From Structure to Function:
Protein Assemblies dissected by Mass Spectrometry

ISBN: 978-90-393-49304

The research described in this thesis was performed in the Biomolecular Mass Spectrometry and Proteomics Group (Bijvoet Center for Biomolecular Research and Utrecht Institute for Chemistry, Utrecht University).

From Structure to Function:
Protein Assemblies dissected by Mass Spectrometry

Intacte eiwitcomplexen bestudeerd
met behulp van native massa spectrometrie

(met en 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 colleges voor promoties in het openbaar te verdedigen op woensdag 10 december 2008 des middags te 12.45uur

door

Kristina Lorenzen

geboren op 31 januari 1978, te Braunschweig, Duitsland

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

This work was financially supported by the Netherlands Proteomics Centre
The printing of the thesis was financially supported by Waters Europe B.V. and Albert J.R. Heck

Introduction

Research

Summary and outlook

From Structure to Function

Protein Assemblies dissected by Mass Spectrometry

FUNDAMENTALS OF NATIVE MASS SPECTROMETRY

Kristina Lorenzen
Biomolecular Mass Spectrometry and Proteomics Group, Bijvoet Center for Biomolecular Research and Utrecht Faculty of Science, Utrecht University, Sorbonnelaan 16, 3584 CA Utrecht, the Netherlands

CONTEXT
In the last couple of years interaction networks of cells have been mapped by different high-throughput techniques. While this data gives an idea of the proteins that are connected to each other, it lacks information on exact protein complex composition, structure and dynamics. New developments in mass spectrometry have enabled the analysis of intact protein (sub)-complexes, giving rise to a method that might be bridging the gap between high resolution structures (generated by methods such as NMR, X-ray crystallography, electron microscopy) and in these interactome networks. Following is an introduction into the fundamentals of native mass spectrometry.

General introduction into mass spectrometry

Since the discovery of cathode rays by J. Thomson in 1897 (Thomson, 1913), mass spectrometry has come a long way in its development and applications. A few years after Thomson's discovery, Aston built the first prototype of what later became the basis of magnetic/sector field mass spectrometers. After the new technique helped to elucidate and determine isotopes of the elements in the 20th century (Aston, 1919), the next major instrumental development took until the middle of the last century. In the late 1940s and early fifties the basics were laid for the instrumentation used today.
The general schematic setup of a mass spectrometer is displayed in Figure 1. It consists out of a source where the ions are generated from the analyte substance. Different ways of ionization are: Bombardment of the sample with electrons (EI), Atoms or Ions (FAB) or Photons (LDI). The ions can also be generated by spraying the sample from an aqueous solution in an electric field (ESI) (Fenn et al., 1989; Hoffmann and Stroobant, 2002). The ions are then transferred into the mass analyzer where they can be focused and/ or separated. Some possibilities of mass analysers and detectors are: magnetic fields and electric fields (sector instruments), high frequency fields of quadrupoles (quadrupole instruments), electrical ion traps (ion trap) or magnetic ion traps (ioncyclotron resonance (ICR) (Hoffmann and Stroobant, 2002). The last step is the detection of the ions that come out of the mass analyzer. There are basically four different kinds of detectors, it has to be mentioned that the detection of ions in an ICR instrument is characteristically different from the ones mentioned here. The most commonly used detector is

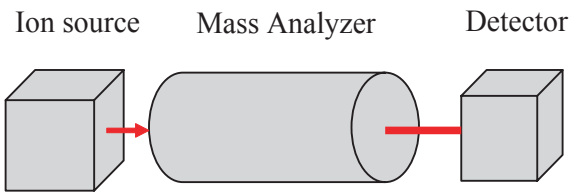

Figure 1 Schematic of a mass spectrometer. The line indicates the trajectory of the ions through the instrument

the electron multiplier. It consists out of about ten to twenty electrodes (dynodes). When an incoming ion hit the first electrode it caused the release of multiple secondary ions. These ions then hit the next dynode and each one causes again the release of multiple tertiary ions. This goes on from dynode to dynode. The increase in current is in the order of 10^4 to 10^8. Other methods of detection where ions also are directed in a sequential manner onto the detector are using photomultipliers or a Farraday cup. A different method is the detection via various focal plane detectors. Here the ions are dispersed simultaneously onto a plane and detected by an array device.
Newer developments in the field of mass spectrometry enabled the expansion beyond the widely used analysis on small molecules like proteomics and drug analysis basis opening up avenues into the structural analysis of proteins and protein complexes as large as about 6 MDa (Loo et al., 2005; Uetrecht et al., 2008; Videler et al., 2005). The advantages over other structural techniques like NMR, X-ray crystallography and cryo electron microscopy (EM) are mainly the low sample consumption, the high accuracy and resolution in mass combined with a fast analysis over a high variability of masses. This success has its origin in mainly two developments. First the invention of nano electrospray ionization (ESI) and second in the further development of the instruments enabling them to measure up to an theoretically unlimited mass range. For completeness it should be noted here that matrix assisted laser desorption (MALDI) is also suitable to measure proteins with high masses. It finds growing interest especially in the direct analysis of tissues (Maxence Wisztorski, 2008) in the so called MALDI imaging technique. However it has major drawbacks in the analysis of complexes and protein flexibility, since it requires the preservation of the sample of interest in a matrix. For reviews see (Andersen et al., 1996; Chaurand et al., 1999; Kenny et al., 2006).

Generation of ions from solutions by ESI

Today the most commonly used ionization method in pro-

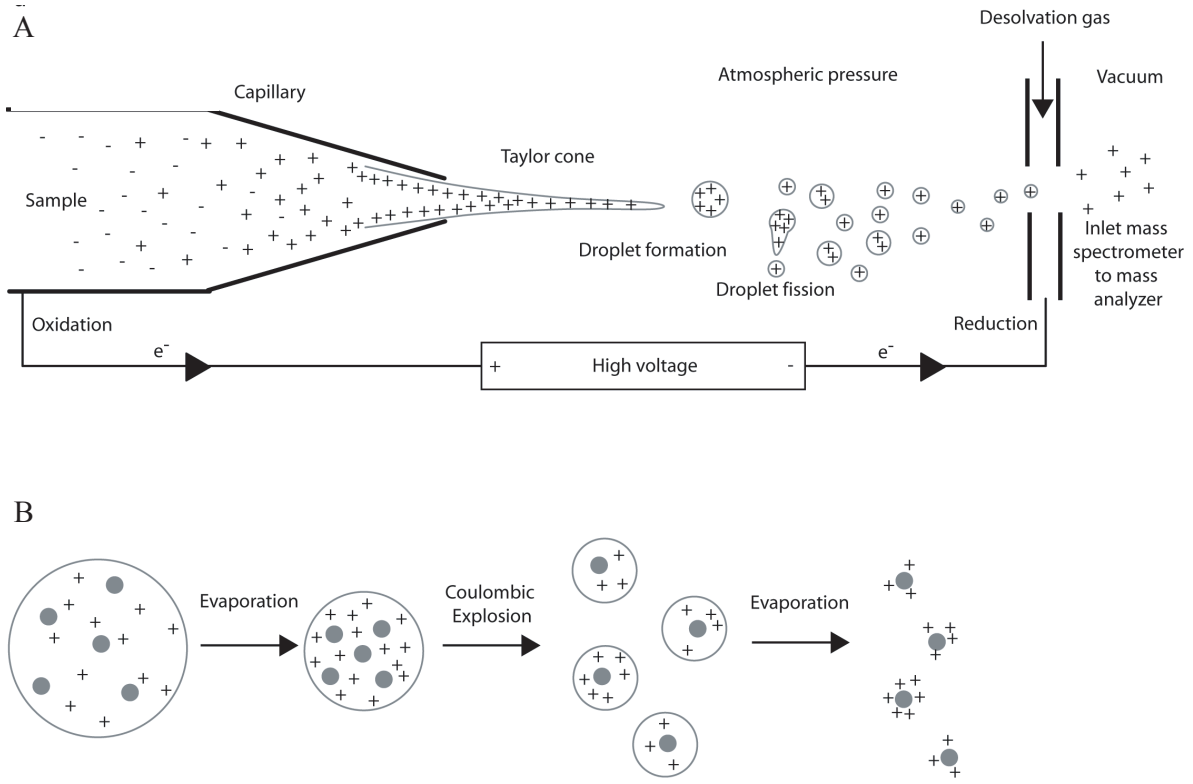

Figure 2 A Schematic of the nano-flow ESI process in the positive ion mode. An electric potential is applied on the capillary containing the analyte that leads to the formation of an elongated meniscus (Taylor cone) arising to highly charged droplets. Evaporation of the solvent results in smaller droplets until they are devoid of solvent molecules. In the ion evaporation model small molecular ions can directly evaporate from the surface of the formed droplet. The IEM model is thought to best describe the formation of small molecules and peptidic ions in the gas phase. B Depicted here is the charge residue model (CRM). During the evaporation process charges in the droplet come closer together. Smaller droplets are then formed by sequential coulombic explosions until only one analyte (protein, grey circle) is present. The CRM is generally accepted to explain the formation of large ions such as folded globular proteins and protein complexes. Adapted from (de la Mora, 2000).

tein mass spectrometry is probably electrospray. Originally used to coat surfaces, Dole and co workers laid the basis of the method how it is applied today in the 1960s (Dole et al., 1968). Still it took another two decades until John Fenn (Fenn et al., 1989; Yamashita and Fenn, 1984) modified the technique and opened up new applications for mass spectrometry. Examples hereof are the analysis of large protein complexes and assemblies (Heck and Van Den Heuvel, 2004; Loo, 1997; Robinson, 2002; Sobott and Robinson, 2002), but also mass spectrometry based proteomics (Chowdhury et al., 1990; Mann et al., 2001). The ability of ESI to ionize macromolecules and thereby maintaining covalent interactions made it possible t study stoichiometry, dynamics and structure of proteins and their interactions with binding partners.

In ESI the substance or protein of interest is sprayed from an aqueous solution though a metal or gold coated glass capillary. For measurements on intact protein (complexes) this generally is a volatile buffer like ammonium acetate, ammonium bicarbonate or tri ethyl ammonium bicarbonate (TEAB) at a near neutral pH. An electric potential (usually between 1 kV and 3 kV) is applied to the capillary (in the case of glass capillaries referred to as needle). On the tip of the capillary a Taylor cone is formed, from which small droplets emerge. These droplets contain an excess of charge which are evenly spread out over the surface, while the interior of the drop is neutral (Fenn et al., 1997). There are mainly two hypothesis about how the multiply charged ions are formed. One was suggested by Iribane and Thomson (Iribarne and Thomson, 1976). It is called ion evaporation mode (IEM) and suggests that the analyte ions can be ejected from the surface of an intact highly charged droplet. While this seems to apply to small and low charged ions, the experimental data for larger ions show that proteins measured by ESI have a higher charge state than would be expected by this model (Iavarone and Williams, 2003). The other hypothesis, which is in agreement with most experimental and computational studies for proteins (Heck and Van Den Heuvel, 2004; Konermann, 2007; Robinson et al., 1996; Tesic et al., 2007) is the charged residue model (CRM) (Dole et al., 1968). Here it is assumed that the solvent keeps evaporating until columbic repulsion happens at the Rayleigh limit:

$$q=\sqrt{(8\pi^2 \varepsilon_0)}\ \gamma r^3$$

Where ε_0 is the permittivity of the vacuum, γ is the surface tension of the liquid and r is the radius of the droplet. The droplets become instable and smaller droplets are formed by jet fission (Duft et al., 2003; Gomez and Tang, 1994). The CRM suggests that proteins can retain significant aspects of their solution structure in the gas phase making it a fast and sensitive method to probe protein conformational changes (Grandori, 2002; Konermann et al., 1997; Pan et al., 1997; Robinson et al., 1994; Robinson et al., 1998; Yan et al., 2004) (see also chapter 3) (Figure 2). During the ESI process, when exiting the needle and entering the first vacuum stages in the mass spectrometer, the ions experience an acceleration of several hundred meters per second, simply due to the pressure difference between the atmosphere in the laboratory and the vacuum in the instrument. This means that larger ions (like proteins or protein complexes) gain energies in the electron volt range, since the energy gained is proportional to the mass at the same velocity (Chernushevich and Thomson, 2004). These high energy ions have extended oscillation movement in axial (leads to lower resolutions and ions missing the detector) and radial (leads to lower transmission since the ions will collide with the orifices or rods in the mass spectrometer) direction, thus they might get lost in the source without ever reaching the detector. Douglas and French (Douglas and French, 1992) proposed the use of quadrupole ion guides in connection with gas molecule collisions to prevent the large ions to get too much kinetic energy. This is referred to as collisional cooling. It reduces the radial and axial movement of the ions leading to better transmission throught the mass spectrometer. Especially when handling protein complexes one needs to be careful that the kinetic energy gained during the transition into the mass spectrometer does not exceed the binding energy of the protein protein or protein ligand bond. Otherwise the interaction will already dissociate during the ESI process and the information about their interaction lost. There are different ways to cool the ions in the first steps in the mass spectrometer. The one typically used today and also applied in this work is to increase of pressure in the source region (the first vacuum stage of the instrument) by throttling the valve of the turbo pump (Tahallah et al., 2001) and insertion of a sleeve around the first hexapole (van den Heuvel et al., 2006).

When a protein gets ionized by ESI it results in a series of consecutive ion signals at different m/z forming a Gaussian distribution, the charge envelope. In positive ESI the charges of the ions will be caused by different amount of protons at the surface of the protein, where each proton will add one charge to the analyte of interest. To calculate the mass from an m/z, one has simply to multiply the m/z value by the charge state and deduct the mass of protons (H)

$$m=(m/z)*z - nH$$

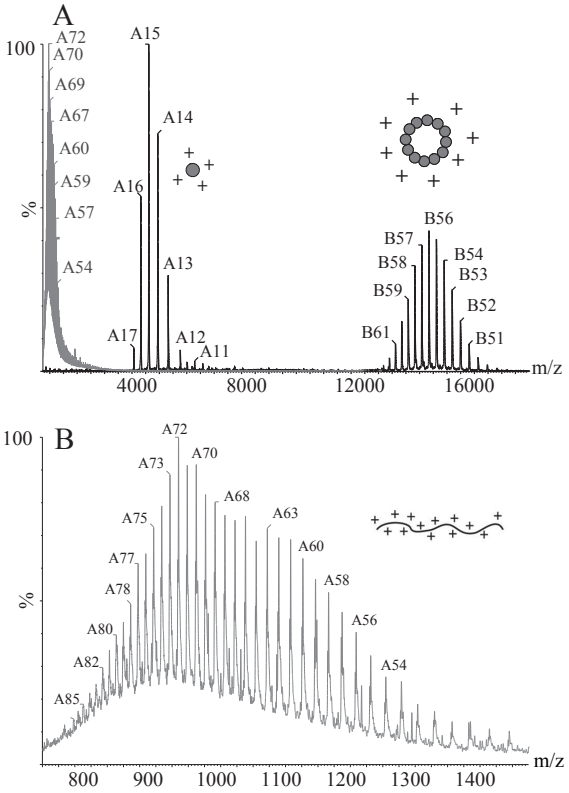

Figure 3 Illustration of different conformations and stoichiometries of a protein (complex) detected by ESI mass spectrometry. (A) shows a native mass spectrum of gp1 (black). In the high mass range the dodecamer can be seen ranging from 13,000 to 16,000 m/z (with an average charge state of 56^+ and labelled with B61 to 51), while folded monomers populate the lower mass range around 5000 m/z (average of 15 charges and labelled with A17 to A11). In dark grey a mass spectrum of gp1 is inserted but this time the protein was denatured with anorganic solvent and acid. (B) shows the zoom in on the unfolded monomer revealing the much higher charge states (around 72^+ and labelled with A85 to A54) pinpointing the open conformation.

Where m is the mass, z the charge, n the number of charges. However the charge state is initially unknown. The distance between two successive peaks can be used to pinpoint the charge:

$$n=((m/z)_{(n+1)}-H)/((m/z)_n - (m/z)_{(n+1)})$$

Figure 3 shows two examples of ESI mass spectra of the same protein complex and illustrates how conformational changes can be tracked by ESI.

The trajectory of the ions through a mass spectrometer

After the ions are formed they are guided through a skimmer further into the mass spectrometer. Here they are first focused before they are guided either directly to a detector (in the case of an LCT) or first through a quadrupole,

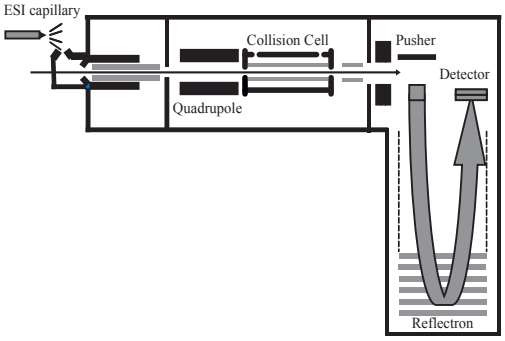

Figure 4 Schematic representation of a Q ToF mass spectrometer. The trajectory of the ions is indicated as a black line until the pusher and a thick grey line in the ToF region.

through a collision cell and then after another focussing step to the detector (for example a Q ToF, Figure 4). Quadrupole analyzers consist out of four parallel rods, where always the opposing ones are electrically connected. The rods are separated by the radius $2r_o$ and a DC voltage (U) is applied with a superimposed radio frequency (RF) potential ($V\cos(\Omega t)$) (Figure 5). When, for example when a positively charged ion enters the quadrupole region it gets attracted to the negatively charged rod. Before the ion can reach the rod and discharge the potential changes and the ion gets repelled again. Due to the oscillating frequency of the rods the ion will enter a stable x/y trajectory. The movement of the ion in the z direction is caused by a potential offset between the source or entrance of the quadrupole and the analyzer. A stable trajectory of the ion is dependent on the m/z and U (Figure 6). If U is zero there is no resolution, but all ions retain a stable trajectory as long as V is within the limits of the area where they are stable. If the quadropole analyzer is coupled with a ToF detector this loss of resolution in the quadrupole region is negligible because the mass resolution is gained in the ToF. If U = 0 then the instrument measures in the so called RF only mode and all ions above a certain threshold get additionally focused there. The focussing of the ion is important to keep them on a stable trajectory after the collision with gas molecules. This effect of Quadrupole analyzers is independent on the m/z and largely independent of the internal kinetic energy of the ion. The m/z range a quadrupole can handle is depending on the distance between the rods and the RF frequency as well as amplitude. Typical quadrupole analyzers can measure ions up to 4000 m/z. In 2002 Sobott et al. described a Q ToF setup that allows to measure ions of much higher molecular weight by changing the RF frequency of the analyzer from 832 kHz to 300 kHz (Sobott et al., 2002). For further improvements of the instrument see (van den Heuvel et al., 2006) and chapter 2 (Lorenzen et al., 2007).

Time of Flight

The concept of Time of Flight (ToF) analyzers was first published by Stephens in 1946 (Stephens, 1946), the first prototype was published in 1948 (Cameron and Eggers, 1948) and a couple of years later Wiley and McLaren published the design of the first commercial linear ToF (Wiley and McLaren, 1955). The principle of ToF is comparable to an ion race, where all ions get the same push at the start and have to reach a finish line in a defined distance. The time they need to finish will be measured. More scientifically spoken: In the instrument the ions, which have been focused or bundled, are accelerated (pushed) by a potential V_S and have to travel a specific distance d before they hit the detector. Thus ions with the same energy but different m/z will have different velocities and will take different time to pass through the defined distance (d). In this way the m/z ratios are determined by the time the ions need to move through the field free region from the pusher to the detector.

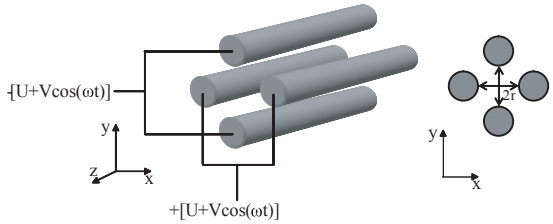

Figure 5 Schematic representation of a quadrupole mass analyzer. Four parallel rods in a radial orientation with opposite rods carrying the voltage. An oscillating electric field drives the ions in a spiral motion through the quadrupole. Only ions with a stable trajectory will reach the detector, others collide with the rods and are lost.

$$t^2=(m/z)/(d^2/(2V_S\ e))$$

Since $(d^2/(2V_S\ e))$ is given by the instrument hardware and is constant, the m/z is dependent on the flight time. Small ions will reach the detector faster than larger ions. Today most ToF´s are set up in the orthogonal mode, while the first ToF instruments were build in one line with the source, the orthogonal ToF is set up in a 90° angle to the incoming ions. The ions get trapped at the first stage of the ToF before they get a defined pulse and are thus pushed in the direction of the detector (Giles et al., 2004). This has the advantage that it uncouples the source region from the ToF and makes it applicable to use continuous ionisation methods. This makes the method very suitable for ESI applications and especially for measuring protein (complexes) since it practically has no upper mass limit, is dynamic and fast. The ToF is rather low in resolution but has high transmission efficiency. The poor resolution has been a major problem especially in the beginning of ToF mass spectrometry. A common definition for the resolution is the m/z divided by the peak width at half height. Basically it gives an indication which masses the instrument is able to separate from each other. The main problem causing the low resolution in the first ToFs was the different initial kinetic energy of the ions. A setup of the ToF orthogonal to the source reduces the problem. An additional solution to this was to decelerate the ions and make the ones with a

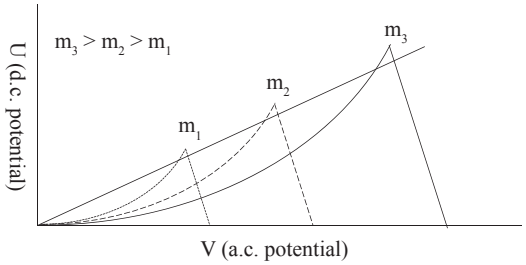

Figure 6 Stability areas as a function of U and V for ions with different masses (m). Changing U linearly as a function of V a straight line can be obtained that shows the setting to allow observing ions successively. The steeper the line, the better the resolution, as long as the line passes through the stability area of the ion. If U is put to zero, all ions will have a stable trajectory with zero resolution (Hoffmann and Stroobant, 2002).

higher energy fly a longer path. It has been achieved by inserting an "ion mirror" (reflectron) into the ToF (Mamyrin et al., 1973), at about half distance. Mamyrin said that he got the idea of playing kids whom were competing in throwing balls into the sky as high as possible (Mamyrin, 2001). The principle can also be compared with balls rolling up a hill, the fastest one will roll up higher, covering the longer distance, but if the other properties of the balls are identical this will lead to them appearing at the starting point again at the same time (Figure 7): The ions with higher energies will penetrate the reflectron deeper before they get repelled, symmetrically restoring their initial energy when leaving the reflectron again. Consequently this is giving them a longer flight path, compared to the ones with lower energies. A reflectron will generally improve the resolution by about one to two orders of magnitude.

Collision induced dissociation

Often the information of purely the mass of a substance is not sufficient for the interpretation of the analysis. Additional structural information can be obtained by tandem mass spectrometry, where specific ions at a narrow m/z window are selected and activated. The analysis of protein complexes in the gas phase requires special instrumentation (as described above) which limits the possibilities of activation devices. The most commonly used method is collision induced dissociation (CID), but there have been publications on blackbody infrared radiative dissociation (BIRD) (Felitsyn et al., 2001), electron capture dissociation (ECD) (Geels et al., 2006) and surface induced dissociation (Jones et al., 2006).

In CID the analyte (ion) of interest enters a gas filled collision cell. During its way through the cell the ion will then undergo collisions with the gas molecules in the cell. The process can generally be described in two steps. First during the collision with the neutral gas molecule a fraction of the translational energy will be converted into internal kinetic energy of the ion. This first process is very fast (10^{-14}-10^{-16} s). Many collisions will eventually lead to the decomposition of the ion. This second process is slow compared to the first (Hoffmann and Stroobant, 2002). The number of collisions (n) an ion will undergo in the CID process depends on the collision cross section (Ω) (size) of the ion, the radius of the collision gas molecules (R_g), the gas pressure (p), the length of the collision cell (l) and the temperature (T).

$$n = c(\Omega lp/T)$$

where c is a specific, instrument dependent, constant (Benesch et al., 2007). Often the collision cross section of the ion is unknown and will be estimated assuming it to be a sphere. It has been shown that there is a correlation between the mass of an ion and its charge state. The amount of charges depends mainly on the accessible surface area of the ion and is less much influenced by charged amino acids on the surface (Kaltashov and Mohimen, 2005; Konermann, 2007). It has been reported that the average amount of charges a globular protein or protein complex obtains during the electrospray process under native conditions can be estimated by the formula:

$$z = 0.078\sqrt{m}$$ (de la Mora, 2000).

The reports on the average density of this assumed sphere for proteins vary (Benesch et al., 2007; de la Mora, 2000; Hautreux et al., 2004; Kaufman et al., 1998). Based on this correlation the collision cross section of a protein can be estimated by the following formula:

$$\Omega = \pi(2\ (3/4\ m_p/\pi\rho + R_g))$$

where ρ is the assumed density for the protein and m_p the mass of the protein. The amount of collisions a molecule encounters increases with its size ranging from hundreds of collisions for small proteins like Myoglobin or Cytochrome c to tens of thousands of collisions for large protein complexes like GroEL (Benesch and Robinson, 2006). The time an ion spends in the collision cell is, apart from its mass, dependent on the acceleration voltage Vs and the gas

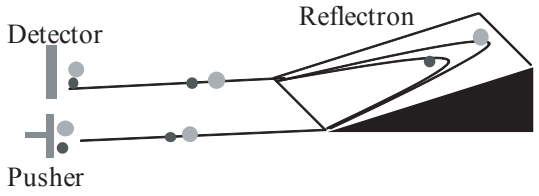

Figure 7 Schematic representation of the reflectron principle. The grey circle shows an ion with a given mass and the correct kinetic energy. The black circle indicates an ion with the same mass but with a kinetic energy that is too low. The black ion will reach the reflectron later but will not penetrate it as deep as the grey ion resulting in both ions reaching the detector simultaneously.

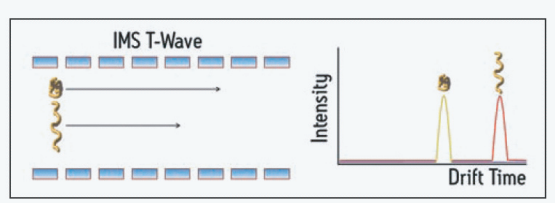

Figure 8 Principle of the ion mobility cell (T Wave, Waters, UK). Assuming the ions have the same mass and charge state the ion with the compact form (light gray line) will traverse the cell faster than the ion with the elongated shape (dark grey line) and hence have a shorter drift time.

pressure in the cell. Higher gas pressures and heavier collision gases favour increased energy conversion. As already mentioned above each encounter with a gas molecule will cause the internal energy of the ion to increase. The energy increase after n collisions can be estimated by:

$$\Delta E_{int} = zV_S\,(1\text{-}2(m_p^2+m_g^2)/(m_p+m_g))^n$$

(Douglas, 1982; Gill et al., 2000). This principle of collision activation and dissociation is of course also applicable without the selection of a specific m/z value in the quadrupole. For large protein complexes the desolvation in the source region is often not sufficient and buffer molecules and water are still attached to the protein (complex) resulting in peak broadening and mass increase (McKay et al., 2006). A low amount of collisional activation will lead to the dissociation of these "bulk" molecules, since the activation energy to dissociate them is generally lower than the energy needed to disrupt the non covalent protein (complex) interactions. As a result the peaks get sharper, the resolution increases and the mass accuracy improves.

When the internal energy exceeds a certain point, then the protein will get decomposed. Generally the non covalent interactions break easier than the covalent bonds, mostly resulting eventually in the ejection of a monomer from a protein complex.

$$C^{(z+)} \rightarrow M^{(x+)}\text{-}[C\text{-}M]^{(x\text{-}z)+}$$

.

Where C is the protein complex, M is a monomeric subunit, z is the initial selected charge of the complex and x is the charge the monomer takes along when leaving the complex. It should be noted that no charges can be gained or lost during this process. There is still uncertainty how exactly the mechanism works. The splitting of the charge distribution between the complex and the dissociating subunits was found to be generally asymmetric (Jurchen et al., 2004; Jurchen and Williams, 2003; Versluis et al., 2001). Generally the process seems to occur as following. During the ESI process an ionized complex is formed, it is assumed that the charges are spread out evenly over its surface. This complex is then accelerated into the collision cell where it collides with the neutral gas molecules taking up more and more energy resulting in local unfolding of parts of the complex. The unfolding on the surface of the complex results in an increase of idem causing the charges to rearrange onto the newly exposed surface area. This leads to further destabili-

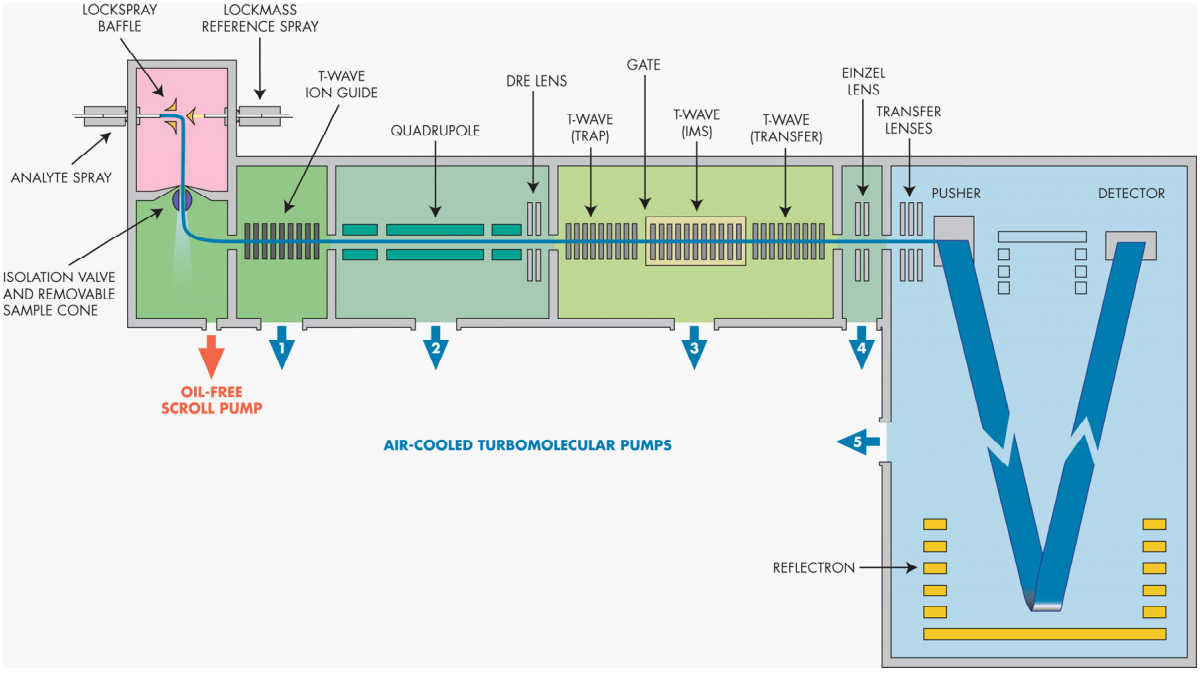

Figure 9 Instrument design of a commercial ion mobility Q ToF mass spectrometer (Waters, UK). The setup is basically similar to the Q ToF in Figure 3, the hexapole guides have been changed to travelling waveTM devices and next to the collision cell the instrument contains an ion mobility cell. The connection points with the turbo pumps which allow pressure regulation independently in each part of the instrument are indicated with grey arrows.

sation of the bonds in the unfolded areas, further charge rearrangement, consequently weakening the bonds even more. The result is eventually the ejection of an unfolded monomer with a high amount of charges correlated to its surface area (Benesch and Robinson, 2006; de la Mora, 2000; Felitsyn et al., 2001; Jurchen et al., 2004; Jurchen and Williams, 2003). This process can happen sequentially several times.
The mechanism has been compared to the coulombic repulsion of fission droplets in the ESI process (see above) (Schwartz et al., 1995). Robinson and co workers (Benesch et al., 2006) have shown that the amount of charges the ejected monomer takes along is consistent with a structure comprised of a single unfolded subunit. It has been shown that using surface induced dissociation (SID) as an activation method, protein complexes dissociate in a more symmetric manner (Jones et al., 2006). The reason lies likely in the activation technique since in SID the energy translation does not happen gradually but is rather a fast single step process where the internal energy is deposited in a matter of picoseconds compared to 200 - 400 µs in CID (Christen et al., 1998; Meroueh et al., 2002).

Ion Mobility Mass spectrometry

Ion mobility adds an extra dimension to mass spectrometry. It enables the separation of ions based on their mobility through a gas filled chamber. The principle of the technique is often compared to gel filtration in the gas phase (Verbeck et al., 2002), only that the proteins do not move along a liquid flow through a column material, but get accelerated by an electric potential through a linear field in a chamber which is filled with neutral gas molecules (for example He, N_2, or Ar) (Figure 8). The separation primarily depends on the ions volume or cross section (Ω). Larger ions will encounter more collisions with gas molecules which slows them down on their way through the ion mobility cell compared to small ions. The mobility of an ion is additionally dependent on the charge (z), ions with a higher charge state will be influenced more by the electric field and thus be faster than lower charged ions with the same mass and cross section. The mass of the ion (m_p) also has an impact on the velocity of the ion, since

$$v=\sqrt{((m)/2E)}$$

applies. Thus the heavier an ion, the slower it will be while drifting through the mobility cell. V (E) is not constant here like it is in the field free vacuum of the ToF, as there is an gas filled electric field present here. Additionally the time an ion needs to pass through the drift tube is dependent on the laboratory frame and the instrument settings. The mass of the neutral gas (m_g), the number density of the neutral gas (N), the temperature (T), and the pressure in the cell (p) have an influence on the drift time. When acquiring spectra with an ion mobility mass spectrometer the m/z, the drift time and the signal intensity are generated. To calculate the collision cross section of an ion the following formula can be used (Mason and McDaniel, 1988):

$$\Omega = \left(\frac{\sqrt{18\pi}}{16}\right)\frac{z}{\sqrt{Tk_b}}\left[\frac{1}{m_p}+\frac{1}{m_g}\right]\frac{t_d E}{L}\frac{760}{p}\frac{T}{273.2}\frac{1}{N}$$

where E is the electric field strength, k_B the Bolzman constant and t_D the drift time.
Ion mobility has long been used in combination with mass spectrometry for the analysis of small molecules and small proteins in the gas phase (Hoaglund-Hyzer et al., 1999; Jarrold, 2000; Karasek et al., 1974; von Helden et al., 1995), but only recent instrumental changes made it possible to apply the method also to protein complexes (Giles et al., 2004; Pringle et al., 2007; Ruotolo et al., 2005) (Figure 9). Here the combination of a travelling wave ion mobility cell with ToF mass spectrometry and the possibility to fine tune the pressure in each region of the instrument opened the way to measure large protein complexes. The advantages are a high transmission efficiency combined with simultaneous acquisition of mobility and mass of the analyte ions. In this new instrument it is difficult to precisely determine parameters like the exact gas pressure or temperature or purity. Additionally (E/L) varies between ion species and with the instrumental parameters. Therefore it becomes essential to calibrate the instrument. There is still an ongoing discussion about the best way to calibrate for large protein complexes. There are suggestions to use CsI cluster ions or polymers, but at the moment the commonly used method is to calibrate on small proteins where the cross section has been determined already (Ruotolo et al., 2008).
The formular for the cross section gets reduced to:

$$\Omega = K\, z\, t_D^{\,x}\, \sqrt{(1/m_p + 1/m_g)}$$

where K is the calibration constant. The cross section has been found to be proportional to $t_D^{\,x}$, where x is an empirical parameter dependent on instrument variables as the wave height and velocity of the electric field in the drift cell.
Recently several publications appeared, which use ion mobility mass spectrometry to gain information about structural properties of protein complexes and their behaviour in the gas phase (Kaddis et al., 2007; Loo et al., 2005; Lorenzen et al., 2008; Ruotolo et al., 2008; Ruotolo et al., 2005). It has been observed with ion mobility that proteins in the gas have multiple conformations and especially highly charged ions (of a specific complex) have often several conformations, while low charged ions (of a specific complex) retain more compact conformations that resemble more closely the native fold (Jarrold, 2000). Altogether ion mobility for large protein complexes is just getting established, it already

opens up new and exciting possibilities in the field of structural biology.

REFERENCES

Andersen, J.S., Svensson, B., and Roepstorff, P. (1996). Electrospray ionization and matrix assisted laser desorption/ionization mass spectrometry: powerful analytical tools in recombinant protein chemistry. Nat Biotechnol 14, 449-457.

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

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

Benesch, J.L., and 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., and Robinson, C.V. (2007). Protein complexes in the gas phase: technology for structural genomics and proteomics. Chem Rev 107, 3544-3567.

Cameron, A.E., and Eggers, D.F. (1948). An Ion ``Velocitron''. Rev. Sci. Instrum. 19, 605-607.

Chaurand, P., Luetzenkirchen, F., and Spengler, B. (1999). Peptide and protein identification by matrix-assisted laser desorption ionization (MALDI) and MALDI-post-source decay time-of-flight mass spectrometry. Journal Of The American Society For Mass Spectrometry 10, 91-103.

Chernushevich, I.V., and Thomson, B.A. (2004). Collisional cooling of large ions in electrospray mass spectrometry. Anal Chem 76, 1754-1760.

Chowdhury, S.K., Katta, V., and Chait, B.T. (1990). Electrospray ionization mass spectrometric peptide mapping: a rapid, sensitive technique for protein structure analysis. Biochem Biophys Res Commun 167, 686-692.

Christen, W., Even, U., Raz, T., and Levine, R.D. (1998). Collisional energy loss in cluster surface impact: Experimental, model, and simulation studies of some relevant factors. The Journal of Chemical Physics 108, 10262-10273.

de la Mora, J.F. (2000). Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism. Analytica Chimica Acta 406, 93-104.

Dole, M., Mack, L.L., Hines, R.L., Mobley, R.C., Ferguson, L.D., and Alice, M.B. (1968). Molecular Beams of Macroions. The Journal of Chemical Physics 49, 2240-2249.

Douglas, D.J. (1982). Mechanism of the collision-induced dissociation of polyatomic ions studied by triple quadrupole mass spectrometry. J. Phys. Chem. 86, 185-191.

Douglas, D.J., and French, J.B. (1992). Collisional focusing effects in radio frequency quadrupoles. Journal Of The American Society For Mass Spectrometry 3, 398-408.

Duft, D., Achtzehn, T., Muller, R., Huber, B.A., and Leisner, T. (2003). Coulomb fission: Rayleigh jets from levitated microdroplets. Nature 421, 128.

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

Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., and Whitehouse, C.M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. Science 246, 64-71.

Fenn, J.B., Rosell, J., and Meng, C.K.J. (1997). In electrospray ionization, how much pull does an ion need to escape its droplet prison? . J. Am. Soc. Mass Spectrom. 8, 1147-1157.

Geels, R.B., van der Vies, S.M., Heck, A.J., and Heeren, R.M. (2006). Electron capture dissociation as structural probe for noncovalent gas-phase protein assemblies. Anal Chem 78, 7191-7196.

Giles, K., Pringle, S.D., Worthington, K.R., Little, D., Wildgoose, J.L., and Bateman, R.H. (2004). Applications of a travelling wave-based radio-frequency-only stacked ring ion guide. Rapid Commun Mass Spectrom 18, 2401-2414.

Gill, A.C., Jennings, K.R., Wyttenbach, T., and Bowers, M.T. (2000). Conformations of biopolymers in the gas phase: a new mass spectrometric method. International Journal of Mass Spectrometry 196, 685-697

Gomez, A., and Tang, K. (1994). Charge and fission of droplets in electrostatic sprays. Physics of Fluids 6, 404-414.

Grandori, R. (2002). Detecting equilibrium cytochrome c folding intermediates by electrospray ionization mass spectrometry: Two partially folded forms populate the molten-globule state. Protein Science 11, 453-458.

Hautreux, M., Hue, N., Kerdaniel, A.D.F.d., Zahir, A., Malec, V., and Laprévote, O. (2004). Under non-denaturing solvent conditions, the mean charge state of a multiply charged protein ion formed by electrospray is linearly correlated with the macromolecular surface International Journal of Mass Spectrometry 231, 131-137

Heck, A.J., and Van Den Heuvel, R.H. (2004). Investigation of intact protein complexes by mass spectrometry. Mass Spectrom Rev 23, 368-389.

Hoaglund-Hyzer, C.S., Counterman, A.E., and Clemmer, D.E. (1999). Anhydrous protein ions. Chem Rev 99, 3037-3080.

Hoffmann, E., and Stroobant, V. (2002). Mass Spectrometry (Paris: John Wiley & Sons).

Iavarone, A.T., and Williams, E.R. (2003). Mechanism of charging and supercharging molecules in electrospray ionization. J Am Chem Soc 125, 2319-2327.

Iribarne, J.V., and Thomson, B.A. (1976). On the evaporation of small ions from charged droplets. The Journal of Chemical Physics 64, 2287-2294.

Jarrold, M.F. (2000). Peptides and proteins in the vapor phase. Annu Rev Phys Chem 51, 179-207.

Jones, C.M., Beardsley, R.L., Galhena, A.S., Dagan, S., Cheng, G., and Wysocki, V.H. (2006). Symmetrical gas-phase dissociation of noncovalent protein complexes via surface collisions. J Am Chem Soc 128, 15044-15045.

Jurchen, J.C., Garcia, D.E., and Williams, E.R. (2004). Further studies on the origins of asymmetric charge partitioning in protein homodimers. J Am Soc Mass Spectrom 15, 1408-1415.

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

Kaddis, C.S., Lomeli, S.H., Yin, S., Berhane, B., Apostol, M.I., Kickhoefer, V.A., Rome, L.H., and 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.

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

Karasek, F.W., Denney, D.W., and DeDecker, E.H. (1974). Plasma chromatography of normal alkanes and its relation to chemical ionization mass spectrometry. Anal. Chem. 46, 970-973.

Kaufman, S.L., Kuchumov, A.R., Kazakevich, M., and Vinogradov, S.N. (1998). Analysis of a 3.6-MDa hexagonal bilayer hemoglobin from Lumbricus terrestris using a gas-phase electrophoretic mobility molecular analyzer. Anal Biochem 259, 195-202.

Kenny, D.J., Brown, J.M., Palmer, M.E., Snel, M.F., and Bateman, R.H. (2006). A parallel approach to post source decay MALDI-TOF analysis. J Am Soc Mass Spectrom 17, 60-66.

Konermann, L. (2007). A minimalist model for exploring conformational effects on the electrospray charge state distribution of proteins. Journal of Physical Chemistry B 111, 6534-6543.

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

Loo, J.A. (1997). Studying noncovalent protein complexes by electrospray ionization mass spectrometry. Mass Spectrom Rev 16, 1-23.

Loo, J.A., Berhane, B., Kaddis, C.S., Wooding, K.M., Xie, Y., Kaufman, S.L., and 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.

Lorenzen, K., Olia, A.S., Uetrecht, C., Cingolani, G., and 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.

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

Mamyrin, B.A. (2001). Time-of-flight mass spectrometry (concepts, achievements, and prospects). International Journal of Mass Spectrometry 206, 251-266.

Mamyrin, B.A., Karataev, V.I., Shmikk, D.V., and Zagulin, V.A. (1973). The mass-reflectron, a new nonmagnetic time-of-flight mass spectrometer with high resolution. Sov. Phys. JETP 37, 45-48.

Mann, M., Hendrickson, R.C., and Pandey, A. (2001). Analysis of proteins and proteomes by mass spectrometry. Annu Rev Biochem 70, 437-473.

Mason, E.A., and McDaniel, E.W. (1988). Transport Properties of Ions in Gases (New York: John Wiley & Sons).

Maxence Wisztorski, D.C.E.M.I.F.M.S. (2008). Molecular MALDI imaging: An emerging technology for neuroscience studies. Developmental Neurobiology 68, 845-858.

McKay, A.R., Ruotolo, B.T., Ilag, L.L., and Robinson, C.V. (2006). Mass measurements of increased accuracy resolve heterogeneous populations of intact ribosomes. Journal Of The American Chemical Society 128, 11433-11442.

Meroueh, S.O., Roblin, P., Golemi, D., Maveyraud, L., Vakulenko, S.B., Zhang, Y., Samama, J.P., and Mobashery, S. (2002). Molecular dynamics at the root of expansion of function in the M69L inhibitor-resistant TEM beta-lactamase from Escherichia coli. J Am Chem Soc 124, 9422-9430.

Pan, X.M., Sheng, X.R., Yang, S.M., and Zhou, J.M. (1997). Probing subtle acid-induced conformational changes of ribonuclease A by electrospray mass spectrometry. FEBS Lett 402, 25-27.

Pringle, S.D., Giles, K., Wildgoose, J.L., Williams, J.P., Slade, S.E., Thalassinos, K., Bateman, R.H., Bowers, M.T., and 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. International Journal of Mass Spectrometry 261, 1-12.

Robinson, C.V. (2002). Protein complexes take flight. Nature Structural Biology 9, 505-506.

Robinson, C.V., Chung, E.W., Kragelund, B.B., Knudsen, J., Aplin, R.T., Poulsen, F.M., and Dobson, C.M. (1996). Probing the nature of noncovalent interactions by mass spectrometry. A study of protein-CoA ligand binding and assembly. Journal of the American Chemical Society 118, 8646-8653.

Robinson, C.V., Gross, M., Eyles, S.J., Ewbank, J.J., Mayhew, M., Hartl, F.U., Dobson, C.M., and Radford, S.E. (1994). Conformation of GroEL-bound alpha-lactalbumin probed by mass spectrometry. Nature 372, 646-651.

Robinson, C.V., Gross, M., and Radford, S.E. (1998). Probing conformations of GroEL-bound substrate proteins by mass spectrometry. Methods Enzymol 290, 296-313.

Ruotolo, B., Benesch, J.L., Sandercock, A.M., Hyung, S.J., and Robinson, C.V. (2008). Ion mobility-mass spectrometry analysis of large protein complexes. Nature Protocols 3.

Ruotolo, B.T., Giles, K., Campuzano, I., Sandercock, A.M., Bateman, R.H., and Robinson, C.V. (2005). Evidence for Macromolecular Protein Rings in the Absence of Bulk Water. Science 310, 1658-1661.

Schwartz, B.L., Bruce, J.E., Anderson, G.A., Hofstadler, S.A., Rockwood, A.L., and Smitha, R.D. (1995). Dissociation of Tetrameric Ions of Noncovalent Streptavidin Complexes Formed by Electrospray Ionization Journal Of The American Society For Mass Spectrometry 6, 459-465

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

Sobott, F., and Robinson, C.V. (2002). Protein complexes gain momentum. Curr Opin Struct Biol 12, 729-734.

Stephens, W. (1946). Pulsed Mass Spectrometer with Time Dispersion. Physical Review 69, 674.

Tahallah, N., Pinkse, M., Maier, C.S., and 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.

Tesic, M., Wicki, J., Poon, D.K.Y., Withers, S.G., and Douglas, D.J. (2007). Gas phase noncovalent protein complexes that retain solution binding properties: Binding of xylobiose inhibitors to the beta-1,4 exoglucanase from Cellulomonas fimi. Journal Of The American Society For Mass Spectrometry 18, 64-73.

Thomson, J.J. (1913). On the Appearance of Helium and Neon in Vacuum Tubes. Nature 90, 645-647.

Uetrecht, C., Versluis, C., Watts, N.R., Roos, W.H., Wuite, G.J.L., Wingfield, P.T., Steven, A.C., and Heck, A.J.R. (2008). High-resolution mass spectrometry of viral assemblies: Molecular composition and stability of dimorphic hepatitis B virus capsids. Proceedings of the National Academy of Sciences 105, 9216-9220.

van den Heuvel, R.H.H., van Duijn, E., Mazon, H., Synowsky, S.A., Lorenzen, K., Versluis, C., Brouns, S.J.J., Langridge, D., van der Oost, J., Hoyes, J., and Heck, A.J.R. (2006). Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. Analytical Chemistry 78, 7473-7483.

Verbeck, G., Ruotolo, B., Sawyer, H., Gillig, K., and Russell, D. (2002). A fundamental introduction to ion mobility mass spectrometry applied to the analysis of biomolecules. J Biomol Tech. 13, 56–61.

Versluis, C., van der Staaij, A., Stokvis, E., Heck, A.J., and de Craene, B. (2001). Metastable ion formation and disparate charge separation in the gas-phase dissection of protein assemblies studied by orthogonal time-of-flight mass spectrometry. J Am Soc Mass Spectrom 12, 329-336.

