeten, terwijl er bij hoge pH en hoge ionsterktes ook grotere oligomeren gedetecteerd werden. Deze oligomeren zijn waarschijnlijk ontstaan door een initiële dissociatie van $T = 3$ capsiden in dimeren, gevolgd door een herassemblering. De meest voorkomende producten hierbij waren de 60-meren en 80-meren. De overgang van $T = 3$ deeltjes naar dimeren en de andere oligomeren bleek reversibel te zijn. Eerdere EM studies wezen al op de vorming van $T = 1$ capsides die bestaan uit 60 capside eiwiteenheden bij hoge pH. Met behulp van IMMS en AFM konden we bevestigen dat deze 60-meren inderdaad een bolvormige structuur hebben met een gelijke grootte als de $T = 1$ deeltjes bij EM. Een 80-meer is nog niet eerder beschreven. Hoewel het aantal eiwiteenheden niet in overeenstemming is met een icosaëdrische symmetrie, wijzen de resultaten van de IMMS analyse ook op een bolvormige, capside-achtige structuur. De symmetrie van de 80-meer is tot op heden onbekend; verder onderzoek is noodzakelijk voor verdere structuuropheldering.

In de viral struktuurbiologie zijn belangrijke vragen hoe het virus of capside gevormd worden. De oligomeren die de tussenvorm zijn van de bouwstenen en het uiteindelijke capside komen vrijwel niet voor en zijn daarom niet te analyseren met conventionele structuuronderzoek methoden. In Hoofdstuk V maken we gebruik van de hoge gevoeligheid van IMMS voor een structuuranalyse van de geobserveerde oligomeren van het norovirus bij hoge pH en van de HBV-oligomeren, die naast de capsiden bestaan. De reversibele opbouw van NVLP kan duiden op intermediairen, zelfs onder evenwichtcondities. We hebben de vorm van de virale oligomeren vergeleken met bolvormige eiwitcomplexen en dwarsdoorsnedes gemodelleerd. Bij beide virussen vertoonden de potentiële intermediairen uitgestrekte structuren, wat verwacht wordt voor intermediairen; eiwitaggregaten vertonen meer gelijkenis met de afmetingen van bolvormige eiwitcomplexen. Aanvullend onderzoek, waarbij in oplossing gewerkt werd, zodat de vorming van capsiden en capside-achtige deeltjes werd onderdrukt, toonde aan dat het norovirus waarschijnlijk bestaat uit een decameer als kern en dat de HBV capsides een hexameer bevatten. De kern vorming in HBV gaat waarschijnlijk gepaard met een conformatieverandering. We hebben een gemeenschappelijk assemblagepad voor de twee virussen opgesteld gebaseerd op de IMMS reultaten: initieel wordt een vijfvoudige knoop gevormd die vervolgens groeit door de toevoeging van dimeren.

Hoofdstuk VI accentueert dat natieve MS niet alleen toegepast kan worden om de capside stabiliteit te bestuderen maar ook de dynamiek van de capsiden. Theoretische assemblagemodel-

len voorspelde dat dimeren die geïncorporeerd in capsiden zitten, uitgewisseld kunnen worden met vrije dimeren in oplossing. Dit proces wordt ook wel “capside ademhaling” genoemd. We hebben geassembleerde HBV $T = 3$ en $T = 4$ capsiden voor langere tijd geincubeerd met isotoop gelabelde capside eiwit dimeren. De condities werden zo gekozen om de integriteit van zowel de dimeren als de capsiden zeker te stellen; met andere woorden condities waaronder de capsiden niet zouden dissociëren en de dimeren niet zouden assembleren. Op verschillende tijdstippen werd tandem MS selectief aan $T = 3$ of $T = 4$ uitgevoerd, om de hoeveelheid isotoop-gelabeld dimeer die geincorporeerd was vast te stellen in de verder ongelabelde capsiden. Door de hydrofobe aard van de capside eiwit interacties in het capside werd uitwisseling allen waargenomen bij lagere temperaturen. Het gehele verloop was erg langzaam en alleen de $T = 3$ capsiden incorporeerden significant dimeren in de onderzoeksperiode van enige maanden. De uitwisselsnelheid was langzaam over een breed pH gebied, wat de vertraagde response zou kunnen verklaren die gezien wordt bij inductie van een HBV capside dissociatie.

References

Alberts, B. (1998). The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* **92**, 291-294.

Benesch, J. L., Aquilina, J. A., Ruotolo, B. T., Sobott, F. & Robinson, C. V. (2006). Tandem mass spectrometry reveals the quaternary organization of macromolecular assemblies. *Chem Biol* **13**, 597-605.

Benesch, J. L., Ruotolo, B. T., Simmons, D. A., Barrera, N. P., Morgner, N., Wang, L., Saibil, H. R. & Robinson, C. V. (2010). Separating and visualising protein assemblies by means of preparative mass spectrometry and microscopy. *J Struct Biol* **doi**, 10.1016/j.jsb.2010.1003.1004.

Bottcher, B., Vogel, M., Ploss, M. & Nassal, M. (2006). High plasticity of the hepatitis B virus capsid revealed by conformational stress. *J Mol Biol* **356**, 812-822.

Ceres, P. & Zlotnick, A. (2002). Weak protein-protein interactions are sufficient to drive assembly of hepatitis B virus capsids. *Biochemistry* **41**, 11525-11531.

Heck, A. J. (2008). Native mass spectrometry: a bridge between interactomics and structural biology. *Nat Methods* **5**, 927-933.

Konermann, L., Collings, B. A. & Douglas, D. J. (1997). Cytochrome c folding kinetics studied by time-resolved electrospray ionization mass spectrometry. *Biochemistry* **36**, 5554-5559.

Mitragotri, S. & Lahann, J. (2009). Physical approaches to biomaterial design. *Nat Mater* **8**, 15-23.

Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. (2000). Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature* **406**, 752-757.

Singh, P., Gonzalez, M. J. & Manchester, M. (2006). Viruses and their uses in nanotechnology. *Drug Dev Res* **67**, 23-41.

Steinmetz, N. F., Lin, T., Lomonossoff, G. P. & Johnson, J. E. (2009). Structure-based engineering of an icosahedral virus for nanomedicine and nanotechnology. *Curr Top Microbiol Immunol* **327**, 23-58.