Videler, H., Ilag, L.L., McKay, A.R.C., Hanson, C.L., and Robinson, C.V. (2005). Mass spectrometry of intact ribosomes. FEBS Letters 579, 943-947.

von Helden, G., Wyttenbach, T., and Bowers, M.T. (1995). Conformation of Macromolecules in the Gas Phase: Use of Matrix-Assisted Laser Desorption Methods in Ion Chromatography. Science 267, 1483-1485.

Wiley, W.C., and McLaren, I.H. (1955). Time-of-Flight Mass Spectrometer with Improved Resolution. Review of Scientific Instruments 26, 1150-1157.

Yamashita, M., and Fenn, J.B. (1984). Another variation of the jet-free theme. J Chem Phys 88, 4551-4459.

Yan, X., Watson, J., Ho, P.S., and Deinzer, M.L. (2004). Mass spectrometric approaches using electrospray ionization charge states and hydrogen-deuterium exchange for determining protein structures and their conformational changes. Mol Cell Proteomics 3, 10-23.

COMPOSITION AND STRUCTURE OF BACTERIOPHAGE P22

Kristina Lorenzen
Biomolecular Mass Spectrometry and Proteomics Group, Bijvoet Center for Biomolecular Research and Utrecht Faculty of Science, Utrecht University, Sorbonnelaan 16, 3584 CA Utrecht, the Netherlands

CONTEXT

An example on how new techniques in mass spectrometry (MS) can provide novel insight into structural properties of protein complexes is shown for the bacteriophage P22. Native and ion mobility mass spectrometry can monitor conformational changes in the assembly process of proteins as shown in chapter 3 of this thesis. Below is given a general introduction on the biophysical properties of P22 and related phages.

An introduction into the bacteriophage P22

Bacteriophages have been discovered at the start of the last century by Twort and d´Herelle (d´Herelle, 1917; Twort, 1915). They identified them as filterable and transmittable components of bacterial lysis. In the 1940s new interest arose in these organisms as the basics of today's modern techniques in biology like DNA cloning, targeted gene therapy, and phage display of novel proteins emerged. Bacteriophages are also interesting for therapy in human diseases, veterinary medicine and agriculture. The benefit of phage therapy is that bacteriophages can be much more specific than commonly used drugs, since they do not affect the patient or other beneficial bacteria living in it. P22 gained importance in the 1950s when it was shown that about one out of a thousand P22 phages does actually incorporate the host genome, instead of its own, when assembling new phages. The bacteriophage still recognizes the appropriate bacteria, attaches, and injects the DNA. The result is transfer of DNA from one host to another. P22 can thus be used in the laboratory to move genes from one bacterium to another, and therefore is of practical importance for example in the study of Salmonella enterica serova typhimurium genetics, which is the host of P22. From the 1970s until 2000 the phage work mostly found its application in biological methods and the interest in the organism itself has only revived again in the last few years.

Although viruses are extremely diverse in their life cycles and infectious mechanisms, they share a remarkable amount of common features. Generally they are consisting of an inner core of condensed nucleic acid enclosed within a protein capsid. Many copies of only few, relatively small proteins make up a phage. The capsid consists out of several copies of the so called coat/capsid protein. Most capsids have a hexagonal (isocahedral) symmetry, exceptions from this are for example the human immunesuppressing virion (HIV), the Tabaccomosaicvirus (TMV) and the M13 bacteriophage. Assuming the binding interface between the coat proteins is the same then the only combinations possible exist of sixty coat proteins or multiple of this. Caspar and Klug (Caspar and Klug, 1962) introduced the triangulation number (T) to indicate how many copies of the coat protein make up the capsid of a phage or virus. In addition many bacteriophages have a tail structure with which they attach to the host organism.

Phage structure and morphology

P22 (Figure 1) belongs to the family of the Podoviridae (short tailed bacteriophages) and is a relative of the famous coliphage λ. P22 incorporates double stranded DNA into its capsid. The capsid is a T=7 construction (Casjens, 1979; Earnshaw, 1979) and is about 60 nm in diameter.

The starting point of the assembly pathway is a closed shell of coat protein that does not contain DNA. These shells, called procapsids, include proteins that are not present later in the infectious phage. These proteins are the so called scaffolding proteins which assemble with the coat protein subunits to form a structure with the scaffolding inside and the coat protein on the outside. The initial formation of the procapsid is triggered by the assembly of the portal protein

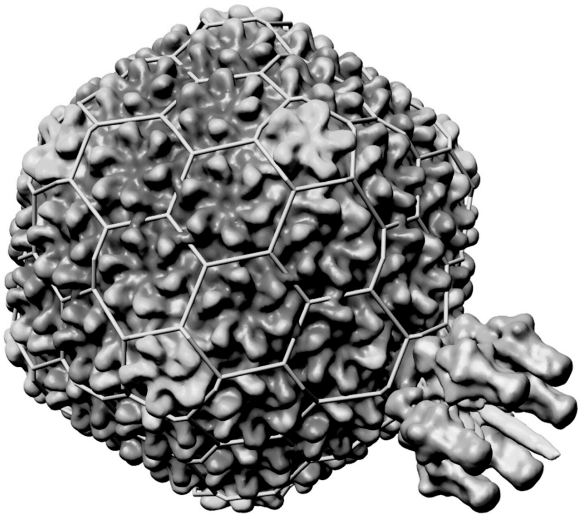

Figure 1 Three dimensional EM reconstruction of the bacteriophage P22. The icosahedral arrangement of the coat protein in the capsid is indicated with the light grey cage. One vertex, consisting out of five coat proteins, is displaced by the tail machinery. Dominantly visible are the six fold symmetry of gp9 and the protruding needle (gp26). (Adapted from (Lander et al., 2006))

(gp1) together with the scaffold protein. Around the portal scaffold complex then the coat protein and more scaffold proteins attach. When the DNA is packed into the procapsid all scaffolding proteins are removed from within the shell (Greene and King, 1994; Thuman-Commike et al., 1998), but stay intact and get reused building the next procapsid. The DNA is pumped into the capsid in a process that involves two virus-encoded proteins and ATP, up to a concentration of about 500 mg/ml (Earnshaw and Casjens, 1980; Smith et al., 2001). During these events the capsid undergoes conformational changes which expand the capsid, resulting in its mature structure. P22 has a linear dsDNA chromosome within its virion that is about 43,500 bp in length with blunt ends and a circular genetic map. The complete genome sequence is available online (accession number AF217253).

The connection between the capsid and the tail

Gp1 or portal protein assembles in a circular structure, where twelve copies are occupying a pentameric vertex. This portal ring is the connection between capsid and tail. Insertion of the portal ring with its twelve fold axis into the five fold axis of the coat protein is a symmetry mismatch (Hendrix, 1978; Simpson et al., 2000) that is common in all studied tailed phages. It has been suggested that this mismatch facilitates rotational movements during the packaging of the DNA into the capsid. There have been attempts to crystallize the portal ring, but the resolution did not exceed 7Å (Cingolani et al., 2002). There are publications, however, on portal structures of the Φ29 bacteriophages (Guasch et al., 2002b; Simpson et al., 2000). The portal protein of Φ29 has a much shorter sequence than gp1 of P22 but the two are supposed to share a certain degree of structural homology which gets obvious from the EM structures of gp1 (see also Figure 2). EM structures of the portal ring are reported assembled in vitro and in vivo (Lander et al., 2006; Tang et al., 2005). Tang et al. reported the EM structure of the truncated portal, ending with amino acid 602 (Figure 2A). The upper two thirds of the structure are connected by a thin density, which to their opinion implies the flexibility of the ring. The top part has been shown to have a regulatory role during

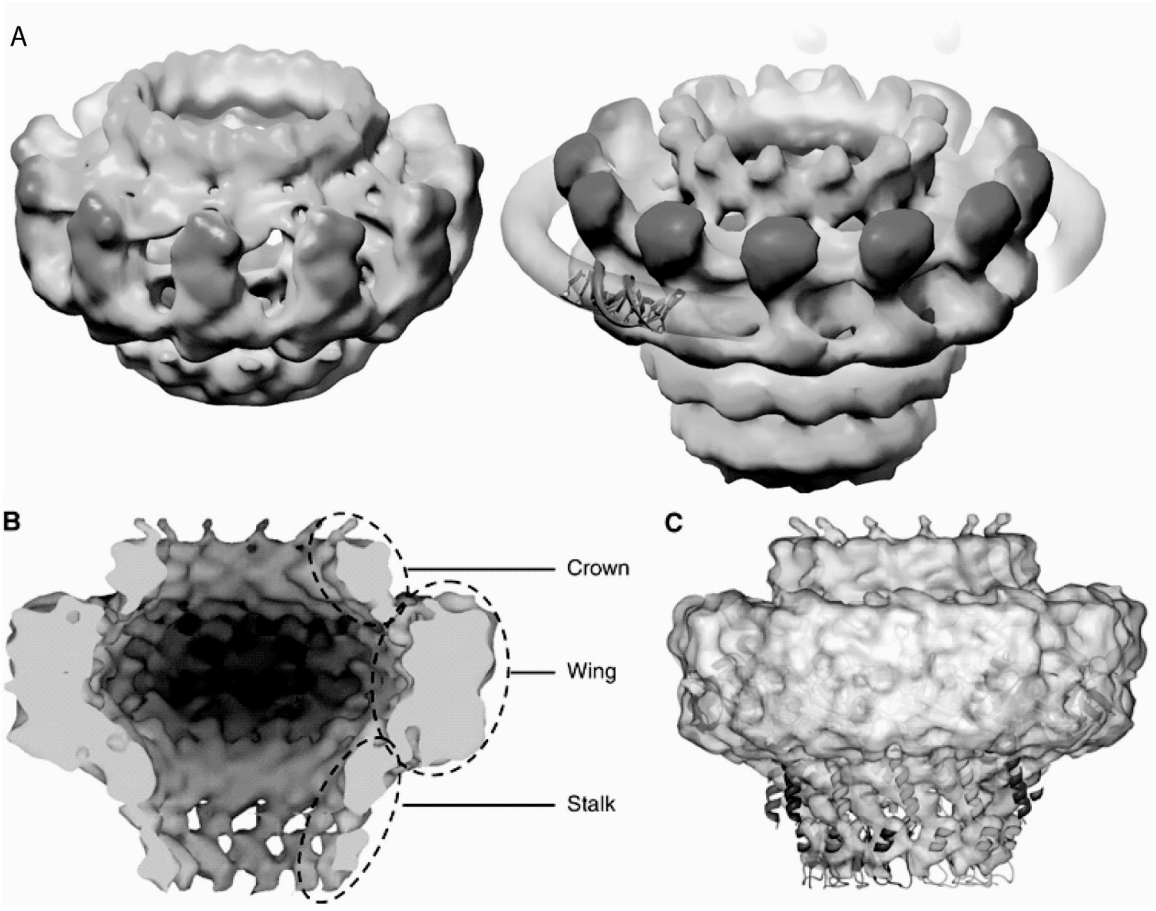

Figure 2 Cyro EM reconstructions of the full portal ring in vivo and in vitro (A) and the truncation mutant gp1602 in vitro (B,C). (A) shows a comparison between the in vitro assembled portal (left) and the density that was extracted from the asymmetric reconstruction of the phage (right). (B) shows the surface view of the x ray structure of the Φ29 (1ijg). (C) shows the portal ring assembled in vitro from the truncation mutant of Gp1602. Into this assembly is fitted the X ray structure of Φ29 (visible are the gray helices on the bottom). The figures are adapted (Lander et al., 2006; Tang et al., 2005) respectively.

DNA packaging and injection (Guasch et al., 2002a; Orlova et al., 2003; Simpson et al., 2000). The central region of the P22 portal sequence, which is predicted to consist of conserved helical modules, is involved in inter-subunit interactions upon assembly (Rodriguez-Casado et al., 2001). The lower half of the EM density shows similarities between P22 and Φ29 that might suggest a similar mechanical role in DNA packaging and ejection (Serwer, 2003). A recent structural study compared results of hydrogen/deuterium exchange mass spectrometry (Kang et al.) with the crystal structure of the Φ29 portal. It provides additional evidence that the subunit interactions in the two complexes are similar (Guasch et al., 2002b; Simpson et al., 2000). Other structural studies of the portal protein have shown that replacement of cysteine 283 with serine results in a decreased ability of the portal to oligomerize and decreased infectivity. Lander et al (Lander et al., 2006) have shown that the portal ring when overexpressed and assembled in vitro has a different conformation than in vivo and suggest that the portal changes its conformation upon the pressure caused by the DNA inside the capsid at the end of the packaging process. They based their findings on a ring of DNA that is located around the portal ring in the readily assembled phage.

Name	Function	Mass (Da)	Copy number
gp1	DNA Packing	82611	12
gp4	Head Completion	18025	12
gp10	Head Completion	52457	6
gp26	Head Completion	24603	3
gp9	Receptor binding	17857	6x3

The C-terminal domain has been putatively assigned to the part of the portal ring that faces into the capsid (Kang et al.) but the 140 C-terminal residues are not critical for the assembly of the portal ring (Bazinet et al., 1988; Tang et al., 2005).

When the packaging process finishes, two additional proteins (gp2 and gp3) are required for the termination of DNA packaging (Lander et al., 2006). The expression of the P22 DNA in the host cell results in an endless DNA strand, which therefore needs to be cleaved to finish the packaging process. Three non structural proteins, gp7, 16 and 20, which are required for successful injection of the DNA into the host (Botstein et al., 1973; Casjens and King, 1974; Hoffman and Levine, 1975; Israel, 1977), are present in the infectious form of P22 and are ejected into the host cell after attachement.

The tail machinery

Podoviridae have a short tail, which attaches to the portal (Figure 3). It is specialized and crucial for infection and assembly. The assembly of this tail is initiated by the portal ring together with the incooperation of the three minor non structural proteins into the procapsid (Bazinet et al., 1988). These minor proteins are required for the infectivity but not for the assembly of P22 (Botstein et al., 1973). The assembly of the tail structure is suggested to be sequential and highly cooperative (King et al., 1973; Olia et al., 2006; Olia et al., 2007a; Olia et al., 2007b). Five different proteins, arranged in a twelve, six and three fold symmetry form the tail of P22 which consists of fifty one subunits in total (see Table 1). After the capsid is fully packed with the dsDNA, gp4, 10 and 26, also referred to as portal closure or head completion proteins, assemble to the portal ring. It has been shown that the loss of one of these proteins causes the DNA to leak out of the capsid (King et al., 1973; Lenk et al., 1975; Strauss and King, 1984), which indicates that they are not involved in DNA packaging but are crucial to retain the DNA inside. The first tail accessory factor, or head completion protein, to assemble to the portal ring is gp4. It is an elongated monomer in solution and has a helical C-terminal region that is important for binding to the portal. Mutation experiments have shown that deletions in the region after amino acid 76, which lacks the C-terminal helical motive results in the loss of the ability to bind to the portal. Deletion after amino acid 126 produces a mutant of gp4 that is still able to bind, but with a much lower affinity than the full length protein (Olia et al., 2006). At conditions similar to the ones in the host

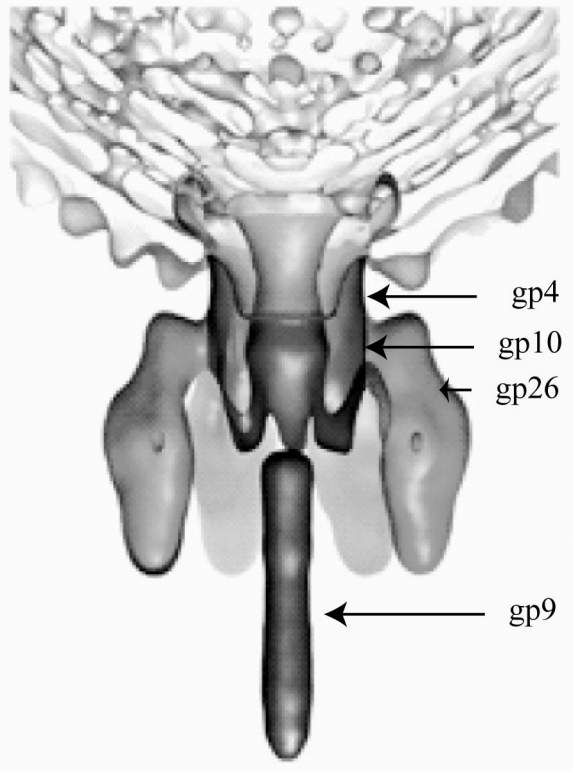

Figure 3 Open side view of the assembled tail machinery of P22. The different subunits involved in the assembly are indicated on the right. Adapted from (Tang et al., 2005).

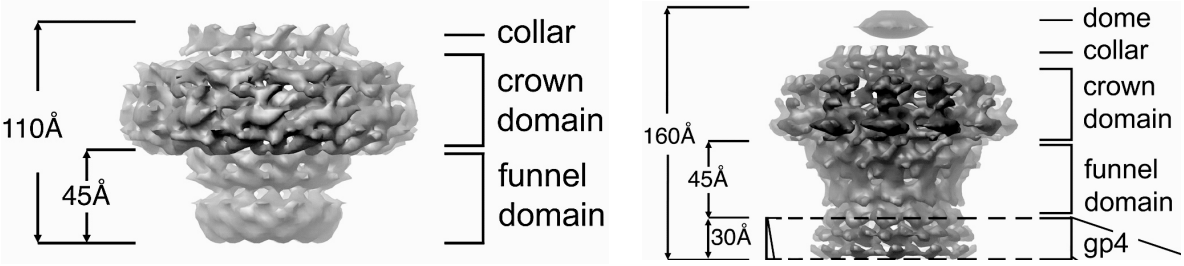

Figure 4 Conformational changes in the portal ring ($gp1_{602}$) upon binding of the tail accessory factor gp4. On the left side is the unbound portal ring with the conformation as suggested to be in the procapsid. The structure is about 110 Å in height and the smallest inner diameter lies in the funnel with 35 Å. When gp4 binds to the ring, the conformation changes to a structure with a length of 160 Å, where 30 Å are originating from gp4 bound to the lower part of the funnel domain. The overall diameter of the crown domain remains unchanged. The funnel domain increased slightly in inner diameter to 40 Å, being the tightest point of the channel Adapted from (Zheng et al., 2008)

cell, gp4 is unstable. Thus the C-terminal domain of gp4 is dispensable for binding, while the N-terminal domain is essential. The melting temperature of the protein lies at 34 °C. Olia et al. suggest that in its natural environment gp4 spontaneously unfolds if not bound to the portal immediately. In addition gp4 has been shown to possess murein hydrolase activity (Moak and Molineux, 2004) which is needed to digest the peptidoglycan layer in the periplasm of Salmonella. Work that was published simultaneously to chapter 3 of this thesis revealed cryo EM structures of the truncated portal protein ($gp1_{602}$) and $gp1_{602}$ with gp4 bound at a resolution of 17 Å (Zheng et al., 2008) (Figure 3). Using the EM structure the portal ring could be separated into three distinct regions: collar, crown and funnel (see Figure 3A). A comparison with Lander et al. (Lander et al., 2006) shows that the missing C-terminal amino acids are corresponding to the exterior "spokes" in the crown domain that can be seen in Figure 2A. Gp4 binds in a stoichiometric manner to the portal, forms two concentric rings and thereby induces dramatic conformational changes in the portal ring. The portal structure changes and on top a dome like lid closes. This dome, which is visible in the in vivo structure as a doughnut like shape, might be important in recruiting the injection proteins. The structure might bind DNA via a conserved ^{286}KRRR^{289} motive that is disordered in the procapsid but becomes ordered upon binding of gp4.

After gp4 induces the conformational change in the portal, it provides the binding interface for gp10. Gp10 presents the link to gp26 and gp9, which bind independently (Botstein et al., 1973; Poteete et al., 1979; Strauss and King, 1984). In solution gp10 is a ß sheet rich protein, which forms monomer, dimer, trimer, tetramer, pentamer and eventually hexameric assemblies. The hexamers then form the portal ring-gp4 structure (Olia et al., 2007a). Contrary to gp4, gp10 is more thermo stable in solution and the melting point of the protein increases with the oligomeric state. With the assembly of gp10 to the nascent tail the twelve fold symmetry changes to six fold symmetry (Lander et al., 2006; Olia et al., 2007a; Tang et al., 2005).

The last of the three tail accessory factors is gp26, also called needle because of the ~220Å elongated structure. Gp26 is necessary to close off the channel created by the portal, gp4 and gp10 and retain the DNA inside the capsid (King et al., 1973; Lenk et al., 1975; Strauss and King, 1984). In the cryo EM reconstruction it is visible as a thin density sticking out of the lower part of the tail (see Figure 1). The elongated structure and the protrusion of gp26 out of the tail have led to the assumption that gp26 is not only necessary to keep the DNA inside the capsid but also plays a role in the penetration of the host cell. This is supported by early findings that gp26 is among the first proteins that are ejected from the capsid into the host cell (Israel, 1977). It has been shown that the first 26 N-terminal residues are involved in the binding of the needle to gp10. The following residues up to amino acid 140 fold into a "highly stable and self assembling" α helical core (Bhardwaj et al., 2007) whereas the C-terminal part of gp26 forms a more flexible and less stable structure that might be involved in penetrating the host lipid bilayer during infection.

The last protein to attach to the tail is gp9, called the tailspike. It has a mass of 72 kDa and forms a trimer. The connection to the rest of the tail is made by the N-terminal part which makes contact with gp4 and gp10 (Tang et al., 2005), where six trimers bind to the tail. The C-terminal receptor part of gp9 binds to the O-antigen polysaccharide receptor on the outer membrane of the host cell (Steinbacher et al., 1997a; Steinbacher et al., 1997b). Gp9 has endorhamnosidase activity allowing it to cleaves the O-antigen chain (Baxa et al., 2001; Steinbacher et al., 1996) which is then followed by the injection of proteins and the DNA into the host cell. It might be that this binding and cleavage of O-antigen allows the virion to burrow its way to the surface of the bacteria's outer membrane.

REFERENCES

Baxa, U., Cooper, A., Weintraub, A., Pfeil, W., and Seckler, R. (2001). Enthalpic barriers to the hydrophobic binding of oligosaccharides to phage P22 tailspike protein. Biochemistry 40, 5144-5150.

Bazinet, C., Benbasat, J., King, J., Carazo, J.M., and Carrascosa, J.L. (1988). Purification and organization of the gene 1 portal protein required for phage P22 DNA packaging. Biochemistry 27, 1849-1856.

Bhardwaj, A., Olia, A.S., Walker-Kopp, N., and Cingolani, G. (2007). Domain Organization and Polarity of Tail Needle GP26 in the Portal Vertex Structure of Bacteriophage P22. Journal of Molecular Biology 371, 374-387.

Botstein, D., Waddell, C.H., and King, J. (1973). Mechanism of head assembly and DNA encapsulation in Salmonella phage p22. I. Genes, proteins, structures and DNA maturation. J Mol Biol 80, 669-695.

Casjens, S. (1979). Molecular organization of the bacteriophage P22 coat protein shell. J Mol Biol 131, 1-14.

Casjens, S., and King, J. (1974). P22 morphogenesis. I: Catalytic scaffolding protein in capsid assembly. J Supramol Struct 2, 202-224.

Caspar, D., and Klug, A. (1962). Physical principles in the construction of regular viruses. Cold Spring Harbor Symposia On Quantitative Biology. 27, 1-24.

Cingolani, G., Moore, S.D., Prevelige, P.E., and Johnson, J.E. (2002). Preliminary crystallographic analysis of the bacteriophage P22 portal protein. Journal of Structural Biology 139, 46-54.

d´Herelle, F. (1917). Sur un microbe invisible antagoniste des bacilles dysentériques. Comptes rendus Acad Sciences 165, 373-375.

Earnshaw, W.C. (1979). Modeling the small-angle x-ray diffraction arising from the surface lattices of phages lambda and P22. J Mol Biol 131, 14-19.

Earnshaw, W.C., and Casjens, S.R. (1980). DNA packaging by the double-stranded DNA bacteriophages. Cell 21, 319-331.

Greene, B., and King, J. (1994). Binding of scaffolding subunits within the P22 procapsid lattice. Virology 205, 188-197.

Guasch, A., Aloria, K., Perez, R., Avila, J., Zabala, J.C., and Coll, M. (2002a). Three-dimensional structure of human tubulin chaperone cofactor A. J Mol Biol 318, 1139-1149.

Guasch, A., Pous, J., Ibarra, B., Gomis-Ruth, F.X., Valpuesta, J.M., Sousa, N., Carrascosa, J.L., and Coll, M. (2002b). Detailed architecture of a DNA translocating machine: the high-resolution structure of the bacteriophage phi29 connector particle. J Mol Biol 315, 663-676.

Hendrix, R.W. (1978). Symmetry Mismatch and DNA Packaging in Large Bacteriophages. Proceedings of the National Academy of Sciences 75, 4779-4783.

Hoffman, B., and Levine, M. (1975). Bacteriophage P22 virion protein which performs an essential early function. I. Analysis of 16-ts mutants. J Virol 16, 1536-1546.

Israel, V. (1977). E proteins of bacteriophage P22. I. Identification and ejection from wild-type and defective particles. J Virol 23, 91-97.

Kang, S., Poliakov, A., Sexton, J., Renfrow, M.B., and Prevelige Jr, P.E. Probing Conserved Helical Modules of Portal Complexes by Mass Spectrometry based Hydrogen/deuterium Exchange. Journal Of Molecular Biology In Press, Accepted Manuscript.

King, J., Lenk, E.V., and Botstein, D. (1973). Mechanism of head assembly and DNA encapsulation in Salmonella phage P22. II. Morphogenetic pathway. J Mol Biol 80, 697-731.

Lander, G.C., Tang, L., Casjens, S.R., Gilcrease, E.B., Prevelige, P., Poliakov, A., Potter, C.S., Carragher, B., and Johnson, J.E. (2006). The structure of an infectious P22 virion shows the signal for headful DNA packaging. Science 312, 1791-1795.

Lenk, E., Casjens, S., Weeks, J., and King, J. (1975). Intracellular visualization of precursor capsids in phage P22 mutant infected cells. Virology 68, 182-199.

Moak, M., and Molineux, I.J. (2004). Peptidoglycan hydrolytic activities associated with bacteriophage virions. Molecular Microbiology 51, 1169-1183.

Olia, A.S., Al-Bassam, J., Winn-Stapley, D.A., Joss, L., Casjens, S.R., and Cingolani, G. (2006). Binding-induced stabilization and assembly of the phage P22 tail accessory factor gp4. Journal of Molecular Biology 363, 558-576.

Olia, A.S., Bhardwaj, A., Joss, L., Casjens, S., and Cingolani, G. (2007a). Role of Gene 10 Protein in the Hierarchical Assembly of the Bacteriophage P22 Portal Vertex Structure. Biochemistry 46, 8776-8784.

Olia, A.S., Casjens, S., and Cingolani, G. (2007b). Structure of phage P22 cell envelope-penetrating needle. Nat Struct Mol Biol.

Orlova, E.V., Gowen, B., Droge, A., Stiege, A., Weise, F., Lurz, R., van Heel, M., and Tavares, P. (2003). Structure of a viral DNA gatekeeper at 10 A resolution by cryo-electron microscopy. EMBO J 22, 1255-1262.

Poteete, A.R., Jarvik, V., and Botstein, D. (1979). Encapsulation of phage P22 DNA in vitro. Virology 95, 550-564.

Rodriguez-Casado, A., Moore, S.D., Prevelige, P.E., Jr., and Thomas, G.J., Jr. (2001). Structure of bacteriophage P22 portal protein in relation to assembly: investigation by Raman spectroscopy. Biochemistry 40, 13583-13591.

Serwer, P. (2003). Models of bacteriophage DNA packaging motors. J Struct Biol 141, 179-188.

Simpson, A.A., Tao, Y., Leiman, P.G., Badasso, M.O., He, Y., Jardine, P.J., Olson, N.H., Morais, M.C., Grimes, S., Anderson, D.L., et al. (2000). Structure of the bacteriophage phi29 DNA packaging motor. Nature 408, 745-750.

Smith, D.E., Tans, S.J., Smith, S.B., Grimes, S., Anderson, D.L., and Bustamante, C. (2001). The bacteriophage straight phi29 portal motor can package DNA against a large internal force. Nature 413, 748-752.

Steinbacher, S., Baxa, U., Miller, S., Weintraub, A., Seckler, R., and Huber, R. (1996). Crystal structure of phage P22 tailspike protein complexed with Salmonella sp. O-antigen receptors. Proc Natl Acad Sci U S A 93, 10584-10588.

Steinbacher, S., Miller, S., Baxa, U., Budisa, N., Weintraub, A., Seckler, R., and Huber, R. (1997a). Phage P22 tailspike protein: Crystal structure of the head-binding domain at 2.3 angstrom, fully refined structure of the endorhamnosidase at 1.56 angstrom resolution, and the molecular basis of O-antigen recognition and cleavage. Journal of Molecular Biology 267, 865-880.

Steinbacher, S., Miller, S., Baxa, U., Weintraub, A., and Seckler, R. (1997b). Interaction of Salmonella phage P22 with its O-antigen receptor studied by X-ray crystallography. Biological Chemistry 378, 337-343.

Strauss, H., and King, J. (1984). Steps in the stabilization of newly packaged DNA during phage P22 morphogenesis. J Mol Biol 172, 523-543.

Tang, L., Marion, W.R., Cingolani, G., Prevelige, P.E., and Johnson, J.E. (2005). Three-dimensional structure of the bacteriophage P22 tail machine. Embo J 24, 2087-2095.

Thuman-Commike, P.A., Greene, B., Malinski, J.A., King, J., and Chiu, W. (1998). Role of the scaffolding protein in P22 procapsid size determination suggested by T = 4 and T = 7 procapsid structures. Biophys. J. 74, 559-568.

Twort, F. (1915). An investigation on the nature of ultra-microscopic viruses. Lancet 189, 1241-1243.

Zheng, H., Olia, A.S., Gonen, M., Andrews, S., Cingolani, G., and Gonen, T. (2008). A conformational switch in bacteriophage p22 portal protein primes genome injection. Mol Cell 29, 376-383.

STRUCTURAL ARCHITECTURE OF RNA POLYMERASE II AND III

Kristina Lorenzen
Biomolecular Mass Spectrometry and Proteomics Group, Bijvoet Center for Biomolecular Research and Utrecht Faculty of Science, Utrecht University, Sorbonnelaan 16, 3584 CA Utrecht, the Netherlands

CONTEXT

RNA polymerases (Pols) are maybe the most essential enzymes in the cell cycle. They are responsible for the transcription of DNA into RNA and thus not only the starting point of protein synthesis but also the target of various regulation mechanisms. Figure 1 shows a general scheme of the eukaryotic transcription apparatus and the central role of Pol II therein.

In chapter 4 of this thesis two protein complexes are focused upon: RNA polymerase II and III. They are both localized in the nucleoplasma. While Pol II synthesizes messenger RNA and small nucleotide RNA, Pol III transcribes the DNA encoding 5s ribosomal RNA, 6U RNA and transfer RNA. Mass spectrometry was used to investigate these intact large (above 500kDa) protein machineries, providing new insights into the structure of RNA polymerase II and III. The mass spectrometric results for the RNA polymerase II were in accordance with the published crystal structure. The data acquired for the related RNA polymerase III, provided unprecedented insights into the topology of this enzyme, which may guide future biochemical and structural studies. Here a short overview is given on Pol II and III, focusing primarily on the structural aspects.

POL II

Structural overview of Pol II

The crystal structure of the Pol II is known, but when it first was published (Cramer et al., 2000), two of its components were missing. The protein purifications for the crystals were made from endogenously expressed yeast cultures, but heterogeneity that was caused by two subunits namely Rpb4 and Rpb7 made the crystals difficult to analyze. For transcription initiation these two proteins are crucial although for transcription elongation they are not essential. The initial crystal structure consisted out of the ten so called "core" subunits (Cramer et al., 2000), later a structure of the complete twelve subunit enzyme was published (Armache et al., 2003) (Figure 2). Rpb1 and 2, which make up the main mass of the protein complex, have a big cleft between them. A helix of Rpb1 transverses on one side of the cleft while the C-terminal region of Rpb2 extends into the opposite side of the cleft. The active center of Pol II lies at the bottom of the cleft. Here, two magnesium ions are important in binding nucleotides in the transcription process. The DNA and the DNA-RNA hybrid are located at the floor of the cleft. The floor of this cleft is very thin and exposes the nucleotides to the space below. A helix forms a bridge at the bottom of the cleft creating two pores. One pore lies beneath the active center (pore 1) and the other one beneath the downstream DNA (pore 2). Both pores lie at the end of a cavity called the funnel. As the cleft is occupied by the DNA and the DNA-RNA hybrid the nucleotides cannot enter the cleft from above but are entering the active site through the funnel and pore 1. The ten other subunits are arranged around these two. On one side they are anchored by Rpb 3/10/11/12. Rpb5 and 9 as well as a part of Rpb1 build up another dominant structure of Pol II called jaw. It is located on the opposite side of the cleft and consists of the upper and lower jaw. The upper jaw, made up by Rpb5, is likely mobile (Miyao and Woychik, 1998). The lower jaw consists of parts of Rpb1 and Rpb9, where Rpb9 has zinc binding domains. Mutations in this subunit lead to altered transcription start sites (Furter-Graves et al., 1994). One side of the cleft is formed by the clamp. This substructure might be involved in "clamping nucleic acids in the cleft" (Fu et al., 1999). The clamp is made up by the N-terminus of Rpb1, C-terminus of Rpb2 and Rpb6. At the bottom of the clamp about three to twelve residues downstream of the active site, 42 charged residues of Rpb6 form a DNA binding site.

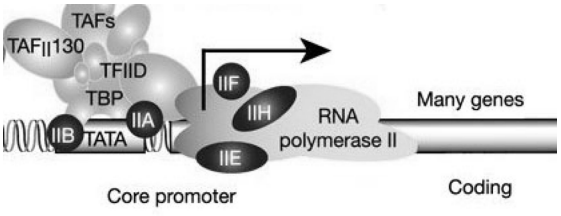

Figure 1 Overview of the eukaryotic transcriptional apparatus. It can be subdivided into classes of multi-subunit ensembles of which in this work are discussed: Pol II core complex and associated general transcription factors (TFIIA, -B,-D,-E,-F and -H) (adapted from (Levine and Tjian, 2003)).

The DNA-RNA hybrid that is formed in the active site cannot pass on straight through the enzyme because it is blocked by a structural element of Rpb2 called the wall. This forces the hybrid to tilt. There are two groves that can accommodate a single nucleic acid strand. One winds around the base of the clamp (likely RNA binding site) and the second (DNA binding site) along the lower part of the wall and Rpb1 continues downward between Rpb1 and Rpb11.

Rpb1, the largest subunit of Pol II has a unique C-terminal domain (CTD) consisting out of a 26 time repeat of seven

Table 1 Comparison between Pol II and Pol III subunits. (Jasiak et al., 2006)

Polymerase Part	Pol III Subunit	Pol II Subunit	Subunit Type	Sequence Identity (%)	Conserved Fold (%)
Core	C160	Rpb1	Homolog	28.4	83.2
	C128	Rpb2	Homolog	35.8	87.2
	AC40	Rpb3	Homolog	25.8	60.2
	AC19	Rpb11	Homolog	20.8	81.6
	Rpb5	Rpb5	Common	100	100
	Rpb6	Rpb6	Common	100	100
	Rpb8	Rpb8	Common	100	100
	Rpb10	Rpb10	Common	100	100
	Rpb12	Rpb12	Common	100	100
Rpb4/7 sub complex	C17	Rpb4	Homolog	7.2	50
	C25	Rpb7	Homolog	25.2	81.3
Upstream sub complex	C82/34/31		Specific	-	-
Downstream sub complex	C53/37/11[a]	Rpb9[a]	Unclear	-	-
Eleven subunit Pol III model	–	–	–	39.4	83.4

[a] Subunit C11 shows homology to Rpb9 and TFIIS. Rpb9 was previously defined as a part of the Pol II core.

amino acids (YSPTSPS) within them three serines. Serine at position two and serine at position five of the repeat get phosphorylated and dephosphorylated hereby regulating the transcription cycle (Meinhart et al., 2005). Another subunit which is known to be phoshporylated is Rpb6 (Kolodziej et al., 1990), however the phosphorylation of Rpb6 is not as well characterized as those of the CTD of Rpb1.

Preinitiation complexes (PIC)

The starting point for the assembly of Pol II are the subunits Rpb3 and Rpb11, where the C-terminal domain of Rpb11 plays a crucial role (Benga et al., 2005). It has been reported that Rpb3 and Rpb11 together interact with the mediator complex and thus have an important role in transcription initiation and regulation (Davis et al., 2002). Mutations on Rpb3 and Rpb11 lead to a terminator read through indicating that those subunits are essential for termination of transcription (Steinmetz et al., 2006). Upon the end of the transcription cycle it is likely that the enzyme falls apart and the subunits get reassembled at a new transcription initiation site.

The process of transcription of messenger RNA requires more than Pol II alone. Before the transcription can start some general transcription factors are required to bind to the TATA box, a common promoter element. TFII (general transcription factor II) B, E, F, H and D and the TBP (TATA binding protein) assemble to Pol II and the promoter element, which lies usually 20 - 30 base pairs upstream of the transcription start side (Bleichenbacher et al., 2003; Pugh, 2000). TFIIB C-terminal domain also binds to the Pol II "dock" domain (Buratowski and Zhou, 1993; Hisatake et al., 1993) and the N-terminal domain is located at the cleft and wall (Chen and Hahn, 2004). A part of the region between the C and the N-terminus extends into the active center of Pol II (Bushnell et al., 2004). The binding site of TFIIB to Pol II overlaps with the site where the generated RNA strand is supposed to exit the enzyme. At the start of transcription after the first eight base pairs, the exiting RNA might cause TFIIB to leave the complex and lead to promotor clearance (Chen and Hahn, 2003) (Figure 3).

Another transcription factor that is required to direct the intitiaton complex to the promoter is TFIIF, consisting of three subunits, which forms a tight complex with Pol II (Tyree et al., 1993). This complex extends along the cleft and interacts with the Rpb4/7 sub complex and links to Rpb2 (Chen et al., 2007). TFIIF interacts with DNA and FCP1 (which is involved in the phosphorylation of the C-terminal domain of Rpb1) through two winged helix domains (Cramer, 2004). Tfg1, which is the largest subunit in TFIIF and Tfg2, the second largest subunit have been shown to heterodimerize via their N-terminal regions. When bound to Pol II the smaller subunit extends along the cleft and the large subunit is in close proximity to the Rpb4/7 heterodimer (Chung et al., 2003). The third subunit is Tfg3. While the first two subunits are essential for viability, the third subunit is dispensable. Mutational studies have shown that the binding side of TFIIF might be in close proximity to Rpb9 and changes in the amino acids of Rpb9, Rpb2 and TFIIF have caused the transcription start site to shift further upstream (Chen et al., 2007).

After the assembly of the polymerase with TFIIB, TFIIF, TBP and DNA two other factors come into play. TFIIE and TFIIH are required for the "DNA melting" the separation of the DNA double strand. The N-terminus of TFIIE forms an unusual helix turn helix domain which might be involved in its recruitment to the complex and binding of DNA (Cramer, 2004).

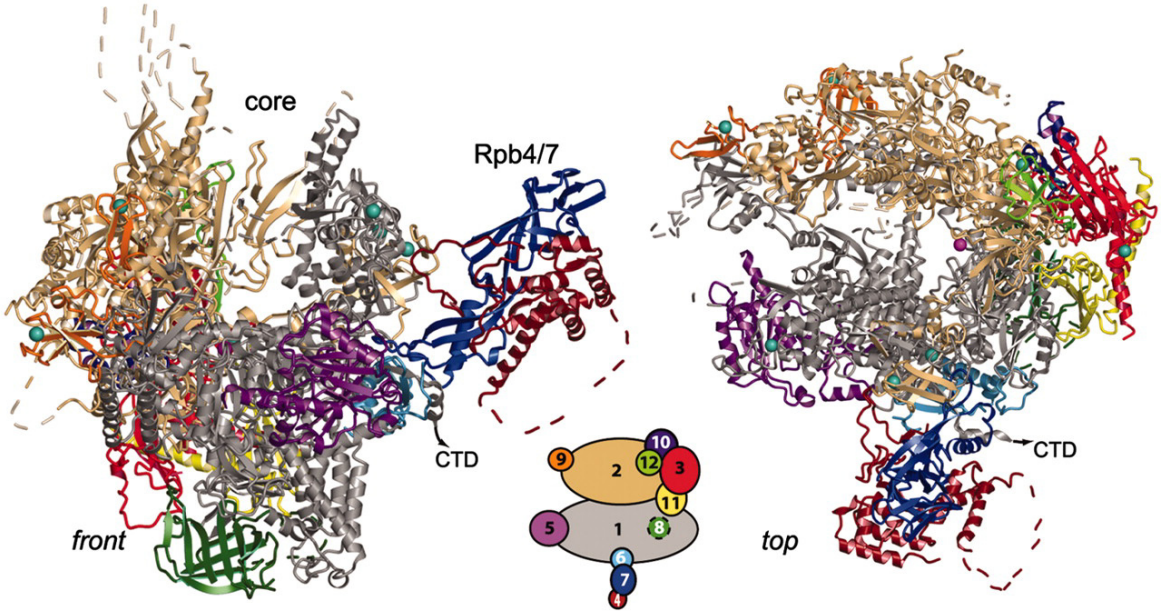

Figure 2 Ribbon diagram of complete twelve subunit Pol II. The subunits are colored as indicated in the scheme in the middle. The dashed lines represent disordered regions. The pink and cyan spheres represent magnesium and zinc ions, respectively. The left hand of the figure shows the front view and the right hand side the top view. Adapted from (Armache et al., 2005).

Transcription elongation and the incorporation of Nucleotide Triphosphates (NTPs)

The two large subunits Rpb1 and Rpb2 form the central mass of Pol II (Cramer et al., 2001). They build the positively charged cleft. One side of the cleft is formed by the mobile clamp region, which adopts an open conformation when there is no DNA substrate present. As mentioned above a bridge helix spans the cleft and this helix and the wall that is formed by the end of the cleft make up the active center of the enzyme. Beneath this active center a pore creates an inverted funnel to the outside. In the elongation phase the DNA-RNA duplex is bound above the pore, and the clamp adopts a closed state (Gnatt et al., 2001). Two metal ions (Mg^{2+}) in the active center are important for nucleotide binding and incorporation. They are coordinated by specific residues and in literature referred to as metal A and B. It has been suggested that polymerization and RNA cleavage requires metal A (Westover et al., 2004a), but different coordination of metal B is important in both processes. Metal A is coordinated by residues D481, D483 and D485 of Rpb1 as well as D836 of Rpb2. The selection of the correct nucleotide for the polymerization process is supposed to be a two step mechanism. Initial binding to the entry site beneath the active center of Pol II is followed by a rotation of the nucleotide for pairing with the DNA template strand (Westover et al., 2004a). If the incorrect nucleotide enters the active center, the coordination of the metal changes and therefore there is no specific binding. In that case metal B coordinates the phosphates of the nucleotide while the phosphate and sugar are in interaction with L752 of Rpb1 and R766, Y769, L987, S1019 and R1020 of Rpb2. As a result the subsequent rotation does not take place.

The general process of polymerization is more complex and involves, besides the selection and incorporation of the correct nucleotide, the translocation of DNA and RNA, the unwinding and rewinding of the helix and the separation of the DNA-RNA hybrid. The downstream DNA enters at the jaw and extends along the cleft towards the active site. The DNA-RNA hybrid formed at the active center then emerges towards the wall at a right angle with the downstream DNA. After the incorporation of the nucleotide the hybrid gets separated and the DNA exits above the so called arch, while the RNA exits opposite through the already mentioned pore beneath the arch (Kettenberger et al., 2004).

The site where the nucleotide is added to the DNA template strand is termed i+1. From i-2, the new RNA and the DNA template are connected over 7 8 base pairs till i-8. The unwinding of the DNA double strand starts at i+6. The residues L1109 and D1110 of Rpb1 (at the floor of the active center cleft) interact with i+5 and play a role in the separation. Rpb1 Y836 interacts with i+2 and i+3. Regarding the RNA DNA hybrid, the lid of Rpb1, (residues 246 264), the rudder of Rpb1 (residues 310 324) and the fork loop1 of Rpb2 (residues 461 480) play a role in strand separation (the rudder and fork loop1 extend as loops from opposite sides of the cleft). In the original free Pol II structure (Cramer et al., 2001) the rudder and fork loop 1 are disordered and likely mobile, however, in the elongation complexes they are ordered and in contact with each other. The lid (a prominent loop that protrudes from the edge of the clamp) drives

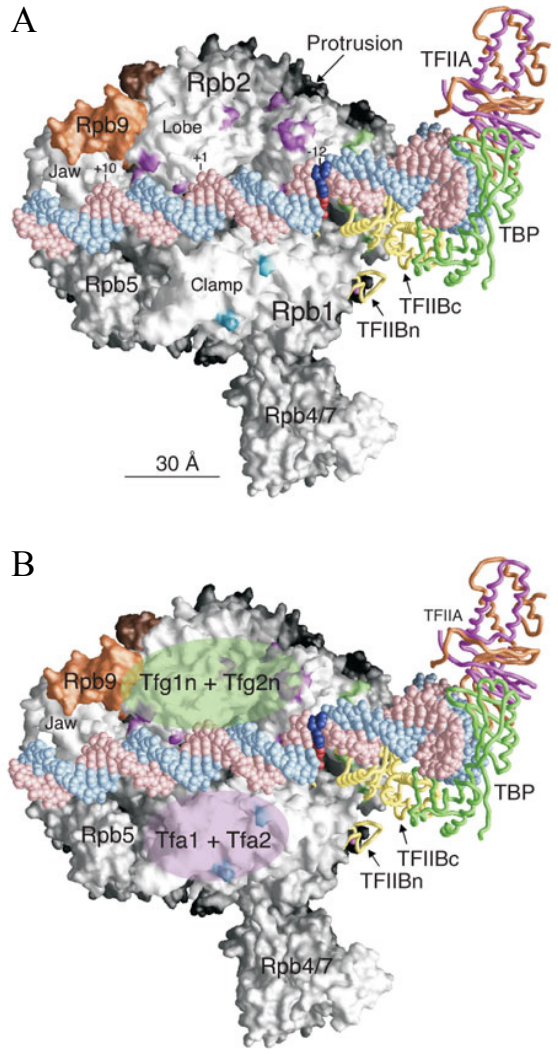

Figure 3 (A) Model of complex showing TFIIB core domain (TFIIBc) binding to Pol II. TFIIA, TBP and DNA are fit into the complex on the basis of the crystal structures of human TFIIB core (TFIIBc)–TBP–TATA box ternary complex (PDB 1C9B) and yeast TFIIA–TBP–TATA box complex (PDB 1RM1). TFIIB ribbon domain is modeled onto the Pol II surface on the basis of the Pol II–TFIIB cocrystal (PDB 1R5U): magenta sphere, zinc atom; yellow and green backbones, TFIIB and TBP, respectively; orange and magenta backbones, Toa1 and Toa2 subunits of TFIIA, respectively. Nontemplate and template strands are colored pink and light blue, respectively; red and dark blue highlight DNA base pair i-12, the site of DNA strand melting for open complex formation. On Pol II surface, Rpb9 is colored orange and Bpa positions cross-linked to TFIIB, TFIIF and TFIIE are colored green, purple and light blue, respectively. (B) Structural model as in A. Green and pink ovals mark locations of TFIIF dimerization domain (Tfg1n + Tfg2n) and TFIIE (Tfa1 + Tfa2), respectively. (Adapted from (Chen et al., 2007).

the RNA strand apart by interacting with the bases at i-8 till i-10 and thus forming a sterical barrier to keep the RNA apart and guide it along the exit path. F252 of Rpb1 splits the DNA-RNA hybrid at i-10 contacting the DNA base with the plane of the aromatic side chain. Additionally F264 might play a role here by contacting i-10/11. The rudder prevents the reassociation of the hybrid by interaction of residues S318, R320 with the phosphate at the sugar backbone of the DNA at i-10. It also has a stabilizing effect on the unwound DNA. Additionally E1403, 1404 and 1407 repel the DNA strand, while R326, L330 and R331 pull the template strand away from the duplex axis. The fork loop1 makes contact with the RNA at i-6/7/8. Rpb2 residues L471 and R476 are in close proximity to the RNA phosphate group, preventing the unwinding of the hybrid past i-8. Another part of fork loop 2 sterically blocks the reassociation of the DNA double helix by interfering with the nontemplate at i+3. The three structural regions interact with each other, the lid with the rudder and the rudder with the fork loop. This mechanism seems to be independent of the sequence of the DNA template, since crystal structures with different DNA templates suggest the same way of functioning (Kettenberger et al., 2003; Westover et al., 2004a). Taking into account the important roles of the four structural elements in the transcription process it might be somewhat surprising that the sequence conservation between the polymerases is rather low.

The role of the Rpb4/7 heterodimer

Rpb4 and Rpb7 form a heterodimer that easily dissociates from the Pol II in vitro. In the crystal structure of Pol II this heterodimeric sub complex is located at a strategic region of the enzyme (see Figure 2 and 4). It lies near the transcript exit groove and in close proximity to the linker of the CTD of Rpb1 (Armache et al., 2003). There are two exit paths suggested for RNA out of the Pol II. One goes around the side of the dock domain and towards Rpb4/7 and the second one curves around Rpb8. In the bacterial Pol only the latter mechanism seems possible (Westover et al., 2004b). New results based on cross linking of nucleic acids to the Pol II show that in the human system the RNA exits via the first path (Ujvari and Luse, 2006). This study also showed the capability of Rpb4/7 to bind nucleic acids.

The interaction of Rpb4/7 to the Pol II core structure is made via two contact areas, one at Rpb4 and one at Rpb7. The more important one is the connection of the conserved N-terminal tip of Rpb7 with three regions of Rpb1, one region of Rpb2 and a region of Rpb6. Mutations in glutamine 100 of Rpb6 (this lies in a conserved helix region of the Rpb6 subunit) have a major impact on the binding stability of the heterodimer to the rest of the polymerase (Armache et al., 2003; Bushnell and Kornberg, 2003). This point of interaction does not require the presence of Rpb4 (Sheffer et al., 1999). The N-terminal region of Rpb4 makes the second contact with the N-terminal region of Rpb1. Except for these two contacts with the ten subunit Pol II core the Rpb4/7 heterodimer faces away from the overall complex. This makes it accessible for other factors to interact. Interactions have been reported for the CTD dephosphorylating enzymes Fcp1 and Ess1, which associate with the CTD of

Rpb1 and hereby influences the transcription process (Kamenski et al., 2004; Kimura et al., 2002; Wu et al., 2003). The presence of Rpb4/7 leads to a closed clamp and thus a closed cleft (Gnatt et al., 2001). The two proteins are not essential for the conformational change itself; they lock the clamp in place during transcription. The closed clamp probably hinders the double stranded promoter DNA to enter the cleft so the DNA only enters the active center of Pol II after the melting process. The complete twelve subunit Pol II has a higher affinity to the TBP TFIIB promotor complex than the Pol II core alone. In the (PIC) the heterodimer is in close proximity to TFIIB and interacts with TFIIF (Bushnell and Kornberg, 2003; Chung et al., 2003). Rpb4/7 are required for promoter dependent transcription initiation (Edwards et al., 1991; Orlicky et al., 2001). They are not required though for a stable association of Pol II to the PIC (Orlicky et al., 2001) or for the elongation process.

The functional heterodimer likely evolved before archea and eukaryotes separated. One indication for this are the archeal homologues RpoE and RpoF. There are homologues of the heterodimeric complex also in the other yeast polymerases. In Pol I the counterpart of Rpb4/7 interacts with Rrn3, a general initiation factor (Peyroche et al., 2002; Yuan et al., 2002) and there is also a homologous complex in Pol III. The homologue of Rpb4 Rpc17 interacts with Brf1, the complex that fulfills the TFIIB function for Pol III (Ferri et al., 2000).

The sequence, structure and function of Rpb7 is highly conserved in several organisms (40 70 % structural homology). Deletion of rpb7 is lethal, but Rpb7 homologues from S. pombe, C. albicans, D. melanogaster and H. sapiens can replace the original protein in S. serevisiae (Khazak et al., 1995; Singh et al., 2004; Zhou and Lee, 2001). Especially the RNA binding folds are highly conserved and Rpb7 is essential for cell viability under any conditions (McKune et al., 1993; Mitsuzawa et al., 2003). In contrast to the high conservation throughout all species, Rpb4 is the least conserved protein in the Pol II complex. Some of the homologues are much longer than Rpb4 itself with extensions at N and C-terminus (Choder, 2004). This might hint at a role of Rpb4 in the interaction and recruitment of other proteins to the complex. While Rpb7 is essential, Rpb4 is dispensable under optimal growth conditions, but essential under certain stress conditions (Choder and Young, 1993; Farago et al., 2003; Khazak et al., 1995; Miyao et al., 2001; Sheffer et al., 1999; Tan et al., 2000; Woychik and Young, 1989). Rpb4 is present in an excess over all other Pol II subunits (Rosenheck and Choder, 1998) which might hint at some independent function. While all Rpb7 homologues have the ability to bind RNA, not all Rpb4 homologues are able to. It seems that Rpb4 and 7 do not contribute equally to the function of the heterodimer, but that Rpb7 plays a key part and Rpb4 enables, enhances or modifies its functions. Additionally it was shown that Rpb4 also plays a role in

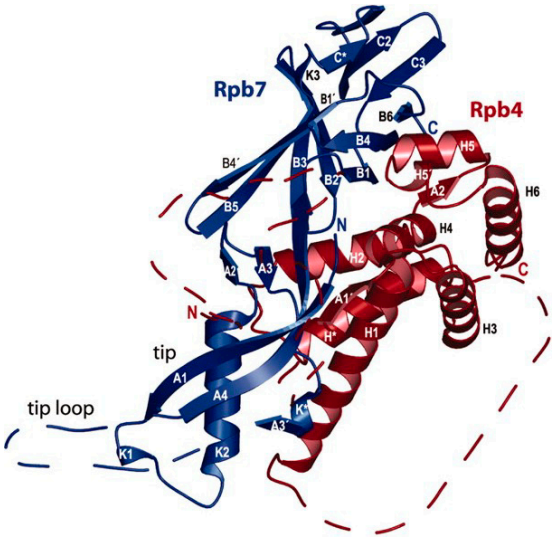

Figure 4 Ribbon diagram of yeast Rpb4/7. Rpb7 is in blue, and Rpb4 in red. Disordered in the Rpb4/7 structure are the Rpb7 tip loop (residues 57–68) and Rpb4 residues 35–46, 77–81, and 101–118. Rpb4 residues 1–34 and 82–100 were not present in the crystallized variant. Adapted from (Armache et al., 2005).

mRNA export and in the mRNA decay pathway (Farago et al., 2003; Lotan et al., 2005). The mRNA export can happen via several distinct pathways one of those being a stress dependent one. The only two proteins which were shown to be essential for mRNA export under stress are Nup42 and Rpb4. Rpb4 and 7 have also been shown to have additional specific functions in regulating transcription coupled repair (Li and Smerdon, 2002).

In proliferating yeast cells only 20 % of the Pol II contains the heterodimer (Choder and Young, 1993), while in starving cells it binds in a stoichiometric manner. Neither of the two subunits seems to be phosphorylated, but modifications on the Pol II core under starving conditions might be responsible for the attachment of the sub complex (Rosenheck and Choder, 1998).

The two subunits are able to shuttle out into the cytoplasm and back into the nucleus (Selitrennik et al., 2006). The kinetics of the process are similar for both, Rpb4 and 7. This suggests that they might shuttle as a heterodimer. There seem to be two distinct shuttling mechanisms that are generally dependent on transcription. The exception to this is severe heat stress where both subunits can shuttle independent from the transcription process. While the heterodimer seems to leave the nucleus the ten subunit core of the Pol II stays there under all conditions (Farago et al., 2003).

Transcriptional arrest and DNA repair

During the elongation process (the actual transcription of the single stranded DNA into mRNA) the polymerase can encounter sequence parts that will cause reverse movement of the enzyme (so called backtracking). This results in 3´RNA entering the above described pore and lead to arrest of the

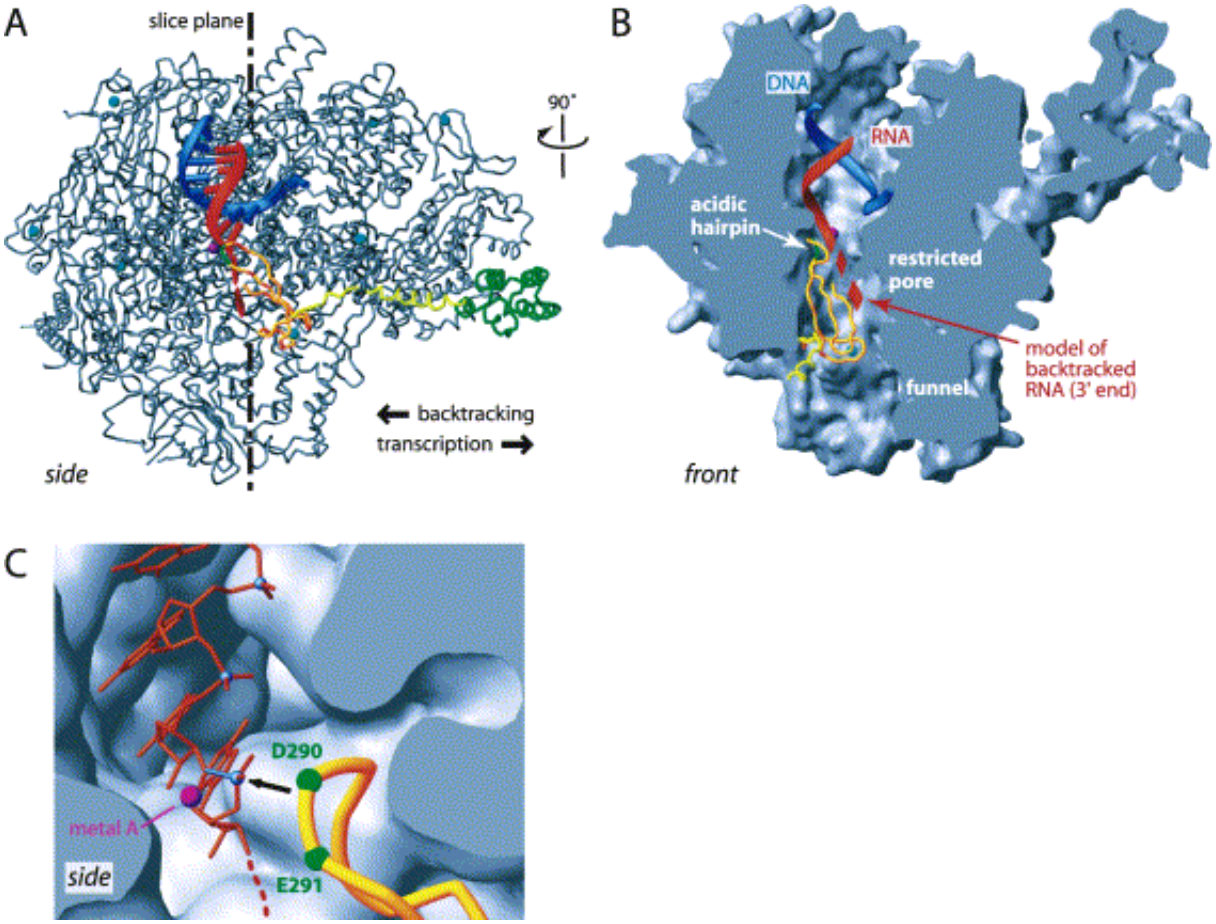

Figure 5 (A) Model of a Pol II-TFIIS-Nucleic Acid Complex. The DNA template strand (blue) and the RNA transcript (red) are displayed. The presumed location of backtracked RNA is indicated as a dashed red ribbon. The arrows indicate movement of Pol II relative to the nucleic acids. (B) Cut-away view of the model in (A) from the front. TFIIS and nucleic acids are shown as ribbon models on the molecular surface of Pol II. The coloring code is the same as in (A). The backtracked portion of RNA would be cut at the active site during TFIIS-induced RNA cleavage. (C) Proximity of the TFIIS acidic hairpin to the potential scissile RNA phosphodiester bond. The view is as in (A). RNA was placed according to the location in the Pol II elongation complex structure (Gnatt et al., 2001), and is shown as a stick model with phosphorous atoms highlighted as blue spheres. The black arrow indicates the direction of a possible cut of the scissile bond (blue).

transcription process. Before transcription can continue the RNA needs to be cleaved. TFIIS is necessary for this task (Fish and Kane, 2002). TFIIS consists out of three domains. A weakly conserved N-terminal domain (I), consisting out of a four helix bundle a central domain (II) formed by a three helix bundle, and a C-terminal domain (III), three antiparallel ß sheets that form a zinc ribbon. While I seems not to be essential for transcription elongation, domain II and III are required for binding to Pol II. II binds to the outer jaw, extends over the funnel and III binds to the pore. The hairpin of II sits in the active center and has two acidic residues that approach the RNA backbone near metal A and could bind metal B adjacently (Kettenberger et al., 2003) (Figure 5). The zinc ribbon of III is essential for the cleavage of the RNA. The metal B in the active center of Pol II can activate a water molecule for a nucleophilic attack on a phosphorous atom, for example the easily cleavable RNA phosphodiester bond. This suggests a mechanism which is similar to the well known function of the Klenov DNA polymerase (Beese and Steitz, 1991; Joyce and Steitz, 1994; Sosunov et al., 2003). The group of Kornberg (Westover et al., 2004a) is suggesting that the Rpb1/5 interface might work as a binding pocket for the backtracking RNA. The distance between the active center and the interface of the two subunits is about nine bases, which is the most common size of RNA cleavage products generated by TFIIS.

Pol III

Structural overview of Pol III

Pol III is the largest and most complex of the Polymerases (Chedin et al., 1998a; Geiduschek and Kassavetis, 2001; Schramm and Hernandez, 2002). It consists out of 17 subunits with a mass of nearly 700 kDa. Pol III has five additional subunits compared to Pol II (C82, 53, 37, 34 and 31) (Table 1). As mentioned above five subunits are the same as in Pol II, namely (Rpb5/6/8/10 and 12). Those are common among all three polymerases while the remaining seven sub-

units share some homology in sequence and function. C18 and 27 were suggested to be a counterpart of the Rpb4/7 heterodimer in Pol II (Hu et al., 2002; Sadhale and Woychik, 1994; Siaut et al., 2003). All genes but the one encoding C37 have been shown to be essential (Chedin et al., 1998a). C160, C128, AC40, AC19 and ABC23 are evolutionally related to the *E. coli* polymerase core subunits.

Yeast two hybrid screening established an initial protein-protein interaction map of Pol III (Flores et al., 1999) (Figure 6). Nine subunits define a structural core, evolutionarily related to the other nuclear RNA polymerases. While the sequence identity of the shared and homologous subunits of Pol II and Pol III lies at only 39.4 %, the structural similarity between the two enzymes is about 83.4 % (Jasiak et al., 2006). The only two domain folds that are not the same in Pol II are the Rpb1 jaw domain and the Rpb2 external domain 1 (Cramer et al., 2001) indicating that most differences occur on the surface of the proteins. Large conserved surface areas in the elongation complexes are only found in the cleft, around the binding site for the triphosphate substrate, the DNA-RNA hybrid and the RNA exit tunnel (Gnatt et al., 2001; Jasiak et al., 2006; Kettenberger et al., 2004; Westover et al., 2004a). This underlines the conservation of the basic transcription mechanism among the different polymerases.

Solving the structure of Pol II and finding that the structural conservation between Pol II and the bacterial polymerase is rather high, gave rise to a comparison between the Pol II structural core being a basis for modeling the structure and composition of Pol III. There is no X-ray structure available of Pol III, but there has been a publication of a cryo electron microscopy structure. Fernandez Tornero et. all report that the Pol III has an overall size of 170 x 150 x 130 $Å^3$ (Fernandez-Tornero et al., 2007). They fitted the known Pol II structure of the nine core subunits (Armache et al., 2005; Cramer et al., 2001) into the cryo EM pictures and found that the main structural differences lie at the downstream end where the clamp, lobe and jaw region show additional densities and extend the downstream region of the enzyme for about another 30 Å. The total volume those additional densities take up is about 190 kDa which approximately make up for the weight of the 5 additional subunits in the Pol III compared to Pol II. Antibody labeling of C82 and C34 combined with negative staining EM revealed the position of the Pol III specific sub complex next to the clamp region. This is in agreement with previous mutagenesis studies which showed mutations on the C160 to affect the binding of the C82/34/31 sub complex (Werner et al., 1992). Furthermore they suggest that the C53/37 sub complex lies next to the subunit C11 in the electron density map at the outer end of the DNA binding cleft. This would allow the sub complex to play a role in sensing the incoming DNA stand (Landrieux et al., 2006).

Besides these functional studies there has been work going on to find out more about the composition and protein protein interactions of Pol III. These approaches mainly involved

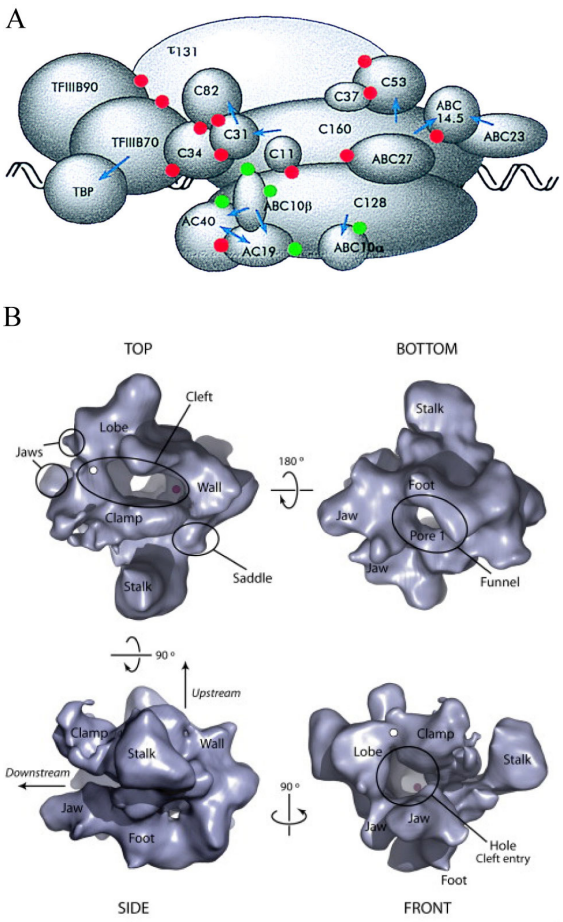

Figure 6 (A) model of the Pol III transcription initiation complex. Protein-protein contacts observed by using the two-hybrid system are indicated by dark dots. Light dots indicate the interaction between AC40, AC19, ABC10α, and ABC10β with A190 and A135 Pol I subunits homologous to C160 and C128. Genetic interactions observed by using multicopy suppression experiments of thermosensitive mutations are indicated by arrows. The arrowhead points toward the subunit harboring the mutation that was suppressed. For the sake of simplicity, only the τ131 and τ138 subunits of TFIIIC are represented.(Flores et al., 1999) (B) Cryo-EM Structure of Pol III (A) Surface representation of the Pol III EM density after reconstruction. DNA binding cleft, jaws, clamp, wall, lobe, stalk, saddle, foot, funnel, and NTP entry pore (pore 1) initially identified for Pol II (Cramer et al., 2001) are indicated on the different views. The presumed position of the active center is marked with a magenta dot. Straight arrows indicate the putative direction of upstream and downstream DNA on the side view (Adapted from (Fernandez-Tornero et al., 2007)).

yeast two hybrid screening and over expression suppressor analysis (Chedin et al., 1998a; Ferri et al., 2000; Flores et al., 1991). Additionally, there have been several approaches to solve the problem of protein interaction within Pol III by cross linking experiments (Bartholomew et al., 1994; Bartholomew et al., 1993; Lannutti et al., 1996; Persinger and Bartholomew, 1996; Tate et al., 1998). There is evidence that C34 in the possible C82, 34, 31 subassembly cross links to DNA at the furthest upstream end of the transcription initiation complex. C31 interaction with C160 was shown by

multi copy suppression analysis. Yeast two hybrid experiments indicated that C31, 17 and 11 may interact with each other. The interaction of C82 and 31 has been shown to be localized just upstream of the transcription start site in close proximity of the non transcribed strand. C82 probably occupies a relatively large space on the surface of Pol III. The suggested sub complex of C82, 34 and 31 lies on the C160 side of the enzyme. C37 has been suggested to interact with C53 as well as C160.

C19/25, the Rpb4/7 like sub complex

C19 and C25 have been suggested to share a similar function as the homologous counterparts Rpb4 and 7 in Pol II. Jasiak et al. reported the crystal structure of the C19/25 sub complex and made a comparison with its Pol II counterpart (Jasiak et al., 2006) (Figure 7). The overall domain folds are similar between the two sub complexes. The interaction area between the two subunits of the heterodimeric complexes is conserved among different species but not among the different polymerases. The connection of the sub complex to the core is made by residues that are conserved between Rpb7 and C25 (P15, F18 and G64). C17, as Rpb7 in Pol II, is essential for the viability in *S. cerevisiae*, in contrast to its counterpart in Pol II Rpb4. C17 interacts with Brf1 and C31 (Ferri et al., 2000). Gel shift assays show that the isolated sub complex of C17/25 binds to single stranded RNAs, while there was no binding observed for double stranded DNA. The most specific binding was observed for tRNA (μM range).This indicates that this heterodimer is involved in transcription initiation.

Preinitiation complexes (PIC)

Functionally Pol III is mainly responsible for the transcription of "housekeeping proteins" and thus does not require heavy transcriptional regulation as Pol II, since those genes are transcribed most of the time. However, there are three different types of promoters that Pol III assembles to. They consist out of three recognition sequence elements called A, B and C box. These can appear as type 1 (A box, intermediate element and C box), type 2 (A and B box) or type 3 (TATA box and A box). A and B boxes are recognized by TFIIIC or C2 followed by a recruitment of TFIIIB with TBP and two other proteins, after this recuitment Pol III will bind to that PIC. Type 3 promoters contain an additional proximal sequence element (PSE) that is recognized by its specific binding protein (PSB) and transcription factor (PTF). The TATA box is recognized by a part of TFIIIB which is the key factor in the transcription by Pol III (Schramm and Hernandez, 2002). Additionally, there is a third factor, TFIIIA which is specific for 5s ribosomal RNA encoding genes. TBP is required for the transcription process while inactivation of TBP showed defects on transcription for all three polymerases (Cormack and Struhl, 1992; Schultz et al., 1992). Contrary to Pol II where TFIIH needs ATP to form the transcription bubble no ATP is necessary for this process in Pol III.

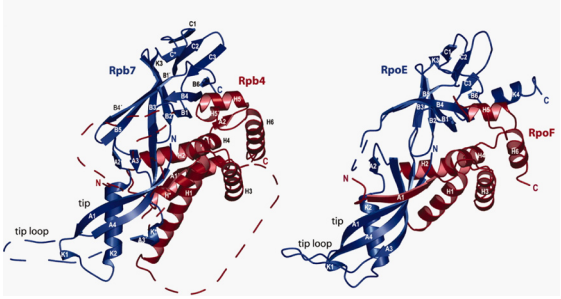

Figure 7 Comparison between the structure of Pol II Rpb4/7 and Pol III C17/25 C25/Rpb7 are in blue and C17/Rpb4 are in red, with the HRDC domain in light red. Disordered in the C17/25 structure are the C25 loop B4-B5 (residues K59–K90), the C17 loop H*-H2 (residues K38–N47), and the C17 linker between the tip-associated domain and the HRDC domain (residues N69–G94). Adapted from (Jasiak et al., 2006)

Ten essential genes were shown to encode for the transcription factors. TFIIIB has three subunits and was shown to bind and melt DNA in the Pol III PIC (Kassavetis et al., 1990; Kassavetis et al., 1992). It can either bind directly to the promoter and Pol III or is mediated by TFIIIF as mentioned above. The interaction point with Pol III is C34 (Brun et al., 1997). It has also been show to interact with TBP. The interaction with the DNA and subunit C34 is surprising because the interfaces lie at opposite sides of the Pol III structure. TFIIIC consists out of six subunits and binds to the dsDNA at the promoter boxes. It also can bind the Pol III subunit C53 and ABC10α (Dumay et al., 1999; Flores et al., 1999). The specific subunits of Pol III, namely C82, 34 and 31 are required for transcription initiation and can be dissociated from the core (Wang and Roeder, 1997; Werner et al., 1993). It was implicated that C11 has a (indirect) role in nascent RNA hydrolysis coupled to retraction of the enzyme along the DNA template (Kassavetis et al., 1990). C11 shares homologies with the subunit Rpb 9 of Pol II. The second largest subunit of Pol III, C128, has been shown to have a direct effect on the RNA chain retraction (Bobkova et al., 1999). The lack of subunit C11 leads to incapability of Pol III to repair nucleotide miss incorporation (Chedin et al., 1998b) and additionally decreases the termination efficiency strongly. This subunit shows functional similarity to TFIIS for it possesses a zinc ribbon domain with two acidic residues that are essential for viability (Chedin et al., 1998a).

REFERENCES

Armache, K.J., Kettenberger, H., and Cramer, P. (2003). Architecture of initiation-competent 12-subunit RNA polymerase II. Proceedings Of The National Academy Of Sciences Of The United States Of America 100, 6964-6968.

Armache, K.J., Mitterweger, S., Meinhart, A., and Cramer, P. (2005). Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. J Biol Chem 280, 7131-7134.

Bartholomew, B., Braun, B.R., Kassavetis, G.A., and Geiduschek, E.P. (1994). Probing close DNA contacts of RNA polymerase III transcription complexes with the photoactive nucleoside 4-thiodeoxythymidine. J Biol Chem 269, 18090-18095.

Bartholomew, B., Durkovich, D., Kassavetis, G.A., and Geiduschek, E.P. (1993). Orientation and topography of RNA polymerase III in transcription complexes. Mol Cell Biol 13, 942-952.

Beese, L.S., and Steitz, T.A. (1991). Structural basis for the 3'-5' exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism. EMBO J 10, 25-33.

Bell, S.D., and Jackson, S.P. (1998). Transcription in Archaea. Cold Spring Harb Symp Quant Biol 63, 41-51.

Benga, W.J., Grandemange, S., Shpakovski, G.V., Shematorova, E.K., Kedinger, C., and Vigneron, M. (2005). Distinct regions of RPB11 are required for heterodimerization with RPB3 in human and yeast RNA polymerase II. Nucleic Acids Res 33, 3582-3590.

Bleichenbacher, M., Tan, S., and Richmond, T.J. (2003). Novel interactions between the components of human and yeast TFIIA/TBP/DNA complexes. J Mol Biol 332, 783-793.

Bobkova, E.V., Habib, N., Alexander, G., and Hall, B.D. (1999). Mutational analysis of the hydrolytic activity of yeast RNA polymerase III. J Biol Chem 274, 21342-21348.

Brun, I., Sentenac, A., and Werner, M. (1997). Dual role of the C34 subunit of RNA polymerase III in transcription initiation. EMBO J 16, 5730-5741.

Buratowski, S., and Zhou, H. (1993). Functional domains of transcription factor TFIIB. Proc Natl Acad Sci U S A 90, 5633-5637.

Bushnell, D.A., and Kornberg, R.D. (2003). Complete, 12-subunit RNA polymerase II at 4.1-angstrom resolution: Implications for the initiation of transcription. Proceedings Of The National Academy Of Sciences Of The United States Of America 100, 6969-6973.

Bushnell, D.A., Westover, K.D., Davis, R.E., and Kornberg, R.D. (2004). Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Angstroms. Science 303, 983-988.

Chedin, S., Ferri, M.L., Peyroche, G., Andrau, J.C., Jourdain, S., Lefebvre, O., Werner, M., Carles, C., and Sentenac, A. (1998a). The yeast RNA polymerase III transcription machinery: A paradigm for eukaryotic gene activation. Cold Spring Harbor Symposia On Quantitative Biology 63, 381-389.

Chedin, S., Riva, M., Schultz, P., Sentenac, A., and Carles, C. (1998b). The RNA cleavage activity of RNA polymerase III is mediated by an essential TFIIS-like subunit and is important for transcription termination. Genes Dev 12, 3857-3871.

Chen, H.T., and Hahn, S. (2003). Binding of TFIIB to RNA polymerase II: Mapping the binding site for the TFIIB zinc ribbon domain within the preinitiation complex. Mol Cell 12, 437-447.

Chen, H.T., and Hahn, S. (2004). Mapping the location of TFIIB within the RNA polymerase II transcription preinitiation complex: a model for the structure of the PIC. Cell 119, 169-180.

Chen, H.T., Warfield, L., and Hahn, S. (2007). The positions of TFIIF and TFIIE in the RNA polymerase II transcription preinitiation complex. Nat Struct Mol Biol 14, 696-703.

Choder, M. (2004). Rpb4 and Rpb7: subunits of RNA polymerase II and beyond. Trends Biochem Sci 29, 674-681.

Choder, M., and Young, R.A. (1993). A portion of RNA polymerase II molecules has a component essential for stress responses and stress survival. Mol Cell Biol 13, 6984-6991.

Chung, W.H., Craighead, J.L., Chang, W.H., Ezeokonkwo, C., Bareket-Samish, A., Kornberg, R.D., and Asturias, F.J. (2003). RNA polymerase II/TFIIF structure and conserved organization of the initiation complex. Mol Cell 12, 1003-1013.

Cormack, B.P., and Struhl, K. (1992). The TATA-binding protein is required for transcription by all three nuclear RNA polymerases in yeast cells. Cell 69, 685-696.

Cramer, P. (2004). RNA polymerase II structure: from core to functional complexes. Curr Opin Genet Dev 14, 218-226.

Cramer, P., Bushnell, D.A., Fu, J., Gnatt, A.L., Maier-Davis, B., Thompson, N.E., Burgess, R.R., Edwards, A.M., David, P.R., and Kornberg, R.D. (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. Science 288, 640-649.

Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. Science 292, 1863-1876.

Davis, J.A., Takagi, Y., Kornberg, R.D., and Asturias, F.A. (2002). Structure of the yeast RNA polymerase II holoenzyme: Mediator conformation and polymerase interaction. Mol Cell 10, 409-415.

Dumay, H., Rubbi, L., Sentenac, A., and Marck, C. (1999). Interaction between yeast RNA polymerase III and transcription factor TFIIIC via ABC10alpha and tau131 subunits. J Biol Chem 274, 33462-33468.

Edwards, A.M., Kane, C.M., Young, R.A., and Kornberg, R.D. (1991). Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro. J Biol Chem 266, 71-75.

Farago, M., Nahari, T., Hammel, C., Cole, C.N., and Choder, M. (2003). Rpb4p, a subunit of RNA polymerase II, mediates mRNA export during stress. Mol Biol Cell 14, 2744-2755.

Fernandez-Tornero, C., Bottcher, B., Riva, M., Carles, C., Steuerwald, U., Ruigrok, R.W.H., Sentenac, A., Muller, C.W., and Schoehn, G. (2007). Insights into Transcription Initiation and Termination from the Electron Microscopy Structure of Yeast RNA Polymerase III. Molecular Cell 25, 813.

Ferri, M.-L., Peyroche, G., Siaut, M., Lefebvre, O., Carles, C., Conesa, C., and Sentenac, A. (2000). A Novel Subunit of Yeast RNA Polymerase III Interacts with the TFIIB-Related Domain of TFIIIB70. Mol. Cell. Biol. 20, 488-495.

Fish, R.N., and Kane, C.M. (2002). Promoting elongation with transcript cleavage stimulatory factors. Biochim Biophys Acta 1577, 287-307.

Flores, A., Briand, J.F., Gadal, O., Andrau, J.C., Rubbi, L., Van Mullem, V., Boschiero, C., Goussot, M., Marck, C., Carles, C., et al. (1999). A protein-protein interaction map of yeast RNA polymerase III. Proc Natl Acad Sci U S A 96, 7815-7820.

Flores, O., Lu, H., Killeen, M., Greenblatt, J., Burton, Z.F., and Reinberg, D. (1991). The small subunit of transcription factor IIF recruits RNA polymerase II into the preinitiation complex. Proc Natl Acad Sci U S A 88, 9999-10003.

Fu, J., Gnatt, A.L., Bushnell, D.A., Jensen, G.J., Thompson, N.E., Burgess, R.R., David, P.R., and Kornberg, R.D. (1999). Yeast RNA polymerase II at 5 A resolution. Cell 98, 799-810.

Furter-Graves, E.M., Hall, B.D., and Furter, R. (1994). Role of a small RNA pol II subunit in TATA to transcription start site spacing. Nucleic Acids Res 22, 4932-4936.

Geiduschek, E.P., and Kassavetis, G.A. (2001). The RNA polymerase III transcription apparatus. Journal Of Molecular Biology 310, 1-26.

Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution. Science 292, 1876-1882.

Hisatake, K., Roeder, R.G., and Horikoshi, M. (1993). Functional dissection of TFIIB domains required for TFIIB-TFIID-promoter complex formation and basal transcription activity. Nature 363, 744-747.

Hu, P., Wu, S., Sun, Y.L., Yuan, C.C., Kobayashi, R., Myers, M.P., and Hernandez, N. (2002). Characterization of human RNA polymerase III identifies orthologues for Saccharomyces cerevisiae RNA polymerase III subunits. Molecular And Cellular Biology 22, 8044-8055.

Jasiak, A.J., Armache, K.J., Martens, B., Jansen, R.P., and Cramer, P. (2006). Structural biology of RNA polymerase III: Subcomplex C17/25 X-ray structure and 11 subunit enzyme model. Molecular Cell 23, 71-81.

Joyce, C.M., and Steitz, T.A. (1994). Function and structure relationships in DNA polymerases. Annu Rev Biochem 63, 777-822.

Kamenski, T., Heilmeier, S., Meinhart, A., and Cramer, P. (2004). Structure and mechanism of RNA polymerase II CTD phosphatases. Mol Cell 15, 399-407.

Kassavetis, G.A., Braun, B.R., Nguyen, L.H., and Geiduschek, E.P. (1990). S-Cerevisiae Tfiiib Is The Transcription Initiation-Factor Proper Of Rna Polymerase-Iii, While Tfiiia And Tfiiic Are Assembly Factors. Cell 60, 235-245.

Kassavetis, G.A., Joazeiro, C.A., Pisano, M., Geiduschek, E.P., Colbert, T., Hahn, S., and Blanco, J.A. (1992). The role of the TATA-binding protein in the assembly and function of the multisubunit yeast RNA polymerase III transcription factor, TFIIIB. Cell 71, 1055-1064.

Kettenberger, H., Armache, K.J., and Cramer, P. (2003). Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. Cell 114, 347-357.

Kettenberger, H., Armache, K.J., and Cramer, P. (2004). Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. Mol Cell 16, 955-965.

Khazak, V., Sadhale, P.P., Woychik, N.A., Brent, R., and Golemis, E.A. (1995). Human RNA polymerase II subunit hsRPB7 functions in yeast and influences stress survival and cell morphology. Mol Biol Cell 6, 759-775.

Kimura, M., Suzuki, H., and Ishihama, A. (2002). Formation of a carboxy-terminal domain phosphatase (Fcp1)/TFIIF/RNA polymerase II (pol II) complex in Schizosaccharomyces pombe involves direct interaction between Fcp1 and the Rpb4 subunit of pol II. Mol Cell Biol 22, 1577-1588.

Kolodziej, P.A., Woychik, N., Liao, S.M., and Young, R.A. (1990). RNA polymerase II subunit composition, stoichiometry, and phosphorylation. Mol Cell Biol 10, 1915-1920.

Landrieux, E., Alic, N., Ducrot, C., Acker, J., Riva, M., and Carles, C. (2006). A subcomplex of RNA polymerase III subunits involved in transcription termination and reinitiation. Embo Journal 25, 118-128.

Lannutti, B.J., Persinger, J., and Bartholomew, B. (1996). Probing the protein-DNA contacts of a yeast RNA polymerase III transcription complex in a crude extract: solid phase synthesis of DNA photoaffinity probes containing a novel photoreactive deoxycytidine analog. Biochemistry 35, 9821-9831.

Lee, T.I., and Young, R.A. (2000). Transcription of eukaryotic protein-coding genes. Annu Rev Genet 34, 77-137.

Levine, M., and Tjian, R. (2003). Transcription regulation and animal diversity. Nature 424, 147-151.

Li, S., and Smerdon, M.J. (2002). Rpb4 and Rpb9 mediate subpathways of transcription-coupled DNA repair in Saccharomyces cerevisiae. Embo J 21, 5921-5929.

Lotan, R., Bar-On, V.G., Harel-Sharvit, L., Duek, L., Melamed, D., and Choder, M. (2005). The RNA polymerase II subunit Rpb4p mediates decay of a specific class of mRNAs. Genes Dev 19, 3004-3016.

McKune, K., Richards, K.L., Edwards, A.M., Young, R.A., and Woychik, N.A. (1993). RPB7, one of two dissociable subunits of yeast RNA polymerase II, is essential for cell viability. Yeast 9, 295-299.

Meinhart, A., Kamenski, T., Hoeppner, S., Baumli, S., and Cramer, P. (2005). A structural perspective of CTD function. Genes Dev 19, 1401-1415.

Mitsuzawa, H., Kanda, E., and Ishihama, A. (2003). Rpb7 subunit of RNA polymerase II interacts with an RNA-binding protein involved in processing of transcripts. Nucleic Acids Res 31, 4696-4701.

Miyao, T., Barnett, J.D., and Woychik, N.A. (2001). Deletion of the RNA polymerase subunit RPB4 acts as a global, not stress-specific, shut-off switch for RNA polymerase II transcription at high temperatures. J Biol Chem 276, 46408-46413.

Miyao, T., and Woychik, N.A. (1998). RNA polymerase subunit RPB5 plays a role in transcriptional activation. Proc Natl Acad Sci U S A 95, 15281-15286.

Orlicky, S.M., Tran, P.T., Sayre, M.H., and Edwards, A.M. (2001). Dissociable Rpb4-Rpb7 subassembly of rna polymerase II binds to single-strand nucleic acid and mediates a post-recruitment step in transcription initiation. J Biol Chem 276, 10097-10102.

Persinger, J., and Bartholomew, B. (1996). Mapping the contacts of yeast TFIIIB and RNA polymerase III at various distances from the major groove of DNA by DNA photoaffinity labeling. J Biol Chem 271, 33039-33046.

Peyroche, G., Levillain, E., Siaut, M., Callebaut, I., Schultz, P., Sentenac, A., Riva, M., and Carles, C. (2002). The A14-A43 heterodimer subunit in yeast RNA pol I and their relationship to Rpb4-Rpb7 pol II subunits. Proc Natl Acad Sci U S A 99, 14670-14675.

Pugh, B.F. (2000). Control of gene expression through regulation of the TATA-binding protein. Gene 255, 1-14.

Rosenheck, S., and Choder, M. (1998). Rpb4, a Subunit of RNA Polymerase II, Enables the Enzyme To Transcribe at Temperature Extremes In Vitro. J. Bacteriol. 180, 6187-6192.

Sadhale, P.P., and Woychik, N.A. (1994). C25, an Essential Rna-Polymerase-III Subunit Related to the Rna-Polymerase-Ii Subunit Rpb7. Molecular and Cellular Biology 14, 6164-6170.

Schramm, L., and Hernandez, N. (2002). Recruitment of RNA polymerase III to its target promoters. Genes & Development 16, 2593-2620.

Schultz, M.C., Reeder, R.H., and Hahn, S. (1992). Variants of the TATA-binding protein can distinguish subsets of RNA polymerase I, II, and III promoters. Cell 69, 697-702.

Selitrennik, M., Duek, L., Lotan, R., and Choder, M. (2006). Nucleocytoplasmic shuttling of the Rpb4p and Rpb7p subunits of Saccharomyces cerevisiae RNA polymerase II by two pathways. Eukaryot Cell 5, 2092-2103.

Sheffer, A., Varon, M., and Choder, M. (1999). Rpb7 can interact with RNA polymerase II and support transcription during some stresses independently of Rpb4. Mol Cell Biol 19, 2672-2680.

Siaut, M., Zaros, C., Levivier, E., Ferri, M.L., Court, M., Werner, M., Callebaut, I., Thuriaux, P., Sentenac, A., and Conesa, C. (2003). An Rpb4/Rpb7-like complex in yeast RNA polymerase III contains the orthologue of mammalian CGRP-RCP. Mol Cell Biol 23, 195-205.

Singh, S.R., Rekha, N., Pillai, B., Singh, V., Naorem, A., Sampath, V., Srinivasan, N., and Sadhale, P.P. (2004). Domainal organization of the lower eukaryotic homologs of the yeast RNA polymerase II core subunit Rpb7 reflects functional conservation. Nucleic Acids Res 32, 201-210.

Sosunov, V., Sosunova, E., Mustaev, A., Bass, I., Nikiforov, V., and Goldfarb, A. (2003). Unified two-metal mechanism of RNA synthesis and degradation by RNA polymerase. EMBO J 22, 2234-2244.

Steinmetz, E.J., Ng, S.B., Cloute, J.P., and Brow, D.A. (2006). cis- and trans-Acting determinants of transcription termination by yeast RNA polymerase II. Mol Cell Biol 26, 2688-2696.

Tan, Q., Li, X., Sadhale, P.P., Miyao, T., and Woychik, N.A. (2000). Multiple mechanisms of suppression circumvent transcription defects in an RNA polymerase mutant. Mol Cell Biol 20, 8124-8133.

Tate, J.J., Persinger, J., and Bartholomew, B. (1998). Survey of four different photoreactive moieties for DNA photoaffinity labeling of yeast RNA polymerase III transcription complexes. Nucleic Acids Res 26, 1421-1426.

Tyree, C.M., George, C.P., Lira-DeVito, L.M., Wampler, S.L., Dahmus, M.E., Zawel, L., and Kadonaga, J.T. (1993). Identification of a minimal set of proteins that is sufficient for accurate initiation of transcription by RNA polymerase II. Genes Dev 7, 1254-1265.

Ujvari, A., and Luse, D.S. (2006). RNA emerging from the active site of RNA polymerase II interacts with the Rpb7 subunit. Nat Struct Mol Biol 13, 49-54.

Wang, Z.X., and Roeder, R.G. (1997). Three human RNA polymerase III-specific subunits form a subcomplex with a selective function in specific transcription Initiation. Genes & Development 11, 1315-1326.

Werner, M., Chaussivert, N., Willis, I.M., and Sentenac, A. (1993). Interaction between a complex of RNA polymerase III subunits and the 70-kDa component of transcription factor IIIB. J Biol Chem 268, 20721-20724.

Werner, M., Hermannledenmat, S., Treich, I., Sentenac, A., and Thuriaux, P. (1992). Effect Of Mutations In A Zinc-Binding Domain Of Yeast Rna Polymerase-C (Iii) On Enzyme Function And Subunit Association. Molecular And Cellular Biology 12, 1087-1095.

Westover, K.D., Bushnell, D.A., and Kornberg, R.D. (2004a). Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. Cell 119, 481-489.

Westover, K.D., Bushnell, D.A., and Kornberg, R.D. (2004b). Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. Science 303, 1014-1016.

Woychik, N.A., Liao, S.M., Kolodziej, P.A., and Young, R.A. (1990). Subunits shared by eukaryotic nuclear RNA polymerases. Genes Dev 4, 313-323.

Woychik, N.A., and Young, R.A. (1989). RNA polymerase II subunit RPB4 is essential for high- and low-temperature yeast cell growth. Mol Cell Biol 9, 2854-2859.

Wu, X., Rossettini, A., and Hanes, S.D. (2003). The ESS1 Prolyl Isomerase and Its Suppressor BYE1 Interact With RNA Pol II to Inhibit Transcription Elongation in Saccharomyces cerevisiae. Genetics 165, 1687-1702.

Yuan, X., Zhao, J., Zentgraf, H., Hoffmann-Rohrer, U., and Grummt, I. (2002). Multiple interactions between RNA polymerase I, TIF-IA and TAF(I) subunits regulate preinitiation complex assembly at the ribosomal gene promoter. EMBO Rep 3, 1082-1087.

Zhou, H., and Lee, K.A. (2001). An hsRPB4/7-dependent yeast assay for trans-activation by the EWS oncogene. Oncogene 20, 1519-1524.

FROM A,B,C TO X,Y,Z

Kristina Lorenzen
Biomolecular Mass Spectrometry and Proteomics Group, Bijvoet Center for Biomolecular Research and Utrecht Faculty of Science, Utrecht University, Sorbonnelaan 16, 3584 CA Utrecht, the Netherlands

CONTEXT
While native mass spectrometry can provide insights into structural features of a protein complex it provides little information on the protein sequence and possible modifications within. Here peptide mass spectrometry comes into play. It is able to supply information about modifications in the primary protein sequence and thereby complement the information gained by native mass spectrometry especially on the level of modifications influencing the regulation or localization of the protein (complex) of interest. The work presented in this thesis (chapter 5) improves the detection of the modification sites, enabling further insight into RNA polymerase II and III.

Protein identification by Mass Spectrometry

One of the most useful research applications of mass spectrometry is still the field of proteomics, which aims to understand biological processes and how they are controlled. Therefore one has to obtain as much information as possible about the composition, sequence, splice variants and post translational modifications from proteins that are often low abundant within a highly complex mixture like a cell lysate (Mann et al., 2001; Yates, 2004). To obtain this information masses of proteins are not enough, therefore the proteins are proteolytically digested by enzymes at specific sites in their amino acid sequence. One common example for such an enzyme is trypsin, which specifically cuts at the C-terminal side of lysines and arginines hereby creating well defined peptide fragments. These created peptides can then be subsequently be analyzed in a mass spectrometer. The masses of the peptides are specific for the amino acid sequence they contain and thus reflect a part of the sequence of the protein they originate from. Based on the specific cleavage products created by the enzyme of choice each protein will give rise to a typical range of masses (like a protein fingerprint). The mass spectrum of the cleaved protein can then be compared to a database that contains the theoretically derived peptide masses (James et al., 1993; Mann et al., 1993; Pappin et al., 1993). This comparison should then identify the original protein. It is possible that the information gained by peptide mass fingerprinting is not sufficient. This is for example often the case in very complex mixtures when especially small proteins do not provide enough peptides to confidently identify them or spectra become too difficult to interpret. Another possibility is that the protein of interest is not in the database used to compare the spectra with. Additionally, the mass of a peptide is not enough to identify the site of a specific post translational modification like a phosphorylation. In this case the amino acid sequence of the peptide must be obtained. To determine the primary sequence of a peptide, the protonated peptide undergoes CID at low fragmentation energies (like in a Q ToF or iontrap) or at high fragmentation energies (like in a ToF ToF). In low energy collision induced fragmentation, the peptide enters a gas filled collision cell and undergoes multiple collisions with an inert gas (typically He or N). A part of the translational energy from the collision is then transferred into internal kinetic energy of the peptide. If the energy gets high enough, the peptide can then break at three different parts of the amino acid backbone: NH-CH, CH-CO or CO-NH (see Figure 1). The nomenclature widely used today for the annotation of the resulting spectra was introduced by Roepstoff and Fohlman in the early 1980s (Roepstorff and Fohlman, 1984) and altered later by Johnson et al. (Johnson et al., 1987). When one of the bonds breaks, two new fragment species arise. If the peptide was singly charged, then the charge will stay one either of the two fragments and one charged ion and one neutral are formed, since there can be no new charges gained in the process. Only the charged species can then be detected by the mass spectrometer. Consequently there are six possible combinations of fragment ions for each amino acid reside. If the charge stays at the N Terminal part of the peptide the ions are labelled as a, b and c, if the charge is at the C Terminal part of the peptide x, y and z (see Figure 1). The most common fragments detected in CID are b and y ions since the CO NH bond is the weakest and breaks already at low fragmentation energy. The mass difference between two subsequent y ions (or b ions) provides the mass of the amino acid residue at that position.

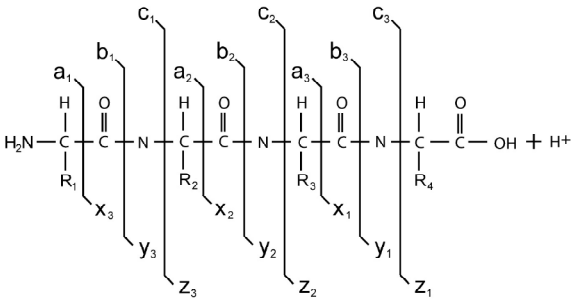

Figure 1 Roepstorff and Fohlmann nomenclature for peptide fragmentation. Fragmentation occurs mainly at the amide bond. A-, b- and c-ions arise from the N-terminus, whereas x-, y- and z-ions originate from the C-terminus.

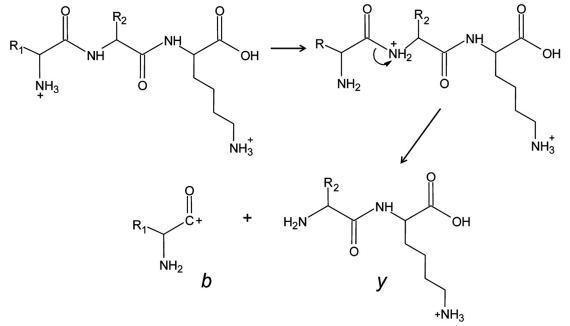

Figure 2 Fragmentation scheme for production of b- and y-type ions by CAD of a multiply protonated peptide. Adapted from (Syka et al., 2004)

Modifications of Proteins determined by Mass Spectrometry

After the transcription and translation process, a protein is not necessarily ready to fulfil its biological function. Chemical modification on amino acid side chains or the N- or C-terminal end of proteins by covalent modification, events like proteolytic cleavage or addition of a modifying group on one or more amino acids can change protein properties. PTMs are needed for proper molecular function, determine the activity, turnover, localization and their interaction with other proteins (Cohen, 2000; Tyers and Jorgensen, 2000). It has been demonstrated that mutations or alterations of PTMs can lead to various diseases (Crane et al., 2004; Liu et al., 2004; Musicki et al., 2005) which underlines their importance.