Stockley, P. G., Rolfsson, O., Thompson, G. S., Basnak, G., Francese, S., Stonehouse, N. J., Homans, S. W. & Ashcroft, A. E. (2007). A simple, RNA-mediated allosteric switch controls the pathway to formation of a *T*=3 viral capsid. *J Mol Biol* **369**, 541-552.

van Duijn, E., Bakkes, P. J., Heeren, R. M., van den Heuvel, R. H., van Heerikhuizen, H., van der Vies, S. M. & Heck, A. J. (2005). Monitoring macromolecular complexes involved in the chaperonin-assisted protein folding cycle by mass spectrometry. *Nat Methods* **2**, 371-376.

Appendix I

Supplementary data

Chapter II

High-resolution mass spectrometry of viral assemblies: Molecular Composition and stability of dimorphic Hepatitis B virus capsids

Table S1: Peaks in the charge state distributions of $T = 3$ and $T = 4$
Comparison of detected peaks for geometries $T = 3$ and $T = 4$ of $3^{C\rightarrow A}$ and 61^{C} from the mass spectra in Chapter II Fig. 2 with the calculated *m/z* for capsids built by 90, 89, 120 and 119 dimers. The monomeric masses of $3^{C\rightarrow A}$ and 61^{C} as based on the sequence are 16.674 kDa and 16.706 kDa, respectively. Complexes do not lose mass, but rather retain some buffer molecules when subjected to native mass spectrometry. [Benesch *et al.*, 2007] Therefore, detected peaks always appear at a higher *m/z* than calculated. In all cases analysed here, the calculated *m/z* for capsids consisting of 90 and 120 dimers was closer to the *m/z* of the corresponding detected peak, thus proving the capsid stoichiometry. The average difference between calculated and detected *m/z* was at least 1.5 fold higher for $T = 4$ capsids with 119 dimers than with 120 dimers. For $T = 3$ capsids the average difference in *m/z* was at least 3.5 fold higher for capsids of 89 dimers than of 90 dimers.

Benesch, J. L., Ruotolo, B. T., Simmons, D. A. & Robinson, C. V. (2007). Protein Complexes in the Gas Phase: Technology for Structural Genomics and Proteomics. *Chem Rev* **107,** 3544-3567.

$3^{C\rightarrow A}$					
$T = 3$ cp149			$T = 4$ cp149		
m/z_{detected}	$m/z_{\text{calculated}}$ 90 dimers	$m/z_{\text{calculated}}$ 89 dimers	m/z_{detected}	$m/z_{\text{calculated}}$ 120 dimers	$m/z_{\text{calculated}}$ 119 dimers
20567	20557	20469	24305	24253	24198
20709	20699	20611	24459	24401	24346
20854	20843	20755	24610	24551	24496
21000	20988	20901	24765	24702	24649
21148	21136	21050	24917	24856	24803
21301	21286	21200	25076	25011	24959
21457	21438	21352	25242	25168	25117
21610	21592	21507	25403	25328	25277
21768	21749	21664	25575	25489	25439
21932	21908	21823	25742	25652	25603
22093	22069	21985	25902	25818	25769
22259	22232	22149	26088	25986	25937
22425	22398	22316	26251	26155	26108
22598	22566	22485	26422	26327	26281
22769	22737	22656	26617	26502	26456
22946	22911	22831	26797	26679	26634
23127	23087	23008	26985	26858	26814
23306	23266	23187	27157	27039	26996
23492	23448	23370	27364	27223	27181
23680	23633	23555	27559	27409	27368
23864	23820	23744			

61[C]					
T = 3 cp149			T = 4 cp149		
$m/z_{detected}$	$m/z_{calculated}$ 90 dimers	$m/z_{calculated}$ 89 dimers	$m/z_{detected}$	$m/z_{calculated}$ 120 dimers	$m/z_{calculated}$ 119 dimers
20073	20046	19956	24049	24007	23951
20206	20181	20091	24191	24152	24096
20344	20317	20228	24340	24298	24243
20483	20455	20367	24492	24447	24392
20622	20595	20507	24640	24597	24542
20766	20737	20649	24795	24748	24695
20909	20881	20794	24949	24902	24849
21055	21027	20940	25105	25058	25005
21206	21176	21089	25266	25215	25163
21355	21326	21239	25423	25375	25324
21508	21478	21392	25587	25537	25486
21663	21633	21547	25751	25700	25650
21821	21789	21704	25920	25866	25817
21980	21948	21864	26087	26034	25986
22140	22110	22026	26260	26204	26157
22306	22274	22190	26428	26377	26330
22471	22440	22357	26611	26551	26505
22641	22608	22527	26785	26728	26683
22814	22780	22699	26975	26908	26864
22989	22954	22873	27150	27089	27046
23165	23130	23051	27337	27274	27232
23345	23309	23231	27523	27461	27419
23525	23492	23414	27714	27650	27610
23762	23677	23599	27908	27842	27803
23902	23864	23788			

Chapter III

Stability and shape of Hepatitis B virus capsids *in vacuo*

Materials and methods

Proteins

HBV capsids of cp149 $3^{C\rightarrow A}$ and cp149 61^{C} for mass spectrometry were purified as described previously. [Zlotnick *et al.*, 1996]

Mass spectrometry

High resolution tandem mass spectra were recorded on a modified Q-ToF 1 instrument (Waters, UK). [Benesch *et al.*, 2007; van den Heuvel *et al.*, 2006] The voltages and pressures were optimised for large noncovalent protein complexes. [Sobott *et al.*, 2002] Ion mobility measurements were performed on a Synapt HDMS (Waters, UK). [Pringle *et al.*, 2007] Separation in the ion mobility device was adjusted to measure proteins below and above 100 kDa. To correlate the data, GroEL single ring (SR) and PCNA were analysed with both settings. In all cases, the ion mobility chamber contained nitrogen. For proteins below 100 kDa argon was used as collision gas, whereas xenon was preferred for larger proteins. [Lorenzen *et al.*, 2007] Xenon was also used to measure HBV capsids with the modified Q-ToF 1. The measured drift times were corrected for the dead time before the pusher. The collision cross section (Ω) is obtained from the absolute drift time according to equation (1) with calibration on known Ω [Clemmer, ; Pringle *et al.*, 2007; Wildgoose *et al.*, 2006]:

$$\Omega = K \times t_D^{0.52} \times z \times \sqrt{\frac{1}{M_{ion}} + \frac{1}{m_{gas}}} \quad (1)$$

Where, K is the calibration constant, t_D is the absolute drift time, *i.e.* corrected for the dead time before the pusher, z represents the charge of an ion, M_{ion} and m_{gas} are the masses of the ion and the gas used in the ion mobility device. Mass and charge of an ion were determined from the corresponding mass spectra. Ω were calculated for each charge state and then averaged.

A list of the analysed proteins and their average Ω is given in Table S3. The Ω were determined from three independent measurements and for most proteins with two different separation settings, a fixed and ramped wave height in the ion mobility chamber. The fixed wave height resulted in larger Ω for high mass proteins, but no systematical error could be observed in the dataset. The Ω of fixed and ramped wave height show a good correlation (Fig. S1) indicating both settings might be used for determination of Ω. Drift time peaks were narrower and the standard deviations in Ω smaller with a ramped wave height, especially for proteins which left the ion mobility chamber at the end of a duty cycle. Therefore, we used these settings for the large viral capsids. The deviation in Ω over all charge states for small proteins is sometimes quite large because of the large difference between the conformers. Ω of an individual charge state deviated usually less than 1% indicating the good reproducibility of the results. The calibration constant and correlation coefficient for the small and large datasets showed a relative standard deviation of less than 4%. The performance of the instruments was tested with CsI and a calibration was applied where necessary.

The HBV capsids formed of cp149 $3^{C\rightarrow A}$ and cp149 61^{C} were introduced to the mass spectrometer at 6 - 8 µM monomeric concentrations in 200 mM ammonium acetate, pH 6.8. [Uetrecht et al., 2008] All other proteins were sprayed from 50 mM ammonium acetate (pH 6.8) at a concentration between 2 and 10 µM regarding the monomers. Capillaries for electrospray ionisation were prepared in house from borosilicate glass tubes of 1.2 mm OD and 0.68 mm ID with filament (World Precision Instruments, USA) using a P-97 micropipette puller (Sutter Instruments, USA) and gold-coated using an Edwards Scancoat six Pirani 501 sputter coater (Edwards Laboratories, USA). Capillary tips were opened on the sample cone of the instrument. For ion mobility, capillaries purchased from Waters (Manchester, UK) were used.

Benesch, J. L., Ruotolo, B. T., Simmons, D. A. & Robinson, C. V. (2007). Protein complexes in the gas phase: technology for structural genomics and proteomics. *Chem Rev* **107**, 3544-3567.
Clemmer, D. E., Cross Section Database.
Lorenzen, K., Versluis, C., van Duijn, E., van den Heuvel, R. H. H. & Heck, A. J. R. (2007). Optimizing macromolecular tandem mass spectrometry of large non-covalent complexes using heavy collision gases. *Int J Mass Spectrom* **268**, 198-206.
Pringle, S. D., Giles, K., Wildgoose, J. L., Williams, J. P., Slade, S. E., Thalassinos, K., Bateman, R. H., Bowers, M. T. & Scrivens, J. H. (2007). An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *Int J Mass Spectrom* **261**, 1-12.
Sobott, F., Hernandez, H., McCammon, M. G., Tito, M. A. & Robinson, C. V. (2002). A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Anal Chem* **74**, 1402-1407.
Uetrecht, C., Versluis, C., Watts, N. R., Roos, W. H., Wuite, G. J. L., Wingfield, P. T., Steven, A. C. & Heck, A. J. R. (2008). High-resolution mass spectrometry of viral assemblies: Molecular composition and stability of dimorphic hepatitis B virus capsids. *Proc Natl Acad Sci USA* **105**, 9216-9220.
van den Heuvel, R. H., van Duijn, E., Mazon, H. & other authors (2006). Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. *Anal Chem* **78**, 7473-7483.
Wildgoose, J. L., Giles, K., Pringle, S. D., Bateman, R. H., Koeniger, S., Clemmer, D. E. & Valentine, S. (2006).A Comparison of Travelling Wave and Drift Tube Ion Mobility Separations. In *ASMS*. Seattle, Wa.
Zlotnick, A., Cheng, N., Conway, J. F., Booy, F. P., Steven, A. C., Stahl, S. J. & Wingfield, P. T. (1996). Dimorphism of hepatitis B virus capsids is strongly influenced by the C-terminus of the capsid protein. *Biochemistry* **35**, 7412-7421.

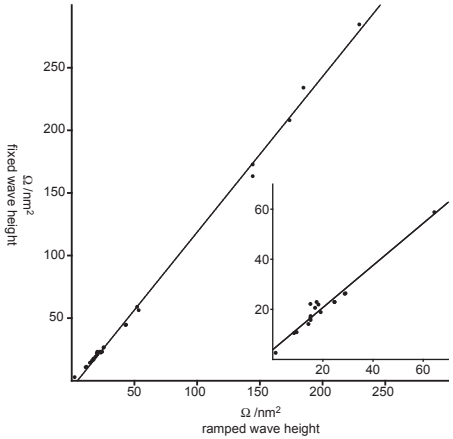


Fig. S1: *Ω* determined with a fixed or ramped wave height are comparable.
They show a linear correlation. No systematical errors are observed. *Ω* of large proteins measured with a fixed wave height are somewhat larger, as the *Ω* for GroEL single ring, which was used for correlation of the small and large datasets, was already greater than in the ramped wave height dataset. The inset shows a zoom in on proteins below 100 kDa which were analysed with identical fixed and ramped wave height settings.

Table S2: Comparison of capsid conformers from ion mobility
Relative amounts of small conformer and difference in *Ω* (Δ*Ω*) between large and small conformers for the different variants (in %). Values are averaged over all detected charge states at particular collision voltages (cv) and the standard deviation is given. The standard deviation for the amount of the geometry $T = 4$ small conformer at 200 V is high, due to overlap with fragmentation products of $T = 3$ capsids. In the case of $T = 3$ at 200 V the standard deviation is high, because highly charged ions fragment easier than less highly charged ions. As the stabilities of the small and large conformer differ, their ratio becomes charge state dependent upon fragmentation. This is also evident from the amount of the small conformer in $T = 3$ capsids missing a dimer. While increasing the collision energy from 175 V to 200 V the amount of small conformer over all charge states is reduced from 61 ± 9% to 43 ± 11% indicating that the small conformer has already lost another dimer.

cp149	cv /V	Small /%		ΔΩ /%	
		$T = 3$	$T = 4$	$T = 3$	$T = 4$
$3^{C\rightarrow A}$	75	79 ± 4	82 ± 4	4.4 ± 0.4	4.5 ± 0.3
	100	80 ± 5	79 ± 6	4.3 ± 0.5	4.6 ± 0.3
	150	74 ± 6	81 ± 5	4.4 ± 0.6	4.5 ± 0.3
61^{C}	100	75 ± 3	82 ± 4	4.1 ± 0.5	4.4 ± 0.5
	150	74 ± 2	82 ± 2	4.3 ± 0.4	4.4 ± 0.3
	175	69 ± 4	79 ± 4	4.4 ± 0.6	4.5 ± 0.6
	200	60 ± 11	77 ± 10	4.5 ± 0.4	4.2 ± 0.6