After determining the contents and interaction networks within cells (Butland et al., 2005; Gavin et al., 2006; Gavin et al., 2002; Ho et al., 2002; Krogan et al., 2006), the new challenge of proteomics now is to determine structural and functional properties of complex protein samples like sub cellular fractions or cells. (Aebersold and Mann, 2003). The perquisite to this is the achievement of high protein sequence coverage, because each missed residue might contain a PTM or information that is important. There have been some approaches to tackle these difficulties (Mann et al., 2002; McLachlin and Chait, 2001; Sickmann and Meyer, 2001) but there are several problems when analysing PTMs. First they are often only present in substoichiometric abundance and a mass spectrometer is only capable to select about 3 - 20 peptides per second and sequence them. Considering the hundreds of thousands peptides that get generated by digestion of a cell lysate it is likely that the low abundant modified peptides will be missed. Therefore enrichment steps are introduced, either biochemically via e.g. immunoprecipitations (Gronborg et al., 2002; Pandey et al., 2000) and gel separation (Soskic et al., 1999; Yamagata et al., 2002). or on the peptide level via fractionation techniques (Liu et al., 2007; Sui et al., 2008; Yen et al., 2006). Specific enrichment of the modification e.g. by metal columns for phosphorylation (Andersson and Porath, 1986; Pinkse et al., 2004; Posewitz and Tempst, 1999; Stensballe et al., 2001) is also an option.

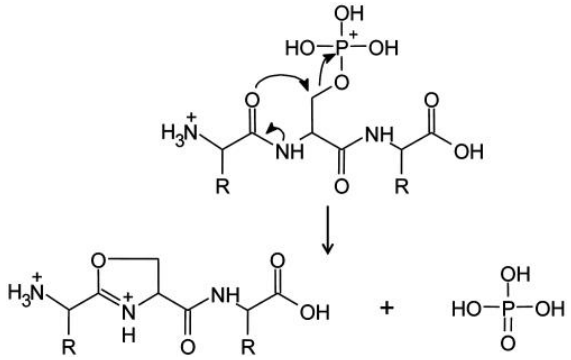

Figure 3 Fragmentation scheme for loss of phosphoric acid from a multiply protonated phosphopeptide by CAD. Adapted from (Syka et al., 2004)

Another possibility is to chemically modify the PTM of interest (Oda et al., 2001; Steen and Mann, 2002; Zhou et al., 2001). It should be considered that modifications can lead to missed cleavages since the enzyme often will not recognize the cleavage site in the protein sequence anymore if a residue has been modified. This can lead to very large ions that can not be detected by most instruments. These missed cleavages will also lead to a higher amount of basic amino acid residues in the peptide, leading to fragmentation only at specific sites of the backbone because random protonation is prevented.

One has to distinguish between modifications that are stable, like acetylation or modification of arginine through methyation and labile ones like phosphorylation or glycosylation. Often the mass spectrum can already hint at the type of modification present on a peptide, since each modification has a specific mass: An additional 42 Da hints at an acetylation or a mass gain of 80 Da at phosphorylation; on the other hand a loss of 98 Da in the instrument is an indication for a phosphorylation of a serine or threonine. Tandem mass spectrometry is used to validate the exact location of the PTM in the sequence. The advantages and drawbacks of the two most commonly used fragmentation methods will be discussed below.

Using electron capture and collision activated dissociation for the detection of phosporylated peptides

As already mentioned the most commonly used fragmentation method is collision activated dissociation (CAD). Here the amide ions of a peptide are kinetically excited through the collisions with an inert gas (He, N2 or Ar). Each collsion is converted into vibrational energy followed by rapid distribution of energy through the covalent bonds of the peptide backbone. The timescale for this reaction to take place is about a picosecond (Wysocki et al., 2000). When the internal energy overcomes the activation energy barrier, that is needed to break a specific bond, fragment ions are formed (Figure 2). Successful identification of the sequence requires a high amount of different fragment ions originating from

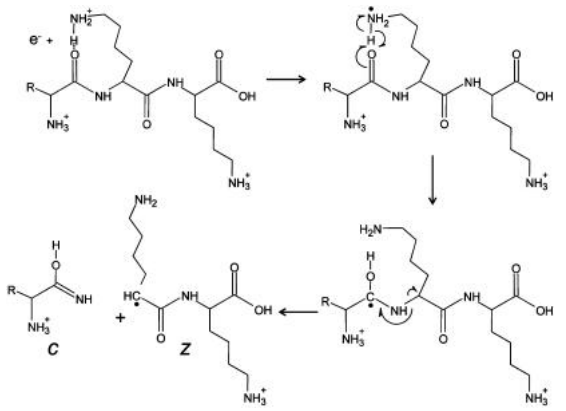

Figure 4 Fragmentation scheme for production of c- and z-type ions after reaction of a low-energy electron with a multiply protonated peptide Adapted from (Syka et al., 2004).

the random protonation of the peptide by the ESI process (Mikesh et al., 2006; Syka et al., 2004). This mechanism causes various problems when analysing phosphopeptides by mass spectrometry. If the random protonation of the backbone is inhibited by, for example multiple arginine residues, it will cause the peptide to break only at a few specific sites. Due to the principle of the reaction, it is most likely that the weakest bond will break first in the fragmentation process. While this is no problem with stable PTMs, phosphorylations especially at serines and threonines, have a lower activation barrier and are thus the preferred site of fragmentation. The peptide looses phosphoric acid, resulting in (peptide – H3PO4) being the most abundant ion in the MS/MS spectrum. This prohibits the gain of sequence information as well as the determination of the exact location of the phosphorylation site (Figure 3).

Another possibility to gain sequence information by peptide backbone fragmentation is electron transfer dissociation (ETD). The reaction principle was introduced as electron capture dissociation (ECD) 1998 by McLafferty and co workers (Zubarev et al., 1998). They used a FTICR ion trap to capture peptides and then let them react with near thermal electrons. The electron then gets captured by the peptide and causes the peptide backbone, after redistribution of the energy, to fragment on the C_{α}-N linkage (Figure 4). The result is a fragmentation of the peptide into c and z ions. There have been two pathways suggested for the mechanism of ECD. One involves the addition of a proton to the peptide backbone carbonyl group and the second possibility is the addition of a proton to an amide N, which then results in dissociation of CO and formation of a and y ions (for an more detailed view on the process see (Zubarev, 2003)). The great benefit of this reaction compared to CID is the independency on the backbone protonation and the short timescale of the reaction that will allow it to fragment also multiply protonated peptides and peptides containing labile PTMs fragment more randomly (Mirgorodskaya et al., 1999; Shi et al., 2001; Stensballe et al., 2001). The main reason why ECD stayed in the scientific background and was only the method of choice for a few groups for some years, was the expensive and not readily available instrumental equipment (Charlebois et al., 2003; Cooper et al., 2003; Fagerquist et al., 2003; Ge et al., 2003a; Ge et al., 2003b). In the last years the principle has been converted for the use on more economy priced instruments like ion traps. Here the transfer of the electron to the peptide is facilitated by an anion radical but the reaction mechanism of peptide fragmentation is similar. More and more publications show that ETD emerges from a niche existence to the method of choice when analysing PTMs.

REFERENCES

Aebersold, R., and Mann, M. (2003). Mass spectrometry-based proteomics. Nature 422, 198-207.

Andersson, L., and Porath, J. (1986). Isolation of phosphoproteins by immobilized metal (Fe3+) affinity chromatography. Anal Biochem 154, 250-254.

Butland, G., Peregrin-Alvarez, J.M., Li, J., Yang, W., Yang, X., Canadien, V., Starostine, A., Richards, D., Beattie, B., Krogan, N., et al. (2005). Interaction network containing conserved and essential protein complexes in Escherichia coli. Nature 433, 531-537.

Charlebois, J.P., Patrie, S.M., and Kelleher, N.L. (2003). Electron capture dissociation and 13C,15N depletion for deuterium localization in intact proteins after solution-phase exchange. Anal Chem 75, 3263-3266.

Cohen, P. (2000). The regulation of protein function by multisite phosphorylation--a 25 year update. Trends Biochem Sci 25, 596-601.

Cooper, H.J., Case, M.A., McLendon, G.L., and Marshall, A.G. (2003). Electrospray ionization Fourier transform ion cyclotron resonance mass spectrometric analysis of metal-ion selected dynamic protein libraries. J Am Chem Soc 125, 5331-5339.

Crane, R., Gadea, B., Littlepage, L., Wu, H., and Ruderman, J.V. (2004). Aurora A, meiosis and mitosis. Biol Cell 96, 215-229.

Fagerquist, C.K., Hudgins, R.R., Emmett, M.R., Hakansson, K., and Marshall, A.G. (2003). An antibiotic linked to peptides and proteins is released by electron capture dissociation fourier transform ion cyclotron resonance mass spectrometry. J Am Soc Mass Spectrom 14, 302-310.

Gavin, A.C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L.J., Bastuck, S., Dumpelfeld, B., et al. (2006). Proteome survey reveals modularity of the yeast cell machinery. Nature 440, 631-636.

Gavin, A.C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.M., Cruciat, C.M., et al. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 415, 141-147.

Ge, Y., El-Naggar, M., Sze, S.K., Oh, H.B., Begley, T.P., McLafferty, F.W., Boshoff, H., and Barry, C.E., 3rd (2003a). Top down characterization of secreted proteins from Mycobacterium tuberculosis by electron capture dissociation mass spectrometry. J Am Soc Mass Spectrom 14, 253-261.

Ge, Y., Lawhorn, B.G., ElNaggar, M., Sze, S.K., Begley, T.P., and McLafferty, F.W. (2003b). Detection of four oxidation sites in viral prolyl-4-hydroxylase by top-down mass spectrometry. Protein Sci 12, 2320-2326.

Gronborg, M., Kristiansen, T.Z., Stensballe, A., Andersen, J.S., Ohara, O., Mann, M., Jensen, O.N., and Pandey, A. (2002). A mass

spectrometry-based proteomic approach for identification of serine/threonine-phosphorylated proteins by enrichment with phospho-specific antibodies: identification of a novel protein, Frigg, as a protein kinase A substrate. Mol Cell Proteomics 1, 517-527.

Ho, Y., Gruhler, A., Heilbut, A., Bader, G.D., Moore, L., Adams, S.L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., et al. (2002). Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature 415, 180-183.

James, P., Quadroni, M., Carafoli, E., and Gonnet, G. (1993). Protein identification by mass profile fingerprinting. Biochem Biophys Res Commun 195, 58-64.

Johnson, R.S., Martin, S.A., Biemann, K., Stults, J.T., and Watson, J.T. (1987). Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. Anal Chem 59, 2621-2625.

Krogan, N.J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A.P., et al. (2006). Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature 440, 637-643.

Liu, F., Iqbal, K., Grundke-Iqbal, I., Hart, G.W., and Gong, C.X. (2004). O-GlcNAcylation regulates phosphorylation of tau: a mechanism involved in Alzheimer's disease. Proc Natl Acad Sci U S A 101, 10804-10809.

Liu, X., Valentine, S.J., Plasencia, M.D., Trimpin, S., Naylor, S., and Clemmer, D.E. (2007). Mapping the human plasma proteome by SCX-LC-IMS-MS. J Am Soc Mass Spectrom 18, 1249-1264.

Mann, M., Hendrickson, R.C., and Pandey, A. (2001). Analysis of proteins and proteomes by mass spectrometry. Annu Rev Biochem 70, 437-473.

Mann, M., Hojrup, P., and Roepstorff, P. (1993). Use of mass spectrometric molecular weight information to identify proteins in sequence databases. Biol Mass Spectrom 22, 338-345.

Mann, M., Ong, S.E., Gronborg, M., Steen, H., Jensen, O.N., and Pandey, A. (2002). Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. Trends Biotechnol 20, 261-268.

McLachlin, D.T., and Chait, B.T. (2001). Analysis of phosphorylated proteins and peptides by mass spectrometry. Curr Opin Chem Biol 5, 591-602.

Mikesh, L.M., Ueberheide, B., Chi, A., Coon, J.J., Syka, J.E.P., Shabanowitz, J., and Hunt, D.F. (2006). The utility of ETD mass spectrometry in proteomic analysis. Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics 1764, 1811-1822.

Mirgorodskaya, E., Roepstorff, P., and Zubarev, R.A. (1999). Localization of O-glycosylation sites in peptides by electron capture dissociation in a Fourier transform mass spectrometer. Anal Chem 71, 4431-4436.

Musicki, B., Kramer, M.F., Becker, R.E., and Burnett, A.L. (2005). Inactivation of phosphorylated endothelial nitric oxide synthase (Ser-1177) by O-GlcNAc in diabetes-associated erectile dysfunction. Proc Natl Acad Sci U S A 102, 11870-11875.

Oda, Y., Nagasu, T., and Chait, B.T. (2001). Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. Nat Biotechnol 19, 379-382.

Pandey, A., Podtelejnikov, A.V., Blagoev, B., Bustelo, X.R., Mann, M., and Lodish, H.F. (2000). Analysis of receptor signaling pathways by mass spectrometry: identification of vav-2 as a substrate of the epidermal and platelet-derived growth factor receptors. Proc Natl Acad Sci U S A 97, 179-184.

Pappin, D.J., Hojrup, P., and Bleasby, A.J. (1993). Rapid identification of proteins by peptide-mass fingerprinting. Curr Biol 3, 327-332.

Pinkse, M.W., Uitto, P.M., Hilhorst, M.J., Ooms, B., and Heck, A.J. (2004). Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. Anal Chem 76, 3935-3943.

Posewitz, M.C., and Tempst, P. (1999). Immobilized gallium(III) affinity chromatography of phosphopeptides. Anal Chem 71, 2883-2892.

Roepstorff, P., and Fohlman, J. (1984). Proposal for a common nomenclature for sequence ions in mass spectra of peptides. Biomed Mass Spectrom 11, 601.

Shi, S.D., Hemling, M.E., Carr, S.A., Horn, D.M., Lindh, I., and McLafferty, F.W. (2001). Phosphopeptide/phosphoprotein mapping by electron capture dissociation mass spectrometry. Anal Chem 73, 19-22.

Sickmann, A., and Meyer, H.E. (2001). Phosphoamino acid analysis. Proteomics 1, 200-206.

Soskic, V., Gorlach, M., Poznanovic, S., Boehmer, F.D., and Godovac-Zimmermann, J. (1999). Functional proteomics analysis of signal transduction pathways of the platelet-derived growth factor beta receptor. Biochemistry 38, 1757-1764.

Steen, H., and Mann, M. (2002). A new derivatization strategy for the analysis of phosphopeptides by precursor ion scanning in positive ion mode. J Am Soc Mass Spectrom 13, 996-1003.

Stensballe, A., Andersen, S., and Jensen, O.N. (2001). Characterization of phosphoproteins from electrophoretic gels by nanoscale Fe(III) affinity chromatography with off-line mass spectrometry analysis. Proteomics 1, 207-222.

Sui, S., Wang, J., Yang, B., Song, L., Zhang, J., Chen, M., Liu, J., Lu, Z., Cai, Y., Chen, S., et al. (2008). Phosphoproteome analysis of the human Chang liver cells using SCX and a complementary mass spectrometric strategy. Proteomics 8, 2024-2034.

Syka, J.E., Coon, J.J., Schroeder, M.J., Shabanowitz, J., and Hunt, D.F. (2004). Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc Natl Acad Sci U S A 101, 9528-9533.

Tyers, M., and Jorgensen, P. (2000). Proteolysis and the cell cycle: with this RING I do thee destroy. Curr Opin Genet Dev 10, 54-64.

Wysocki, V.H., Tsaprailis, G., Smith, L.L., and Breci, L.A. (2000). Mobile and localized protons: a framework for understanding peptide dissociation. J Mass Spectrom 35, 1399-1406.

Yamagata, A., Kristensen, D.B., Takeda, Y., Miyamoto, Y., Okada, K., Inamatsu, M., and Yoshizato, K. (2002). Mapping of phosphorylated proteins on two-dimensional polyacrylamide gels using protein phosphatase. Proteomics 2, 1267-1276.

Yates, J.R., 3rd (2004). Mass spectral analysis in proteomics. Annu Rev Biophys Biomol Struct 33, 297-316.

Yen, C.Y., Russell, S., Mendoza, A.M., Meyer-Arendt, K., Sun, S., Cios, K.J., Ahn, N.G., and Resing, K.A. (2006). Improving sensitivity in shotgun proteomics using a peptide-centric database with reduced complexity: protease cleavage and SCX elution rules from data mining of MS/MS spectra. Anal Chem 78, 1071-1084.

Zhou, H., Watts, J.D., and Aebersold, R. (2001). A systematic approach to the analysis of protein phosphorylation. Nat Biotechnol 19, 375-378.

Zubarev, R.A. (2003). Reactions of polypeptide ions with electrons in the gas phase. Mass Spectrometry Reviews 22, 57-77.

Zubarev, R.A., Kelleher, N.L., and McLafferty, F.W. (1998). Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. J. Am. Chem. Soc. 120, 3265-3266.

OPTIMIZING MACROMOLECULAR TANDEM MASS SPECTROMETRY

Kristina Lorenzen, Cees Versluis, Esther van Duijn, Robert H.H. van den Heuvel, and Albert J. R. Heck
Biomolecular Mass Spectrometry and Proteomics Group, Bijvoet Center for Biomolecular Research and Utrecht Faculty of Science, Utrecht University, Sorbonnelaan 16, 3584 CA Utrecht, the Netherlands

SUMMARY
We modified and optimized a first generation quadrupole time-of-flight (Q-ToF) to perform tandem mass spectrometry on macromolecular protein complexes. The modifications of the Q-ToF 1 include the introduction of 1) a flow-restricting sleeve around the first hexapole ion bridge, 2) a low frequency ion-selecting quadrupole, 3) a high pressure hexapole collision cell, 4) high transmission grids in the multi-component ion lenses and 5) a low repetition rate pusher. We then evaluated and optimized conditions for tandem mass spectrometry on macromolecular protein complexes, varying the gas pressure inside the hexapole collision cell, and studied the influence of using different collision gases, e.g. argon, krypton and xenon. These variations affected ion transmission through the instrument and dissociation efficiency, with in particular the high mass fragment ions being generated more readily and transmitted more efficiently using the heavier xenon collision gas. We studied in detail the influence the binding of one and two gp5 substrate proteins had on the gas-phase dissociation of the chaperonin complex constructing break-down diagrams. We observed that the binding of substrate polypeptides had a significant effect on the gas-phase stability of the chaperonin complex, with the complex containing one and two substrate molecules being less susceptible towards dissociation than the substrate free GroEL. From all data acquired, we conclude that the heavier xenon is the preferred collision gas for tandem mass spectrometry on very large macromolecular complexes.

INTRODUCTION

For structural biology electrospray ionization (ESI) mass spectrometry has become a key method to study protein complexes of increasing size and complexity (Benesch et al., 2006; Borch et al., 2005; Heck and Van Den Heuvel, 2004; Loo, 1997; McKay et al., 2006; Poliakov et al., 2007; Ruotolo and Robinson, 2006; van Duijn et al., 2005) culminating recently in the analysis of intact endogenously expressed heterogeneous protein complexes, such as the in RNA degradation involved exosome complex (Hernandez et al., 2006; Synowsky et al., 2006). The analysis of proteins and protein complexes by mass spectrometry (native mass spectrometry) has become possible because of the relatively gentle ionization procedure ESI, which can retain non-covalent interactions in protein complexes (Ashcroft, 2005). The mass analysis of larger proteins and protein complexes is not a routine technique, since careful optimization of operating conditions of the mass spectrometer is required, which so far was difficult to realize using commercially available mass spectrometers.

Despite the theoretically unlimited mass range of ToF analyzers, most instruments have detection problems when the m/z values exceed 4,000. For the analysis of a single protein, mass measurement is often sufficient and can identify e.g. protein modifications and oligomerization state. However, when more complex biomolecules, such as protein complexes or protein-DNA complexes are analyzed mass measurement alone is insufficient to delineate all possible combinations of proteins and their ligands. For these more complex samples tandem mass spectrometry can be used to dissect complexes in the gas-phase in a sequential fashion to identify the building blocks, and potentially also complex topology, quaternary structure and stability (McCammon et al., 2004). In tandem mass spectrometry an ion is selected on the basis of the m/z value in the first quadrupole mass analyzer after which the selected ion is dissociated in a gas-filled collision cell. The subsequent ToF analyzer then allows the detection of the product ions. Studies have shown that the asymmetric separation of mass and charge when dissecting protein complexes in the gas phase is a common phenomenon (Heck and Van Den Heuvel, 2004; Sobott et al., 2005)). The common pathway of gas-phase collision induced protein complex dissociation inside the mass spectrometer can be described as follows (Equation 1):

$$C^z \rightarrow M^x + [C\text{-}M]^{(z-x)}$$

Where C is the protein complex, M is dissociated monomeric subunit, z is the number of charges of the complex and x is the number of charges carried by the dissociated monomeric subunit.

Tandem mass spectrometry of large protein complexes is a relatively new field in native mass spectrometry. This is due to the fact that most commercial quadrupoles can transmit ions to a certain limit, which is dependent on the RF amplitude, frequency and diameter of the rod assembly. The transmission limit (M_{max}) is given by (Equation 2):

$$M_{max} = 7(10^6 V_m)/(f^2 r_0^2)$$

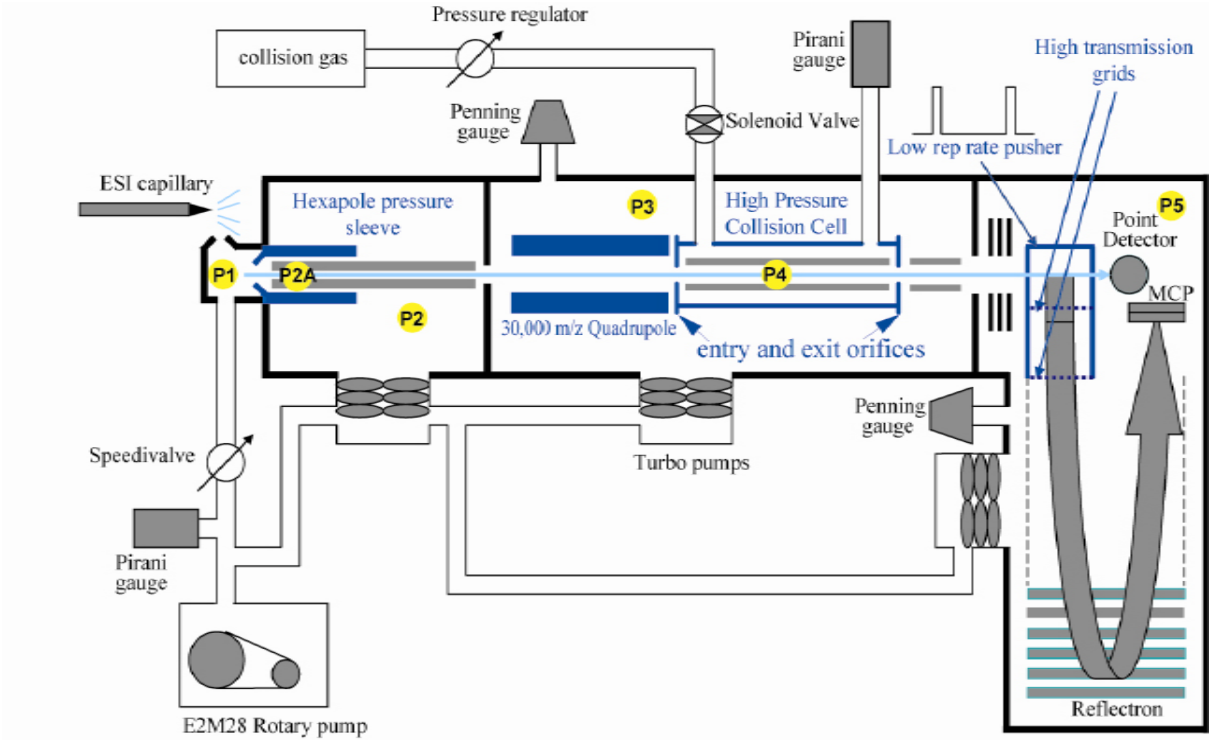

Figure 1 Schematic layout of the modified first generation Q-ToF 1 instrument (Micromass, U.K.). For details of modifications see Experimental section. Items in dark blue are modifications relative to the standard Q-ToF 1 configuration. The position of the entrance and exit orifices of the hexapole collision cell, which sizes were varied in the current experiments, are indicated as well as the pressure read-outs mentioned in the manuscript.

With Vm cos(2ft) as the RF voltage applied between the rods ($2V_m$ is the peak-to-peak amplitude; f, the frequency; and t the time) and r_0 as the inner radius between the rods. A standard first generation Q-ToF 1 (Micromass) operates at a frequency of 832 kHz setting the transmission limit to m/z 4,190. By replacing the standard RF generator by a RF generator that operates at a lower frequency the upper transmission limit can be extended (Loo et al., 2005; Sobott et al., 2002).

These procedures are quite well known for small molecules, but the collisional activation of macromolecular complex ions is rather different. Ions of macromolecular complexes, such as the 800 kDa GroEL chaperone machinery, may attain tens of charges (approximately 70 for the chaperone complex GroEL), and with the voltage over the hexapole collision cell reaching up to a maximum of 200 V this leads to laboratory frame collision energies in the order of 15,000 eV. However, typically these macromolecular ions are subjected to collisions with relatively static, small argon atoms in the gas-filled hexapole collision cell. When such macromolecular ions with high translational energies undergo inelastic collisions with neutral rare gas atoms, part of the translational energy may be converted into internal energy, leading to decomposition. To simplify the description of such a process, one needs to consider the center-of-mass (com) framework instead of the laboratory reference frame. The maximum available energy for the transfer of kinetic energy to internal energy, in the first collision event, is the relative energy (E_{com}) and depends on the collision partners' masses. Conversion from laboratory to center of mass collision energies is given by (Sleno and Volmer, 2004) (Equation 3):

$$E_{com}=m_t/(m_p+m_t)\ E_{lab}$$

where E_{lab} is the ion's kinetic energy and m_t and m_p represent the masses for the neutral target gas and precursor ion (800 kDa for GroEL), respectively. Using argon as a collision partner in this equation it is evident that 15,000 eV in the laboratory framework reduces to approximately 0.75 eV in the center-of-mass framework, hardly sufficient to break any chemical bond. Therefore, it is generally expected that successful native tandem mass spectrometry depends on multiple collision events, as happening in the hexapole collision cell of the Q-ToF mass spectrometer. Considering this simple equation it is evident that with a choice for a heavier collision partner such as krypton or xenon considerably more energy may be transferred into internal energy.

One of the most important findings has been that the pressures in the first and second vacuum chambers have to be increased to analyze large protein complexes (Chernushevich and Thomson, 2004; Krutchinsky et al., 1998; Sanglier et al., 2003; Schmidt et al., 2001; Sobott et al., 2002; Tahallah et al., 2001). This effect is attributed to the collisional cooling and focusing of large ions in the ion guides and,

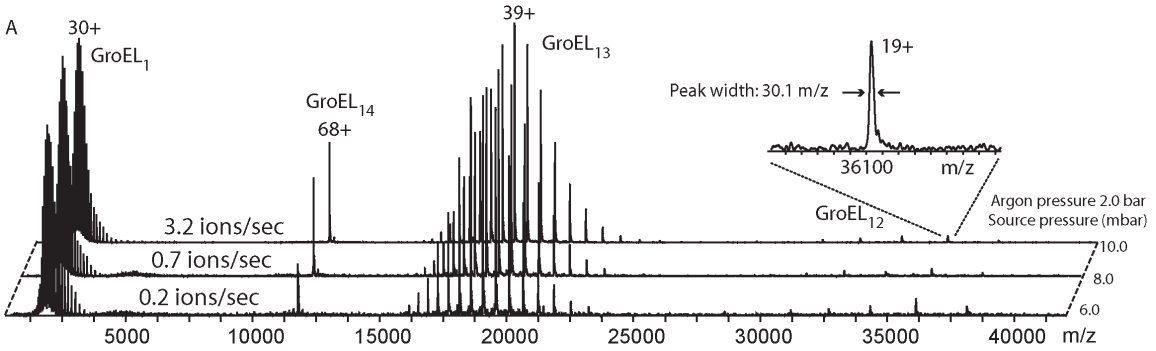

Figure 2 Gas-phase dissociation of GroEL14 complex. Tandem mass spectrometry was performed on 0.36 µM GroEL14 in 50 mM ammonium acetate pH 6.8. Source pressure dependent tandem mass spectra of the 68+ ion of GroEL14 (11,780 m/z) at an acceleration voltage of 200 V. The argon pressure was kept constant at 2.0 bar (collision cell pressure of 1.5 10^{-2} mbar). The parent ion peak of the tetradecameric complex is almost completely dissociated into GroEL13 and monomeric GroEL ions. At high collision energies (200 V) the produced tridecameric complex is, in turn, partly dissociated into GroEL12 and monomeric GroEL products. These two processes take place in sequential order. The inset shows the FWHM of the 19^+ GroEL12 ion at m/z 38,142.

therefore, improved transmission through the ion guides and ToF (Chernushevich and Thomson, 2004). Although ions of macromolecular complexes attain a large number of charges in ESI (Heck and Van Den Heuvel, 2004), making their kinetic energy high, their velocity inside the mass spectrometer is relatively low, which makes them susceptible to off-axis drifting. This may be even more true for ions formed in highly energetic dissociation processes such as collision induced dissociation, whereby fragment ions are released with significant off-axis kinetic energies.

In this study, we have modified and optimized a first generation Q-ToF 1 instrument for native tandem mass spectrometry. The described low cost modifications can be applied on all standard Q-ToF machines. We report on the use of the collision gases argon, krypton and xenon for native tandem mass spectrometry. We show that the pressure in the collision cell has a significant effect on the transmission of precursor and fragment ions, with in particular the lowly charged high mass macromolecular fragment ions being negatively discriminated at lower gas pressures. Additionally, we show that the heavier target gases result in more efficient fragmentation and fragment ion detection, making xenon the collision gas of choice in native mass spectrometry.

RESULTS AND DISCUSSION

Adaptation of a Q ToF 1 to measure protein complexes

To analyze intact ions in vacuo from large complexes in solution conditions have to be optimized. Insufficient axial cooling makes ions miss the detector, whereas insufficient radial cooling results in poor transmission of ions through the different apertures in the mass spectrometer. We acquired tandem mass spectra at different source pressures (Figure 1) (P1) between the sample and extraction cone to by tuning the speedivalve (Chernushevich and Thomson, 2004; Krutchinsky et al., 1998; Sanglier et al., 2003; Schmidt et al., 2001; Tahallah et al., 2001). The pressures at the other stages in the mass spectrometer were kept constant. Reducing the pumping efficiency of the rotary pump up to 10 mbar clearly led to an enhanced detection of the GroEL ions. At a source pressure of 6.0 mbar only 0.2 ions per second were detected after ToF analysis, whereas at a pressure of 10 mbar 3.2 ions per second were detected (Figure 2). Thus, the maximal achievable pressure of 10 mbar seems to be the optimal value for the transmission of the GroEL product ions. As expected, the relative abundance of the different species did not depend on the source pressure.

We also introduced a metal cylinder around the first half of the hexapole ion lens (length 100 mm) to increase the pressure locally (P2A). It was estimated that the initial pressure in the sleeve is three-fold higher than in the rest of the hexapole and that the pressure in the sleeve decreases linearly. The experimental design was theoretically tested by simulating the ion trajectories of a GroEL14 ion (801 kDa; 70^+ ion) within the modified hexapole ion guide using a similar method as Chenushevich and Thomson (Chernushevich and Thomson, 2004) (Figure 2 and 3). This figure clearly shows that in order to avoid significant GroEL14 ion losses on the

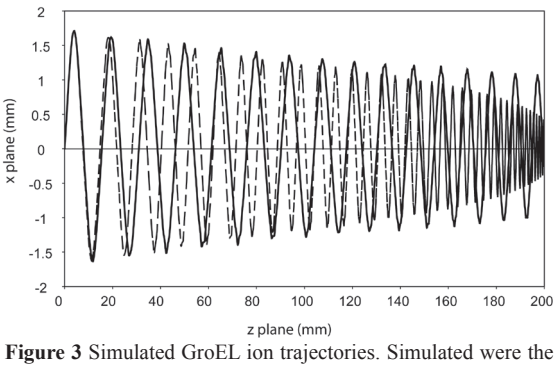

Figure 3 Simulated GroEL ion trajectories. Simulated were the 68^+ GroEL ion trajectories in the hexapole ion lens in the xz plane, at gas pressures of 4 10^{-3} mbar (solid line) and 4 10^{-3} mbar including the metal sleeve around the first half of the hexapole lens (dashed line). The pressure is estimated to be on average two-fold higher than in the remaining of the hexapole ion bridge. The ion has an initial velocity of 185 m/s and is aligned 30 degrees off the optic axis.

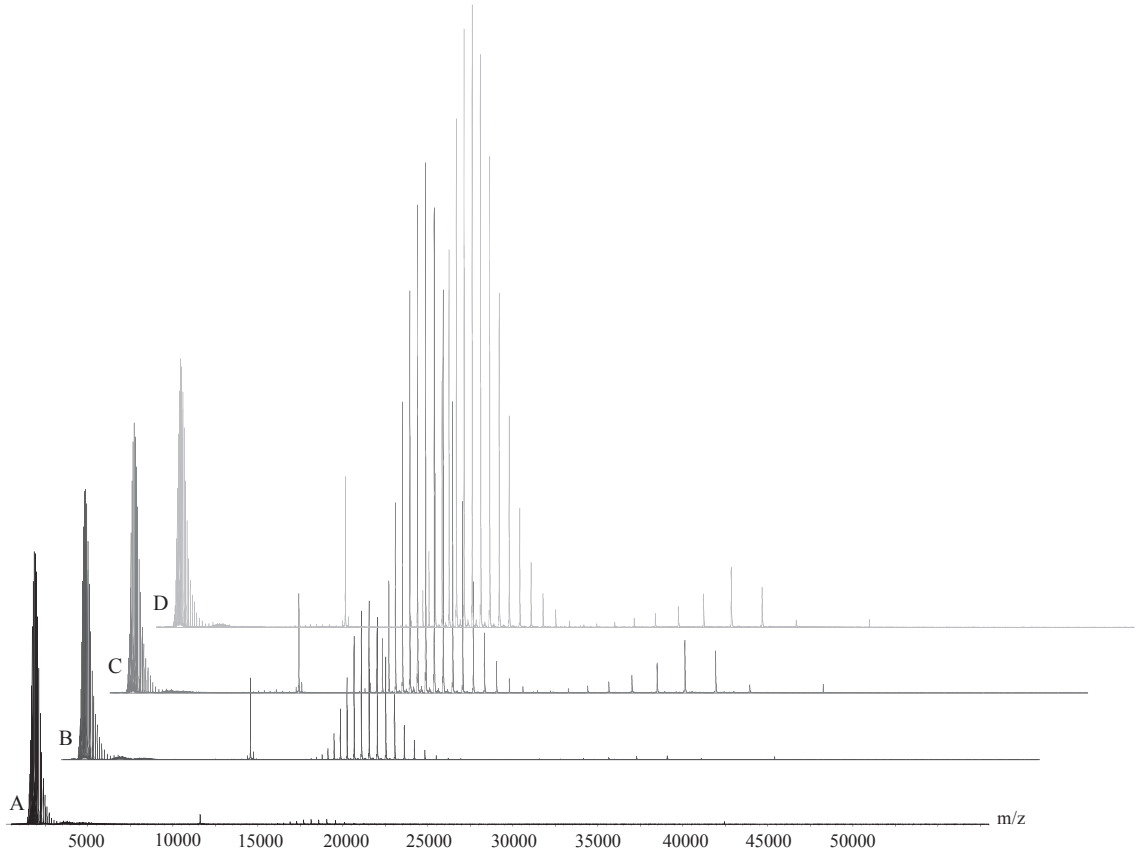

Figure 4 Gas-phase dissociation of GroEL varying the collision gas pressure. Tandem mass spectrometry was performed on GroEL in 50 mM ammonium acetate pH 6.8 at different P4 argon pressures in the collision cell. The 69^{+} ion of GroEL (11,640 m/z) was selected and dissociated at a collision voltage of 200 V. The parent ion peak of the tetradecameric complex dissociates into tridecameric GroEL (15,000-22,000 m/z; average charge state 40^{+}) and monomeric GroEL (1,500-2,200 m/z; average charge state 28^{+}). At higher gas pressure also the dodecameric GroEL (28,600-38,200 m/z; average charge state 20^{+}) is present. Figure A, B, C and D show collision gas pressures of $1x10^{-2}$ mbar, $1.5x10^{-2}$ mbar, $2x10^{-2}$ mbar and $3x10^{-2}$ mbar respectively.

different apertures a combination of an increased pressure of 4 10^{-3} mbar in the hexapole vacuum chamber (P2) and a locally increased pressure around the first part of the hexapole lens was required (P2A). Besides the source and hexapole ion bridge pressure conditions also the argon cylinder head pressure, and thus the collision cell pressure, influences the transport of the ions to the transport hexapole and the ToF analyzer (Sobott et al., 2002). We and others have noticed that increasing the collision gas-pressure in the collision cell supports dissociation and ion transmission of high m/z ions. The entrance and exit orifices of the hexapole collision cell (see Figure 1 for their location) in our instrument have standard apertures of 2.0 mm. When using argon and increasing the P4 pressure above 1.5 x 10^{-2} mbar the transmission of the ions decreased strongly (the signal was 20% lower when changing from 1.5 to 2 x 10^{-2} mbar). The pressure in the P3 region increased to approximately 6.7 x 10^{-4} mbar, which is at the limit of being tolerated by the instrument. As we hypothesized that increasing the collision cell gas pressure might further assist the dissociation and ion transmission, we evaluated smaller entrance and exit orifices of 1.5 mm as the gas leakage out of the cell is expected to decrease using these smaller apertures allowing us to achieve a higher pressure P4 in the collision cell. We first tested whether ion transmission was affected using the different orifices and noticed decreases in total ion transmission in the order of 25 % for ions with m/z values around 10,000, using the smallest apertures, a loss we took for granted. Using the smaller apertures of 1.5 mm we noticed that the collision gas pressure P4 could even be increased close to 3 x 10^{-2} mbar (using argon). Now a P4 pressure in the collision cell of 2 x 10^{-2} mbar resulted in a pressure P3 of 1.2 x 10^{-4} mbar is resulting in a decrease of the leakage effect of gas from the collision cell into the quadruople region which is 6 times lower with the 1.5 mm orifices.

Next we evaluated what the effect of increasing the collision cell pressure (using argon) had on the tandem mass spectrometry spectra of the 14-mer GroEL chaperone machinery. For these measurements GroEL complexes in 50 mM ammonium acetate pH 6.8, were used at a monomer concentration of 6 µM. The capillary voltage was typically set at 1,500 V, the cone voltage at 175 V and the ion energy at 2.0 V. Produced ions were mass isolated in the quadrupole analyzer and accelerated into the target gas-filled collision cell. In Figu-

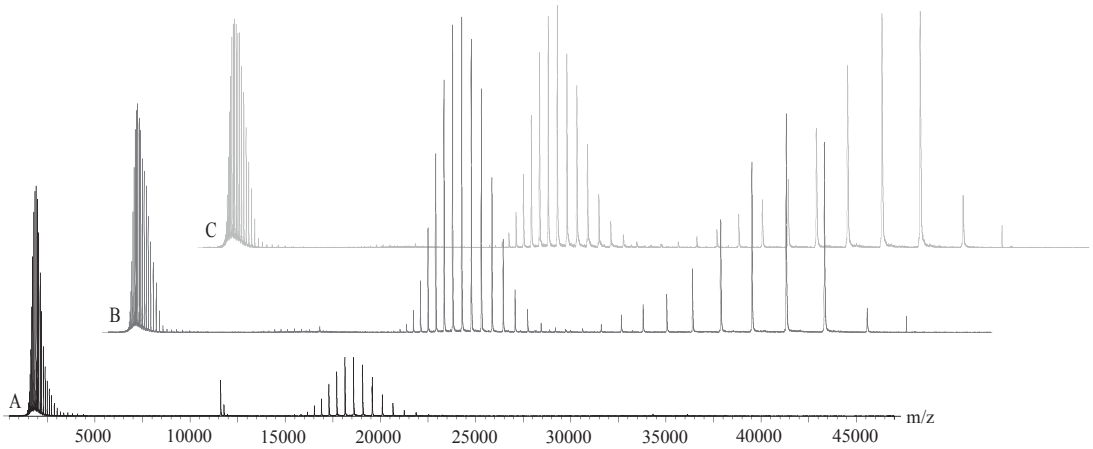

Figure 5 Gas-phase dissociation of GroEL varying the collision gas identity. Tandem mass spectrometry on 69^+ charge state of GroEL complex. All spectra were measured at a collision voltage of 200 V using different collision gases at a constant pressure of $1.5x10^{-2}$ mbar. Argon for figure A, Krypton for figure B and Xenon for figure C. As in Figure 2 the precursor ions dissociate into monomer (1,500-2,200 m/z; average charge state 29^+), tridecamer (15,000-22,000 m/z; average charge state 40^+) and dodecamer (28,600-38,200 m/z; average charge state 20^+).

re 4 the tandem mass spectra are shown of the mass selected 69^+ ion of GroEL (at an m/z value of 11,640) measured under identical conditions, with the collision voltage at the maximum of 200 V. The argon collision gas pressure was varied from 1 x 10^{-2} mbar, 1.5 x 10^{-2} mbar, 2 x 10^{-2} mbar, to 3 x 10^{-2} mbar. In all these cases an equal amount of mass spectra were averaged. In Figure 4 the spectra are normalized on the intensities of the monomeric fragment ions. The mechanism of the dissociation of subunits from protein complexes in tandem mass spectrometry is still a relatively poorly understood process, however, it is generally accepted that multi-component protein complexes dissociate via similar pathways, i.e. expulsion of the smallest monomeric subunit (Heck and Van Den Heuvel, 2004) attaining an inequitable high amount of the charges. Recently, some interesting exceptions have been reported notably with surface induced dissociation (Jones et al., 2006), but also with specific "dimer-of-dimers" tetramers (van den Heuvel et al., 2006). The gas-phase dissociation pathway of unliganded GroEL (800 kDa) has been well documented (Sobott and Robinson, 2004; van Duijn et al., 2006b) and follows the generally observed trends described above. Upon isolation of the 69^+ ion of GroEL at an m/z value of 11,640 a monomeric GroEL subunit becomes ejected from the tetradecameric complex. The monomers attain during the dissociation a large amount of the charges and are observed around m/z 2,000 (28^+), whereas the tridecameric ions are observed around m/z 18,000 (41^+). At higher collision gas pressure a second GroEL monomer is dissociated from the tridecameric complex, resulting in the formation of dodecameric GroEL product ions. The dodecameric GroEL obtains an average charge state of only 20^+. Therefore, these ions appear at a high m/z values above ~30,000. The spectra shown in Figure 4 reveal nicely the effect of increasing the collision gas pressure. In the range studied the total detected ion intensities of the monomer fragment ions were only changing about 10 % with increasing pressure. Therefore, we used these ions to normalize the spectra. The higher m/z ions show a significant increase in transmission. The ion intensities of the high mass ions rise with increasing gas pressure, the precursor ions increases 1.5 times, the tridecameric fragment ions 4 times and the dodecameric fragment ions 10 times comparing the collision cell pressures of 1.5 x 10^{-2} mbar and 3 x 10^{-2} mbar. In the ideal case the summed intensity of the high-mass fragment ions should be identical to the summed monomeric fragment ion intensities, which seems to be the case from a gas pressure of 2 x 10^{-2} mbar and upwards in the collision cell. We attribute the observed positive influence of increasing the gas pressure only marginally to increased dissociation of the precursor ions, but mainly too the improved transmission of the high m/z precursor and fragment ions. We evaluated additionally a collision gas pressure in the hexapole of 6 x 10^{-3} mbar, but then only monomeric fragment ions were observed, at an ion intensity reduced to 30 %, compared to the measurements at 1 x 10^{-2} mbar. Therefore, reducing the gas pressure below 1 x 10^{-2} mbar deteriorated the ion transmission too much. We tested the ion transmission also for krypton and xenon. While for krypton the presence of monomeric to oligomeric ions is equal starting at a pressure of 0.5 x 10^{-2} mbar, Xenon showed the same transmission efficiency from from 1 x 10^{-3} mbar to 2 x 10^{-2} mbar. So the observed effects are very significant when using the commonly used argon as collision gas, but they seem to diminish when using krypton and especially xenon. As the results with the aperture of 1.5 mm seem to be optimal for native mass spectrometry, we performed all later described experiments with these entrance and exit orifices.

Energy resolved gas-phase dissociation of GroEL chaperone complexes using argon, krypton and xenon collision gases

Following the optimization of the collision cell entrance and exit apertures we investigated the effect of the identity of the collision gas. It is well-known that the identity of the target gas may influence the dissociation efficiency, although such data is largely limited to the analysis of relatively small molecules/peptides (Curcuruto and Hamdan, 1993; Leggett et al., 1990; Sleno and Volmer, 2004). Using identical experimental conditions, selecting the same precursor ions of GroEL, we performed tandem mass spectrometry using sequentially argon (40 Da), krypton (average mass 83.8 Da) and xenon (average mass 131.3 Da) as collision gas. For illustrative purposes we have depicted in Figure 5 tandem mass spectra of mass selected 69+ ions of GroEL measured under identical conditions, with the collision voltage at the maximum of 200 V, only varying the nature of the collision gas from argon, krypton to xenon (bottom to top in Figure 5). With all these three collision gases we recorded spectra with varying collision voltages, and calculated the ion intensities by averaging over all charge states of a particular ionic species. These ion intensities do evidently not correspond directly to species abundance, as therefore several instrument response factors (transmission and detection) have to be taken into account. Here we just use the integrated ion intensities for relative comparisons. In Figure 6 A, B and C these "break-down" diagrams are depicted, plotting the measured ion intensities of the GroEL precursor ions, the monomeric, the tridecameric and dodecameric fragment ions versus the collision voltage for argon, krypton and xenon as collision gases, respectively. In all these

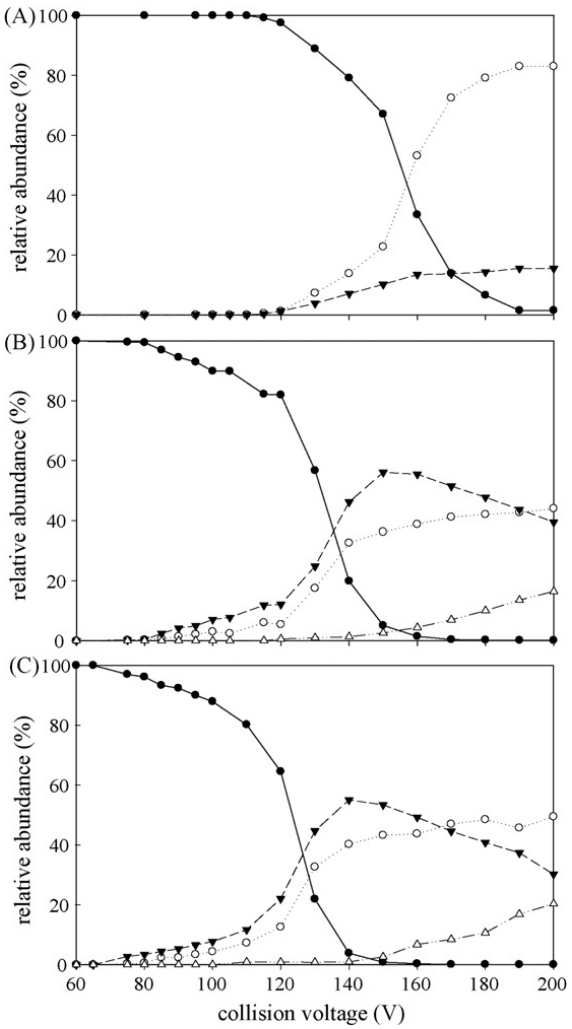

Figure 6 Break down graphs of GroEL using different collision gases. Ion intensities of the GroEL monomers, tetradecamers, tridecamers and dodecamers at increasing collision energies for Argon (A), Krypton (B) and Xenon (C). The different ion intensities were integrated and plotted in relative abundance to the total ion intensity. The pressure of the collision gas was 1.5×10^{-2} mbar for all measurements. ● represent the tetradecamer selected for dissociation (69^+ at 11640 m/z), ○ the dissociated monomer, ▼the tridcamer and Δ the dodecameric GroEL complexes.