Table S3: Absolute Ω of the proteins and protein complexes analysed under near "native" *in vacuo* conditions

Given are the protein name, the number of subunits (*N*), the determined mass (*m*) and collision cross section (Ω) with standard deviation (± SD) as determined with fixed (fix) and ramped (ramp) wave height in the ion mobility chamber. HBV Ω are averaged over all measurements regardless of the applied acceleration energy. Synapt HDMS settings were different for small ([a]) and large ([b]) proteins and the Ω are therefore correlated on basis of the Ω of GroEL SR. nd – not detected

Protein	*N*	*m* /kDa	Ω /nm^2			
			Fix	± SD	Ramp	± SD
GroEL SR, *E.coli* [b]	7	399.8	163.1	9.6	144.6	8.7
Holomyoglobin, horse heart [a]	1:1	17.6	22.9	3.1	20.7	4.4
Holomyoglobin, horse heart [a]	2:2	35.2	nd	nd	32.1	4.4
Apomyoglobin, horse heart [a]	1:1	17.0	20.5	4.8	20.5	3.7
Lysozyme, hen egg white [a]	1	14.3	14.1	2.1	14.8	2.3
Lysozyme, hen egg white [a]	2	28.7	nd	nd	20.3	0.9
Carbonicanhydrase II, bovine [a]	1	29.1	26.3	4.3	25.8	3.5
Ubiquitin, bovine [a]	1	8.6	10.5	2.1	11.3	2.6
Im9, *E.coli* [a]	1	9.6	10.8	1.6	12.0	2.0
Im9, *E.coli* [a]	2	19.2	18.8	1.2	19.2	0.9
E9, *E.coli* [a]	1	15.1	16.4	3.3	17.7	4.4
E9 (Zn^{2+}), *E.coli* [a]	1:1	15.2	17.2	3.5	17.6	3.9
E9 (Ni^{2+}), *E.coli* [a]	1:1	15.1	15.7	2.9	16.4	3.5
E9 (Ni^{2+}):Im9, *E.coli* [a]	1:1:1	24.7	22.9	2.9	22.6	3.6
E9:Im9, *E.coli* [a]	1:1	24.7	22.8	3.0	24.5	2.4
PCNA, human [b]	3	86.4	56.0	7.2	53.7	4.3
PCNA, human [a]	1	28.8	26.2	3.2	25.7	3.4
Hemoglobin, bovine [a]	2:2:4	64.5	58.7	5.2	52.4	5.5
Heme [a]	1	1.3	nd	nd	2.7	$2.1e^{-2}$
Heme-Na^+ [a]	1:1	1.2	2.5	$1.4e^{-2}$	2.6	$1.4e^{-2}$
P22 gp4 [a]	1	18.3	21.9	3.1	20.3	3.6
GroEL, *E.coli* [b]	14	801.0	284.3	5.4	229.5	5.0
VAO, *Penicillium simplicis*. [b]	8	508.5	207.7	4.9	173.9	5.4
V_1-ATPase (subunit A), yeast [b]	1	67.6	44.4	2.6	43.6	3.3
V_1-ATPase III, yeast [b]	11	427.9	172.4	3.7	144.6	5.7
V_1-ATPase I, yeast [b]	13	592.8	233.7	6.1	185.0	5.9
P22 gp1 [b]	1	69.2	44.2	4.1	43.1	3.5
$T = 3\ 3^{C\rightarrow A}$ (small conformer) [b]	180	3006.5			683.6	6.3
$T = 3\ 3^{C\rightarrow A}$ (large conformer) [b]	180	3006.5			713.1	6.4
$T = 4\ 3^{C\rightarrow A}$ (small conformer) [b]	240	4016.2			869.0	6.2
$T = 4\ 3^{C\rightarrow A}$ (large conformer) [b]	240	4016.2			908.6	6.5
HBV cp149 $3^{C\rightarrow A}$ [b]	1	16.7			17.8	3.4
$T = 3\ 61^C$ (small conformer) [b]	180	3012.4			687.4	8.4
$T = 3\ 61^C$ (large conformer) [b]	180	3012.4			717.3	9.0
$T = 4\ 61^C$ (small conformer) [b]	240	4018.9			875.4	5.7
$T = 4\ 61^C$ (large conformer) [b]	240	4018.9			913.5	6.8
HBV cp149 61^C [b]	2	33.4			30.0	3.4

Chapter IV

Norwalk virus assembly and stability monitored by mass spectrometry

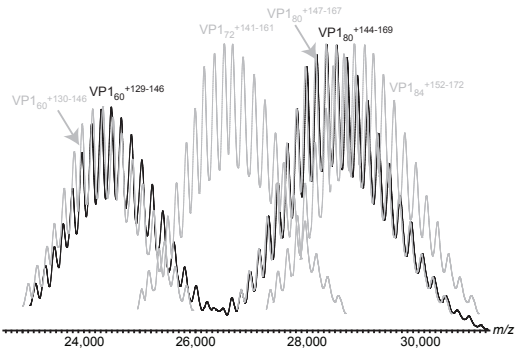


Fig. S2: Representative ESI mass spectra at high desolvation energy (200 V) of a solution containing 30 µM VP1 in a 250 mM ammonium acetate buffer (pH 9). Overlaid with this spectrum (light grey) are the expected charge state distributions for the $VP1_{60}$, $VP1_{80}$, $VP1_{72}$, and $VP1_{84}$ oligomers.

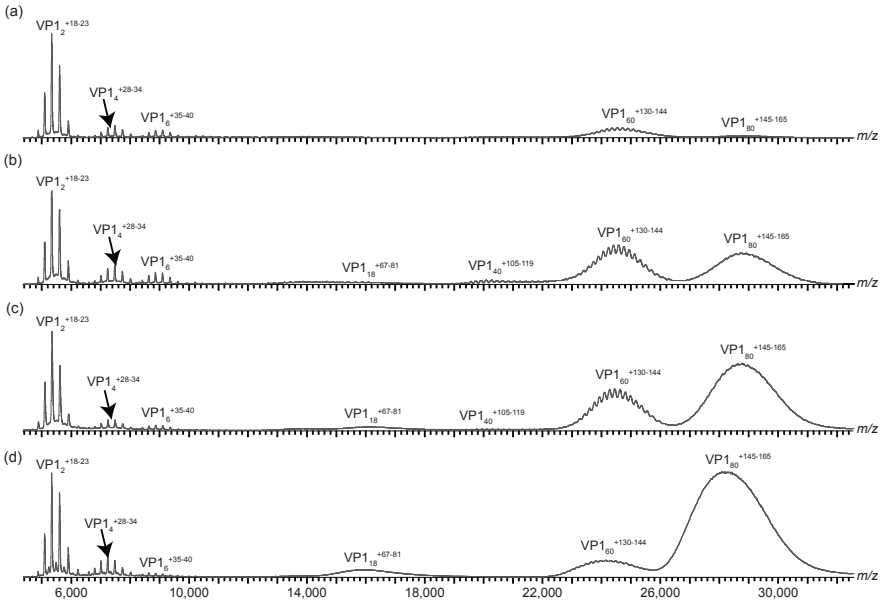


Fig. S3: Representative ESI mass spectra demonstrating the effect of protein concentration on the oligomeric state of VP1. Spectra were obtained from aqueous solutions of VP1 in a 250 mM ammonium acetate buffer (pH 9) at concentrations of (a) 16 µM, (b) 23 µM, (c) 40 µM and (d) 80 µM.

Table S4: Experimental and theoretical *m/z* values for the charge state distribution corresponding to $VP1_{80}$. To confirm that the stoichiometry was assigned correctly, theoretical *m/z* values for oligomers corresponding to ± 1 VP1 dimer are also included. The difference between the experimental and theoretical *m/z* are given in parentheses. The corresponding masses assuming the various stoichiometries were calculated for each *m/z* value. The average masses for $VP1_{82}$, $VP1_{80}$ and $VP1_{78}$ were found to be 4.620 ± 0.005 MDa, 4.4780 ± 0.0005 MDa and 4.364 ± 0.004 MDa, respectively. The lowest standard deviation in mass was assuming a stoichiometry of 80 VP1 monomers, indicating that this is the correct assignment. Due to incomplete desolvation during the ESI process, theoretical *m/z* values must always be less than the experimental *m/z* values, thus a negative deviation indicates an incorrect assignment.

Experimental *m/z*	$VP1_{80}$ theoretical	$VP1_{82}$ theoretical	$VP1_{78}$ theoretical
26812	26808 (< 0.1%)	26680 (0.5%)	26779 (0.1%)
26972	26970 (< 0.1%)	26836 (0.5%)	26945 (0.1%)
27136	27133 (< 0.1%)	26993 (0.5%)	27112 (< 0.1%)
27302	27299 (< 0.1%)	27153 (0.5%)	27282 (< 0.1%)
27471	27466 (< 0.1%)	27315 (0.4%)	27453 (< 0.1%)
27641	27636 (< 0.1%)	27478 (0.4%)	27627 (< 0.1%)
27813	27807 (< 0.1%)	27644 (0.4%)	27803 (< 0.1%)
27987	27981 (< 0.1%)	27811 (0.4%)	27981 (< 0.1%)
28164	28157 (< 0.1%)	27981 (0.4%)	28162 (< 0.1%)
28342	28335 (< 0.1%)	28153 (0.7%)	28344 -(<0.1%)
28523	28516 (< 0.1%)	28326 (0.7%)	28530 -(<0.1%)
28706	28698 (< 0.1%)	28502 (0.7%)	28717 -(<0.1%)
28892	28884 (< 0.1%)	28681 (0.7%)	28907 -(<0.1%)
29080	29071 (< 0.1%)	28861 (0.8%)	29100 -(<0.1%)
29271	29261 (< 0.1%)	29044 (0.8%)	29295 -(<0.1%)
29465	29454 (< 0.1%)	29229 (0.8%)	29493 -(0.1%)
29659	29649 (< 0.1%)	29416 (0.8%)	29694 -(0.1%)
29859	29846 (< 0.1%)	29606 (0.8%)	29897 -(0.1%)
30060	30047 (< 0.1%)	29798 (0.9%)	30104 -(0.1%)
30263	30250 (< 0.1%)	29993 (0.9%)	30313 -(0.1%)

Table S5: Experimental and theoretical *m/z* values for the charge state distribution corresponding to $VP1_{60}$. To confirm that the stoichiometry was assigned correctly, theoretical *m/z* values for oligomers corresponding to ± 1 VP1 dimer are also included. The difference between the experimental and theoretical *m/z* are given in parentheses. The corresponding masses assuming the various stoichiometries were calculated for each *m/z* value. The average masses for $VP1_{62}$, $VP1_{60}$ and $VP1_{58}$ were found to be 3.501 ± 0.005 MDa, 3.3783 ± 0.0008 MDa, and 3.255 ± 0.004 MDa, respectively. The lowest standard deviation in mass was assuming a stoichiometry of 60 VP1 monomers, indicating that this is the correct assignment. Due to incomplete desolvation during the ESI process, theoretical *m/z* values must always be less than the experimental *m/z* values, thus a negative deviation indicates an incorrect assignment.

Experimental *m/z*	$VP1_{60}$ theoretical	$VP1_{62}$ theoretical	$VP1_{58}$ theoretical
23126	22998 (0.6 %)	22978 (0.6%)	23020 (0.5%)
23302	23157 (0.6 %)	23131 (0.7%)	23184 (0.5%)
23464	23318 (0.6 %)	23286 (0.8%)	23351 (0.5%)
23622	23481 (0.6 %)	23444 (0.8%)	23520 (0.4%)
23785	23646 (0.6 %)	23603 (0.8%)	23692 (0.4%)
23956	23814 (0.6 %)	23765 (0.8%)	23866 (0.4%)
24127	23984 (0.6 %)	23929 (0.8%)	24043 (0.3%)
24301	24156 (0.6 %)	24095 (0.8%)	24222 (0.3%)
24480	24331 (0.6 %)	24263 (0.9%)	24405 (0.3%)
24659	24509 (0.6 %)	24434 (0.9%)	24589 (0.3%)
24839	24689 (0.6 %)	24608 (0.9%)	24777 (0.2%)
25027	24872 (0.6 %)	24783 (1.0%)	24968 (0.2%)
25214	25058 (0.6 %)	24962 (1.0%)	25161 (0.2%)
25404	25246 (0.6 %)	25142 (1.0%)	25358 (0.2%)
25601	25437 (0.6 %)	25326 (1.1%)	25558 (0.2%)
25795	25632 (0.6 %)	25512 (1.1%)	25760 (0.1%)
25999	25829 (0.7 %)	25701 (1.1%)	25966 (0.1%)
26196	26029 (0.6 %)	25893 (1.2%)	26176 (0.1%)