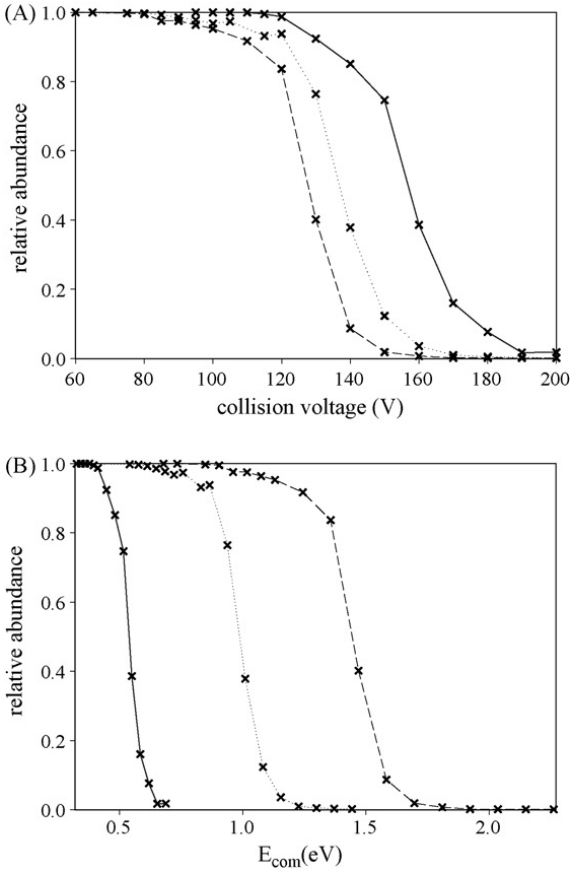

Figure 7 Break down graphs of GroEL precursor ions using different collision gases in the laboratory and center-of-mass frame. To construct relative break-down curves of the GroEL precursor ions the ion intensity of the precursor was divided by the summed ion intensities of the precursor and monomeric fragment ions (Y-axis). Dissociation threshold scheme of the tetradecameric GroEL complex. These ratios are plotted versus collision voltage in (A) and versus center-of-mass collision energy (E_{com}) in B. The lines represent the different gases: – for Argon, ⋯ for krypton and --- for xenon.

graphs a gradual relative decrease in precursor ion intensity is observed (the solid circles), although the onset of this decrease varies with the identity of the collision gas. Interestingly, using xenon and krypton the summed ion intensities of the concomitant fragment ions (i.e. monomeric versus tridecameric fragment ions) seems to be in line, which a priori might be expected as for each monomeric fragment ion one tridecameric fragment ion is formed, indicating that the other instrument parameters influencing the response factors for these two very different ions are quite identical. In contrast, using argon as collision gas the ratio monomer versus tridecameric fragment ions is very much in favor of the monomeric ions, revealing that using argon the transmission of the tridecameric fragment ions is reduced, just as described in the previous section. To determine a relative threshold for dissociation we plotted in Figure 7A the ratio (precursor ions/(precursor ions + monomer ions) versus the collision voltage. These graphs reveal that the collision voltages at which the precursor ion intensity is reduced to 50% are approximately at 157, 137 and 128 Volts, for argon, krypton and xenon, respectively. In Figure 7B these results are re-plotted, but now versus the center-of-mass collision energy. Evidently, in this graph the three curves should overlay, when the simple relationship eq. 3 would be valid for our measurements. However, this seems not to be the case, indicating that eq. 3 either overestimates the amount of collision energy converted into internal energy using argon, or underestimates the amount of collision energy converted into internal energy using xenon. Clearly xenon is found to be a much more efficient collision gas in our measurements. The fact that the simple eq. 3 is not valid may be due to the fact that it describes single collision events, whereas the collisional activation leading to dissociation of the native complexes described here is a multiple collision event, estimated to involve hundreds of collisions. This process is evidently much more efficient in case of xenon as collision gas compared to argon. As expected krypton is found to be a collision gas providing results in between argon and xenon. Mauk et al. (Mauk et al., 2002) suggested a model for multiple collision events based on the dissociation of Cytochrome c. In the paper they find that the conversion to the center-of-mass collision energy does not make up for the expected collision energy converted into internal energy. They describe that the energy transfer of neon, argon and krypton is similar and the reduced energy needed for dissociation with the different gases is due to increased time available for reaction in chase of increased pressures or heavier gases.

Energy resolved gas-phase dissociation of GroEL, GroEL:gp5 and GroEL:gp52 chaperone-substrate complexes using xenon as collision gas

In chaperone assisted folding an unfolded polypeptide substrate is encapsulated in one of the rings of the GroEL chaperone. Binding of a substrate to one of the GroEL rings is generally assumed to have an anti-cooperative effect i.e. binding of the first substrate inhibits the binding of a second substrate in the opposite ring (Falke et al., 2001; Wang and Chen, 2003). In contrast we showed previously by native mass spectrometry that the stoichiometry of substrate binding to the chaperonin complex is dependent on the type of substrate (van Duijn et al., 2005; van Duijn et al., 2006a). For instance, the bacteriophage P22 capsid protein gp5 is able to occupy both GroEL rings simultaneously (van Duijn, Protein Science, in press). To illustrate this in Figure 8A an ESI mass spectrum is shown of a mixture of tetradecameric GroEL to which unfolded gp5 substrate was added in a ratio of 1:2. This rather complex mass spectrum reveals three charge state distributions, clearly originating from GroEL, GroEL:gp5 and GroEL:$(gp5)_2$ as indicated by the convoluted zero-charge mass spectrum in Figure 8B.

Next, we compared the gas phase stability of GroEL to GroEL:gp5 and GroEL:$(gp5)_2$ by measuring "break-down" curves for the dissociation of these species, selectively performing tandem mass spectrometry measurements on precursor ions of the chaperone with zero, one and two substrate polypeptide gp5 bound. Measurements were done with argon, krypton and xenon as collision gases. All three species revealed similar fragmentation reactions, namely the loss of one monomeric subunit at intermediate collision voltages, followed by the loss of a second GroEL monomer at higher collision voltages. Interestingly, the elimination of 57 kDa GroEL subunits is favored over the loss of the 47 kDa gp5 substrate, which may indicate that the substrate is also in the gas-phase still encapsulated within the GroEL ring hampering its diossciation, in line with our previous data on similar GroEL substrate complexes (van Duijn et al., 2006a). In Figure 9 break-down curves are shown obtained with krypton as collision gas, but results with the other gases revealed similar effects. In Figure 8 the break-down curves for loss of one GroEL subunit (solid symbols) and 2

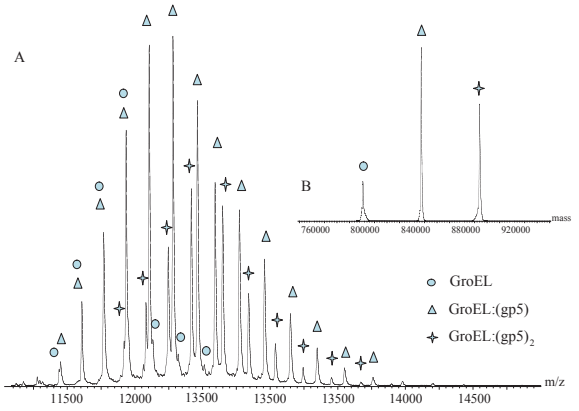

Figure 8 ESI mass spectrum of intact GroEL gp5 substrate complexes. GroEL (14-mer) and gp5 were mixed in a 1:2 ratio. Three different charge state distributions can be observed in these spectra originating from GroEL (○) GroEL:(gp5) (Δ) and GroEL:$(gp5)_2$ (✦). In B the corresponding convoluted zero-charge spectrum is given.

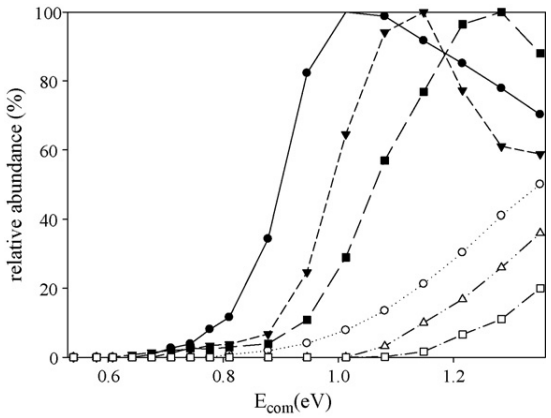

Figure 9 Appearance curves of high mass fragment ions in the tandem mass spectra of GroEL, GroEL:(gp5) and GroEL:$(gp5)_2$. The only high mass fragment ions observed are due to the sequential loss of one and two monomeric GroEL sub-units. The curves depict the observed relative ion intensities in case of GroEL (○,●) GroEL:(gp5) (Δ,▲) and GroEL:$(gp5)_2$ (□,■). The GroEL tridecamer fragment ions are depicted by solid symbols, whereas the dodecamer fragment ions are presented by open symbols. The tridecameric abundances are relative to their highest intensity, the dodecameric ions are relative to the corresponding tridecamer appearance.

GroEL subunits (open symbols) are depicted. We evaluated break-down graphs of different charge states of all three GroEL - GroEL:gp5 - GroEL:$(gp5)_2$ species, but always observed similar trends, the threshold for dissociation increases when one, and even more when two gp5 substrate molecules are bound to the GroEL machinery. We estimated the relative thresholds for the formation of tridecameric species. For GroEL this was 0.73 eV for GroEL:gp5 0.87 eV and for GroEL:(gp5)2 0.91 eV, using krypton as collision gas. In the elimination of the second monomer, and thus the formation of the dodecamer ions, the same trend is observed, with relative thresholds for elimination of the second monomer at 0.86 eV for GroEL, GroEL:gp5 1.01 eV and for GroEL:$(gp5)_2$ 1.06 eV confirming that the gas-phase stability of the complexes is increased by the binding of substrate molecules. This result may be not that surprising as the encapsulated substrate molecules will bind multivalently to the GroEL subunits of the rings, as one substrate molecule can have numerous interactions with multiple subunits of the ring. It is interesting to note that the in solution expected stabilizing effects of substrate binding are mimicked in this case in the gas-phase, providing further evidence that some of the solution-phase structure of larger protein complexes may be retained in the gas-phase (Ruotolo et al., 2005; Ruotolo and Robinson, 2006; van Duijn et al., 2006b).

CONCLUSIONS

Native electrospray ionization (ESI) mass spectrometry has become a viable method to study multimeric protein complexes of increasing size and complexity, which may be used to complement other techniques in structural biology.

In this paper we report on the low-cost modification and optimization of a first generation Q-ToF 1, whereby we enhance the instruments' capabilities to perform gas-phase dissociation experiments of macromolecular systems, such as large proteins and protein complexes. The modifications in the first hexapole ion guide, ion-selecting quadrupole, collision cell and ToF analyzer yielded a tandem mass spectrometer that offers excellent possibilities to study protein stoichiometries and gas-phase stabilities of large macromolecular complexes (Figure 1). The combination of an increased pressure in the ionization chamber and the locally increased pressure in the first stage of the hexapole ion bridge allowed the efficient cooling of large protein ions. We demonstrated experimentally that the low-frequency quadrupole allows ion isolation of ions up to m/z 12,800 (theoretically 30,000) and that the decreased repetition rate of the ToF analyzer allows analysis of ions up to m/z 38,150 (theoretically m/z 58,000). The modifications can be applied on every Q-ToF instrument and extend the possibilities of the instrument for native mass spectrometry significantly. The chemical architecture of such an assembly may be further probed by tandem mass spectrometry, which can dissect even large protein complexes of a few hundreds of kDa in a sequential fashion, helping to identify building blocks, and potentially the complex topology, quaternary structure and stability. Here we show that using a modified Q-ToF 1 such tandem mass spectrometry experiments may be even further optimized by using smaller entrance and exit orifices on the collision cell and heavier collision gas targets. With the standard used lighter argon as collision gas we observed that the dissociation was not efficient, but more noticeably we observed only very low abundant high mass fragment ions in the tandem mass spectra. These effects can be largely attributed to the decreased ability to focus ions of very high m/z before and after collisional activation, which may easily drift off-axis due to their relative low axial velocity. The increased dissociation using different gases is not solely due to better energy transfer with increasing collision gas mass which is shown by the non overlay when converted to center-of-mass collision energy, but also due to increased time available for reaction when using higher pressures and heavier gases (Mauk et al., 2002). We conclude therefore that the heavier xenon is the preferred collision gas for tandem mass spectrometry on very large native complexes.

EXPERIMENTAL SECTION

Chemicals and preparation of proteins

GroEL was overexpressed in Escherichia coli strain MC1009 containing plasmid pSL6. Cells were grown in Luria-Bertani (LB) medium containing 100 mg/l ampicillin and 0.0005 (w/v) arabinose at 37 °C under vigorous aeration. GroEL was purified according to a previously described protocol, slightly modified by the introduction of an acetone precipitation step (van Duijn et al., 2005; van Duijn et al., 2006a). Purified gp5 was a kind gift of C. Teschke (University of Connecticut, USA).

The buffer was exchanged to 50 mM ammonium acetate with a pH of 6.8 using ultra centrifugation filters with a cut-off of 10 kDa

(Millipore, Bedford) GroEL gp5 complexes were formed by first unfolding the gp5 in 8M urea for 30 min at room temperature at a final gp5 concentration of 25 µM. The unfolded substrate was added to 50 mM ammonium acetate buffer containing 28µM monomeric GroEL at ratios varying from 1:0 to 1:5. The resulting maximum concentration of 1.7 M urea did not affect the oligomeric GroEL structures. Excess of urea was removed from the sample by filtration while changing the buffer to 50 mM ammonium acetate. Final concentrations of the GroEL proteins on the basis of their monomeric masses were 6 µM. Cesium iodide, used for calibration of the mass spectrometer, and ammonium acetate were purchased from Sigma (Swijndrecht).

Mass spectrometry

All the mass spectrometry measurements were performed in positive ion mode. Proteins and protein complexes were introduced into the mass spectrometer by using borosilicate needles in the absence of any backpressure. The needles were pulled from borosilicate glass capillaries (Kwik-Fil, World Precision Instruments, Sarasota, FL) on a P-97 puller (Sutter Instruments, Novato, CA), coated with a thin gold layer by using an Edwards Scancoat (Edwards Laboratories, Milpitas, CA) six Pirani 501 sputter. All the mass spectra were calibrated using cesium iodide (5-25 mg/ml) in 50% (v/v) 2-propanol. The instrument used was a modified first generation Q-ToF 1 (Waters, Manchester, UK) equipped with a Z-spray source (see Figure1).

Electrospray ions were produced at atmospheric pressure after which the partly solvated ions were introduced into the first vacuum stage (P1) of the mass spectrometer through a sample cone with an orifice of 400 µm. The pressure in P1 was increased by reducing the pumping efficiency of the rotary pump to 10 mbar (unless stated otherwise in the text). The molecular beam of charged ions that evolves behind the sample cone was then extracted orthogonally through an extractor cone with an orifice of 1 mm into the second pumping stage (P2; 4 10^{-3} mbar). The cone voltage used was dependent on the type of protein complex and varied between 75 and 150 V. The ions that entered the second pumping stage (P2) were focused by a hexapole ion guide. Around the first part of the hexapole a flow-restricting sleeve in the form of a metal tube (25 mm id, 100 mm long) was installed to increase the pressure locally (P2A; estimated average pressure 8 10^{-3} mbar). In a slightly different form this modification has been reported by Chernushevich and Thomson (Chernushevich and Thomson, 2004). The third vacuum chamber (P3; 6.7 10^{-4} mbar) contains a quadrupole, a separate collision cell and a transport hexapole lens. The standard RF generator in the Q-ToF was replaced by a generator with a frequency of operation of 300 kHz. This sets the theoretical accessible transmission limit for a selected ion to 30,000 m/z according to equation 2. When acquiring ToF mass spectra the quadrupole operates in RF-only mode as a broad-band mass filter. However, in the tandem mass spectrometry mode the quadrupole transmits only ions in a narrow m/z window around a set m/z, thus operating as a narrow-band mass filter. Ions are then transferred to the hexapole collision cell, which was modified such that the collision gas pressure can be increased. The entrance and exit orifices of collision cell were varied between 1.5 and 2.0 mm, respectively, as described in detail in the results section. The voltage over the collision cell was varied between 0 and 200 V. To measure protein complexes we used a cylinder head pressure of 1.8 mbar admitted to the collision cell via a needle leak valve with a 100 µm capillary with a length of 1.25 m leading to a typical P4 pressure of around 2x10^{-2} mbar. Using argon the resulting pressure at P3 was 6.7x10^{-4} mbar in case of the 2 mm orifices and 1.5x10^{-4} mbar in case of the 1.5 mm orifices. After activation in the collision cell the ions are transferred to the ToF vacuum chamber (P5; 2 10^{-6} mbar). A multi-component ion lens focuses the ions and two of the meshes were changed from 1,000 to 200 lines/inch for a three-fold enhancement of the sensitivity. When the ions arrive in the pusher region they were pulsed orthogonally for ToF analysis. The ToF repetition rate was decreased to 410 µs in order to increase the maximum attainable m/z, which theoretically could be 58,000. When acquiring in ToF spectra the quadrupole operates in RF-only mode as a broad-band mass filter, however, in the tandem MS mode the quadrupole transmits only ions in a narrow m/z window around the set m/z, thus operating as a narrow-band mass filter. Ions are then transferred to the hexapole collision cell, which was modified such that the collision gas pressure can be increased. The entrance and exit orifices of collision cell were varied between 1.5 and 2.0 mm, respectively, as described in detail in the results section. The voltage over the collision cell was varied between 0 and 200 V. To measure protein complexes we used a cylinder head pressure of 1.8 mbar admitted to the collision cell via a needle leak valve with a 100 µm capillary with a length of 1.25 m leading to a typical P4 pressure of around 2 x 10^{-2} mbar. Using argon the resulting pressure at P3 was 6.7 x 10^{-4} mbar in case of the 2 mm orifices and 1.5 x10-4 mbar in case of the 1.5 mm orifices.

REFERENCES

Ashcroft, A.E. (2005). Recent developments in electrospray ionisation mass spectrometry: noncovalently bound protein complexes. Nat Prod Rep 22, 452-464.

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

Borch, J., Jorgensen, T.J., and Roepstorff, P. (2005). Mass spectrometric analysis of protein interactions. Curr Opin Chem Biol 9, 509-516.

Chernushevich, I.V., and Thomson, B.A. (2004). Collisional cooling of large ions in electrospray mass spectrometry. Anal Chem 76, 1754-1760.

Curcuruto, O., and Hamdan, M. (1993). Effects of target identity and collision energy on the fragmentation of some mass-selected protonated neuropeptides. Rapid Commun Mass Spectrom 7, 989-993.

Falke, S., Fisher, M.T., and Gogol, E.P. (2001). Structural changes in GroEL effected by binding a denatured protein substrate. J Mol Biol 308, 569-577.

Heck, A.J., and Van Den Heuvel, R.H. (2004). Investigation of intact protein complexes by mass spectrometry. Mass Spectrom Rev 23, 368-389.

Hernandez, H., Dziembowski, A., Taverner, T., Seraphin, B., and Robinson, C.V. (2006). Subunit architecture of multimeric complexes isolated directly from cells. EMBO Rep 7, 605-610.

Jones, C.M., Beardsley, R.L., Galhena, A.S., Dagan, S., Cheng, G., and Wysocki, V.H. (2006). Symmetrical gas-phase dissociation of noncovalent protein complexes via surface collisions. J Am Chem Soc 128, 15044-15045.

Krutchinsky, A.N., Chernushevich, I.V., Spicer, V.L., Ens, W., and Standing, K.G. (1998). Collisional damping interface for an electrospray ionization time-of-flight mass spectrometer. Journal Of The American Society For Mass Spectrometry 9, 569-579.

Leggett, G.J., Briggs, D., and Vickerman, J.C. (1990). Collision Target-Gas Effects During the Tandem Secondary-Ion Mass-Spectrometric Analysis of Polymers. Journal of the Chemical Society-Faraday Transactions 86, 1863-1872.

Loo, J.A. (1997). Studying noncovalent protein complexes by electrospray ionization mass spectrometry. Mass Spectrom Rev 16, 1-23.

Loo, J.A., Berhane, B., Kaddis, C.S., Wooding, K.M., Xie, Y., Kaufman, S.L., and 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.

Mauk, M.R., Mauk, A.G., Chen, Y.-L., and Douglas, D.J. (2002). Tandem mass spectrometry of protein-protein complexes: cytochrome c-cytochrome b5. Journal of the American Society for Mass Spectrometry 13, 59-71.

McCammon, M.G., Hernandez, H., Sobott, F., and Robinson, C.V. (2004). Tandem mass spectrometry defines the stoichiometry and quaternary structural arrangement of tryptophan molecules in the multiprotein complex TRAP. Journal Of The American Chemical Society 126, 5950-5951.

McKay, A.R., Ruotolo, B.T., Ilag, L.L., and Robinson, C.V. (2006). Mass Measurements of Increased Accuracy Resolve Heterogeneous Populations of Intact Ribosomes. J. Am. Chem. Soc. 128, 11433-11442.

Poliakov, A., Duijn, E.v., Lander, G., Fu, C.-y., Johnson, J.E., Prevelige, J.P.E., and Heck, A.J.R. (2007). Macromolecular mass spectrometry and electron microscopy as complementary tools for investigation of the heterogeneity of bacteriophage portal assemblies. Journal of Structural Biology 157, 371.

Ruotolo, B.T., Giles, K., Campuzano, I., Sandercock, A.M., Bateman, R.H., and Robinson, C.V. (2005). Evidence for Macromolecular Protein Rings in the Absence of Bulk Water. Science 310, 1658-1661.

Ruotolo, B.T., and Robinson, C.V. (2006). Aspects of native proteins are retained in vacuum. Curr Opin Chem Biol 10, 402-408.

Sanglier, S., Leize, E., Dorsselaer, A., and Zal, F. (2003). Comparative ESI-MS study of approximately 2.2 MDa native hemocyanins from deep-sea and shore crabs: from protein oligomeric state to biotope. J Am Soc Mass Spectrom 14, 419-429.

Schmidt, A., Bahr, U., and Karas, M. (2001). Influence of pressure in the first pumping stage on analyte desolvation and fragmentation in nano-ESI MS. Anal Chem 73, 6040-6046.

Sleno, L., and Volmer, D.A. (2004). Ion activation methods for tandem mass spectrometry. J Mass Spectrom 39, 1091-1112.

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

Sobott, F., McCammon, M.G., Hernandez, H., and Robinson, C.V. (2005). The flight of macromolecular complexes in a mass spectrometer. Philos Transact A Math Phys Eng Sci 363, 379-389; discussion 389-391.

Sobott, F., and Robinson, C.V. (2004). Characterising electrosprayed biomolecules using tandem-MS - the noncovalent GroEL chaperonin assembly. International Journal of Mass Spectrometry 236, 25-32.

Synowsky, S.A., van den Heuvel, R.H., Mohammed, S., Pijnappel, P.W., and 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.

Tahallah, N., Pinkse, M., Maier, C.S., and 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.

van den Heuvel, R.H.H., van Duijn, E., Mazon, H., Synowsky, S.A., Lorenzen, K., Versluis, C., Brouns, S.J.J., Langridge, D., van der Oost, J., Hoyes, J., and Heck, A.J.R. (2006). Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. Analytical Chemistry 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., and 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., and Heck, A.J. (2006a). 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., Simmons, D.A., van den Heuvel, R.H.H., Bakkes, P.J., van Heerikhuizen, H., Heeren, R.M.A., Robinson, C.V., van der Vies, S.M., and Heck, A.J.R. (2006b). Tandem mass spectrometry of intact GroEL-substrate complexes reveals substrate-specific conformational changes in the trans ring. Journal of the American Chemical Society 128, 4694-4702.

Wang, J., and Chen, L. (2003). Domain motions in GroEL upon binding of an oligopeptide. J Mol Biol 334, 489-499.

DETERMINATION OF STOICHIOMETRY AND CONFORMATIONAL CHANGES IN THE FIRST STEP OF THE P22 TAIL ASSEMBLY

Kristina Lorenzen[1], Adam S. Olia[2], Charlotte Uetrecht[1], Gino Cingolani[2] and Albert J.R. Heck[1]
[1]Biomolecular Mass Spectrometry and Proteomics Group, Bijvoet Center for Biomolecular Research, Utrecht Faculty of Science, Utrecht University, Sorbonnelaan 16, 3584 CA Utrecht, the Netherlands
[2]Department of Biochemistry and Molecular Biology, SUNY Upstate Medical University 750, E. Adams Street, Syracuse NY 13210, USA

SUMMARY
Large oligomeric portal assemblies play a central role in the life cycles of bacteriophages and herpesviruses. The stoichiometry of in vitro assembled portal proteins has been a subject of debate for several years. The intrinsic polymorphic oligomerization of ectopically expressed portal proteins makes it possible to form in solution rings of diverse stoichiometry (e.g. 11-mer, 12-mer, 13-mer, etc.) in solution. In this paper, we have investigated the stoichiometry of the in vitro assembled portal protein of bacteriophage P22 and characterized its association with the tail factor gp4. Using native mass spectrometry, we show for the first time that the reconstituted portal protein assembled in vitro using a modified purification and assembly protocol, is exclusively dodecameric. Under the conditions used here twelve copies of tail factor gp4 bind to the portal ring, in a co operative fashion, to form a 12:12 complex of 1.050 MDa. We applied tandem mass spectrometry to the complete assembly and found an unusual dimeric dissociation pattern of gp4, suggesting a dimeric sub-organization of gp4 when assembled to the portal ring. Furthermore, native and ion mobility mass spectrometry reveal a major conformational change in the portal upon binding of gp4. We propose, that the gp4 induced conformational change in the portal ring initiates a cascade of events assisting in the stabilization of newly filled P22 particles, which marks the end of phage morphogenesis.

INTRODUCTION

P22 is a short tailed, double stranded DNA bacteriophage of the Podoviridae family. The mature phage infects the gram - negative Salmonella enterica serovar Typhimurium (Israel et al., 1967), a common human pathogen that lives in the gut. The single particle asymmetric reconstruction of the phage P22 mature virion was recently reported (Chang et al., 2006; Lander et al., 2006). The virus capsid, ~65 nm in diameter, shows an icosahedral organization and is built up by 415 copies of coat protein gp5. A unique pentameric vertex of the icosahedral capsid is replaced by the dodecameric portal protein (also called gp1), which provides a surface for attachment to the tail apparatus. Overall, the P22 tail, also referred to as the portal vertex structure, contains six molecules of the trimeric tailspike protein: gp9 (Goldenberg et al., 1982), and several copies of each of the three tail accessory proteins (or "head completion proteins"): gp4, gp10 and gp26 (Casjens and Huang, 1982; Hartweig et al., 1986; Lander et al., 2006; Tang et al., 2005). The portal protein ring facilitates packaging of DNA into the capsid by use of ATP hydrolysis (Smith et al., 2001). The P22 chromosome is introduced into the virion through the portal ring in an enzymatic reaction that requires the terminase subunits gp2 and gp3 (Casjens et al., 2004). During DNA packaging, the DNA is encapsulated inside the phage in a quasi-crystalline state at concentrations about 500 mg/ml (Earnshaw and Casjens, 1980; Smith et al., 2001).

The outer minor structural proteins, gp4, gp10, and gp26 are required for sealing the capsid as well as may play an important role in penetrating the host cell envelope (Andrews et al., 2005; Bhardwaj et al., 2007; Moak and Molineux, 2004). Deletions in the genes encoding these factors do not affect assembly or packaging, but result in particles that leak their DNA into the surrounding media (Strauss and King, 1984). Therefore, these proteins are not required for DNA packaging, but rather to seal the portal channel. The last protein to add to the nascent tail, the tailspike gp9 (Steinbacher et al., 1997) binds to the lipopolysaccharide protruding out from the outer lipid bilayer of the Salmonella cell envelope. The exact assembly mechanism of phage P22 tail is still not completely understood and it remains difficult to study this molecular machine in virtue of the large mass (~2.8 MDa) and chemical complexity. To gain better insight, it is important to characterize the stoichiometry, oligomerization, and binding interactions between individual tail proteins. Mass spectrometry is an emerging tool for such an analysis. It has become a valuable addition to methods commonly used in structural biology. The coupling of gentle electrospray ionization with a time-of-flight (ToF) detector allows the analysis of non - covalent protein complexes in an environment where the proteins are likely to retain their quaternary structure (Benesch et al., 2006). In the nanoflow electrospray process the molecule of interest is gently transferred from an aqueous solution to the gas phase via protonation at atmospheric pressure (Fenn et al., 1989). Desolvation of the protein assemblies in the ion source interface generates multiply charged ions of the intact complexes prior to analysis by the mass spectrometer. Recently, combinations of quadrupole

and ToF analyzers (QToF) have been modified in such a way that the detectable mass range in electrospray ionization mass spectrometry exceeds several million Da, allowing the analysis of species as big as ribosomes and viruses (Benjamin et al., 1998; Bothner and Siuzdak, 2004; Sobott et al., 2002; van den Heuvel et al., 2006). Mass spectrometric detection of the assemblies has made it possible to obtain accurate information about protein complex stoichiometry, stability and dynamics (Heck and Van Den Heuvel, 2004; Loo, 1997; van den Heuvel and Heck, 2004). Some examples are the folding cycle of the GroEL-gp31 machinery which could be monitored while folding and the bacteriophage T4 capsid protein gp23 (van Duijn et al., 2005). Macromolecular mass spectrometry lends itself as an excellent tool to study protein complex assembly; in particular virus assembly where major questions focus on the early multi - protein intermediates or the stoichiometries of subcomplexes, such as the portal protein.

In this paper, we have used mass spectrometry and other biochemical methods to examine the oligomerization of phage P22 portal complex and its interaction with the tail accessory factor gp4, which is known to play a a role in portal protein closure. We show that the portal protein assembles uniquely into a dodecamer and that twelve gp4 bind to the portal ring, whereby the gp4 exhibits a dimeric sub organization. Additionally, we show with native mass spectrometry and ion mobility separation mass spectrometry that the binding of gp4 to the portal ring induces a major conformational change in the assembly, likely preparing the portal for further assembly with the tail accessory proteins.

RESULTS

Oligomeric state of in vitro reassembled gp1 portal ring

Ectopically expressed and reassembled portal proteins display significant structural polymorphisms in vitro. In case of the phage P22 portal protein it was reported that the recombinant rings form undecamers and dodecamers in a ratio of approximately 70:30 % (Lander et al., 2006; Poliakov et al., 2007). It is plausible that those different quaternary states of assembly are due to the oligomerization procedure used during the purification. Therefore, the oligomeric state of the portal protein in vitro likely reflects both the intrinsic plasticity of the protein as well as the specific expression and purification conditions used during reconstitution. It is possible that the presence of a six-histidine tag used in references (Lander et al., 2006; Poliakov et al., 2007) for quick purification of portal monomers may have interfered with the correct oligomerization process yielding in vitro rings of aberrant stoichiometries (e.g. 11-mers). To ensure complete, proper oligomerization of P22 portal protein, we first expressed an untagged truncation mutant of phage P22 portal protein (residues 1-602) that has been shown to attain relevant physiological properties (Bazinet et al., 1990; Olia et al., 2006). We then devised a novel purification and oligomerization strategy where portal protein monomers were concentrated to >200 mg/ml (2.2 mM) in 60 mM EDTA and subjected to heat shock at 37 °C followed by ultracentrifugation to remove precipitation products. While the EDTA treatment is thought to remove metal ions that potential affect the correct oligomerization process, the heat shock at 37 °C followed by high-speed ultracentrifugation was found to be critical to selectively remove missfolded or aggregated, likely aberrant portal rings. This procedure led to a single, distinct oligomeric state of the portal complex as evidenced by native mass spectrometry. In native mass spectrometry the transfer of the protein into the gas phase via electrospray ionization involves protonation and leads to the presence of the protein in different charge states. From the consecutive ion signals, the charge state and thus the exact mass for the sample of interest can be determined. Figure 1 shows native mass spectra of the portal ring (the charge state is indicated above the corresponding ion signal). We measured samples derived from the old strategy as reported (Lander et al., 2006; Poliakov et al., 2007) (Figure 1A) and purifications treated as described above (Figure 1B). The samples prepared by the original method showed still undecameric portal ring with a mass of 761.2 ± 0.2 kDa. The ratio between the undecamer and the dodecamer in the sample was 40:60 %. The portal protein, assembled as described here, was unambiguously shown to be dodecameric (Figure 1B). The total mass measured for the latter complex was 830.2 ± 0.2 kDa in agreement with twelve copies of 69.2 kDa (predicted from the sequence).

Assembly of gp4 to the portal ring

Gp4, an 18.3 kDa P22 encoded protein, is the first tail factor binding to the nascent tail during virus morphogenesis (Strauss and King, 1984). Olia et al. reported that gp4 exists as a monomer in solution (Olia et al., 2006), which is highly sensitive to temperature denaturation. As detected by mass spectrometry, purified gp4 at a final concentration of 10 µM in 100 mM ammonium acetate is mainly monomeric. About 8 % was detected as a dimer of gp4, but no higher oligomerization states were observed (data not shown). A concentration artifact cannot be ruled out but is unlikely.

To characterize the assembly of recombinant gp4 to dodecameric portal protein, a portal:gp4 complex was formed in the presence of a 3-fold excess of gp4. The excess gp4 was separated from the portal:gp4 complex using gel filtration chromatography. The sample was stored at 4 °C prior to analysis. Homogeneously assembled portal:gp4 complex was buffer exchanged against 100 mM ammonium acetate buffer pH 6.8 and the sample was directly introduced into the mass spectrometer at a concentration of 0.8 µM. The assembly appeared to be very heterogeneous with mass differences among the different species of 2.8 kDa. Since the portal assembly showed a homogeneous dodecamer when analyzed individually (Figure 1B), we concluded that the

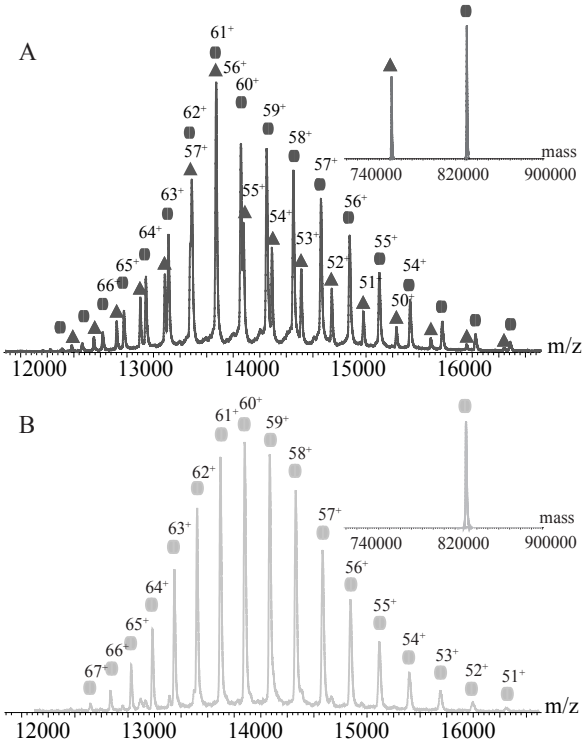

Figure 1 Mass spectrometric analysis of oligomeric P22 portal protein. (A) shows the portal purified and assembled according to previous literature (Lander et al., 2006; Poliakov et al., 2007). The triangles indicate the distribution originating from the undecameric portal ring and the circles the dodecameric portal ring. The charge state is indicated above the corresponding peaks. (B) shows the portal purified as described in this article. The homogeneous distribution originates from dodecameric portal ring. The charge state is indicated above the corresponding peak. The insets show deconvoluted mass spectra, representing the abundance of the different species/distributions.

heterogeneity seen in the portal:gp4 complex originated from degradation products of gp4. These retain the ability to bind to the portal ring, or possibly get degraded after the assembly. To test this hypothesis we first determined the exact masses of the putative gp4 degradation products. This was done by dissolving the portal:gp4 complex in 50 % acetonitrile, 0.2 % acetic acid. Under these conditions the complex falls apart, the proteins unfold and gain more charges during the electrospray process, which results in more and narrower peaks, allowing a more precise determination of the monomeric mass. The measurements under denaturing conditions showed that there were two forms of gp4 present in the complex: the expected complete protein with a mass of 18,383 ± 1 Da and a truncated form with a lower mass of 15,570 ± 1 Da. The difference in mass of 2,813 Da fitted to the difference detected in the assemblies of the portal with the gp4 wild type (wt) (data not shown). With such accurate masses of the monomeric protein, we were able to determine that a unique cleavage event takes place in gp4 between Ala141 and Asp142 (Figure 2A). Thus all complexes observed in the native mass spectrometry measurements where 12:12 portal protein:gp4 assemblies with a mix of full and truncated gp4 binding to the portal ring.

The C - terminal moiety (res. 142-166) of gp4 is dispensable for binding to the portal ring

The idea that the C-terminal tail of gp4 is dispensable for binding to the portal ring was further investigated by engineering two new constructs of gp4. In the first mutant, we sought to reduce the flexibility at position 142 of gp4, which likely causes the protein to be degraded. Alanine 141 was therefore mutated to proline, which is a less flexible amino acid (Huang and Nau, 2003) (gp4 $141^{A\rightarrow P}$). In the second mutant, we deleted the entire C terminus of the protein ending with amino acid 141 (gp4-142^{STOP}). Both gp4 constructs were expressed and purified to homogeneity (Figure B). To test whether the two mutants exhibit similar binding behavior as the gp4 (wt) we titrated the two gp4 mutants to the dodecameric portal ring and analyzed the formation of a portal:gp4 complex by native electrophoresis on agarose gels. Similar to gp4 wt, the two gp4 variants showed uniform and saturable binding to the dodecameric portal ring when 12 equivalents of gp4 were added to a dodecameric portal protein ring (Figure 3A). Interestingly, the shorter deletion construct gp4-142^{STOP} shifted the portal ring to a greater degree than the wt or the gp4-$141^{A\rightarrow P}$ mutant. This incre-

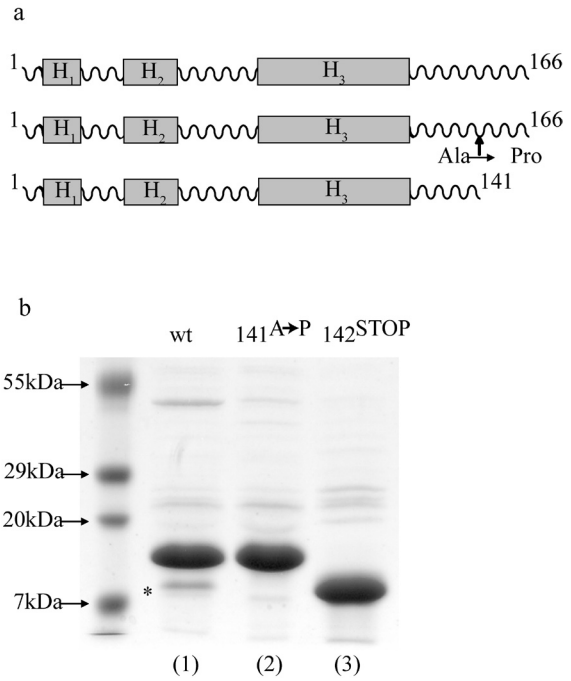

Figure 2 Design of gp4 constructs with altered C-terminal region. (A) Schematic diagram of gp4 predicted structural topology and the two constructs designed in this study. (B) SDS-PAGE of purified gp4 wt and the two constructs gp4-142^{STOP} and gp4-$141^{A\rightarrow P}$. The asterisk in lane 1 indicates the degradation product at position 141, which yields a product of the same size as the construct gp4-142^{STOP}

ased shift could be indicative of either soley a difference in exposed charges (due to the truncated residues), or possibly, the formation of a different final product. To determine the binding behavior more precisely as well as to confirm that the behavior of gp4-142^{STOP} was not due to any differences in binding, we carried out isothermal titration calorimetry (ITC) experiments. Either purified gp4 or gp4-mutants were titrated into a cell containing homogeneous dodecameric portal rings at an oligomer concentration of 6.6 µM (Figure 3B). The results showed no significant differences in binding. Both mutants and gp4 wt show the same bi-modal binding characterized by a first weak interaction followed by a second stronger interaction as reported previously (Olia et al., 2006). This indicated that the increased retardation by gp4-142^{STOP} on the native agarose gel is likely due to the removal of charged residues in the truncated C-terminal.

Stoichiometry of portal protein bound to gp4-141$^{A\rightarrow P}$ and gp4-142^{STOP}

We next investigated the portal:gp4 141$^{A\rightarrow P}$ and the portal:gp4 142^{STOP} assemblies by mass spectrometry. After formation the complexes were stored at 4 °C until measured. Directly before introducing the samples into the mass spectrometer they were buffer exchanged in ultrafiltration units against 100 mM ammonium acetate pH 6.8. The results of the measurements are shown in Figure 4. The portal:gp4 141$^{A\rightarrow P}$ assembly clearly showed binding of twelve gp4 141$^{A\rightarrow P}$ to the dodecameric portal ring. There are two major species visible in the spectrum. One originates from twelve complete gp4-141$^{A\rightarrow P}$ proteins (18,336 ± 3 Da) binding to the dodecameric portal and the second is eleven complete gp4 141$^{A\rightarrow P}$ and one truncated gp4 141$^{A\rightarrow P}$ binding to the dodecameric portal, which is 2.8 kDa smaller. This becomes evident from the two distributions visible in the spectrum. The larger complex has a mass of 1,050.7 ± 0.3 kDa, which coincides with the mass of twelve times portal protein (69.2 kDa) and twelve times gp4 141$^{A\rightarrow P}$. The minor distribution has a mass of 1,047.9 ± 0.2 kDa, which is about 2.8 Da smaller. This mass difference was already encountered when the measurements on the portal assembled with gp4 wt were done.

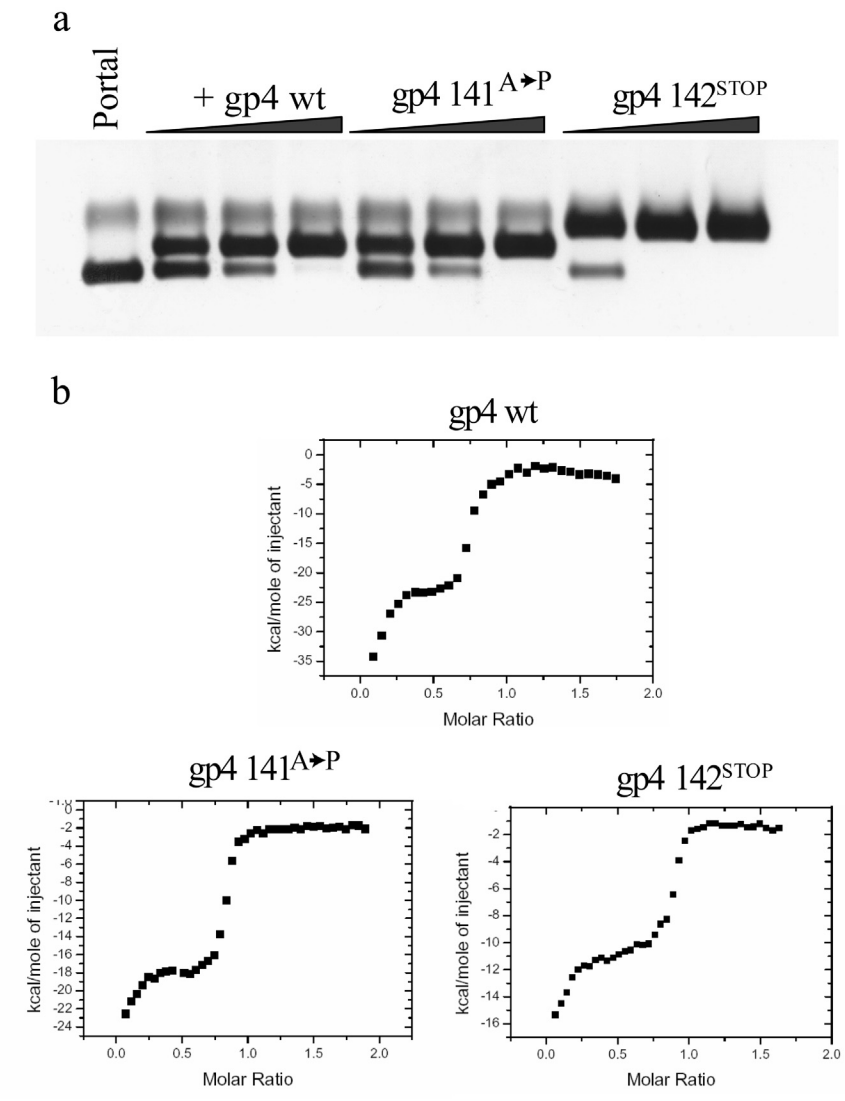

Figure 3: Biochemical analysis of gp4 and gp4 mutants binding to the portal ring. (A) Native agarose gel of different gp4 mutants binding to the portal ring. The portal ring is on the first lane, followed by titrations of the different gp4 constructs as indicated above the lanes. There is a uniform binding for all different gp4 to the portal ring as confirmed by ITC. The slower migration of the portal ring:gp4 142^{STOP} is likely due to fewer charges on the surface of the assembly. (B) Integrated peak heights from ITC isotherms obtained by injecting gp4 wt, gp4 141$^{A\rightarrow P}$, or gp4 142^{STOP} into a solution of dodecameric portal protein. All three binding isotherms clearly demonstrate similar binding reactions, which strengthen the idea that the C terminus of gp4 is dispensable for assembly to the portal protein.

Additional mass spectrometric experiments confirmed that the mass difference was again due to degradation of gp4. However, this sample was less heterogeneous compared to the portal:gp4 wt, which showed up to six truncated gp4 wt bound to the dodecameric portal. Here, only one truncated gp4 $141^{A\rightarrow P}$ was detected. In the mutant, the distribution with the degraded gp4 bound is less than 30 % of the total distribution, indicating that the mutation $141^{A\rightarrow P}$ had effectively reduced the flexibility of gp4, and consequentially, its degree of proteolysis. The spectrum of the portal:gp4 142^{STOP} showed a single distribution with a mass of 1,016.5 ± 0.3 kDa, which is in good agreement with the predicted mass of a dodecamer of portal protein bound to twelve gp4 142^{STOP} (15,498 ± 3 Da). The fact that no degradation was observed in the portal:gp4 142^{STOP} assembly confirms the truncation site of gp4 wt being alanine 141.

Mass spectrometry reveals no intermediate state in the assembly of gp4 to the portal

It has been reported before that the binding of gp4 to the portal ring may at elevated temperatures proceed through formation of an assembly intermediate, which contains 4-6 copies of gp4 bound to the dodecameric portal ring (Olia et al., 2006). This putative $(portal\ protein)_{12}$:$(gp4)_{4\text{-}6}$ assembly intermediate was identified on an agarose gel based on its altered migration and was found to be strongly temperature dependent. We assembled the portal ring with gp4 wt directly before the mass spectrometric measurement in different ratios, starting at a ratio of 1:2 and up to an excess of 5 equivalents of portal protein per copy of gp4. When stoichiometric amounts of gp4 were added to the portal, the favored stoichiometry detected by mass spectrometry was twelve copies of gp4 binding to a dodecamer of portal protein. There were minor distributions visible showing eleven copies of gp4 binding to the portal ring. When sub stoichiometric amounts of gp4 were added to the dodecameric portal ring the main distribution was still the 12:12 (portal protein:gp4) assembly with minor distributions of 12:11. At those concentrations, there was also free portal ring present, however no intermediate with a 12:6 or 12:4 (portal protein:gp4) stoichiometry (Figure 5). Our experiments indicate a high cooperative binding of the gp4 to the portal ring under the conditions used. However under the experimental conditions used here (at laboratory temperature) the putative assembly intermediate might be not significantly populated. The data presented here are not in contrast to the ITC experiments (Olia et al., 2006) which also failed to reveal an intermediate at 20°C.