Table S6: Collision cross sections (*Ω*) of all proteins and protein complexes analysed
The protein name, number of subunits in the assembly (*N*) and the corresponding Protein Data Bank identifier (PDB-ID) are indicated. The masses (*m*) in kDa used in the figures are indicated, whereby the experimental mass was obtained from the mass spectra, the theoretical mass from the protein sequence and in the next column the total mass of all amino acids resolved in the PDB deposited structure (exp, theo and PDB, respectively). The *Ω* values as determined with ramped or fixed wave heights from a representative set of ion mobility experiments are shown together with the respective modelled *Ω* from the PDB deposited structures. Those theoretical *Ω* data were calculated with the projection approximation algorithm implemented in the mobcal software. [Mesleh et al., 1996; Shvartsburg & Jarrold, 1996] Viral intermediates were extracted from the capsid structures assuming 5-fold vertices as initial assemblies. Schemes of the intermediate structures are depicted in the last column whereby each stick represents a dimer. Experimental *Ω* values were averaged over three analysis and all detected charge states. The standard deviation for individual sets of experiments (day to day variation) was usually below 5%. nd – not detected

Protein	*N*	PDB-ID	*m* /kDa			*Ω* /nm²			Schematic structure
			exp	theo	PDB	ramped	fixed	mobcal	
Hemoglobin, bovine	2:2	1HDA	64.6	64.5	60.0	37.2	38.3	32.0	
Avidin, gallus	4	1AVE	63.4	63.3	51.3	33.8	33.5	31.0	
Lysozyme, gallus	1	193L	14.3	14.3	13.5	13.4	13.6	11.6	
Proteinase K, *Tritirachium album*	1	2PRK	29.0	29.0	26.8	21.6	21.3	17.8	
alkaline phosphatase, bovine	2	1EW2[a]	117.2	116.2	96.6	56.1	54.9	46.0	
PCNA, human	1	1VYM	28.8	28.8	26.1	23.6	23.9	20.8	
Exosome core, *Pyrococcus abyssi*	3:3	2BR2[b]	178.9	177.4	152.8	76.1	75.0	66.8	
Exosome core-Csl4, *Pyrococcus abyssi*	3:3:3	2JE6[b]	245.6	243.9	215.4	104.1	104.3	84.2	
Vanillyl-alcohol oxidase, *Penicillium simplicissimum*	4	1VAO	254.6	255.7	233.0	106.4	107.6	86.6	
Vanillyl-alcohol oxidase, *Penicillium simplicissimum*	8	1VAO	509.7	511.3	465.0	185.9	186.9	138.1	
Vanillyl-alcohol oxidase, *Penicillium simplicissimum*, plane (II)	8	1VAO	nd	511.3	465.0	nd	nd	163.5	
Vanillyl-alcohol oxidase, *Penicillium simplicissimum*, rod (III)	8	1VAO	nd	511.3	465.0	nd	nd	170.8	
Myoglobin, equine	1	1DWR	17.6	17.6	16.5	19.2	18.3	13.7	
Cytochrome C, equine	1	1HRC	12.4	12.4	11.5	14.5	14.3	10.6	
Ubiquitin, bovine	1	1UBQ[c]	8.6	8.6	7.9	11.3	11.1	8.9	
NMN/NaMN adenyltransferase, human	6	1KQN	200.1	191.6	152.0	88.8	86.2	66.1	
Elastase, porcine	1	3EST	25.9	25.9	24.3	19.7	19.4	17.7	
Catalase, bovine	6	1TGU	236.0	241.6	216.0	96.4	94.6	73.9	
Norovirus P domain	2	2ZL5	63.8	63.7	57.3	37.0	36.1	32.4	
α-glutamate synthase, *Azospirillum brasilense*	2	1EA0	326.7	326.7	297.6	134.0	132.8	103.2	
α-glutamate synthase, *Azospirillum brasilense*	4	1EA0	654.0	653.4	595.3	234.6	234.9	177.7	
α-glutamate synthase, *Azospirillum brasilense*	6	1EA0	980.9	980.1	892.9	339.7	347.0	227.9	
GroEL SR, *Escherichia coli*[d]	1	1KPO	57.1	57.2	57.1	35.6	35.0	35.1	

Protein	*N*	PDB-ID	*m* /kDa			*Ω* /nm²			Schematic structure
			exp	theo	PDB	ramped	fixed	mobcal	
GroEL SR, *Escherichia coli*[d]	7	1KPO	401.5	400.4	399.7	158.0	155.0	141.0	
GroEL, *Escherichia coli*[e]	14	1KPO	804.2	800.8	799.3	267.6	269.8	219.5	
HBV cp140	2	1QGT	31.7	31.7	31.9	32.8	32.7	22.9	
HBV cp140	4	1QGT	63.4	63.4	63.7	47.8	48.0	40.3	
HBV cp140	6	1QGT	95.3	95.0	95.6	60.9	62.9	57.3	
HBV cp140	8	1QGT	127.1	126.7	127.5	71.9	75.1	73.1	
HBV cp140	10	1QGT	159.0	158.4	159.3	82.9	87.1	85.8	
HBV cp140	12	1QGT	190.9	190.1	191.2	94.2	100.0	97.8	
HBV cp140	14	1QGT	222.8	221.8	223.1	102.5	111.8	109.7	
HBV cp140	16	1QGT	254.8	253.5	254.9	115.2	126.2	121.4	
HBV cp140	24	1QGT	383.1	380.2	382.4	164.1	181.2	176.2	
HBV cp140	28	1QGT	446.9	443.5	446.2	188.2	212.6	198.3	
HBV cp140, collapsed	28	1QGT	nd	443.5	446.2	nd	nd	115.5	
HBV cp140, $T = 3$	180	pseudo-model[f]	2862.8	2851.4	2868.2	809.3	nd	710.6	
HBV cp140, $T = 4$	240	1QGT	3802.5	3801.8	3824.2	1049.6	nd	905.6	
HBV cp142, non-assembly[g]	6	3KXS	96.7	96.2	88.4	50.6	nd	54.0	
HBV cp142, assembled[g], triangular	6	1QGT	96.5	96.2	95.6	59.9	nd	53.4	
HBV cp142, collapsed	6	1QGT	nd	96.2	95.6	nd	nd	47.6	
NV VP1	2	1IHM	112.1	113.2	99.2	58.1	59.8	52.6	