Dissociation of the portal:gp4 complex reveals a dimeric structural organization of gp4

We applied tandem mass spectrometry to the portal ring and the complex of the portal ring with twelve gp4 bound to it. As mentioned above and labeled in Figure 1, the transfer of proteins into the mass spectrometer and the protonation

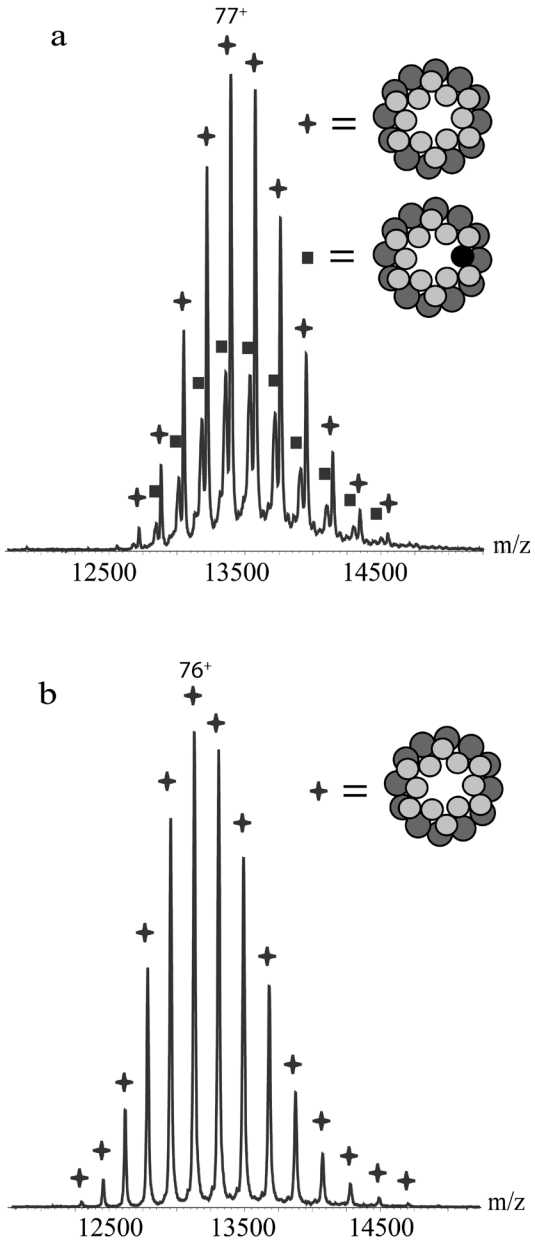

Figure 4 Mass spectra of the portal ring bound to gp4 C-terminal mutants. (A) Distributions of the gp4 $141^{A\rightarrow P}$ bound to the portal ring. Marked with squares are the charge states of the assembly where eleven complete and one truncated gp4 $141^{A\rightarrow P}$ are bound to the dodecameric portal ring. The stars show the distribution of twelve complete gp4 $141^{A\rightarrow P}$ bound to the portal ring. (B) In the assembly of the gp4 142^{STOP} with the portal ring no truncation products of the gp4 142^{STOP} are present. The stars indicate the charge states of the complex with twelve gp4 142^{STOP} bound to the dodecameric portal ring. The charge state of the most intense peak is indicated above each distribution. The dark gray circles in the schematic Figures indicate the portal protein and the light gray circles indicate the gp4 mutants. The black circle indicates the truncated gp4 $141^{A\rightarrow P}$. The schematics do not show a structural but rather only a stoichiometrical representation of the complexes.

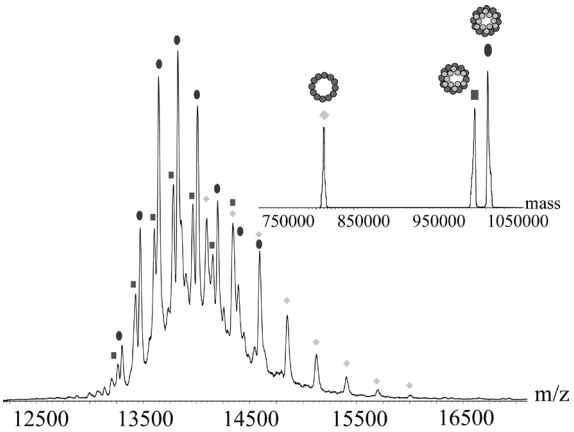

Figure 5 Addition of gp4 to the portal ring in a ratio of 1:3 (gp4:gp1). The mass spectrum shows mainly three distributions belonging to the free portal ring (diamonds), the complex where twelve gp4 are bound to the portal ring (ellipse) and the distribution originating from the assembly where eleven gp4 are bound to the portal ring (squares). The dark gray circles in the schematic Figures indicate portal protein and the light gray circles indicate gp4. The inset shows a deconvoluted mass spectrum. The schematics do not show a structural but rather only a stoichiometrical representation of the complexes.

lead to the presence of the protein in different charge states. Tandem mass spectrometry allows selection of one charge state, which can be collisionally activated in the mass spectrometer, resulting in mass selected dissociation of precursor ions. In these experiments, the precursor ions generally eliminate a single subunit, which is thought to unfold, and thus become highly charged, combined with the formation of the relatively low charged complex lacking the eliminated subunit (Heck and Van Den Heuvel, 2004). Evidently, the summed mass and charge of the two formed fragment ions should add up to the mass and charge of the precursor. When we selected one of the charge states of the dodecameric portal ring as the precursor, we saw the sequential dissociation of up to two monomeric subunits, resulting in the formation of undecameric and decameric portal assemblies, as expected. However, when we applied these measurements to the different assemblies of the portal ring with the gp4 wt and mutants they all showed an uncommon dissociation behavior next to the usual elimination of a single subunit (Figure 6). The known behavior resulted in the elimination of a monomeric subunit (circles) and formation of a complex of 12:11 portal:gp4 142^{STOP} (triangles). Additionally the elimination of a dimer (double circles) from the portal:gp4 142^{STOP} complex and concomitant formation of 12:10 portal:gp4 142^{STOP} (squares) occurred. This finding is very interesting as it gives additional information about the interactions between the involved proteins, indicating that the interaction between the gp4 dimers is stronger than the one of the gp4 dimers with the portal ring. It cannot be excluded that the formation of 12:10 portal:gp4 142^{STOP} additionally originates from loss of two monomers sequentially.

Gp4 binding induces a major conformational change in the dodecameric portal

The amount of charges a protein gathers during the electrospray process, used in native mass spectrometry, is dependent on the buffer conditions, the hydrophobicity and the general conformation of the protein. The amount of charges depends mainly on its accessible surface area and is not so much influenced by charged amino acids on the surface (Kaltashov and Mohimen, 2005; Konermann, 2007; Patriksson et al., 2007). It has been reported that the average amount of charges a globular protein or protein complex obtains during the electrospray process under native conditions can be estimated by the formula:

$$z = 0.078 \sqrt{m} \quad [eq. 1]$$

(de la Mora, 2000), where z is the charge of the complex and m is the mass. Catalina et al. showed the applicability of the theory (Catalina et al., 2005). The formula has its origin in the droplet formation during the electrospray process and thus is directly connected to the protein surface area (Pringle et al., 2007). In Figure 7B the experimental average charge state of a whole range of proteins and protein complexes measured under similar conditions is plotted versus the theoretical surface area of globular proteins/protein complexes. We assumed all proteins are globular with the average density of 0.78 g/cm3 (Benesch et al., 2007). With the mass known we calculated the surface a sphere with that mass and density would have. A list of the proteinscan be found in (Uetrecht et al., 2008). The calculated average charge state of the portal is expected to be 71. The measured charge distribution of the portal is quite broad, but has a most abundant charge state of 60, an unusual large difference of about 11 charges. The average charge state for the three measured portal:gp4 assemblies, is 76 – 77, in fair agreement with the predicted average of 79 – 80. The measured deviation in charge observed for the portal is pronounced, as can be seen in Figure 7B, where it deviates strongly from all other data points. To illustrate this deviation more clearly we overlaid the spectra of GroEL, the portal ring and the portal:gp4 142STOP assembly (Figure 7B). In case of GroEL, which is similar in mass to the portal ring, the measured charge states match the predicted value. As expected (seen in Figure 7B and predicted by eq. 1) the portal:gp4 142^{STOP} assembly attains more charges than GroEL, just by the fact that it is larger. The outlier is the free portal, which attains relatively few charges in the electrospray process, shifting its appearance in the mass spectra to m/z values not only higher than those of GroEL, but even higher than the portal:gp4 142^{STOP} assembly. From this data we conclude that the free portal has a more compact structure and therefore a smaller surface area than would be estimated. Upon binding of gp4, the portal undergoes a major conformational change. Thus next to direct changes in the portal:gp4 interface region we predict

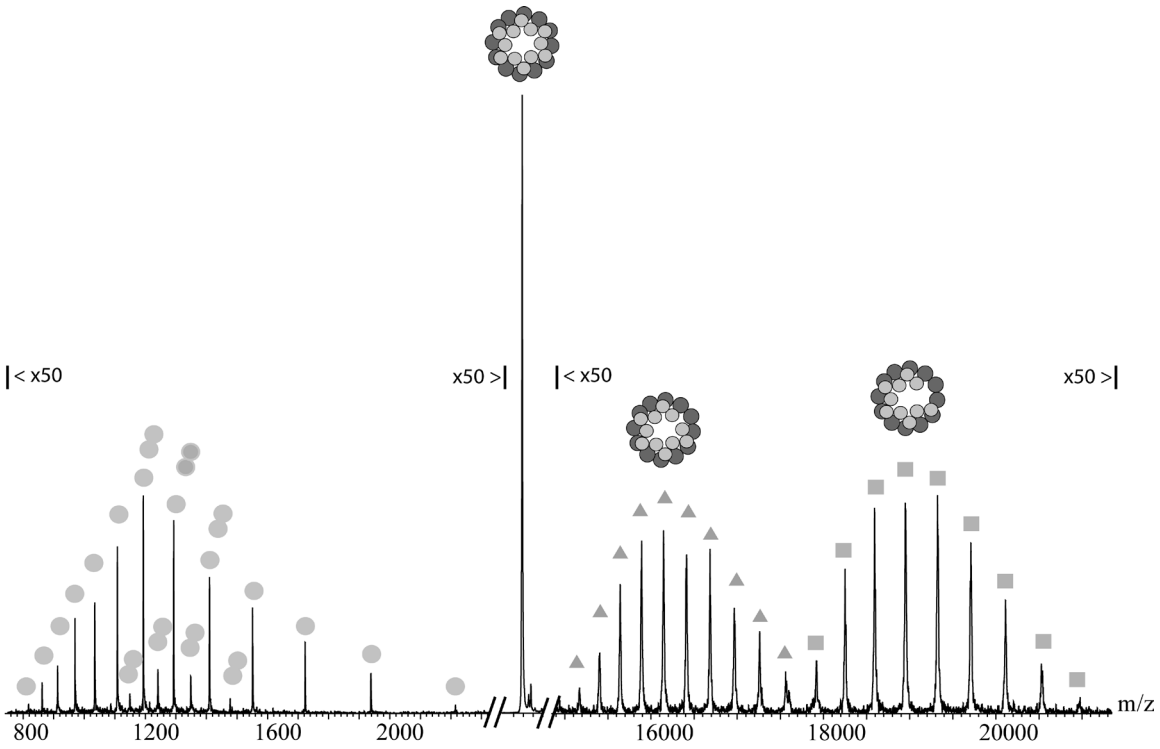

Figure 6 Tandem mass spectrum of the gp4 142^{STOP} bound to dodecameric portal ring (charge state 76+). On the left side of the spectrum the eliminated gp4 142^{STOP} monomer (circles, average charge state 13+) and dimer (double circles, average charge state 25+) are visible; in the middle the precursor is indicated; the right side of the spectrum shows the distributions of the concomitant high mass fragments where either the monomer (triangles, average charge state 63+) or the dimer (squares, average charge state 51+) was eliminated. The high and low m/z areas are magnified fifty fold as indicated. The dark gray circles in the schematic Figures indicate portal protein and the light gray circles indicate gp4 142^{STOP}. The schematics do not show a structural but rather only a stoichiometrical representation of the complexes.

conformational changes in other regions of the portal ring, possibly preparing the portal for further assembly or DNA binding.

We further investigated this conformational change using ion mobility separation mass spectrometry (IMS). In IMS the ions are separated by their mass to charge ratio like in conventional mass spectrometers, but IMS additionally allows separation of the ions dependent on their drift time (Koeniger and Clemmer, 2007; Koeniger et al., 2006; Ruotolo and Robinson, 2006). Molecules with compact shapes or volumes have a shorter drift time than molecules with a large cross section, assuming they have the same mass and charge. Thus, with the drift time of a certain complex, predictions can be made about their (relative) cross sections, if the influence of mass and charge state are taken into account (for more detail see Material and Methods). Approximation of absolute cross sections and volumes of proteins can be obtained upon calibration with a protein of known cross section (Ruotolo et al., 2005). In Figure 7C the mass and charge corrected drift times are depicted for GroEL, the portal and the portal:gp4 142^{STOP} assembly. It is interesting to note that the free portal has a smaller cross section and volume than GroEL, while it is 30 kDa "bigger" in mass, validating a rather compact structure for the free portal. We determined calibrated volumes for a range of proteins and protein complexes (Uetrecht et al., 2008) (for more detail see Material and Methods). In Figure 7D the determined volumes of all these proteins and protein complexes are plotted versus their mass. It can be seen, in agreement with previously reported data (Kaddis and Loo, 2007; Ruotolo et al., 2005), that there is a correlation between the mass and the cross section/ determined volume. However, the portal ring and the portal:gp4 assemblies do not match the correlation. Whereas the volume for the portal ring lies below the trendline, the portal:gp4 142^{STOP} assembly is found above the trendline (see Figure 7D). Both, the unusual behavior during the electrospray process and the deviations in their volumes/ cross sections, strongly suggest a major conformational change occurring in the portal ring upon gp4 binding.

DISCUSSION

Characterizing the assembly of bi-dodecameric portal protein:gp4 complex

Portal proteins are oligomeric molecular machines that have been shown to reassemble into polymorphic rings of diverse stoichiometry in vitro. For instance, phage SPP1 portal is dodecameric in the mature phage, but forms tridecamers when reassembled in vitro (Lurz et al., 2001). Likewise pha-

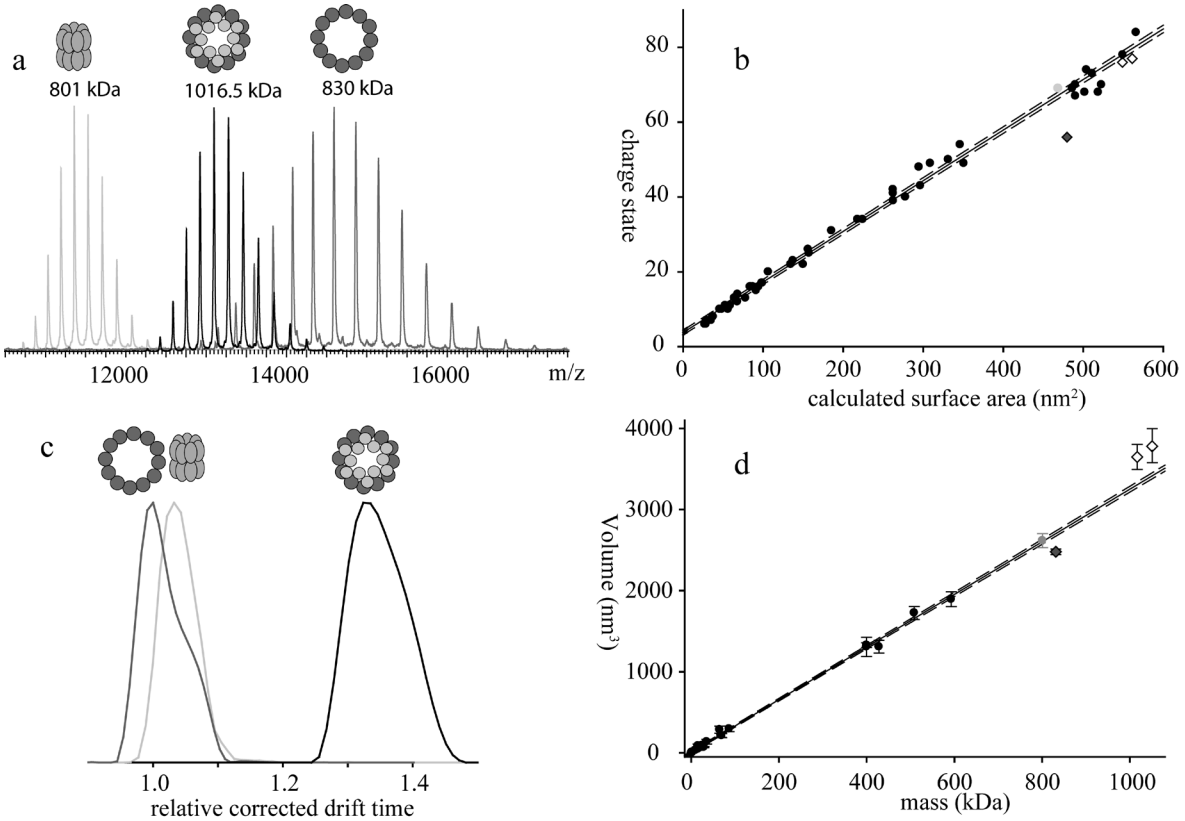

Figure 7 Comparison between the portal ring and the portal:gp4 142^{STOP} assembly. (A) An overlay of mass spectra from GroEL (light gray), dodecameric portal ring (gray) and the assembly of the dodecameric portal ring to gp4 142^{STOP} (black). Even though the mass of the assembly is 186 kDa higher than the portal ring alone, the distribution shifts to a lower m/z indicating a major conformational change. GroEL has a similar mass as the portal ring and represents a typical charge distribution for a protein of that mass. (B) Linear fit representing the calculated surface area of a protein complex against its expected average charge state. The circles are proteins used to generate the fit (with GroEL in grey), the diamonds the portal ring (grey), the portal:gp4 141$^{A\rightarrow P}$ and the portal:gp4 142^{STOP} (white). The solid line represents the fit through the standard proteins, the dashed lines give the 95 % confidence interval based on the fit. The standard deviations between measurements are around 1% of the averaged charge state. The proteins used to generate this graph are available in supplementrary table 1. (C) Mass and charge corrected drift times for the main charge state of GroEL (light grey), the dodecameric portal ring (dark grey) and the portal:gp4 141$^{A\rightarrow P}$ assembly (black). The values are normalized to the portal ring. The absolute cross sections (± standard deviation) are: portal (221 ± 2 nm^2), GroEL (230 ± 5 nm^2) and portal:gp4 142^{STOP} (286 ± 8 nm^2). (D) Volumes determined from the absolute cross sections of proteins and protein complexes (circles, with GroEL in grey), the dodecameric portal ring (grey diamond), portal:gp4 141$^{A\rightarrow P}$ and portal:gp4-142^{STOP} complexes (white diamonds). All these volumes were averaged over all detected charge states. The error bars represent the standard deviation. The solid line represents the fit through the set of standard proteins and protein complexes, the dashed line indicates the 5 % confidence interval based on the fit. The schematics do not show a structural but rather only a stoichiometrical representation of the complexes.

ge P22 portal appears dodecameric both in the asymmetric reconstruction of the mature virion (Lander et al., 2006), and in the crystal of the isolated portal connector (Cingolani et al., 2002), but was found to form undecamer and dodecamer when reassembled in vitro (Poliakov et al., 2007).

Using mass spectrometry, we show that the discrepancy between the in vitro versus the in vivo assembled portal rings is due to the assembly process utilized during purification. Consequently, optimization of the protocol leads to homogeneous dodecameric portal rings.

We have used the physiologically significant dodecamer of the portal mutant (Bazinet et al., 1990) to study the assembly of gp4, the tail factor known to initiate portal protein closure (Strauss and King, 1984). Gp4 is mainly monomeric in solution and binds the portal ring in a 1:1 (or 12:12) stoichiometry. This stoichiometry has been suggested by Olia et al. with ITC experiments. However mass spectrometry is a far more accurate method and we can exclude that there are any other conformations present after the finished assembly. Furthermore we can conclude that gp4, under the conditions used here, displays high binding affinity for the C-terminally deleted portal protein Δ602, which suggests that the C-terminus of portal protein is not required for binding to gp4. This agrees well with the observation that in vivo, the carboxyl terminal end of portal protein is dispensable both for assembly into the virion as well as virus infectivity (Bazinet et al., 1990). In gp4 the helix spanning motive ranging from amino acid 76 to 126 is crucial for binding to the portal ring. A mutated form of gp4 ending at amino acid 126 shows reduced binding to the portal ring (Olia et al., 2006). In this study gp4 assembled to portal protein in a degraded form. A C-terminal deletion construct of gp4 lacking residue

142-166 displays wild type binding to portal protein. This points out an important role of the residues between amino acid 126 and 142 in binding and suggests that the region between 142 and 166 is dispensable. When we assembled gp4 wt directly with the portal ring no degradation products of the gp4 were observed (see also Figure 5). In contrast to this the sample of portal:gp4 wt which was preassembled and stored at 4 °C until measurement in the mass spectrometer showed heavy degradation. This is suggesting that the cleavage of gp4, which we have seen before, happens after the assembly is finished but is not due to the mass spectrometric experiment.

Tandem mass spectrometry revealed the dissociation of a gp4 dimer from the portal:gp4 complex. Elimination of protein dimers is rarely observed in tandem mass spectrometry of protein complexes, since subunits largely unfold upon dissociation. To our knowledge there are only two examples in literature where a dimer elimination from a macromolecular complex has been described (Lorenzen et al., 2007a; van den Heuvel et al., 2006). The observation of gp4 dimers dissociating from the assembly strongly hints at a non uniform binding of gp4 to the portal ring and indicates a dimeric structural organization of gp4 when bound to the portal. The idea of a dimeric gp4:gp4 interface agrees well with the symmetry mismatches observed in the P22 tail, where a hexamer of gp10 attaches to a dodecamer of gp4 (Olia et al., 2007).

It has been suggested that the association of gp4 to dodecameric portal protein proceeds via formation of an assembly intermediate formed by portal ring occupied with 4-6 gp4 equivalents (Olia et al., 2006). In this study, we added gp4 to the assembled portal ring just prior to measurement to determine the stoichiometry of the precomplex. In our experiments, however, we did not see an intermediate binding state. When gp4 was added to the portal ring in substoichiometric amounts only assemblies of 12:11 or 12:12 portal protein to gp4 and free portal ring was present under the conditions used here. We were able to show that under the conditions used here the 1:2 (gp4:gp1) stoichiometry is not a stable intermediate state that occurs when substoichiometric amounts of gp4 are titrated to the portal, but that rather a complete assembly takes place leaving free portal ring behind. The fact that the assembly is complete within a minute even at low substoichiometric amounts of gp4 supports a cooperative form of binding of the tail accessory factor gp4 to the portal ring. We cannot completely exclude that an intermediate binding state exists, since it has been shown that this intermediate is very temperature sensitive and the conditions for the mass spectrometric measurements do not allow to keep the temperature stable at 30 °C, where the intermediate is populated.

Is the gp4-induced conformational change of the portal the missing link in head closure?

It has been shown that the portal ring has a different conformation in the mature capsid compared to its isolated form in vitro (Chang et al., 2006; Lander et al., 2006). The conformational change observed here might explain this difference. Upon packaging of the DNA into the head, the pressure exerted by the DNA highly condensed into the capsid increases accordingly. Gp4 is the first tail factor to initiate portal protein closure, which also requires tail accessory factors gp10 and gp26 (Strauss and King, 1984). Our data show that gp4 induces a conformational change in the portal protein upon assembly, which likely stabilizes a closed conformation of the portal protein, facilitating the sealing of DNA in the capsid and assembly to the other tail factors. Further structural investigation by cryo-electron microscopy and X-ray crystallography will be critical to delve into the nature of the gp4-induced conformational change in portal protein and its functional role in the stabilization of newly filled phage P22 heads.

EXPERIMENTAL PROCEDURES

Expression and purification of recombinant proteins

Recombinant proteins were expressed in E. coli (strain BL21) cells in LB broth supplemented with 2.5 g/l glucose. After growth at 37 °C to an OD600 of 0.6, gp4 or truncated gp1 (physiologically significant portal mutant (Bazinet et al., 1990)) expression was induced with 0.5 mM IPTG and the culture was incubated, shaking at 22 °C for 16 h. For gp4, the cells were collected and lysed by sonication in Lysis Buffer (250 mM NaCl, 20 mM Tris-HCl pH 8.0, 3 mM ß-mercaptoethanol (β-ME)) plus various protease inhibitors. Gp4, fused to an N-terminal-6xHis tag, was purified by metal chelating affinity chromatography using Qiagen Nickel-NTA beads. Typically, one liter of E. coli yielded about 5 mg of pure gp4, which was concentrated to approximately 50 mg/ml using a Millipore-Amicon concentrator (MW cutoff 10 kDa). Freshly eluted and concentrated protein was further purified by gel filtration on a Superdex S-200 column (GE Healthcare) equilibrated in Gel Filtration Buffer (200 mM NaCl, 20 mM Tris-HCl pH 8.0, 3 mM β-ME, 0.1 mM PMSF). Calibration of the S-200 column was carried out using high molecular weight globular protein standards (BioRad). Gp4 C-terminal mutants, gp4 142^{STOP} and gp4 $14^{1A\rightarrow P}$, were made by introduction of a nonsense mutation at position 142 in the case of gp4 142^{STOP} and an alanine to proline mutation at position 142 in the case of gp4 $141^{A\rightarrow P}$ using the Qiagen Quikchange Mutagenesis Kit. All plasmids were sequenced to confirm the fidelity of the gene. Mutated proteins were expressed and purified according to the same protocol as gp4 wt.

Recombinant untagged portal protein was expressed and purified using a protocol modified from (Olia et al., 2006). After induction, cells containing recombinant portal protein were lysed in 300 mM NaCl, 20 mM Tris HCl pH 8.0, 5 mM β-ME. The lysed cells were centrifuged at 20,000 x g and the supernatant was recovered. Solid ammonium sulfate was added to 30 % saturation, and after 1 h incubation, the solution was centrifuged at 18,000 x g. The resulting pellet contained relatively pure portal protein, which was then resuspended and immediately desalted using a Bio-Rad Desalting Column into 100 mM NaCl, 20 mM Tris-HCl pH 8.0, 10 mM β-ME, 60 mM EDTA. The desalted protein was then concentrated to >200 mg/ml and incubated at room temperature for 48 h. This resulted in a dramatic enrichment of higher order oligomers of the portal

protein. The protein was then incubated at 37 ºC for 3 h, during which massive precipitation occurs. The protein was clarified by a single round of ultracentrifugation at 100,000 x g for 35 min. This procedure results in the production of homogeneous dodecameric portal protein, as confirmed by native agarose gel as well as mass spectrometry.

Native Agarose Gel Electrophoresis and Isothermal Titration Calorimetry

Native agarose gel electrophoresis was carried out as described in (Olia et al., 2006). Protein samples and mixtures were allowed to incubate for 1 h at room temperature before analysis on agarose gel. For ITC (Isothermal Titration Calorimetry) experiments, all samples were dialyzed extensively against 200 mM NaCl, 20 mM Tris-HCl pH 8.0, 2 mM ß-ME buffer prior to each experiment. ITC experiments were carried out at 30 °C using a 300 µl syringe with a rotation speed of 295 rpm in the VP-ITC (Microcal, Northampton, MA). Gp4 (wt, gp4 $141^{A\rightarrow P}$ or gp4 142^{STOP}) was injected in 5 µl increments at a concentration of 245 µM into 1.8 ml of portal protein at an oligomer concentration of 6.6 nM with a spacing of 360 s between consecutive injections. Titration data were analyzed using the Origin 7.0 data analysis software (Microcal Software, Northampton, MA).

Mass spectrometry measurements

The sample buffer was exchanged five times to 100 mM ammonium acetate pH 6.8 using centrifugal filter units with a cut-off of 5 kDa (Millipore, UK). The concentration used for the mass spectrometry measurements was ~10 µM (based on monomeric masses). The samples were measured on an LCT for the denaturing conditions and on a modified Q-ToF (Quadrupole Time of Flight) for native mass spectrometry and tandem mass spectrometry (van den Heuvel et al., 2006) (Waters, UK). Nanospray glass capillaries were used to introduce the samples into the Z spray source. The source pressure was increased to 10 mbar to create increased collisional cooling (Krutchinsky et al., 1998; Tahallah et al., 2001). Source temperature was set to 80 °C and sample cone voltage varied from 125 to 175 V. Needle voltage was around 1300 V in case of the LCT and around 1600 V for the Q-ToF. For the tandem mass spectrometry (MS/MS) measurements, precursor ions of specific m/z were selected in the quadrupole. Subsequently, they were accelerated before the collision cell applying voltages from 10 to 200 V over this cell. Xenon was used as collision gas in the MS/MS experiments to increase ion transmission and energy transfer efficiency during the process (Lorenzen et al.). Pressure conditions were 1.5 x 10^{-2} mbar in the collision cell and 2.3 x 10^{-6} mbar in the ToF. The calculation of the calculated surface area in Figure 7B is based on the surface of a sphere with the mass of the protein complex using a density of 0.78 g/cm3 (Benesch et al., 2007). The measurements using the Q ToF did not allow a temperature control at the instrument, thus all experiments were done at laboratory temperature (20°C).

Ion mobility separation (IMS) measurements were carried out on a Synapt HDMS (Waters, UK)(Pringle et al., 2007). The source pressure was set to 6.9 mbar, the pressure in the trap was 2.9 x 10^{-2} mbar, 0.5 mbar in the IMS cell and 2.7 x 10^{-6} mbar in the ToF. The wave height in the IMS cell was ramped from 7 to 30 V and the wave velocity set to 250 m/s. The gas used in the trap was xenon (Israel et al., 1967; Lorenzen et al.; Lorenzen et al., 2007b) and nitrogen in the IMS cell. Needle voltage was 1300 V and cone voltage 175 V. The comparison of the cross sections was based on the following formula:

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

where Ω is the calibrated cross section, K is the calibration constant, t_D is the drift time, z is the charge state, m_{ion} is the mass of the ion and m_{gas} is the mass of the target gas in the IMS cell. T_D was corrected before for the retention time of the ions in the transfer cell and the field free region between transfer cell and pusher. For Figure 7D the volume was deducted from the cross section by assuming the protein complex being a sphere.

All cross sections of the proteins were calibrated using known cross sections (Ruotolo et al., 2005; Uetrecht et al., 2008) (http://www.indiana.edu/~clemmer/Research/cross%20section%20database/Proteins/protein_cs.htm). These cross sections were determined with a ramped wave height in the ion mobility chamber with different settings for small and large proteins. Large proteins (above ~ 100 kDa) were analyzed under the same conditions and instrumental settings as the protein complexes in this study. To obtain absolute cross sections datasets for small and large proteins were correlated taking GroEL single ring as a common point. To rule out systematical errors when using a ramped wave height, measurements were also performed with a fixed wave height. Measurements and calculations of the cross sections on the portal and portal:gp4 142^{STOP} complex obtained with a ramped and fixed wave height lead to the same conclusion of a conformational change. However a ramped wave height leads to lower standard deviations and narrower ion mobility peaks making it more suitable for analyzing large protein complexes.

REFERENCES

Andrews, D., Butler, J.S., Al-Bassam, J., Joss, L., Winn-Stapley, D.A., Casjens, S., and Cingolani, G. (2005). Bacteriophage P22 tail accessory factor gp26 is a long triple-stranded coiled-coil. Journal of Biological Chemistry 280, 5929-5933.

Bazinet, C., Villafane, R., and King, J. (1990). Novel second-site suppression of a cold-sensitive defect in phage P22 procapsid assembly. Journal of Molecular Biology 216, 701-716.

Benesch, J.L., Aquilina, J.A., Ruotolo, B.T., Sobott, F., and 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., and 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., and Dobson, C.M. (1998). Mass spectrometry of ribosomes and ribosomal subunits. Proc Natl Acad Sci U S A 95, 7391-7395.

Bhardwaj, A., Olia, A.S., Walker-Kopp, N., and Cingolani, G. (2007). Domain Organization and Polarity of Tail Needle GP26 in the Portal Vertex Structure of Bacteriophage P22. Journal of Molecular Biology 371, 374-387.

Bothner, B., and Siuzdak, G. (2004). Electrospray ionization of a whole virus: Analyzing mass, structure, and viability. Chembiochem 5, 258-260.

Casjens, S., and Huang, W.M. (1982). Initiation of sequential packaging of bacteriophage P22 DNA. J Mol Biol 157, 287-298.

Casjens, S., Winn-Stapley, D.A., Gilcrease, E.B., Morona, R., Kuhlewein, C., Chua, J.E., Manning, P.A., Inwood, W., and Clark, A.J. (2004). The chromosome of Shigella flexneri bacteriophage Sf6: complete nucleotide sequence, genetic mosaicism, and DNA packaging. J Mol Biol 339, 379-394.

Catalina, M.I., van den Heuvel, R.H.H., van Duijn, E., and Heck, A.J.R. (2005). Decharging of Globular Proteins and Protein Complexes in Electrospray. Chemistry - A European Journal 11, 960-968.

Chang, J., Weigele, P., King, J., Chiu, W., and Jiang, W. (2006). Cryo-EM asymmetric reconstruction of bacteriophage P22 reveals organization of its DNA packaging and infecting machinery. Structure 14, 1073-1082.

Cingolani, G., Moore, S.D., Prevelige, P.E., and Johnson, J.E. (2002). Preliminary crystallographic analysis of the bacteriophage P22 portal protein. Journal of Structural Biology 139, 46.

de la Mora, J.F. (2000). Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism. Analytica Chimica Acta 406, 93-104.

Earnshaw, W.C., and Casjens, S.R. (1980). DNA packaging by the double-stranded DNA bacteriophages. Cell 21, 319-331.

Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., and Whitehouse, C.M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. Science 246, 64-71.

Goldenberg, D.P., Berget, P.B., and King, J. (1982). Maturation of the tail spike endorhamnosidase of Salmonella phage P22. J Biol Chem 257, 7864-7871.

Hartweig, E., Bazinet, C., and King, J. (1986). DNA injection apparatus of phage P22. Biophysical Journal 49, 24-26.

Heck, A.J., and Van Den Heuvel, R.H. (2004). Investigation of intact protein complexes by mass spectrometry. Mass Spectrom Rev 23, 368-389.

Huang, F., and Nau, W.M. (2003). A conformational flexibility scale for amino acids in peptides. Angew Chem Int Ed Engl 42, 2269-2272.

Israel, J.V., Anderson, T.F., and Levine, M. (1967). in vitro Morphogenisis of Page P22 from Heads and Base-Plate Parts. Proc Natl Acad Sci U S A 57, 284-291.

Kaddis, C.S., and Loo, J.A. (2007). Native protein MS and ion mobility large flying proteins with ESI. Anal Chem 79, 1778-1784.

Kaltashov, I.A., and Mohimen, A. (2005). Estimates of protein surface areas in solution by electrospray ionization mass spectrometry. Analytical Chemistry 77, 5370-5379.

Koeniger, S.L., and Clemmer, D.E. (2007). Resolution and structural transitions of elongated states of ubiquitin. Journal of the American Society for Mass Spectrometry 18, 322-331.

Koeniger, S.L., Merenbloom, S.I., and Clemmer, D.E. (2006). Evidence for many resolvable structures within conformation types of electrosprayed ubiquitin ions. Journal of Physical Chemistry B 110, 7017-7021.

Konermann, L. (2007). A minimalist model for exploring conformational effects on the electrospray charge state distribution of proteins. Journal of Physical Chemistry B 111, 6534-6543.

Krutchinsky, A.N., Chernushevich, I.V., Spicer, V.L., Ens, W., and Standing, K.G. (1998). Collisional damping interface for an electrospray ionization time-of-flight mass spectrometer. Journal Of The American Society For Mass Spectrometry 9, 569-579.

Lander, G.C., Tang, L., Casjens, S.R., Gilcrease, E.B., Prevelige, P., Poliakov, A., Potter, C.S., Carragher, B., and Johnson, J.E. (2006). The structure of an infectious P22 virion shows the signal for headful DNA packaging. Science 312, 1791-1795.

Loo, J.A. (1997). Studying noncovalent protein complexes by electrospray ionization mass spectrometry. Mass Spectrom Rev 16, 1-23.

Lorenzen, K., Vannini, A., Cramer, P., and 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., and Heck, A.J.R. (2007b). Optimizing macromolecular tandem mass spectrometry of large non-covalent complexes using heavy collision gases. International Journal of Mass Spectrometry 268, 198-206.

Lurz, R., Orlova, E.V., Gunther, D., Dube, P., Droge, A., Weise, F., van Heel, M., and Tavares, P. (2001). Structural organisation of the head-to-tail interface of a bacterial virus. Journal of Molecular Biology 310, 1027-1037.

Moak, M., and Molineux, I.J. (2004). Peptidoglycan hydrolytic activities associated with bacteriophage virions. Molecular Microbiology 51, 1169-1183.

Olia, A.S., Al-Bassam, J., Winn-Stapley, D.A., Joss, L., Casjens, S.R., and Cingolani, G. (2006). Binding-induced stabilization and assembly of the phage P22 tail accessory factor gp4. Journal of Molecular Biology 363, 558-576.

Olia, A.S., Bhardwaj, A., Joss, L., Casjens, S., and Cingolani, G. (2007). Role of Gene 10 Protein in the Hierarchical Assembly of the Bacteriophage P22 Portal Vertex Structure. Biochemistry 46, 8776-8784.

Patriksson, A., Marklund, E., and van der Spoel, D. (2007). Protein structures under electrospray conditions. Biochemistry 46, 933-945.

Poliakov, A., Duijn, E.v., Lander, G., Fu, C.-y., Johnson, J.E., Prevelige, J.P.E., and Heck, A.J.R. (2007). Macromolecular mass spectrometry and electron microscopy as complementary tools for investigation of the heterogeneity of bacteriophage portal assemblies. Journal of Structural Biology 157, 371.

Pringle, S.D., Giles, K., Wildgoose, J.L., Williams, J.P., Slade, S.E., Thalassinos, K., Bateman, R.H., Bowers, M.T., and 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. International Journal of Mass Spectrometry 261, 1-12.

Ruotolo, B.T., Giles, K., Campuzano, I., Sandercock, A.M., Bateman, R.H., and Robinson, C.V. (2005). Evidence for Macromolecular Protein Rings in the Absence of Bulk Water. Science 310, 1658-1661.

Ruotolo, B.T., and Robinson, C.V. (2006). Aspects of native proteins are retained in vacuum. Current Opinion in Chemical Biology 10, 402-408.

Smith, D.E., Tans, S.J., Smith, S.B., Grimes, S., Anderson, D.L., and Bustamante, C. (2001). The bacteriophage phi 29 portal motor can package DNA against a large internal force. Nature 413, 748-752.

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

Steinbacher, S., Miller, S., Baxa, U., Budisa, N., Weintraub, A., Seckler, R., and Huber, R. (1997). Phage P22 tailspike protein: Crystal structure of the head-binding domain at 2.3 angstrom, fully refined structure of the endorhamnosidase at 1.56 angstrom resolution, and the molecular basis of O-antigen recognition and cleavage. Journal of Molecular Biology 267, 865-880.

Strauss, H., and King, J. (1984). Steps in the stabilization of newly packaged DNA during phage P22 morphogenesis. J Mol Biol 172, 523-543.

Tahallah, N., Pinkse, M., Maier, C.S., and 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.

Tang, L., Marion, W.R., Cingolani, G., Prevelige, P.E., and Johnson, J.E. (2005). Three-dimensional structure of the bacteriophage P22 tail machine. Embo J 24, 2087-2095.

Uetrecht, C., Versluis, C., Watts, N., Wingfield, P., Steven, A.C., and Heck, A.J. (2008). Structure, Stability and Shape of Hepatitis B Virus Capsids. Angew Chem Int Ed Engl. 47, 6247-5.

van den Heuvel, R.H., and 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., Synowsky, S.A., Lorenzen, K., Versluis, C., Brouns, S.J., Langridge, D., van der Oost, J., Hoyes, J., and Heck, A.J. (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., and Heck, A.J.R. (2005). Monitoring macromolecular complexes involved in the chaperonin-assisted protein folding cycle by mass spectrometry. Nature Methods 2, 371-376.

STRUCTURAL BIOLOGY OF RNA POLYMERASE III: MASS SPECTROMETRY ELUCIDATES SUBCOMPLEX ARCHITECTURE

Kristina Lorenzen[1], Alessandro Vannini[2], Patrick Cramer[2], and Albert J.R. Heck[1]
[1]Biomolecular Mass Spectrometry and Proteomics Group, Bijvoet Center for Biomolecular Research and Utrecht Faculty of Science, Utrecht University, Sorbonnelaan 16, 3584 CA Utrecht, the Netherlands
[2]Gene Center Munich and Center for integrated Protein Science CiPSM, Department of Chemistry and Biochemistry, Ludwig-Maximilians-Universität München, Feodor-Lynen-Strasse 25, 81377 Munich, Germany.

SUMMARY

RNA polymerases (Pol) II and III synthesize eukaryotic mRNAs and tRNAs, respectively. The crystal structure of the 12-subunit Pol II is known, but only limited structural information is available for the 17-subunit Pol III. Using mass spectrometry (MS) we correlated masses of Pol II complexes with the Pol II structure. Analysis of Pol III showed that the complete enzyme contains a single copy of each subunit and revealed a 15-subunit form lacking the Pol III-specific subcomplex C53/37. DMSO treatment dissociated the C17/25 heterodimer of Pol III, confirming a peripheral location as its counterpart in Pol II. Tandem MS revealed the Pol III-specific subunits C82 and C34 dissociating as a heterodimer. C11 was retained, arguing against a stable trimeric subcomplex C53/37/11. These data suggest Pol III consists of a 10-subunit Pol II-like core, the peripheral heterodimers C17/25, C53/37, and C82/34, and subunit C31, which bridges between C82/34, C17/25 and the core.

INTRODUCTION

The multisubunit RNA polymerases (Pols) I, II, and III synthesize eukaryotic RNA during gene transcription. Pol I and Pol II synthesize ribosomal and mainly messenger RNA, respectively, and Pol III transcribes small RNAs, including transfer RNAs (tRNAs), 5S ribosomal RNA, and U6 small nuclear RNA. The size and complexity of the polymerases increase from Pol II (12 subunits, 514 kDa) via Pol I (14 subunits, 588 kDa) to Pol III (17 subunits, 693 kDa). Detailed structural information is available for Pol II, including the crystal structures of the 10-subunit core enzyme (Cramer et al., 2000; Cramer et al., 2001), the additional subcomplex Rpb4/7, and the complete 12-subunit enzyme (Armache et al., 2003; Armache et al., 2005; Bushnell and Kornberg, 2003). These structures provide a framework for understanding the subunit architecture of the two other polymerases.

Open questions remain in particular for the subunit architecture of Pol III since it is the most complex enzyme (Chedin et al., 1998; Geiduschek and Kassavetis, 2001; Schramm and Hernandez, 2002). Pol III apparently contains a 10-subunit core homologous to Pol II. The C17/25 subcomplex it comprises is the counterpart of the Rpb4/7 subcomplex in Pol II, and in addition it contains the five additional subunits C82, 53, 37, 34 and 31. The C17/25 subcomplex constitutes a stalk that protrudes from the polymerase core, is involved in the initiation of transcription via recognition of promoter-bound factors, and may bind exiting RNA (Dezelee et al., 1976; Jasiak et al., 2006). In Pol II, Rpb4/7 can dissociate from the core as a heterodimer, but it remained unclear if C17/25 as the homologue counterpart in Pol III shows this behavior as well. Subunits C82, 34 and 31 form a sTable trimeric subcomplex in isolation (Wang and Roeder, 1997; Werner et al., 1992), but it is unknown if and how the C82/34/31 subcomplex can dissociate from the core. C53 and C37, form a heterodimer and are important for terminator recognition (Landrieux et al., 2006a). C53/37 also plays a key role in facilitated reinitiation together with C11. Subunit C11 is involved in the intrinsic RNA cleavage activity of Pol III (Chedin et al., 1998b) and shows sequence similarity with the Pol II subunit Rpb9 and the elongation factor TFIIS. It has been suggested that the Pol III-specific subunits C53 and C37 belong to an autonomous structural module together with C11, based on the fact that mutations in either C37 or C11 lead to the loss of C53, C37, and C11 following Pol III purification (Hu et al., 2002; Landrieux et al., 2006b). This is difficult to reconcile with the idea that C11 is part of a conserved structural core and association of these three subunits in isolation has not been demonstrated. The recent electron microscopic analysis of Pol III indicated that the C82/34/31 subcomplex is located mainly on one side above the central DNA-binding cleft, whereas the C53/37 subcomplex may be situated across the cleft on the side of the enzyme (Fernandez-Tornero et al., 2007). It also remains unclear if all cellular Pol III exists as an intact 17-subunit complex or if other species coexist that lack particular subunits or subcomplexes and possibly differ in their function (Ferri et al., 2000; Huet et al., 1985).

Here we address open questions on the Pol III architecture by analyzing multisubunit RNA polymerases and their subcomplexes by native mass spectrometry. To establish the experimental approach, we analyzed Pol II and correlated the mass spectrometric data with known structural information. Application of this method to Pol III provided insights into the architecture of this enzyme. Our results not only provide insights into the subunit architecture of the largest cellular RNA polymerase, they also indicate that it may become feasible to study functional complexes of endogenously expressed RNA polymerases by mass spectrometry.

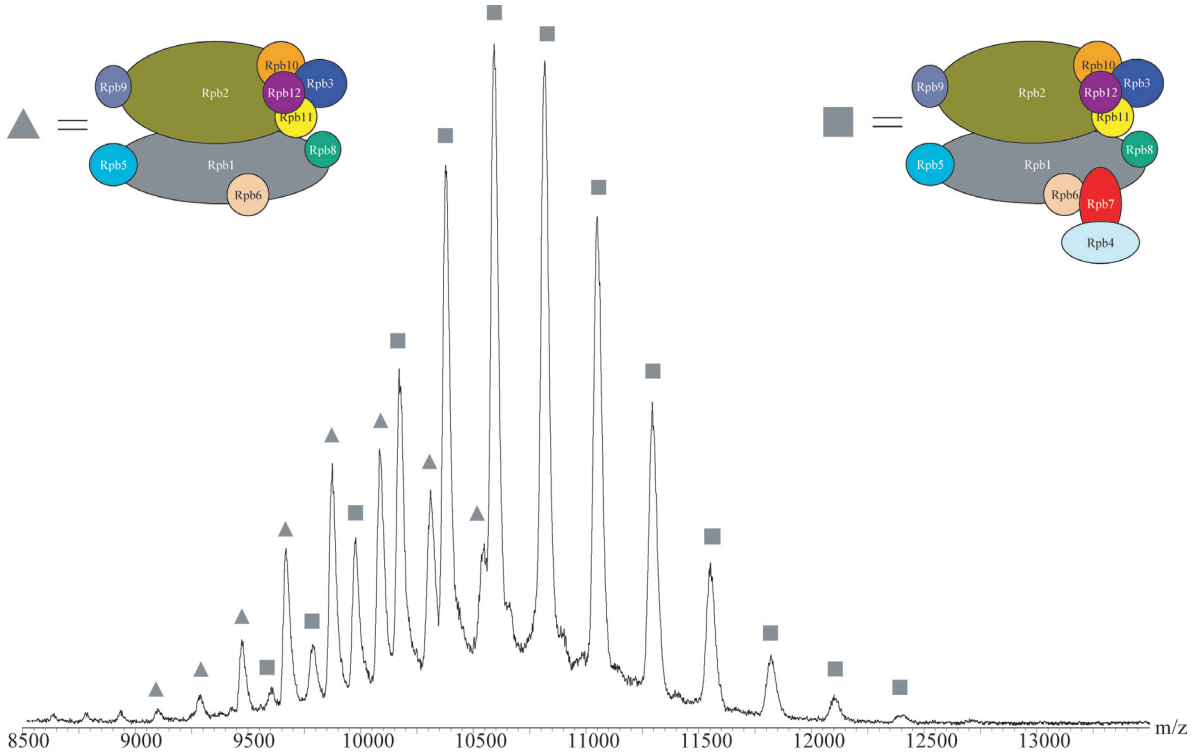

Figure 1 Native mass spectrum of the purified Pol II complexes. The two different charge distributions reveal two distinct complex stoichiometries. The squares indicate the charge distribution of the complete twelve subunit protein complex with a mass of 517 kDa. The triangles indicate the charge distribution of the complex missing Rpb4 and Rpb7 (472 kDa). The insets show structural models of the proposed complexes.