Protein	*N*	PDB-ID	*m* /kDa			*Ω* /nm²			Schematic structure
			exp	theo	PDB	ramped	fixed	mobcal	
NV VP1	4	1IHM	224.2	226.4	198.4	99.6	98.5	89.0	
NV VP1	6	1IHM	336.4	339.5	297.6	13.4	132.8	123.8	
NV VP1	8	1IHM	448.5	452.7	396.7	173.3	nd	155.1	
NV VP1	10	1IHM	561.5	565.9	495.9	207.4	nd	177.8	
NV VP1	12	1IHM	672.5	679.1	595.1	239.5	nd	197.8	
NV VP1	14	1IHM	783.8	792.2	694.3	272.2	nd	217.9	
NV VP1	16	1IHM	896.4	905.4	793.5	306.3	nd	237.7	
NV VP1	18	1IHM	1008.4	1018.6	892.7	340.0	nd	258.1	
NV VP1	20	1IHM	1120.1	1131.8	991.9	365.8	nd	277.6	
NV VP1	40	1IHM	2242.3	2263.6	1983.7	696.6	nd	496.2	
NV VP1, *T* = 1	60		3373.8	3395.3	3327.8	921.2	nd	nd	
NV VP1	80		4482.0	4527.1	4437.0	1172.6	nd	nd	
NV VP1, *T* = 3	180	1IHM	10099.7	10186.0	9983.3	2063.4	nd	1232.9	

[a] structure of human placental alkaline phosphatase
[b] structure of archeal exosome core and exosome core-Csl4 from *Sulfolobus solfataricus*
[c] structure of human ubiquitin (sequence identical)
[d] carrying the following mutations relative to the GroEL used: R452A, E461A, S463A, V464A
[e] carrying an A433E and A397D relative to 1KPO
[f] pseudo model for a *T* = 3 capsid based on the cp149 structure 1QGT
[g] cp142 hexamer under conditions that promote (200 mM ammonium acetate, pH 6.8) or abolish assembly (10 mM ammonium acetate, pH 6.8 or 9.6)

Table S7: Collision cross sections (*Ω*) of HBV and norovirus small intermediates assuming different symmetrical arrangements.
The experimental *Ω* values of the viral oligomers with 6 to 12 subunits (*N*) are compared to calculated *Ω* results of different subunit arrangements observed in the crystal structures (chapter V Fig. 4). The structures were based on initial assembly of dimers into a 5-fold vertex, a quasi-6-fold or a triangular structure. For both norovirus and HBV the 5-fold symmetry arrangement showed the best agreement with the experimental data. This indicates that capsids likely assemble *via* the addition of dimers starting from a nucleus with a 5-fold symmetry axis, *i.e.* a pentameric vertex.

Mesleh, M. F., Hunter, J. M., Shvartsburg, A. A., Schatz, G. C. & Jarrold, M. F. (1996). Structural information from ion mobility measurements: Effects of the long-range potential. *J Phys Chem* **100**, 16082-16086.
Shvartsburg, A. A. & Jarrold, M. F. (1996). An exact hard-spheres scattering model for the mobilities of polyatomic ions. *Chem Phys Lett* **261**, 86-91.

Virus	*N*	*Ω* /nm²				
		ramped	fixed	5-fold	quasi-6-fold	triangular
HBV cp140	6	60.9	62.9	57.3	57.7	53.4
HBV cp140	8	71.9	75.1	73.1	73.7	70.2
HBV cp140	10	82.9	87.1	85.8	89.4	86.1
HBV cp140	12	94.2	100.0	97.8	98.9	101.0
NV VP1	6	137.4	132.8	123.8	161.5	109.8
NV VP1	8	173.3	nd	155.1	151.3	141.5
NV VP1	10	207.4	nd	177.8	178.7	163.1
NV VP1	12	239.5	nd	197.8	198.7	194.7

Appendix II

About author

Curriculum vitae

The author was born in Ostercappeln, Germany, on 28 April 1982. She went to Söderblom Gymnasium in Espelkamp, Germany, after elementary school in 1992. In 2001, she obtained the general qualification for university entrance (Abitur) with majors in physics and biology. Afterwards, she studied biochemistry at the University of Potsdam, Germany. A short internship at the biochemistry of micronutrients group, DIFE, Potsdam-Rehbrücke, Germany, was supervised by Dr. Gaby-Fleur Böl and Prof. Dr. Regina Brigelius-Flohé and involved the study of signalling pathways in mammalian cell cultures. Under supervision of Prof. Dr. Robert Seckler and Dr. Stefanie Barbirz, the major research project leading to the Diplom thesis, "Interaction of HK620 tailspike protein with fragments of lipopolysaccharide", was performed at the physical biochemistry group, University of Potsdam, Germany. The graduation with Diplom biochemist (equivalent to Master of Science) was obtained in October 2006. Shortly thereafter, in November 2006, the author started a PhD project supervised by Prof. Dr. Albert J.R. Heck at the biomolecular mass spectrometry and proteomics group, Utrecht University, the Netherlands. The results of which are presented in this thesis.

Publication list

Barbirz, S., Muller, J. J., Uetrecht, C., Clark, A. J., Heinemann, U. & Seckler, R. (2008). Crystal structure of Escherichia coli phage HK620 tailspike: podoviral tailspike endoglycosidase modules are evolutionarily related. *Mol. Microbiol.* **69**, 303-316.

Fu, C., Uetrecht, C., Kang, S., Morais, M., Heck, A. J., Walter, M. R. & Prevelige, P. E. (2010). A Docking Model Based on Mass Spectrometric and Biochemical Data Describes Phage Packaging Motor Incorporation. *Mol. Cell. Proteomics* **9**, 1764-1773.

Karagöz, G. E., Duarte, A. M. S., Ippel, J., Uetrecht, C., Sinnige, T., van Rosmalen, M., Hausmann, J., Heck, A. J. R., Boelens, R. & Rüdiger, S. G. D. (2010). The N-terminal domain of human Hsp90 triggers binding to the co-chaperone p23. **submitted**.

Lorenzen, K., Olia, A. S., Uetrecht, C., Cingolani, G. & Heck, A. J. (2008). Determination of stoichiometry and conformational changes in the first step of the P22 tail assembly. *J. Mol. Biol.* **379**, 385-396.

Roos, W. H., Gibbons, M. M., Arkhipov, A., Uetrecht, C., Watts, N. R., Wingfield, P. T., Steven, A. C., Heck, A. J., Schulten, K., Klug, W. S. & Wuite, G. J. (2010). Squeezing protein shells: how continuum elastic models, molecular dynamics simulations and experiments coalesce at the nanoscale. *Biophys. J.* **99**, 1175-1181.

Shoemaker, G. K., van Duijn, E., Crawford, S. E., Uetrecht, C., Salumbides-Baclayon, M., Roos, W. H., Wuite, G. J., Estes, M. K., Prasad, B. V. & Heck, A. J. (2010). Norwalk virus assembly and stability monitored by mass spectrometry. *Mol. Cell. Proteomics* **9**, 1742-1751.

Uetrecht, C., Versluis, C., Watts, N. R., Roos, W. H., Wuite, G. J., Wingfield, P. T., Steven, A. C. & Heck, A. J. (2008). High-resolution mass spectrometry of viral assemblies: molecular composition and stability of dimorphic hepatitis B virus capsids. *Proc. Natl. Acad. Sci. USA* **105**, 9216-9220.

Uetrecht, C., Versluis, C., Watts, N. R., Wingfield, P. T., Steven, A. C. & Heck, A. J. (2008). Stability and shape of hepatitis B virus capsids *in vacuo*. *Angew. Chem. Int. Ed. Engl.* **47**, 6247-6251.

Uetrecht, C., Barbu, I. M., Shoemaker, G. K., van Duijn, E. & Heck, A. J. (2010). Interrogating viral capsid assembly with ion mobility mass spectrometry.**submitted**.

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

Uetrecht, C., Watts, N. R., Stahl, S. J., Wingfield, P. T., Steven, A. C. & Heck, A. J. (2010). Subunit Exchange Rates in Hepatitis B Virus Capsids are Geometry- and Temperature-Dependent. *Phys. Chem. Chem. Phys.* **12**, 13368 - 13371.

Tseng, Y.-H., Uetrecht, C., Heck, A. J. R. & Peng, W.-P. (2010). Interpreting the Charge State Assignment in Electrospray Mass Spectra of Bioparticles. **submitted**.

Acknowledgements

First of all, I would like to thank my supervisor, Albert, for constant support and a fruitful, but more importantly, fascinating project. I highly appreciated the ability to visit several conferences abroad throughout the last four years. This was always inspiring and I got to know many researchers from the fields of virology and mass spectrometry. The FASEB in 2008 was a special adventure, as the only person from our group I went to Vermont after a short visit to the NIH, not having met anyone from face-to-face before besides our collaborator, Wouter, and entitled to give a short talk at the conference. In the end, everything worked out perfectly well.

Very high influence on the success of my project had our collaborators, since I was invariably dependent on the samples they supplied. Special thanks go to Alasdair, Norman and Paul for the gorgeous HBV particles, which gave rise to four chapters in this thesis, and fruitful discussions. Many thanks to Gijs and Wouter for having my at the VU in Amsterdam regularly - I owe you a lot of tea. Maybe I still have time to pay it back with cookies. Furthermore, thanks to Mary and Venkataram, Peter and Chi-Yu, Gino and Adam, Diego and Jon, Nenad and Markus, Wen-Ping, Stefan and Elif, Jeroen, Melanie and Rik for interesting samples/projects, some of which already have led to publications and others hopefully will in the near future.

An invaluable contribution to this thesis came from Jan Commandeur, who kept fixing and pimping the QToF as well as the other instruments of the native MS group. Thanks!

This was the easy part. A while ago, I started to make a list of all the people that should be mentioned. Soon I had to realise that the turnover in this group in conjunction with its size, makes complete listing a dangerous and almost impossible task. Therefore, I would like to thank the whole group including the past members for a pleasant time here in Utrecht. Especially, I want to thank the native group: Arjan, Esther, Glen, Ioana, Jessica, Kees, Kristina, Rebecca, Sara, Silvia and Vera. Thanks to my bachelor student Renske, who did the initial work on the Triatoma virus. Thanks to Maarten, Marco, Shabaz and Soenita for assistance with proteomics work and Renske v.G. for writing/polishing the Dutch summary in this thesis. To my past and present office mates, Danny, Gideon, Hongtao, Kristina and Paul: You cannot imagine, how much I appreciated your endurance/tolerance when listening to “music”. Also thank you for the chats. Corine, thanks for your inexhaustible assistance. Serena, I am sure you’ll represent the PhDs well in the staff meeting, thanks for taking over.

For fun, distraction, leisure activities etc. but also scientific input, I would like to thank all the other people in or related to the group. No responsibility is taken for the correctness of this information: Andreas, Arjen, Bas, Christian Jr/Sr, Hanka, Henk, Javier, Jeffrey, Martina, Martje, Mirjam, Nadia, Onno, Pieter, Reinout, Sarah, Sharon, Thin Thin, Thomas, Tieneke, Tom, Wilma…

There are more colleagues even on the phone list in my office, but I leave it here. You are welcome to fill in your name; I left some space for that purpose.

@ Nadia and Thin Thin: Soon I am going to join the Scandinavian connection! ;)

A lot of persons outside Utrecht University had an impact on the development of this thesis as well. My former supervisors Robert, who came up with the idea to go here for my PhD, and Steffi got me interested in viruses. The fellows from Potsdam: Insbesondere dem Stammtisch, Anne, Janine, Julia, Katrin, Ramona, Saskia; sowie Wolle und Kerstin, die mir gen Niederlande

folgten, gilt mein Dank. Schade, dass die Utrecht Kartenspieltruppe nicht mehr beisammen ist. Philipp and Luba also besides the card games, you were great housemates. Peter same holds for you, all the best with your own thesis and the future with Ema in Vienna.

Unendlich dankbar bin ich auch meinen Freunden in und ausserhalb der Metalszene fuer die noetige Ablenkung, aber auch reges Interesse an meiner Arbeit. Vielen Dank für euer Vertrauen. Ich sehe von einer Auflistung hier ab, ihr wisst, dass ihr gemeint seid. Freue mich auf eure Besuche in Schweden, nicht nur um dort die Festivals zu “rocken”. Beinahe hätte ich Robert vergessen. Vielen, vielen Dank für ungezählte Reparaturen, Wartungen und Ferndiagnosen an meinem Volvo, ohne deine Hilfe wäre das Wochenendpendeln unmöglich.

Danken möchte ich auch meiner Familie, insbesondere meinen Eltern, Inge und Gerd. Die Atmosphäre zu Hause mit den intensiven Diskussionen zu praktisch jedem Themengebiet und bei jeder gemeinsamen Mahlzeit (und derer gab es viele) waren der richtige Nährboden für meine Entwicklung. Dank auch an meine Geschwister Karo und Ferdi. Meistens mag ich euch ;) – sogar sehr. Nicht zu letzt möchte ich Jan danken, unter anderem weil es niemals einfach ist, eine Doktorandin im Abgabestress als Freundin zu haben…