RESULTS

Mass spectrometry of Pol II detects the two known forms of the enzyme

Macromolecular mass spectrometry (Heck and Van Den Heuvel, 2004; Ilag et al., 2004; Robinson, 2002; van den Heuvel and Heck, 2004; van Duijn et al., 2005) can be used to detect and measure the masses of large multi-component complexes, such as ribosomes and exosomes (Hernandez et al., 2006a; McKay et al., 2006a; Synowsky et al., 2006), but has not been shown to be applicable to the multisubunit RNA polymerases. Figure 1 shows a mass spectrum of a preparation of endogenous Pol II from the yeast Saccharomyces cerevisiae. In the spectrum two charge distributions can be distinguished in the range from 9000 to 13000 m/z, which allow deduction of the masses of the underlying complexes. The species with the highest mass is most abundant, accounting for 66% of the total ion intensity, and has a mass of 517,320 ± 140 Da. This mass fits well with the expected mass of the intact, complete 12-subunit Pol II (513,540 Da). For very large complexes the masses measured in macromolecular mass spectrometry are typically higher than the expected masses since desolvation is generally incomplete, leaving additional water and buffer molecules attached to the protein complex (McKay et al., 2006b). The second species has a mass of 472,720 ± 120 Da, giving a measured mass difference between the two Pol II species of 44,600 Da. This difference corresponds very well to the sum of the masses of subunits Rpb4 and Rpb7 (44,530 Da). The co-occurrence of this second complex may be expected since Rpb4 and Rpb7 can dissociate from Pol II as a heterodimer. Thus the second species corresponds to the 10 subunit Pol II core enzyme.

Stoichiometry of Pol III and a Pol III form lacking C53/37

We recorded the macromolecular mass spectra of a preparation of endogenous yeast Pol III under similar experimental settings as for Pol II. The mass spectra also displayed more than one species (Figure 2). Better resolved spectra at a higher desolvation energy showed that the main species had a mass of 699,180 ± 120 Da, which corresponds well with the expected mass of the intact complete 17-subunit Pol III if a single copy of each subunit is assumed to be present (693,230 Da). A second, distinct charge distribution revealed a subcomplex with a mass of 620,320 ± 130 Da. The difference in mass between these two species is 78,860 Da, which is consistent with the second complex missing C53 and C37 (expected mass 78,800 Da). To make sure that this was not an artifact due to degraded forms of C53 in the purification that could fall apart under conditions used for mass spectroscopy we checked the purification on a SDS gel (supplementary Figure 6). There were no degraded forms of C53 visible. Additonaly it has been described in previous studies

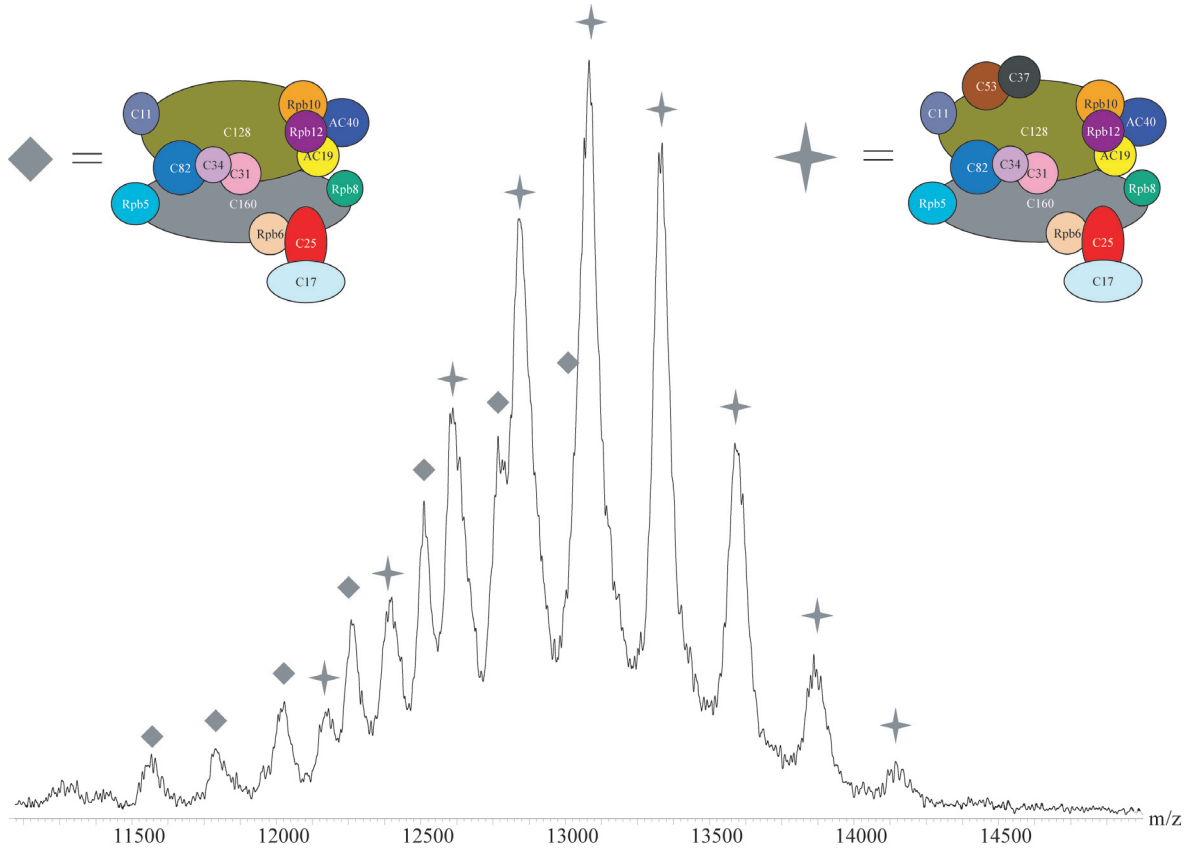

Figure 2 Native mass spectrum of the purified Pol III complexes. The two different charge distributions reveal two complex stoichiometries. The stars indicate the charge distribution of the complete seventeen subunit Pol III complex with a mass of 699 kDa. The diamonds mark the charge distribution of the complex missing the subunits C53 and C37 (620 kDa). The insets show structural models of the proposed complexes.

that subunit C53 occurs at substoichiometric levels upon Pol III purification (Sadhale and Woychik, 1994; Sentenac, 1985). In our measurements the intact Pol III was more abundant in the spectrum accounting for 66% of the ion intensity. At high desolvation energy a low amount of complex could be detected that lost only the subunit C53 and still retained C37 (see supplementary Figure 1). An overview of all measured masses and the protein homologies between Pol II and Pol III is summarized in Table 1. A comparison between the masses detected and theoretical masses derived from the protein sequence is given in supplementary Table 1. To confirm that the different subcomplexes of Pol III are not a mass spectrometric artifact and are not due to the buffer exchange to ammonium acetate, native PAGE, which clearly showed two different bands for the two protein complexes (data not shown) was carried out.

Peripheral subunits in Pol II and Pol III

To probe which subunits are positioned at the surface of Pol II and Pol III, we gently dissociated the polymerase complexes in solution adding low percentages of organic modifiers to the aqueous ammonium acetate electrospray solutions (Hernandez et al., 2006b). Adding from 1 % to 10 % DMSO we observed charge distributions of several subunits and a few subcomplexes in the low mass range (Table 2). For Pol II we detected ions originating from subunits Rpb4, 5, 6, 7, 8, 9 and 12. The mass detected for the Rpb9 was 65 Da higher than expected showing that one of the two zinc ions in Rpb9 was still bound to the subunit. In addition we detected the intact Rpb4/7 heterodimer. All these subunits are located on the Pol II surface in the X-ray structure. Interestingly, subunits Rpb3, 10 and 11 were not dissociated, in agreement with their proposed role as an assembly nucleus that anchors the two large subunits, and consistent with Rpb1, 2, 3, and 11 forming the evolutionarily conserved core also found in the bacterial RNA polymerase. Thus the DMSO dissociation experiments yield valuable insights into subunit architecture and complex stability. The same treatment of the Pol III complex resulted in dissociation of subunits homologous to the dissociating Pol II subunits (Table 2). For C11, the counterpart to Rpb9, a mass was detected that was 130 Da higher than expected, suggesting that two predicted structural zinc ions remained bound to the subunit. The spectra also revealed the C17/25 heterodimer, which corresponds to Rpb4/7 in Pol II, underlining the similarity in the architecture of Pol II and Pol III complexes. Additi-

Table 1 Overview of polymerase subunits and their homologies. Listed next to the subunit is the mass that was detected for this subunit during the experiments. In the bottom of the table the masses of the detected complexes and subcomplexes are indicated.

Polymerase part	Pol III subunit	Masses detected (Da)	Pol II subunit	Masses detected (Da)	Subunit type
Core	C160	-	Rpb1	-	homolog
	C128	-	Rpb2	-	homolog
	AC40	37,603±5	Rpb3	-	homolog
	AC19	16,024±4	Rpb11	13660±3	homolog
	Rpb5	25124±4	Rpb5	25127±5	common
	Rpb6	17902±3	Rpb6	17902±3	common
	Rpb8	16425±4	Rpb8	16425±4	common
	Rpb10	8280±1	Rpb10	8280±1	common
	Rpb12	7692±1	Rpb12	7692±1	common
	C11*	12639±2	Rpb9	14265±3	homolog
Rpb4/7 / C17/25 heterodimer	C17	18530±4	Rpb4	25460±6	counterpart
	C25	24330±4	Rpb7	19062±5	counterpart
C53/37 heterodimer	C53	-	-	-	specific
	C37	32009±5	-	-	specific
Upstream heterotrimer	C82	74075±8	-	-	specific
	C34	36052±5	-	-	specific
	C31	-	-	-	specific
Intact complex		699180±120		517320±140	
subcomplex		620320±130		472720±120	
Rpb4/7 / C17/25 heterodimer		42846±7		44526±10	
C82/34 heterodimer		110130±20			

*Rpb9 was previously defined as a part of the Pol II core. C11 shows homology to Rpb9 and TFIIS.
- not detected

onally we detected signals for C82, 34 and a signal of low intensity assigned to C37. These experiments also revealed a likely abundant single phosphorylation in Rpb6 in both polymerases that may serve a common function; Rpb6 was detected with a mass of 17,902 Da and with a mass that is 80 Da higher (17982 Da) (Table 2).

Tandem mass spectrometry of Pol II and Pol III

To further investigate the architecture of the Pol II and Pol III complexes and their coexisting subcomplexes we employed tandem mass spectrometry (McCammon et al., 2004; van Duijn et al., 2006). During the transfer of proteins into the mass spectrometer, protonation leads to the presence of the protein in different charge states. Tandem mass spectrometry allows selection of one charge state of the subset precursor complexes, which can be collisionally activated in the mass spectrometer, resulting in specific dissociation. In these experiments, the precursor ions generally eliminate a single subunit, which will become highly charged, concomitant with the formation of relatively low charged ions of the precursor complex lacking the eliminated subunit (Heck and Van Den Heuvel, 2004). Evidently, the summed mass and charge of the two formed fragment ions should add up to the mass and charge of the precursor complex. Elimination of different protein subunits may occur parallel and sequentially. We performed tandem mass spectrometry by selecting charge states originating either from the complete Pol II or Pol III, or from their subcomplexes that had lost the Rpb4/7 or C53/37 heterodimer, respectively.

Tandem mass spectrometry patterns reflect the Pol II architecture

Tandem mass spectrometry on the complete Pol II (517 kDa) led primarily to the elimination of subunits Rpb4, 7 and the intact Rpb4/7 heterodimer (Table 3 and supplementary Figure 2). Elimination of protein dimers is rarely observed in tandem mass spectrometry of protein complexes, since subunits often unfold upon dissociation. However, observation of the Rpb4/7 heterodimer is not so surprising since it forms a sTable subcomplex. The concomitant high

Table 2 List of subunits that dissociated from the polymerases as a result of the DMSO treatment. x indicates the detected subunit, - the subunits that were not found.

Pol III subunit		Pol II subunit	
C160	-	Rpb1	-
C128	-	Rpb2	-
AC40	-	Rpb3	-
AC19	-	Rpb11	-
Rpb5	x	Rpb5	x
Rpb6	x	Rpb6	x
Rpb6#	x	Rpb6#	x
Rpb8	x	Rpb8	x
Rpb10	-	Rpb10	-
Rpb12	x	Rpb12	x
C11**	x	Rpb9*	x
C17	x	Rpb4	x
C25	x	Rpb7	x
C53	-		
C37	x		
C82	x		
C34	x		
C31	-		
C17/25 heterodimer	x	Rpb4/7 heterodimer	x

x indicates the subunit detected during the measurement.
Rpb6 was additionally detected with a mass difference of plus 80 Da.
*Rpb9 was detected with a mass difference of an additional 65 kDa.
**C11 was detected with a mass difference of an additional 130 kDa.

mass fragment ions of the complexes could be assigned to the ΔRpb4, ΔRpb7 and ΔRpb4/7 complexes. These results reflect the known biochemical and structural data on the architecture of Pol II. It has been observed before that structures can be partially retained in the gas-phase (Ruotolo and Robinson, 2006; van Duijn et al., 2005). Other single subunits that were detected were Rpb12 and, at higher collision energy, Rpb10. The Pol II complexes ΔRpb12 and ΔRpb10 were not visible, likely their signal was suppressed or too low abundant compared to the strong signal of the ΔRpb4/7 complex. Tandem mass spectrometry at higher energy of the Pol II core complex lacking Rpb4/7 revealed additional dissociation products (supplementary Figure 3). Eliminated were monomeric ions of the surface subunits Rpb5, 8, 9, 10, 11 and 12. For most of the dissociated subunits the concomitant remaining complexes could be assigned. The ΔRpb12 and complexes that lacked Rpb10 were found. The ΔRpb8 complex could be dissociated further, resulting in complexes lacking Rpb8 together with either Rpb11, 5 or 9. Since no Pol II complexes ΔRpb4 or ΔRpb7 were observed, we conclude that indeed the initial Pol II subcomplex corresponds to the 10-subunit core.

The Pol III core architecture resembles that of Pol II

The complete Pol III and the presumed C53/37 complex showed a similar pattern of subunit elimination in tandem mass spectrometry (supplementary Figure 4 and 5). For both complexes, nine out of the 17 Pol III subunits were detected as monomeric fragment ions (Table 3). The common and homologous subunits that were eliminated during the tandem mass spectrometry of the Pol II core complex were all found to dissociate from the Pol III complexes as well, with the exception of Rpb5. All concomitant high mass fragments that were detected for the complete Pol III (699 kDa) still included the mass of C37 and C53, indicating that these subunits are quite stably associated with Pol III and do not dissociate from the complex under the conditions used. The subunits that were found to dissociate from Pol III only leave Rpb5, 6 and C11, 37, 53 and 31 to account for the missing mass between the complete Pol III and its smaller subcomplex. The only combination that fits the experimental mass difference (78,870 Da) between the complete Pol III and the smaller complex of Pol III is the combination of C37 and C53, supporting the notion that a discrete C53/37 complex of Pol III is present in the preparation.

A peripheral C82/34 heterodimer can dissociate from Pol III

A striking difference in tandem mass spectrometry of Pol III when compared to Pol II is the elimination of a Pol III-specific C82/34 heterodimer. We detected fragment ions of a mass of 110,130 Da, corresponding very well to the theoretical (as given in Table 1) combined mass of C82 and C34 (110,130 Da). These fragment ions were observed in tandem mass spectra of both the complete Pol III as well as in the subcomplex lacking C37/53 (Figure 3). The concomitant high mass fragments could also be detected and mass-assigned, confirming dissociation of a C82/34 heterodimer and retention of C31 in the remaining polymerase complex. These data can not completely exclude the existence of a C82/34/31 heterotrimer that was described previously (Hu et al., 2002; Landrieux et al., 2006b). Instead, they indicate that the interaction of C31 with the polymerase in the gas phase is tighter than its interaction with the C82/34 heterodimer. C82 and C34 were also detected as fragment ions, as for the Rpb4/7 heterodimer in Pol II. Another difference between Pol II and Pol III concerns an apparently different stability of the assembly core. The Pol III subunits AC40 and AC19 were detected as fragment ions in tandem mass spectrometry, whereas the Pol II homologues Rpb3 and Rpb11 did not dissociate from Pol II. AC19 dissociates with

Table 3 Subunits eliminated from the different polymerase complexes. At the top of the table is indicated from which complex the subunit dissociated. The masses that were assigned to each subunit are listed in table 1.

Pol III	Pol II				
	699 kDa	620 kDa		517 kDa	472 kDa
Pol III subunit			Pol II subunit		
C160	-	-	Rpb1	-	-
C128	-	-	Rpb2		-
AC40	x	x	Rpb3	-	-
AC19	x	x	Rpb11	-	x
Rpb5	-	-	Rpb5	-	x
Rpb6	-	-	Rpb6	-	-
Rpb8	x	x	Rpb8	-	x
Rpb10	x	x	Rpb10	x	x
Rpb12	x	x	Rpb12	x	x
C11	-	-	Rpb9	-	x
C17	x	x	Rpb4*	x	
C25	x	x	Rpb7*	x	
C53	-	-	-		
C37	-	-	-		
C82	x	x	-		
C34	x	x	-		
C31	-	-	-		
C82/34 subcomplex	x	x	Rpb4/7 subcomplex	x	

x indicates the subunit detected during the measurement.
* due to contamination of the 517 kDa Pol II species in the tandem MS of the 472 kDa subcomplex there were low amounts of Rpb4 and 7 detected in the spectrum.

a mass of 16024 Da and a mass of 16104 Da, revealing a potential phosphorylation site in Pol III.

DISCUSSION

Here we have used, for the first time, native mass spectrometry to analyze the architecture of the multisubunit RNA polymerases Pol II and Pol III. Our results reveal different forms of Pol II and Pol III in solution. With the use of DMSO treatment and tandem mass spectrometry, peripheral subcomplexes could be dissociated from the enzymes and identified. Our results have two major implications. First, they demonstrate the technical feasibility of studying multisubunit RNA polymerases by native mass spectrometry. The validity of this approach is demonstrated by successful correlation of the observed masses for Pol II and its subunits and subcomplexes to the known three-dimensional structure. In the future, it may be possible to study formation and stability of functional complexes of RNA polymerases with mass spectrometry. Second, our results provide important insights into the subunit architecture of Pol III, the largest cellular RNA polymerase, which comprises five additional, specific subunits arranged in specific subcomplexes. The data demonstrates that each of the 17 subunits of Pol III is present as a single copy. Pol III apparently comprises a 10-subunit core that resembles the structure of the Pol II core and apparently includes subunit C11, even though a trimeric subcomplex C53/37/11 as suggested previously can not be completely ruled out. Arrayed around the periphery of the core are three heterodimeric subcomplexes, C53/37, C82/34 and C17/25, the latter one corresponding to Rpb4/7 in Pol II. Subunit C31 binds to the Pol III core and subunit C17 (Chedin et al., 1998a; Geiduschek and Kassavetis, 2001; Schramm and Hernandez, 2002), but has also been shown to form a sTable trimer with C82 and C34 (Wang and Roeder, 1997; Werner et al., 1992), suggesting that it is involved in bridging between C82/34, the core and C17/25. The obtained model for the Pol III subunit architecture will guide future biochemical and structural studies of this central cellular enzyme.

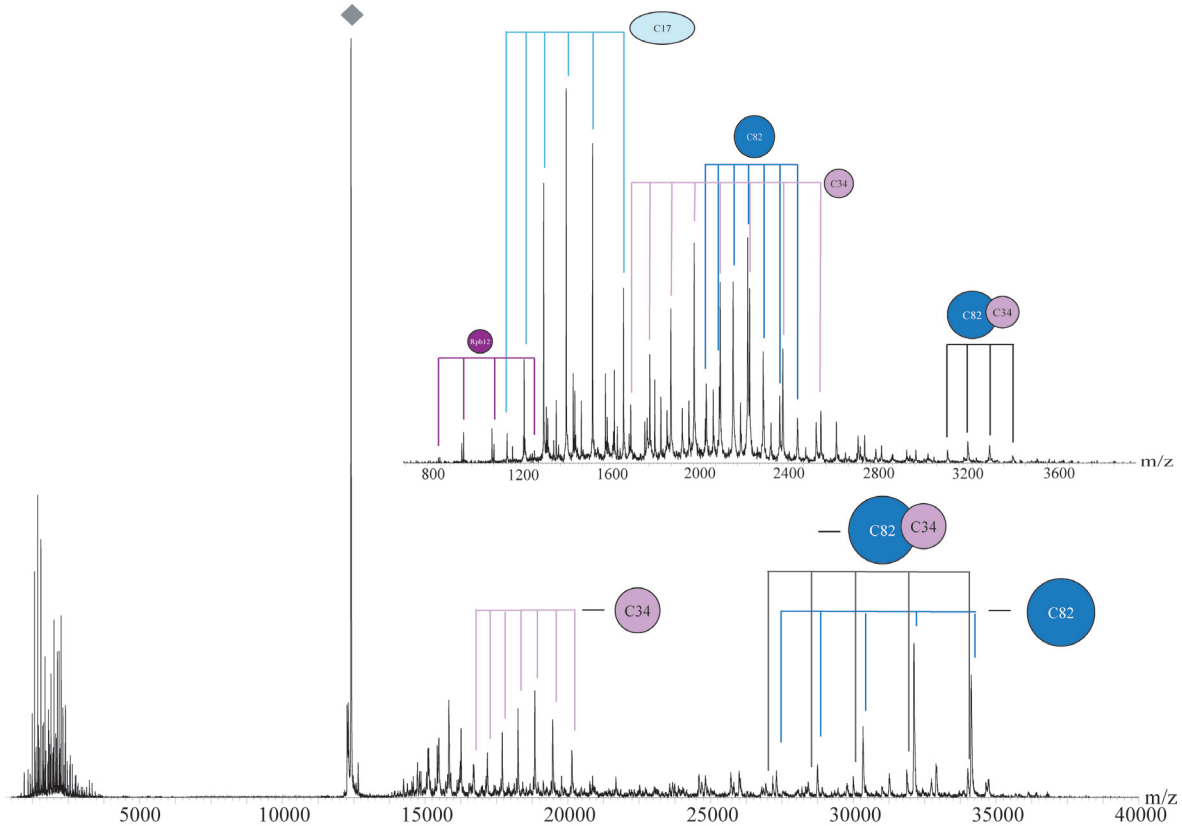

Figure 3. Tandem mass spectrum of the 620 kDa subcomplex of Pol III (the 50+ charge state was selected, indicated with a diamond). The spectrum reveals the dissociation of the C82/34 dimer. Above m/z 12,000 the distributions are labeled representing the subcomplex that eliminated C34, C82 or its heterodimer. The zoom in at the low m/z region reveals subunits that were eliminated from the complex. The dimer of C82/34 is visible.

Supplementary data can be obtained from http://download.structure.org/supplementarydata/foldes/15/10/1237/DC1/mmc1.pdf

EXPERIMENTAL PROCEDURES

Purification of Pol II and Pol III

Pol II was purified as described before (Brueckner et al., 2007). Pol III was purified as described (supplementary Figure 6) (Kassavetis et al., 1990).

Mass spectrometry of Pol II and III

The sample buffer was exchanged sequentially to 200 mM (Pol II) or 400 mM (Pol III) aqueous ammonium acetate pH 6.8 using centrifugal filter units with a cut-off of 100 kDa (Millipore, England). The concentration used for the mass spectrometry measurements was ~2 µM (assuming intact protein complexes). The samples were measured on a LCT electrospray time-of-flight instrument for the DMSO measurements and on a modified Q-ToF electrospray quadrupole time-of-flight instrument for native mass spectrometry and tandem mass spectrometry (van den Heuvel et al., 2006) (Waters, Manchester, UK). Nanospray glass capillaries were used to introduce the samples into the Z-spray source. The source pressure was increased to 10 mbar to create increased collisional cooling (Krutchinsky et al., 1998; Tahallah et al., 2001). Source temperature was set to 80°C and sample cone voltage varied from 125 to 175 V. Needle voltage was around 1300 V in case of the LCT and around 1600 V for the Q-ToF.

For the tandem mass spectrometry measurements precursor ions of specific m/z were selected in the quadrupole. Subsequently, they were accelerated through the collision cell applying voltages from 10 to 200V over this cell. Xenon was used as collision gas in the MS/MS experiments to increase ion transmission and energy transfer efficiency during the process (Lorenzen et al., 2007). Pressure conditions were 1.5 x 10-2 mbar in the collision cell and 2.3 x 10-6 mbar in the ToF. The faults indicated in the measurements are calculated from the variability of the different charge states. For each charge state a mass is calculated and the masses are averaged and the standard deviation is calculated. The masses measured during repetition of the experiments were within these deviations. The sequence masses for supplementary Figure 1 have been taken from the swiss prot database (http://www.expasy.org/).

REFERENCES

Armache, K. J., Kettenberger, H., and Cramer, P. (2003). Architecture of initiation-competent 12-subunit RNA polymerase II. Proc Natl Acad Sci U S A 100, 6964-6968.

Armache, K. J., Mitterweger, S., Meinhart, A., and Cramer, P. (2005). Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. J Biol Chem 280, 7131-7134.

Brueckner, F., Hennecke, U., Carell, T., and Cramer, P. (2007). CPD damage recognition by transcribing RNA polymerase II. Science 315, 859-862.

Bushnell, D. A., and Kornberg, R. D. (2003). Complete, 12-subunit RNA polymerase II at 4.1-angstrom resolution: Implications for the initiation of transcription. Proceedings Of The National Academy Of Sciences Of The United States Of America 100, 6969-6973.

Chedin, S., Ferri, M. L., Peyroche, G., Andrau, J. C., Jourdain, S., Lefebvre, O., Werner, M., Carles, C., and Sentenac, A. (1998a). The yeast RNA polymerase III transcription machinery: a paradigm for eukaryotic gene activation. Cold Spring Harb Symp Quant Biol 63, 381-389.

Cramer, P., Bushnell, D. A., Fu, J., Gnatt, A. L., Maier-Davis, B., Thompson, N. E., Burgess, R. R., Edwards, A. M., David, P. R., and Kornberg, R. D. (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. Science 288, 640-649.

Cramer, P., Bushnell, D. A., and Kornberg, R. D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. Science 292, 1863-1876.

Dezelee, S., Wyers, F., Sentenac, A., and Fromageot, P. (1976). Two forms of RNA polymerase B in yeast. Proteolytic conversion in vitro of enzyme BI into BII. Eur J Biochem 65, 543-552.

Fernandez-Tornero, C., Bottcher, B., Riva, M., Carles, C., Steuerwald, U., Ruigrok, R. W., Sentenac, A., Muller, C. W., and Schoehn, G. (2007). Insights into Transcription Initiation and Termination from the Electron Microscopy Structure of Yeast RNA Polymerase III. Mol Cell 25, 813-823.

Ferri, M.-L., Peyroche, G., Siaut, M., Lefebvre, O., Carles, C., Conesa, C., and Sentenac, A. (2000). A Novel Subunit of Yeast RNA Polymerase III Interacts with the TFIIB-Related Domain of TFIIIB70. Mol Cell Biol 20, 488-495.

Geiduschek, E. P., and Kassavetis, G. A. (2001). The RNA polymerase III transcription apparatus. J Mol Biol 310, 1-26.

Heck, A. J., and Van Den Heuvel, R. H. (2004). Investigation of intact protein complexes by mass spectrometry. Mass Spectrom Rev 23, 368-389.

Hernandez, H., Dziembowski, A., Taverner, T., Seraphin, B., and Robinson, C. V. (2006a). Subunit architecture of multimeric complexes isolated directly from cells. EMBO Rep 7, 605-610.

Hu, P., Wu, S., Sun, Y., Yuan, C. C., Kobayashi, R., Myers, M. P., and Hernandez, N. (2002). Characterization of human RNA polymerase III identifies orthologues for Saccharomyces cerevisiae RNA polymerase III subunits. Mol Cell Biol 22, 8044-8055.

Huet, J., Riva, M., Sentenac, A., and Fromageot, P. (1985). Yeast RNA polymerase C and its subunits. Specific antibodies as structural and functional probes. J Biol Chem 260, 15304-15310.

Ilag, L. L., Westblade, L. F., Deshayes, C., Kolb, A., Busby, S. J. W., and Robinson, C. V. (2004). Mass Spectrometry of Escherichia coli RNA Polymerase: Interactions of the Core Enzyme with [sigma]70 and Rsd Protein. Structure 12, 269-275.

Jasiak, A. J., Armache, K. J., Martens, B., Jansen, R. P., and Cramer, P. (2006). Structural biology of RNA polymerase III: Subcomplex C17/25 X-ray structure and 11 subunit enzyme model. Molecular Cell 23, 71-81.

Kassavetis, G. A., Braun, B. R., Nguyen, L. H., and Geiduschek, E. P. (1990). S-Cerevisiae Tfiiib Is The Transcription Initiation-Factor Proper Of Rna Polymerase-Iii, While Tfiiia And Tfiiic Are Assembly Factors. Cell 60, 235-245.

Krutchinsky, A. N., Chernushevich, I. V., Spicer, V. L., Ens, W., and Standing, K. G. (1998). Collisional damping interface for an electrospray ionization time-of-flight mass spectrometer. Journal of the American Society for Mass Spectrometry 9, 569-579.

Landrieux, E., Alic, N., Ducrot, C., Acker, J., Riva, M., and Carles, C. (2006a). A subcomplex of RNA polymerase III subunits involved in transcription termination and reinitiation. Embo Journal 25, 118-128.

Lorenzen, K., Versluis, C., van Duijn, E., van den Heuvel, R. H. H., and Heck, A. J. R. (2007). Optimizing macromolecular tandem mass spectrometry of large non covalent complexes using heavy collision gases. International Journal of Mass Spectrometry In Press, Accepted Manuscript.

McCammon, M. G., Hernandez, H., Sobott, F., and Robinson, C. V. (2004). Tandem mass spectrometry defines the stoichiometry and quaternary structural arrangement of tryptophan molecules in the multiprotein complex TRAP. J Am Chem Soc 126, 5950-5951.

McKay, A. R., Ruotolo, B. T., Ilag, L. L., and Robinson, C. V. (2006a). Mass measurements of increased accuracy resolve heterogeneous populations of intact ribosomes. J Am Chem Soc 128, 11433-11442.

Robinson, C. V. (2002). Protein complexes take flight. Nat Struct Biol 9, 505-506.

Ruotolo, B. T., and Robinson, C. V. (2006). Aspects of native proteins are retained in vacuum. Curr Opin Chem Biol 10, 402-408.

Sadhale, P. P., and Woychik, N. A. (1994). C25, an Essential Rna-Polymerase-Iii Subunit Related to the Rna-Polymerase-Ii Subunit Rpb7. Molecular and Cellular Biology 14, 6164-6170.

Schramm, L., and Hernandez, N. (2002). Recruitment of RNA polymerase III to its target promoters. Genes Dev 16, 2593-2620.

Sentenac, A. (1985). Eukaryotic RNA polymerases. CRC Crit Rev Biochem 18, 31-90

Synowsky, S. A., van den Heuvel, R. H., Mohammed, S., Pijnappel, P. W., and 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.

Tahallah, N., Pinkse, M., Maier, C. S., and 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.

van den Heuvel, R. H., and 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., Synowsky, S. A., Lorenzen, K., Versluis, C., Brouns, S. J., Langridge, D., van der Oost, J., Hoyes, J., and Heck, A. J. (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., and 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., and 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.

Wang, Z., and Roeder, R. G. (1997). Three human RNA polymerase III-specific subunits form a subcomplex with a selective function in specific transcription initiation. Genes Dev 11, 1315-1326.

Werner, M., Hermann-Le Denmat, S., Treich, I., Sentenac, A., and Thuriaux, P. (1992). Effect of mutations in a zinc-binding domain of yeast RNA polymerase C (III) on enzyme function and subunit association. Mol Cell Biol 12, 1087-1095.

MULTIPLEXED PROTEOMICS MAPPING OF YEAST RNA POLYMERASE II AND III ALLOWS NEAR-COMPLETE SEQUENCE COVERAGE AND REVEALS SEVERAL NOVEL PHOSPHORYLATION SITES

Shabaz Mohammed[1], Kristina Lorenzen[1], Robert Kerkhoven[1], Bas van Breukelen[1], Alessandro Vannini[2], Patrick Cramer[2], and Albert J.R. Heck[1]

[1] Biomolecular Mass Spectrometry and Proteomics Group, Bijvoet Center for Biomolecular Research and Utrecht Faculty of Science, Utrecht University, Sorbonnelaan 16, 3584 CA Utrecht, the Netherlands

[2]Gene Center Munich and Center for integrated Protein Science CiPSM, Department of Chemistry and Biochemistry, Ludwig-Maximilians-Universität München, Feodor-Lynen-Strasse 25, 81377 Munich, Germany.

SUMMARY

The multisubunit RNA polymerases (Pol) II and III synthesize mainly eukaryotic mRNAs and tRNAs, respectively. Pol II and Pol III are protein complexes consisting of 12 and 17 subunits. Here we analyzed both yeast Pol II and Pol III by multiplexed mass spectrometric analysis using various proteases and both collision induced and electron transfer dissociation. The cumulative data obtained from using the various proteases (trypsin, chymotrypsin, Glu-C and Lys-C) and the two peptide fragmentation approaches allowed us to map nearly the complete sequences of all constituents of both Pol II and III. Notably, chymotrypsin behaved equally well as and in certain circumstances better than trypsin in the context of protein coverage. Although, the available high resolution structures have exposed extensive mechanistic insights into transcription, the role of post-translational modification in these processes has been addressed to a lesser extent. In our analysis of Pol II and III we detected 19 phosphorylation sites, of which 12 have not been previously reported. Identified phosphosites were mapped on the Pol II structure which provided indications that they might play a role in regulating the conformation of the clamp region and, as a consequence, interaction of Pol II with nucleic acids. The described multiplexed proteomics approach is generic and reveals that it is possible to map a protein complex to near completion while applying less than 5 ug (approximately 10 pmol) of total starting material.

INTRODUCTION

The multisubunit RNA polymerases (Pols) I, II, and III synthesize eukaryotic RNA during gene transcription. Pol I and Pol II synthesize ribosomal and mainly messenger RNA, respectively, and Pol III transcribes small RNAs, including transfer RNAs, 5S ribosomal RNA, and U6 small nuclear RNA. The size and complexity of the polymerases increase from Pol II (12 subunits, 514 kDa) via Pol I (14 subunits, 588 kDa) to Pol III (17 subunits, 693 kDa). Detailed structural information is available for Pol II, including the crystal structures of the 10-subunit core enzyme (Cramer et al., 2000; Cramer et al., 2001), the additional subcomplex Rpb4/7 and the complete 12-subunit enzyme (Armache et al., 2003; Armache et al., 2005; Bushnell and Kornberg, 2003). For Pol III recent electron microscopy and mass spectrometric data gave important new insight into the structure of the complex (Fernandez-Tornero et al., 2007; Lorenzen et al., 2007). In order to aid further structure and functional elucidation we set out to map, by proteomics technologies, full protein sequences including possible post translational modifications of Pol II and Pol III.

Several strategies are available for performing a comprehensive analysis of a complex protein mixture. Common methods involve tryptic digests and reducing complexity through the use of multidimensional separation techniques such as SCX, (Wu et al., 2003) HILIC (Boersema et al., 2007) or IEF (Cargile et al., 2004; Krijgsveld et al., 2006) at the peptide level or IEX (Pieper et al., 2003), RP(Molina et al., 2007) or SDS page (Krijgsveld et al., 2006; Synowsky et al., 2006) at the protein level followed by digestion of fractions. However, since most current mass spectrometers have optimal m/z ranges for analysis, 500-4000 Th for MALDI and 300-1500 Th for ESI, trypsin digestion may not allow a complete analysis due to certain proteolytic peptides falling outside this optimal window. One way to remove all proteolytic issues is by analyzing at the protein level using a top down procedure involving electron capture dissociation (ECD) based sequencing (Han et al., 2006; Zubarev et al., 1998) with a prior to analysis protein separation and fractionation (Garcia et al., 2007). Another, more easily approachable, possibility is to use a multi-protease strategy where each enzyme will provide complementary protein fragments as well as sequence overlap (Cheeseman et al., 2002; Distler et al., 2006; Fischer and Poetsch, 2006; Gatlin et al., 2000; MacCoss et al., 2002; Molina et al., 2007; Running et al., 2007; Schlosser et al., 2005; Wu et al., 2005). For instance, Schlosser et al. used a multi-enzyme strategy to create a cocktail of peptides that would allow the full sequence of the murine circadian protein period 2 (mPER2) to be examined by nanoLC-MS for the purpose of identifying

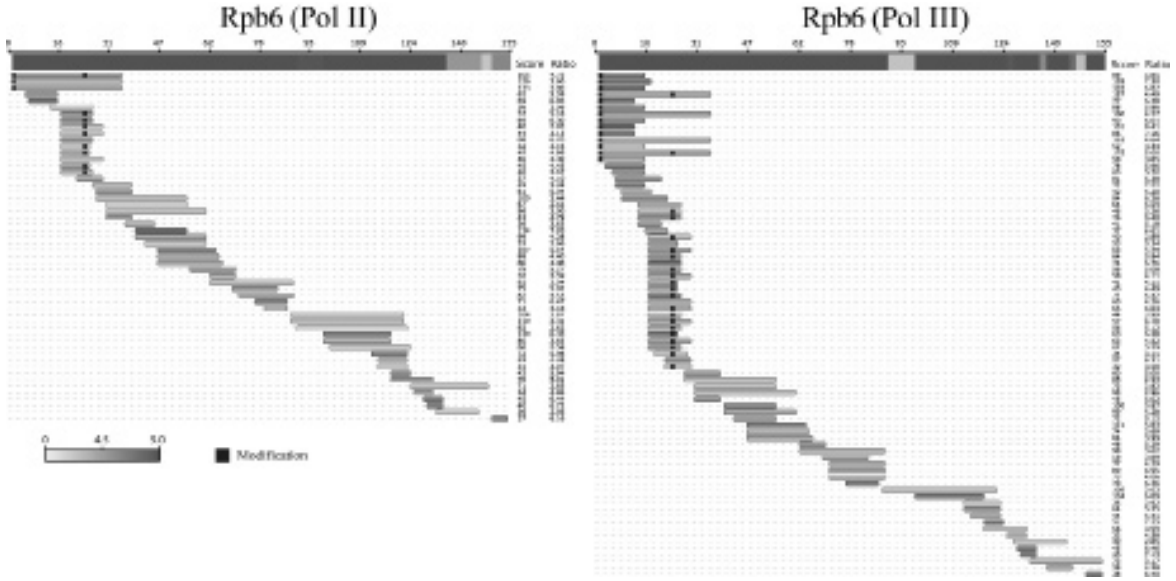

Figure 1 Sequence coverage for Rpb6 from Pol II and Pol III as obtained over all individual experiments using different proteases and activations methods. A white to grey color scheme is applied to highlight confidence in identification. The color of each peptide is determined by the Mascot score divided by the number of amino acids in the identified peptide. White represents a Mascot score of 0 while grey represents a Mascot score >9, per amino acid. For the overall sequence coverage the scores per amino acid in all different peptides as sequenced in all different experiments were summed. Peptides that were found to be phosphorylated or contained an oxidized methionine were considered to be unique.

phosphorylation sites.(Schlosser et al., 2005) MacCoss et al. not only used a multi-protease strategy but combined it with MuDPIT. Initial experiments focused on model proteins where near complete sequence coverages were obtained alongside several post translational modifications.(MacCoss et al., 2002) A more daunting sample which consisted of Cdc2p-TAP complexes, was also subjected to the same strategy where they identified over 200 proteins with 20 proteins attaining more than 40% sequence coverage. Finally, Cheeseman et al. applied the MacCoss et al. methodology on each subcomplex of the kinetochore which allowed the identification of 18 phosphorylation sites on 7 of the 28 proteins.(Cheeseman et al., 2002)

Recently electron transfer dissociation (ETD) has emerged as a new method for peptide sequencing (Good and Coon, 2006; Syka et al., 2004), possessing complementary features to collision induced dissociation (CID). ETD prefers higher charge states and therefore more basic peptides. The technique shows signs of maturity with elegant phosphopeptide sequencing data generated (Chi et al., 2007; Molina et al., 2007). The latter study also used multiple enzymes for a more comprehensive analysis although emphasis was placed on phosphorylation with all peptides being subjected to TiO_2 based phosphopeptide enrichment (Larsen et al., 2005; Pinkse et al., 2004).

Here we report a multiple-protease strategy, resulting in four complex peptide digests that are separated by reversed phase (RP) nanoLCMSMS, whereby each of the peptides is subjected to ETD with a supplemental collisional activation step (ETcaD) and CID sequencing by a linear ion trap mass spectrometer, leading to eight individual analyses. For comparison, we also analyzed the same samples by nanoLC FT-ICR using only CID MS/MS. An in-house tool was built that allows filtering on peptide scores followed by displaying the observed sequence coverage with a heat map indicating the confidence of the proteomics mapping of each of the amino acids in every protein of both Pol II and III as obtained in each of the eight individual experiments and combinations thereof.

By combining all these experiments, not consuming more than 5 µg of sample in total, we were able to achieve near to 100% sequence coverage of all individual protein constituents involved in the Pol II and Pol III complexes. Not only did we reach our goal of achieving comprehensive analysis, we identified a significant number of novel phosphorylation sites. The sites of Pol II were mapped on its structure, suggesting a role in regulating the conformation of the clamp region and, as a consequence, interaction of Pol II with nucleic acids.

RESULTS AND DISCUSSION

General results and assimilation of data

Initial focus was on the quality of RNA Pol II and III purifications and choices of proteases. Therefore, we analyzed Pol II and Pol III tryptic digests on an LTQ-FT mass spectrometer followed by a search against the Swis-Prot database. These results indicated a realtively clean purification where all 12 and 17 components of the Pol II and Pol III complexes, respectively were confidently identified (see scaffold file for full list of protein identifications found using the Swis-Prot database, https://bioinformatics.chem.uu.nl/supplementary/mohammed_RNAPol/ftcid_sprot_search). After applying a

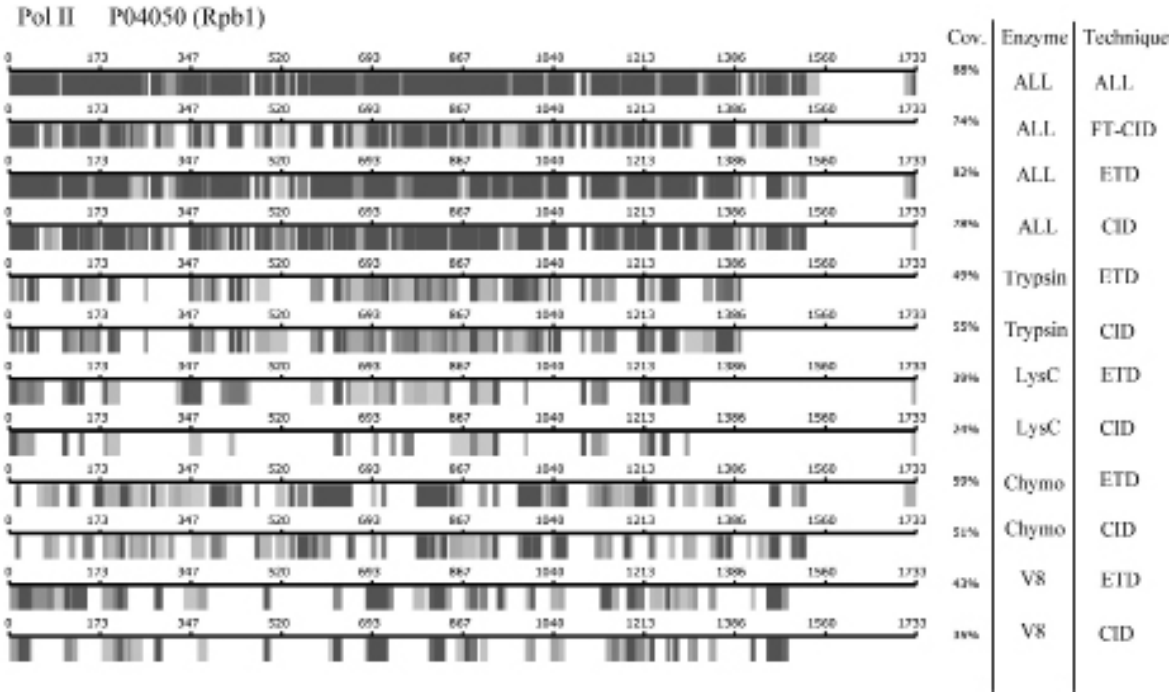

Figure 2 Sequence coverage achieved for Pol II Rbp1 using the indicated enzymes and activation methods for peptide sequencing. A white to grey colour scheme is applied to highlight confidence in peptide/amino acid identification. Each unique peptide (from all analyses) is broken down to its constituent amino acids and an equal fraction of the total peptide Mascot score is applied to each amino acid. If certain amino acids are present in more than one peptide with unique sequence then the amino acid scores are summed. White represents a Mascot score of 0 while grey represents a Mascot score >9, per amino acid

Mascot score cut-off of 30 ($p<0.01$), the average sequence coverage was 39%, although certain proteins were covered by not more than 10% (Supplementary Figure 1). Although the analyte mixtures represent reasonably complex digests where under-sampling by the mass spectrometer can be an issue there are other potential obstacles that may hamper obtaining higher sequence coverage including inappropriate size of the tryptic peptide, incomplete digestion, poor MS responses and poor CID fragmentation.

In order to tackle such issues a strategy was designed using a number of alternative enzymes to create complementary pools of proteolytic peptides. Moreover, sequential analysis by CID and electron transfer dissociation (ETD) was applied to allow a broader range of peptides to be successfully sequenced. In present-day proteomics trypsin has been the enzyme of choice for two main reasons. First trypsin digestion leads to highly selective protein hydrolysis and secondly tryptic peptides will contain a basic residue at the C-terminus, which is beneficial for sequencing by CID.(Huang et al., 2005; Wysocki et al., 2000) However, the introduction of ETD, which has a slightly different set of criteria for optimal peptide sequencing, may require a re-analysis of preferred enzyme for protein identification. In order to add to the protease discussion and to achieve comprehensive sequence coverage we chose and applied a set of proteases; trypsin, chymotrypsin, V8 (primarily a Glu-C) and Lys-C. Two microgram aliquots of sample were digested with one of the above mentioned proteases and approximately 0.5 ug (approximately 1 pmol) of material was subjected to nanoLC-MS-MS analysis, where each peptide ion was subjected to both CID and ETD peptide sequencing. To accept a peptide as being identified we applied a minimum MASCOT score of 30 per peptide and a MASCOT score/amino acid value of 2.5. The latter criterion was applied to filter out large peptides with high peptide MASCOT scores that possessed poor spectra with dubious annotation (see experimental).

A program was developed to filter and assimilate multiple mascot results files (*.dat), providing graphical outputs for peptide identifications and protein sequence coverage. Figure 1 represents such graphical output files for the identical proteins Rpb6 from Pol II (on the left) and Rpb6 from Pol III (on the right). In these files each unique peptide identified is aligned and mapped on the protein sequence, whereby the color of each amino acid indicates its identification confidence core. At the top the full protein sequence is shown, where the color indicates the confidence of identification which is obtained by summing the results over all parallel analyses (for details see experimental). As can be observed a similar proteome map is achieved for Rpb6 originating from Pol II or III. Identical sections of Rpb6 from Pol II and Pol III are observed to have similar levels of sequencing success. An alternative way of analyzing the data is represented in Figure 2 which shows the breakdown of the contribution for each enzyme and activation technique as obtained by each of the eight individual experiments for Rpb1 (Pol II) highlighting the complementary nature of the

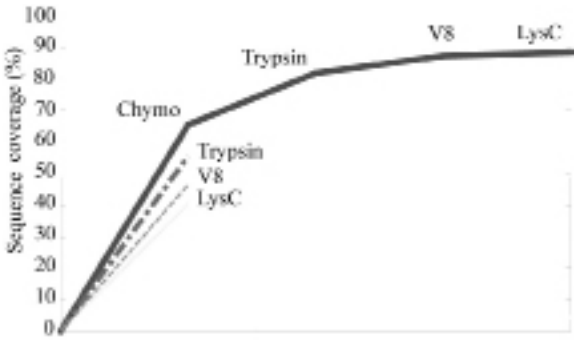

Figure 3 Graph indicating the shortest route to optimal sequence coverage for both Pol II and III (thick continuous line) with the four enzymes used: Chymotrypsin, Trypsin,V8 and LysC. In a single experiment Chymotrypsin performed best (first segment continuous line), closely followed by Trypsin (dot line). Applying these two enzymes alone, the coverage was on average 82%. As discussed earlier, LysC contributes the least to the coverage.

analyses. This Figure also emphasizes the need to perform multiple analyses to achieve a comprehensive analysis. Similar diagrams for the other Pol proteins are shown in the supplementary Figure 2. Additionally, supplementary dataset 1 displays the individual protein sequence coverage for each enzyme using ETD, CID and FT-CID).

An unanticipated finding of our multi-enzyme, multi-activation approach was that chymotrypsin rivals the commonly used protease trypsin in terms of peptide identifications and protein coverage for both CID and ETD. CID spectra of chymotryptic peptides are usually not as easily interpretable as tryptic peptides since they lack the characteristic dominant y –series. However, our data with chymotryptic peptides indicate that database sequence matching incomplete b and y series can be equally sufficient for confident peptide identifications. V8 produced a slightly lower number of peptide identifications for Pol II and Pol III for both CID and ETD, but did attain the highest number of phosphopeptides identifications. Lys-C provided a slightly different picture with peptides possessing three charges or more being more dominant – possibly ideal for ETD. However, the overall number of peptides identified with Lys-C was lower relative to the other enzymes. A possible cause may be found in the fact that Lys-C produces larger peptides, where non-covalent interactions can hamper peptide fragmentation and thus successful sequencing.(Hakansson et al., 2003; Swaney et al., 2007) The combination of ETD on an orbitrap or other high-resolution mass spectrometer would likely improve our Lys-C results, for instance through the use of initial charge state screening and an appropriate 'supplemental' activation(Swaney et al., 2007) to allow dissociation of non-covalently attached ETD product ions.

Figure 3, shows the average sequence coverage per protein analyzed for each of the used proteases. In a single experiment we found chymotrypsin most successful. If two experiments are performed we found, somewhat unexpectedly, that trypsin is the best partner to achieve highest sequence coverage. Figure 4 highlights that it is necessary to employ all enzymes to reach near complete sequence coverage, although in our experiment Lys-C did not provide a significant addition in sequence coverage. Combining all experimental data we obtain a near complete proteome sequence map-

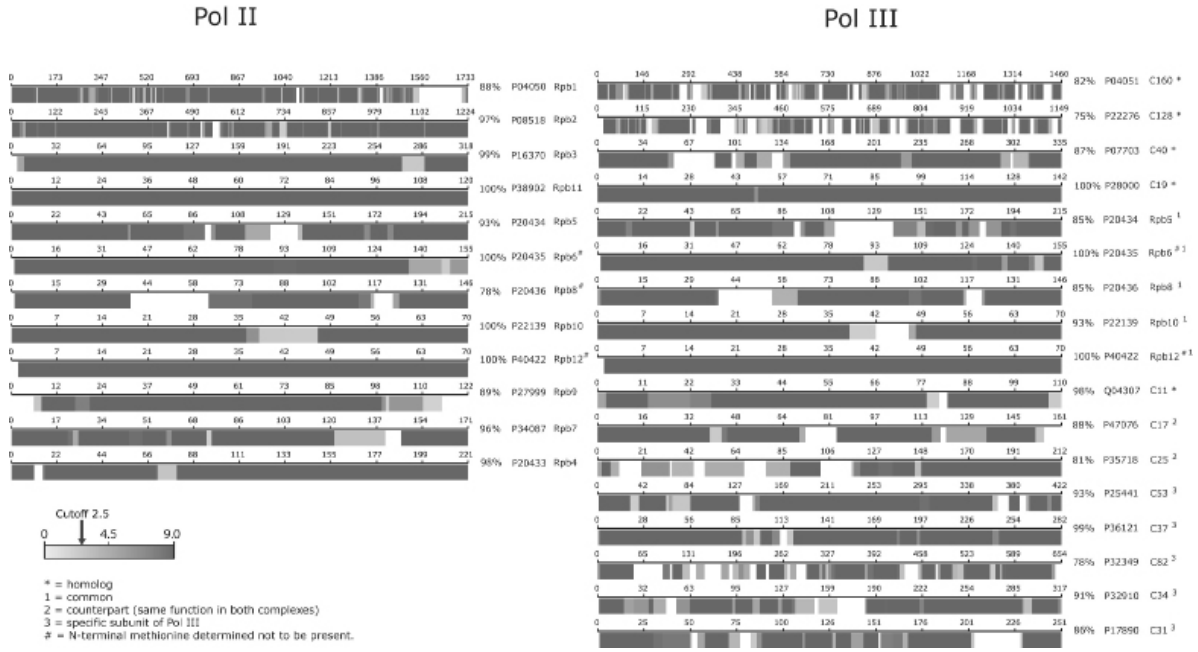

Figure 4 Graphical representation of the total sequence coverage achieved for each protein in Pol II and Pol III. A similar to Figure 3 white to red colour scheme is applied to highlight confidence in identification. * denotes homologous proteins in Pol II and Pol III. 1 denotes identical proteins that in Pol II and Pol III. 2 denotes counterpart proteins in Pol II and Pol III i.e. same function in both complexes. 3 denotes specific subunit of Pol III. # denotes N-terminal methionine determined not to be present.

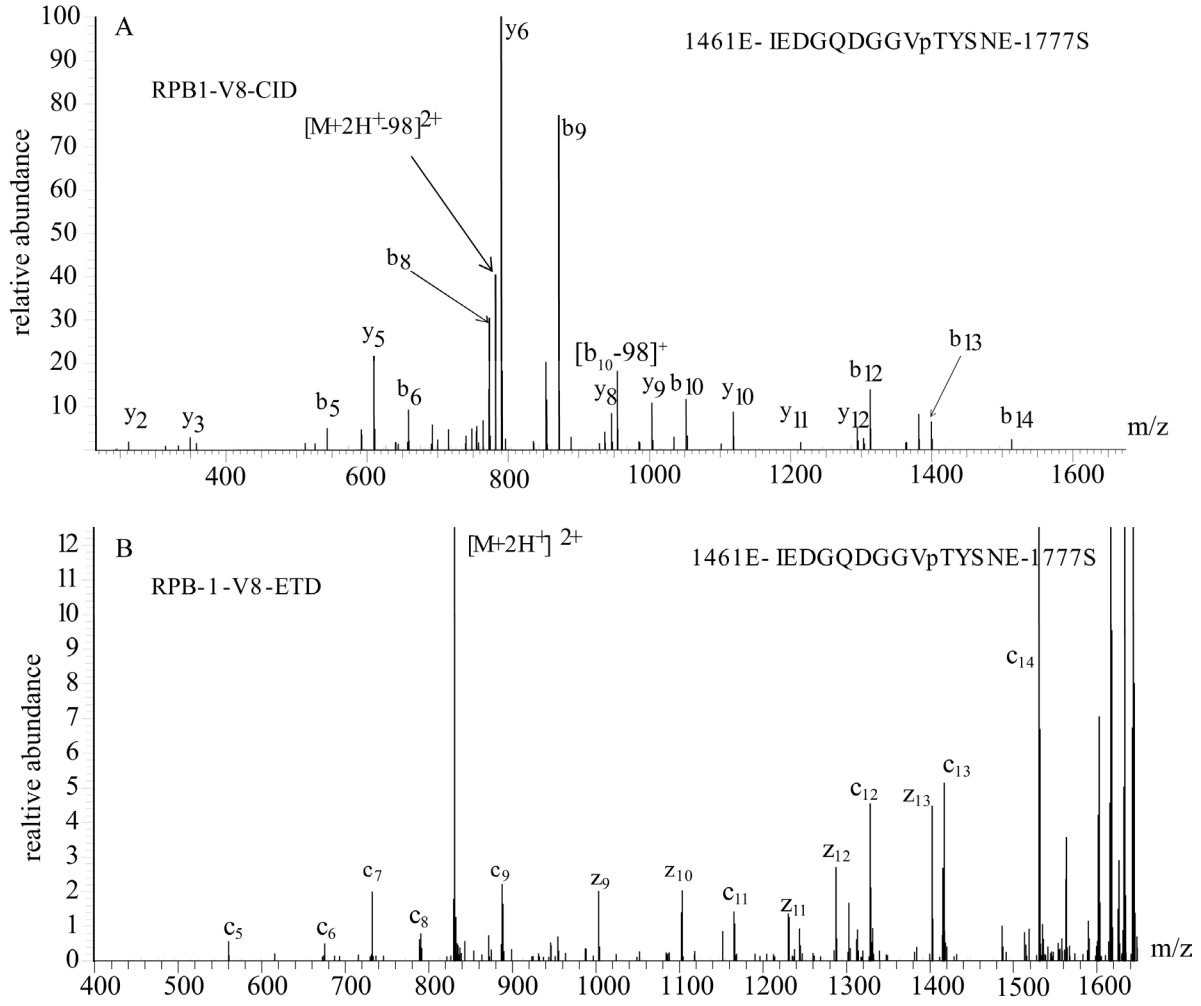

Figure 5 Tandem MS spectrum of the phosphopeptide IEDGQDGGVpTYSNE (Rpb1) originating from a V8 digest of the Pol II complex sequenced by using (A) CID, mascot score 106 and (B) ETD, mascot score 79. Both spectra were recorded on the LTQ ion trap, whereby the latter was obtained using supplemental collision activation.

ping of all protein subunits of Pol II and Pol III as visualized in Figure 4. A few proteins still show some gaps in the sequence coverage. For instance, the Rpb1 sequence coverage map contains a gap in its C-terminal domain (CTD) that is formed by 26-27 heptapeptide repeats of the sequence YSPTSPS spanning over 180 amino acids not targeted by any of the four used proteases. Such a peptide will be difficult to detect and sequence considering its size and its lack of basic residues. Interrogation of MS data did not indicate any candidate unidentified peaks that could suggest the presence of the CTD region, which may mean it was lost during sample preparation or introduction into the MS. In C31 an unobserved stretch of sequence consists almost exclusively of acidic residues which seem to be, by the here used proteases and conditions, non-cleavable. Included in all analyses was a search for phosphorylated peptides and acetylated protein N-termini. Most phosphorylated peptides were identified in more than one of the individual experiments. All identified and validated phosphosites are listed in Table 1 where it states the site location within the protein and in which proteolytic peptide it was detected that had the best Mascot score. Corresponding annotated spectra, additional confirmatory annotated spectra are available in supplemental dataset 2. Below, we focus in more detail on these phosphorylations and provide more detailed context for two Pol proteins; Rpb1 of Pol II and Rpb6, which is present and identical in Pol II and Pol III.

Rpb1

Rpb1 has a mass of approximately 192 kDa and represents the biggest subunit of Pol II. It is one of the ten core proteins and is homologous to C160 in Pol III. By applying different enzymes and activation methods we were able to obtain a near complete sequence coverage with the exception of the CTD domain (see above). Rpb1 represented a protein for which chymotrypsin provided the highest sequence coverage obtained by a single enzyme (59%). However, typical for the total dataset, the majority of the phosphorylation sites in Rpb1 were identified by using V8 as the enzyme for digestion. Rpb1 has been the subject of intensive research both in structural and biochemical ways. Its structure is known (Armache et al., 2003) and the phosphorylation status of the CTD governs the transcription cycle (Dahmus, 1995). However, despite extensive studies on the phosphorylation of the CTD of Rpb1 and other large scale phosphorylation studies (Chi et al., 2007; Li et al., 2007; Ptacek et al., 2005) we still detected three previously unidentified phosphorylation sites in Rpb1 that do not lie within the CTD; S348, T351 and T1471. Figure 5 shows CID and ETD spectra originating from V8 peptides used to identify T1471. This phosphopeptide and all the other phospho- identifications were mapped onto the structural model of Pol II as shown in Figure 6. T1471 lies in a region that links the core of Rpb1 to the CTD. Of particular interest, both phosphosites S348 and T351 are located adjacent to the so-called switch 2 at the base of the polymerase clamp, a region that undergoes conformational changes in the transition to a transcribing complex and contact DNA-RNA hybrids in the active site. Therefore, the phosphorylation status of these sites can play a direct role in regulating the conformation of the clamp region and, as a consequence, interaction of Pol II with nucleic acids (Armache et al., 2003; Armache et al., 2005; Bushnell and Kornberg, 2003).

Rpb6

Common to both polymerases, Rbp6 has been shown to play a role in the binding of the Rpb4/7 heterodimer in Pol II (Tan et al., 2003). Nevertheless, the regulation mechanism of the possible release or attachment of the two additional subunits is unknown. Rpb4 and Rpb7 are counterparts to C17 and C25 in Pol III. The heterodimer remained partly intact in dissociation experiments on both polymerases in native tandem mass spectrometry, confirming that Rpb4 and Rpb7 are located closely to each other, and probably strongly physically bound (Lorenzen et al., 2007). In those experiments Rpb6 was detected with two masses, separa-

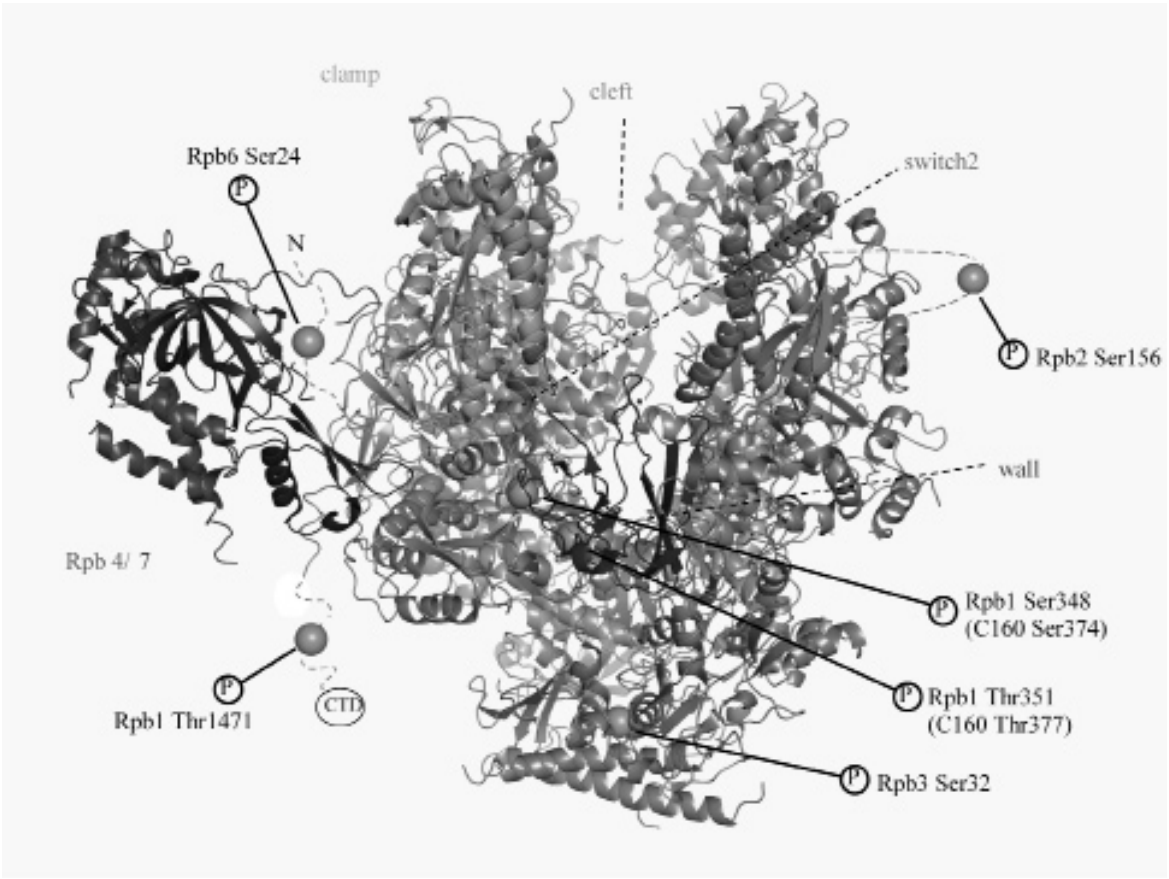

Figure 6 Structural model of Pol II highlighting the locations of the identified phosphosites.

Table 1 Identified Pol II and Pol III phosphorylation sites. In the last column it is indicated whether the site was identified earlier in large scale phosphoproteomics studies and if there is an interaction with a kinase reported.

Subunit	Phosphosite	Mascot Score	Method	Remark
Pol II Rbp1 (B220)	Ser 348	81	Tryp – LTQ (CID)	Not reported S1293 reported in (1) and T621 in (3)
Pol II Rbp1 (B220)	Thr 351	81	Tryp – LTQ (CID)	Not reported S1293 reported in (1) and T621 in (3)
Pol II Rbp1 (B220)	Thr 1471	110	Chym– LTQ-FT (CID)	
Pol II Rbp2 (B150)	Ser 156	49	V8– LTQ-FT (CID)	Not previously reported. Rpb2 has been shown to interact with RCK1 (4)
Pol II Rbp3 (B3)	Ser 32	53	Tryp – LTQ (ETD)	
Pol II Rbp6 (ABC 23)	Ser 24	69	V8– LTQ-FT (CID)	Not reported. Stoichiometrically phosphorylated, conserved between Pol II and Pol III, Rpb6 interacts with CMK1 (4)
Pol III C160	Thr 9	53	V8– LTQ (ETD)	
Pol III C53	Ser 119	58	Tryp – LTQ-FT (CID)	Not reported. S182,228, 232 & T323 reported (1,2,3)
Pol III C53	Ser 178	51	Tryp– LTQ-FT (CID)	Reported (1)
Pol III C53	Ser 224	62	Tryp – LTQ (ETD)	Reported (1,2)
Pol III C53	Thr 347	47	Lys-C– LTQ (ETD)	
Pol III C82	Ser 394	81	Lys-C – LTQ (ETD)	Reported (2,3), additionally S392,394 reported (2,3)
Pol III C37	Ser 52	71	V8 – LTQ (CID)	
Pol III C37	Thr 61	61	V8 – LTQ (ETD)	Reported (3)
Pol III AC19	Thr 15	48	Tryp– LTQ (CID)	Reported (3)
Pol III AC19	Thr 33	115	Tryp – LTQ (CID)	Reported (1)
Pol III C31	Thr 101	98	Lys-C– LTQ (ETD)	
Pol III C31	Ser 189	107	Lys-C– LTQ (ETD)	Reported (1,3) T190 reported in (3)
Pol III Rpb6 (ABC 23)	Ser 24	62	V8 – LTQ-FT (CID)	Not reported. Stoichiometrically phosphorylated, conserved between Pol II and Pol III, Rpb6 interacts with CMK1 (4)

(1) Li X., Gerber S.A., Rudner A.D., Beausoleil S.A., Haas W., Villen J., Elias J.E., Gygi S.P.;"Large-scale phosphorylation analysis of alpha-factor-arrested Saccharomyces cerevisiae."; J. Proteome Res. 6:1190-1197 (2007).
(2) Chi A., Huttenhower C., Geer L.Y., Coon J.J., Syka J.E.P., Bai D.L., Shabanowitz J., Burke D.J., Troyanskaya O.G., Hunt D.F.„Analysis of phosphorylation sites on proteins from Saccharomyces cerevisiae by electron transfer dissociation (ETD) mass spectrometry."; Proc. Natl. Acad. Sci. U.S.A. 104:2193-2198 (2007).
(3) Smolka M.B., Albuquerque C.P., Chen S.H., Zhou H.; „Proteome-wide identification of in vivo targets of DNA damage checkpoint kinases."; Proc. Natl. Acad. Sci. U.S.A. 104:10364-10369 (2007).
(4) Ptacek J, Devgan G, Michaud G, Zhu H, Zhu X, Fasolo J, Guo H, Jona G, Breitkreutz A, Sopko R, McCartney RR, Schmidt MC, Rachidi N, Lee SJ, Mah AS, Meng L, Stark J, Stern DF, De Virgilio C, Tyers M, Andrews B, Gerstein M, Schweitzer B, Predki PF, Snyder M. "Global analysis of protein phosphorylation in yeast" Nature. 438(7068):679-84 (2005).

ted by 80 Da, in a ratio of about 2:3, most likely caused by an abundant single phosphorylation on Rpb6. Through our multi-enzyme approach we were able to determine the site as being S24. This phosphosite was identified using different enzymes and with both activation methods for Rpb6 in both Pol II and Pol II complexes.

Further RNA pol phosphorylation

Including the phosphorylation sites mentioned above we found in total 19 sites in the two Pol complexes excluding the CTD of Rpb1, the largest subunit of Pol II. Seven of them were found in Pol II and thirteen in Pol III (see Table 1). Six of these sites have been reported previously in large scale phosphoproteomics studies (Chi et al., 2007; Gerber et al., 2008; Li et al., 2007; Ptacek et al., 2005). Interestingly, most of the phosphorylation sites are in proteins, which are unique for Pol II or Pol III and only one was found in a subunit that is shared by both Pol II and Pol III (i.e. the above mentioned site in Rpb6).

The level of phosphorylation is dependent on many factors like cell cycle growth condition, activity in transcription etc. Evidently, due to the purification and sample handling as well as detection methods some phosphorylation sites may still be missed. However, all but one of the amino acids in identical protein subunits were detected as unphosphorylated in this analysis. Protein phosphorylation often has a regulatory function; yet, the two protein complexes fulfill different roles in the transcription machinery and share the same environment in the nucleus. The locations of the identical subunits are similar in both Pol II and Pol III and are thus likely to be susceptible to phosphorylation at similar sites as found for the subunit Rpb6. Since Pol II and Pol III transcribe different classes of genes, it is unlikely that phosphorylation on these identical/common proteins would al-

low this differential behavior and, perhaps, this is a reason why little phosphorylation is observed since these subunits are unlikely to be kinase targets. As mentioned above, the exception is Rpb6 where an abundant, almost stoichiometric, phosphorylation is observed. Rpb6 has been shown to play a role in the binding of a homologue heterodimer (Rpb4 and Rpb7 which are homologue to C17 and C25 in Pol III) to the two polymerases. The available crystal structure (Armache et al., 2003) reveals that the N-terminal tail of Rpb6, where the observed phosphosite resides, is not resolved and would be expected to lie on the surface of the complex. These features underline the N-terminus as a possible target for a kinase.

Further analysis of the remaining identified phosphorylation sites in the homologous proteins of the two polymerases was performed with the primarily tool being sequence comparison using the BLAST algorithm (Altschul et al., 1997). S156 is located in a flexible loop and can likely be a target of kinase activity. All other sites were not deemed to be conserved.

CONCLUSION

Here we used a multiplexed proteomics approach using four different proteases and two different peptide activation methods to attempt a complete mapping of all proteins of the yeast RNA polymerase II and III. Our approach allowed a near complete sequence coverage of each protein, consuming in total less than 5 ug (approximately 10 pmol) of starting material, and provided informative insights in the strength and weaknesses of different proteases and activation methods. As an additional benefit, our near comprehensive mapping of the two Pol complexes enabled us to reveal several novel phosphorylation sites. Mapping these new sites onto the crystal structure suggests that, for some, a role may exist in regulating the conformation of the clamp region and, as a consequence, interaction of RNA with nucleic acids.

EXPERIMENTAL PROCEDURES

Purification of Pol II and Pol III

Pol II and Pol III core were purified as described in references 29 and 7, respectively. The additional subunits Rpb4 and 7 were over expressed and purified as described (Sakurai et al., 1999). The RNA pol purifications were of very high purity as evidenced by the fact that all RNA pol subunits were detected at the top of the list of identified proteins. Next to RNA pol subunits a few elongation factor components were detected as well as some of the usual background, abundant, heat shock proteins. A list of all identified proteins (and their identified peptides) is given in the supplementary data presented in the scaffold file .

Mass spectrometry

Two microgram aliquots of the purified protein complex were resuspended in 50 mM NH4HCO3 pH 8.0. Reduction and subsequent alkylation were performed with 45 mM DTT for 30 min at 56 °C and 100 mM iodoacetamide for 30 min at room temperature in the dark, respectively. Subsequently, 80 ng (1/25 w/w) of protease (endoproteinases Lys-C, Trypsin, Glu-C, Chymotrypsin (all Roche-Diagnostics, Netherlands)), was added and digestion was performed overnight at 37 °C. The sample was then acidified using 5% formic acid. The digestions were subsequently analysed by nanoLC-LTQ-FT-MS (Thermo, San Jose, CA) and by nanoLC-LTQ-XL-MS (Thermo, San Jose, CA) at a material level of 0.5 µg. An Agilent 1100 series LC system was equipped with an 20 mm Aqua C18 (Phenomenex, Torrance, CA) trapping column (packed in-house, i.d., 100 µm; resin, 5 µm) and a 250 mm ReproSil-Pur C18-AQ (Dr. Maisch GmbH, Ammerbuch, Germany) analytical column (packed in-house, i.d., 50 µm; resin, 3 µm). Trapping was performed at 5 µL/min for 10 min and washed with solvent A (0.6% acetic acid in water) and elution was achieved with a gradient of 0-32% B (0.6% acetic acid in 80/20 acetonitrile/water) in 60 min, 32-40% B in 5 min, 40-100% B in 2 min and 100% B for 2 min leading to a total analysis time of 90 minutes. The flow rate was passively split from 0.4 ml/min to 100 nl/min when performing the elution analysis. Nanospray was achieved using a distally coated fused silica emitter (New Objective, Cambridge, MA) (o.d., 360 µm; i.d., 20 µm, tip i.d. 10 µm) biased to 1.8 kV. In the case of the LTQ-FT, the mass spectrometer was operated in the data dependent mode to automatically switch between MS and MS/MS. Survey full scan MS spectra were acquired from m/z 350 to m/z 1500 in the FT-ICR with a resolution of R=100,000 at m/z 400 after accumulation to a target value of 2,000,000 in the linear ion trap. The two most intense ions were fragmented in the linear ion trap using collisionally induced dissociation at a target value of 10,000. In the case of the LTQ-XL, the mass spectrometer was operated in the data dependent mode to automatically switch between MS and MS/MS ETcaD and MS/MS CAD. Survey full scan MS spectra were acquired from m/z 350 to m/z 1500 in the LTQ after accumulation to a target value of 30,000 in the linear ion trap. The two most intense ions were fragmented in the linear ion trap at a target value of 10,000.

Spectra were processed with Bioworks 3.3 (Thermo, Bremen, Germany) and the subsequent data analysis was carried out using the Mascot (version 2.1.0) software platform (Matrix Science, London, UK). For the initial screen, the LTQ-FT analyses were searched against the swiss protein database (version 51.6) with the appropriate enzyme allowing 2 missed cleavages, carbamidomethyl (C) as fixed modification and oxidation (M), phosphorylation (ST) and N-acetylation (protein N terminus) as variable modifications. The peptide tolerance was set to 20 ppm and the MS/MS tolerance to 0.9 Da. The proteins identified were used to create a new database to be used for analysis of the LTQ-XL ETcaD and CAD Data. The LTQ results were searched against the new database with the appropriate enzyme allowing 2 missed cleavages and appropriate instrument configuration. For Glu-C the number of miscleavages allowed was 5 due to the enzyme possessing poor efficiency under these digestion conditions. Carbamidomethyl (C) was used as a fixed modification and oxidation (M), phosphorylation (ST) and N-acetylation (protein N terminus) as variable modifications. The peptide tolerance was set to 3 Da and the MS/MS tolerance to 1.5 Da. All data was compiled into a scaffold database which can be downloaded (https://bioinformatics.chem.uu.nl/supplementary/mohammed_RNAPol/) and freely interrogated using the scaffold viewer.

Software and visualization

A software program was developed implementing a graphical user interface (GUI, JAVA 1.6 Netbeans IDE 5.5) for the integration and filtering of the peptide identifications from the multiple parallel LC/MS/MS experiments. Multiple Mascot output files (i.e. *.dat-files) together with the corresponding protein sequences database (i.e. fasta-files) were used as input. Peptide identifications were combined and mapped onto the proteins followed by calculating the sequence coverage per protein. Before mapping, peptides can be and were filtered based on several criteria, such as, the mascot peptide score. As we found that the MASCOT peptide score filter was not very reliable when considering larger peptides (i.e. above 2500 Da)

we chose to introduce an extra filter by dividing the peptide score by the number of amino acids present in the peptide. The latter value enabled us to filter out large peptides which tend to have better scores by chance but are often ambiguous peptide identifications. The final filter values used were; Mascot peptide score >30 with at least a score contribution of 2.5 per amino acid. As an additional filter all identical peptides within a single experiment were removed, accepting only the peptide with the highest Mascot score. To attain the overall sequence coverage picture the scores of the same amino acid in different unique sequences are algebraically summed. Peptides that are phosphorylated or contain an oxidized methionine at a particular site are considered to be unique. In the heat maps produced protein coverage is now indicated per individual amino acid using the score/amino acid ratios described above for all the underlying peptides. Since phosphosite localization by Mascot is still an imprecise process, the amino acid is only considered 'detected' with no consideration placed on its phosphorylation status and thus, the heat maps do not consider or display phosphorylation sites. All phosphosites suggested by Mascot are manually evaluated and the results are presented in table 1 and supplementary dataset 2. These values were transformed into a white to dark grey color gradient in linearized RGB space interpolating between 0 and 9.0, thereby creating a confidence heat map. All scores above 9.0 are shown in dark grey. From the program images as well as tabular data with detailed peptide information can be exported. A stand-alone version of the software is available and can be downloaded from https://bioinformatics.chem.uu.nl/supplementary/mohammed_RNAPol/.

REFERENCES

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J.H., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25, 3389-3402.

Armache, K.J., Kettenberger, H., and Cramer, P. (2003). Architecture of initiation-competent 12-subunit RNA polymerase II. Proceedings Of The National Academy Of Sciences Of The United States Of America 100, 6964-6968.

Armache, K.J., Mitterweger, S., Meinhart, A., and Cramer, P. (2005). Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. J Biol Chem 280, 7131-7134.

Boersema, P.J., Divecha, N., Heck, A.J.R., and Mohammed, S. (2007). Evaluation and optimization of ZIC-HILIC-RP as an alternative MudPIT strategy. Journal of Proteome Research 6, 937-946.

Bushnell, D.A., and Kornberg, R.D. (2003). Complete, 12-subunit RNA polymerase II at 4.1-angstrom resolution: Implications for the initiation of transcription. Proceedings Of The National Academy Of Sciences Of The United States Of America 100, 6969-6973.

Cargile, B.J., Bundy, J.L., Freeman, T.W., and Stephenson, J.L. (2004). Gel based isoelectric focusing of peptides and the utility of isoelectric point in protein identification. Journal of Proteome Research 3, 112-119.

Cheeseman, L.M., Anderson, S., Jwa, M., Green, E.M., Kang, J.S., Yates, J.R., Chan, C.S.M., Drubin, D.G., and Barnes, G. (2002). Phospho-regulation of kinetochore-microtubule attachments by the aurora kinase Ipl1p. Cell 111, 163-172.

Chi, A., Huttenhower, C., Geer, L.Y., Coon, J.J., Syka, J.E.P., Bai, D.L., Shabanowitz, J., Burke, D.J., Troyanskaya, O.G., and Hunt, D.F. (2007). Analysis of phosphorylation sites on proteins from Saccharomyces cerevisiae by electron transfer dissociation (ETD) mass spectrometry. Proceedings of the National Academy of Sciences of the United States of America 104, 2193-2198.

Cramer, P., Bushnell, D.A., Fu, J., Gnatt, A.L., Maier-Davis, B., Thompson, N.E., Burgess, R.R., Edwards, A.M., David, P.R., and Kornberg, R.D. (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. Science 288, 640-649.

Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. Science 292, 1863-1876.

Dahmus, M.E. (1995). Phosphorylation of the C-terminal domain of RNA polymerase II. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression 1261, 171-182.

Distler, A.M., Kerner, J., Peterman, S.M., and Hoppel, C.L. (2006). A targeted proteomic approach for the analysis of rat liver mitochondrial outer membrane proteins with extensive sequence coverage. Analytical Biochemistry 356, 18-29.

Fernandez-Tornero, C., Bottcher, B., Riva, M., Carles, C., Steuerwald, U., Ruigrok, R.W.H., Sentenac, A., Muller, C.W., and Schoehn, G. (2007). Insights into Transcription Initiation and Termination from the Electron Microscopy Structure of Yeast RNA Polymerase III. Molecular Cell 25, 813.

Fischer, F., and Poetsch, A. (2006). Protein cleavage strategies for an improved analysis of the membrane proteome. Proteome Science 4.

Garcia, B.A., Pesavento, J.J., Mizzen, C.A., and Kelleher, N.L. (2007). Pervasive combinatorial modification of histone H3 in human cells. Nature Methods 4, 487-489.

Gatlin, C.L., Eng, J.K., Cross, S.T., Detter, J.C., and Yates, J.R. (2000). Automated identification of amino acid sequence variations in proteins by HPLC/microspray tandem mass spectrometry. Analytical Chemistry 72, 757-763.

Gerber, J., Reiter, A., Steinbauer, R., Jakob, S., Kuhn, C.D., Cramer, P., Griesenbeck, J., Milkereit, P., and Tschochner, H. (2008). Site specific phosphorylation of yeast RNA polymerase I. Nucleic Acids Res 36, 793-802.

Good, D.M., and Coon, J.J. (2006). Advancing proteomics with ion/ion chemistry. Biotechniques 40, 783-789.

Hakansson, K., Chalmers, M.J., Quinn, J.P., McFarland, M.A., Hendrickson, C.L., and Marshall, A.G. (2003). Combined electron capture and infrared multiphoton dissociation for multistage MS/MS in a Fourier transform ion cyclotron resonance mass spectrometer. Analytical Chemistry 75, 3256-3262.

Han, X.M., Jin, M., Breuker, K., and McLafferty, F.W. (2006). Extending top-down mass spectrometry to proteins with masses greater than 200 kilodaltons. Science 314, 109-112.

Huang, Y.Y., Triscari, J.M., Tseng, G.C., Pasa-Tolic, L., Lipton, M.S., Smith, R.D., and Wysocki, V.H. (2005). Statistical characterization of the charge state and residue dependence of low-energy CID peptide dissociation patterns. Analytical Chemistry 77, 5800-5813.

Krijgsveld, J., Gauci, S., Dormeyer, W., and Heck, A.J.R. (2006). In-gel isoelectric focusing of peptides as a tool for improved protein identification. Journal of Proteome Research 5, 1721-1730.

Larsen, M.R., Thingholm, T.E., Jensen, O.N., Roepstorff, P., and Jorgensen, T.J.D. (2005). Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. Molecular & Cellular Proteomics 4, 873-886.

Li, X., Gerber, S.A., Rudner, A.D., Beausoleil, S.A., Haas, W., Villen, J., Elias, J.E., and Gygi, S.P. (2007). Large-scale phosphorylation analysis of alpha-factor-arrested Saccharomyces cerevisiae. J Proteome Res 6, 1190-1197.

Lorenzen, K., Vannini, A., Cramer, P., and Heck, A.J.R. (2007). Structural biology of RNA polymerase III: mass spectrometry elucidates subcomplex architecture. Structure 15, 1237-1245.

MacCoss, M.J., McDonald, W.H., Saraf, A., Sadygov, R., Clark, J.M., Tasto, J.J., Gould, K.L., Wolters, D., Washburn, M., Weiss, A., et al. (2002). Shotgun identification of protein modifications from protein complexes and lens tissue. Proceedings of the National Academy of Sciences of the United States of America 99, 7900-7905.

Molina, H., Horn, D.M., Tang, N., Mathivanan, S., and Pandey, A. (2007). Global proteomic profiling of phosphopeptides using electron transfer dissociation tandem mass spectrometry. Proceedings of the National Academy of Sciences of the United States of America 104, 2199-2204.

Pieper, R., Gatlin, C.L., Makusky, A.J., Russo, P.S., Schatz, C.R., Miller, S.S., Su, Q., McGrath, A.M., Estock, M.A., Parmar, P.P., et al. (2003). The human serum proteome: Display of nearly 3700 chromatographically separated protein spots on two-dimensional electrophoresis gels and identification of 325 distinct proteins. Proteomics 3, 1345-1364.

Pinkse, M.W.H., Uitto, P.M., Hilhorst, M.J., Ooms, B., and Heck, A.J.R. (2004). Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-nanoLC-ESI-MS/MS and titanium oxide precolumns. Analytical Chemistry 76, 3935-3943.

Ptacek, J., Devgan, G., Michaud, G., Zhu, H., Zhu, X., Fasolo, J., Guo, H., Jona, G., Breitkreutz, A., Sopko, R., et al. (2005). Global analysis of protein phosphorylation in yeast. Nature 438, 679-684.

Running, W.E., Ravipaty, S., Karty, J.A., and Reilly, J.P. (2007). A top-down/bottom-up study of the ribosomal proteins of Caulobacter crescentus. Journal of Proteome Research 6, 337-347.

Sakurai, H., Mitsuzawa, H., Kimura, M., and Ishihama, A. (1999). The Rpb4 Subunit of Fission Yeast Schizosaccharomyces pombe RNA Polymerase II Is Essential for Cell Viability and Similar in Structure to the Corresponding Subunits of Higher Eukaryotes. Mol. Cell. Biol. 19, 7511-7518.

Schlosser, A., Vanselow, J.T., and Kramer, A. (2005). Mapping of phosphorylation sites by a multi-protease approach with specific phosphopeptide enrichment and nanoLC-MS/MS analysis. Analytical Chemistry 77, 5243-5250.

Swaney, D.L., McAlister, G.C., Wirtala, M., Schwartz, J.C., Syka, J.E.P., and Coon, J.J. (2007). Supplemental activation method for high-efficiency electron-transfer dissociation of doubly protonated peptide precursors. Analytical Chemistry 79, 477-485.

Syka, J.E.P., Coon, J.J., Schroeder, M.J., Shabanowitz, J., and Hunt, D.F. (2004). Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proceedings of the National Academy of Sciences of the United States of America 101, 9528-9533.

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

Tan, Q., Prysak, M.H., and Woychik, N.A. (2003). Loss of the Rpb4/Rpb7 subcomplex in a mutant form of the Rpb6 subunit shared by RNA polymerases I, II, and III. Mol Cell Biol 23, 3329-3338.

Wu, C.C., MacCoss, M.J., Howell, K.E., and Yates, J.R. (2003). A method for the comprehensive proteomic analysis of membrane proteins. Nature Biotechnology 21, 532-538.

Wu, S.L., Kim, J., Hancock, W.S., and Karger, B. (2005). Extended Range Proteomic Analysis (ERPA): A new and sensitive LC-MS platform for high sequence coverage of complex proteins with extensive post-translational modifications-comprehensive analysis of beta-casein and epidermal growth factor receptor (EGFR). Journal of Proteome Research 4, 1155-1170.

Wysocki, V.H., Tsaprailis, G., Smith, L.L., and Breci, L.A. (2000). Special feature: Commentary - Mobile and localized protons: a framework for understanding peptide dissociation. Journal of Mass Spectrometry 35, 1399-1406.

Zubarev, R.A., Kelleher, N.L., and McLafferty, F.W. (1998). Electron capture dissociation of multiply charged protein cations. A nonergodic process. Journal of the American Chemical Society 120, 3265-3266.

SUMMARY AND OUTLOOK

Kristina Lorenzen
Biomolecular Mass Spectrometry and Proteomics Group, Bijvoet Center for Biomolecular Research and Utrecht Faculty of Science, Utrecht University, Sorbonnelaan 16, 3584 CA Utrecht, the Netherlands

SUMMARY

While technologies in native mass spectrometry are still under development, it already allows research on complete proteins and protein complexes up to a seemingly unlimited size. This would not have been possible without the technical developments in all related fields, for example ionization, instrumentation and sample preparation and handling. An example for further instrumentation development is given in chapter two of this thesis. Here we attempted to improve one of the instruments frequently used in native mass spectrometry, the Q ToF, by modifying the pressures, nature of collision gas and orifices of the collision cell. These changes on the Q ToF allowed us to obtain the results that are described in chapter 2 to 4 of this thesis.

Nowadays mass spectrometry aims to bridge between data of high throughput proteomic screens of protein networks in the cell and high resolution structural data obtained by methods like x ray, cryo EM and NMR. The most direct result obtained by native mass spectrometry for protein complexes is often the stoichiometry. While with homogeneous complexes this is visible at first sight most of the times, heterogeneous complexes often need to be dissociated by tandem mass spectrometry. In chapter 3 we determined the stoichiometry of the purified portal ring and the assembly of the portal ring with the first tail accessory factor gp4 of bacteriophage P22. Besides determining the stoichiometry we were also able to make conclusions about the binding behavior and structural changes that occur upon the assembly. Charge state distributions generated in the ESI process arouse the suspicion that the portal might undergo a major conformational change upon binding of gp4. A new technique called ion mobility mass spectrometry confirmed this structural change. These results gave new insights into the maturation process of P22.

Several interactome studies have shown that it is possible to establish interaction networks of endogenously expressed proteins in yeast by affinity pulldowns combined with mass spectrometry. While these studies have shown that proteins almost always work in interaction with other proteins they fail to give information about the structure and stoichiometry of the proteins. Native mass spectrometry can provide new and complementary information here. We have analyzed the architecture of the multisubunit RNA polymerases Pol II and Pol III. The measurements presented in chapter 4 resulted in insights into the subunit architecture of Pol II and III. We here compared the results gained from Pol II, where detailed structural properties are known, with Pol III. The data demonstrates that each of the 17 subunits of Pol III is present as a single copy. Our analysis provided new information about the heterodimeric subcomplexes in Pol III. The obtained model for the Pol III subunit architecture might guide future biochemical and structural studies of this central cellular enzyme.

Native mass spectrometric analysis of the polymerases has shown modifications on some of the subunits, for example the binding of Zn^{2+} to C11 or the stoichiometric phosphorylation of Rpb6. However, native mass spectrometry fails to pinpoint the exact site of such modifications. In chapter 5 we revealed several new modifications on Pol II and III. In vivo those post translational modifications (PTM) are needed for proper molecular function, to determine the activity, turnover, localization and interaction of proteins with other proteins. The perquisite to find PTMs with mass spectrometry is high protein sequence coverage, because each missed residue might contain a PTM or information that is important. Different ways of cutting the proteins into small pieces (i.e. peptides) combined with different fragmentation methods in the mass spectrometer allowed us to map nearly the complete sequences of all constituents of both Pol II and III and revealed several new phosphorylation sites.

CHALLENGES

This thesis demonstrates some of the possibilities mass spectrometry can provide to gain new insight into structure and function of protein complexes. However native mass spectrometry is still far from being a routine technique, because optimization and careful treatment of each sample measured is required. The major bottleneck at the moment is possibly coupling biochemistry to mass spectrometry. This begins with the sample preparation for mass spectrometry is very sensitive to adducts produced in the ESI process by regularly used buffers in protein chemistry. Sodium, potassium or other salts as well as non volatile substances like detergents or glycerol lead to adducts on the protein of interest, hampering the determination of charges states and thus masses. On the other hand these buffer substances are often necessary to keep the proteins in solution and maintain their conformation. While for easily soluble proteins it is mostly unproblematic to exchange the buffer to a mass spectrometry compatible substance like ammonium acetate, especially membrane proteins remain a challenge. Proteins that are endogenously expressed and then purified, often yield only micrograms of material and normally need to be in a near

physiological buffer containing substances not compatible with mass spectrometry as just mentioned above. Thus there is the need to concentrate the sample and exchange the buffer, this often results in the loss of the sample due to aggregation or degradation. Additionally the time consuming sample handling can lead to the loss of transient or weak interactions of protein complexes. Besides the heterogeneity caused by salts and detergents also heterogeneity in the sample can be a critical point. One cause for this may be post translational modifications. Especially glycosylation with the multiple and differently branched sugars lead to a magnitude of charge states, so close to each other that a mass determination is not possible anymore. Another cause can be RNA or DNA of different sizes bound to the protein or protein complex of interest. In protein assemblies degradation products of even only one of the assembly partners might also give rise to spectra which are no longer interpretable.

OUTLOOK

While in the last decade biomolecular mass spectrometry has mainly been a method for peptide sequencing and stoichiometric determination of protein oligomers, it now more and more is used to give important insights into the structural properties of proteins and protein complexes. Main advantages in comparison to other structural techniques are the rather low sample consumption, principally no size limit and fast analysis. The drawback is the rather low resolution obtained by native mass spectrometry. Even though there are still major improvements necessary, native mass spectrometry is in principle compatible with high throughput purification techniques. Solutions here might be more selective tagging, up scaling of the amounts purified or finding new buffers and reagents compatible with both, retaining the protein intact and in its native state during the purification and the requirements of the mass spectrometric analysis. Foreseeable instrumentation development will result in faster and more sensitive mass spectrometers, leading to even lower sample consumption. Very recent developments on the modified Q ToF as described in chapter 2 allow detection up to even higher m/z values (approximately 90.000) than have been possible so far. Adjustments in the instruments also allow now twice as high fragmentation voltages. This combination enables the fragmentation of large protein assemblies which has not been possible before. With the high mass technology available for Quadrupols it might be interesting to adapt a triple quadrupole instrument for high mass ions, allowing parent ion scanning and neutral loss detection. Even more radical innovations can be expected. For instance there have been attempts to couple a mass spectrometer directly to electron microscopy by introducing a EM grid in the time of flight region of a Q ToF. The invention of ion mobility coupled with mass spectrometry has opened a new avenues. Combined with computational modeling increased insights will be gained on the dynamics, structural properties and behavior of proteins in the gas phase. Combined with energy and solvents ion mobility mass spectrometry and computational modeling this might also offer possibilities for the determination of folding characteristics of proteins. Evidently native mass spectrometry should not be the only technique taken into account when researching protein (complex) properties. Methods like peptide mass spectrometry, chemical cross linking, hydrogen deuterium exchange and of course (cryo) EM, NMR and X ray combined with computational modeling can be used to obtain complementary insights into the protein complex of interest.

NEDERLANDSE SAMENVATTING

Kristina Lorenzen, Esther van Duijn-Keizer
Biomolecular Mass Spectrometry and Proteomics Group, Bijvoet Center for Biomolecular Research and Utrecht Faculty of Science, Utrecht University, Sorbonnelaan 16, 3584 CA Utrecht, the Netherlands

Terwijl de techniek natieve massaspectrometrie (natieve MS) continue onder ontwikkeling is, behoort het al tot de mogelijkheden om intacte eiwitten en eiwit complexen te meten tot een wat haast lijkt ongelimiteerde grootte. Dit zou niet mogelijk zijn geweest zonder technologische ontwikkelingen in alle aanverwante onderzoeksvelden, bijvoorbeeld op het gebied van de ionisatie, de massaspectrometers zelf, maak ook wat betreft de monster voorbereiding.
Hoofdstuk 2 beschrijft slechts een voorbeeld van mogelijke technologische ontwikkelingen binnen de natieve MS. Een veel gebruikte massaspectrometer, de Q-ToF is hier aangepast waardoor dit instrument nu beter geschikt is voor metingen aan zeer grote eiwitcomplexen. De drukken in de verschillende ruimtes van de massaspectrometer zijn geoptimaliseerd maar ook het botsingsgas is veranderd en de grootte van de openingen van de botsingscel. Deze mogelijkheden met de Q-ToF die hierdoor zijn ontstaan worden beschreven in hoofdstuk 2 en 4 van dit proefschrift.

De natieve massaspectrometrie richt zich erop het informatie-gat te dichten dat tot op heden bestaat tussen de data gegenereerd uit proteomics experimenten die eiwit netwerken in de cel kunnen ophelderen en de hoge resolutie structuur informatie verkregen met technieken zoals kristallografie, elektronen microscopie en NMR. Met natieve MS kan soms direct de stoichiometrie van eiwit complexen worden bepaald. Voor homo-oligomere complexen is dit vaak gelijk te zien, terwijl voor meer ingewikkelde hetero-oligomere complexen tandem-massaspectrometrie nodig is om de stoichiometrie vast te stellen. In hoofdstuk 3 hebben we de stoichiometrie bepaald van een gezuiverd complex uit bacteriofaag P22. Dit complex vormt een ring (portal) en zorgt voor een opening in het faag-hoofd waardoor het genoom getransporteerd kan worden. De portal interacteert met gp4, het eerste eiwit betrokken bij de connectie van de ring aan de staart van de faag. Ook de binding van de portal aan gp4, en de effecten hiervan op de vorm van het complex zijn onderzocht met massaspectrometrie. De verkregen ladingstoestanden van zowel de vrije portal als deze geassocieerd aan gp4, wezen erop dat de binding van gp4 grote gevolgen had voor de structuur van het complex. Met een nieuwe techniek in het natieve MS veld, ion-mobiliteit MS kon dit worden bevestigd. Met onze resultaten zijn nieuwe inzichten in het functioneren van de P22 faag verkregen.

Verschillende studies hebben laten zien dat het mogelijk is informatie te verkrijgen over eiwit-interactie netwerken aanwezig in de cel. Dit kan onder andere door endogeen tot expressie gebrachte eiwitten met behulp van affiniteittechnieken te isoleren en te analyseren met massaspectrometrie. Deze studies laten bijna altijd zien dat eiwitten niet alleen functioneren maar juist samenwerken door eiwitcomplexen te vormen. Helaas gaat bij dit type onderzoek alle informatie over de samenstelling van deze complexen, de stoichiometrie en de structuur verloren. Met natieve MS kan dit wel worden bekeken. Wij hebben de samenstelling van het RNA polymerase II en III onderzocht. De resultaten, beschreven in hoofdstuk 4, geven inzicht in de opbouw van deze complexen. We hebben de gevonden resultaten van Pol II , waarvan al gedetailleerde structuur informatie bestond, vergeleken met die van Pol III. Hierdoor werd het duidelijk dat alle 17 verschillende eiwitten die aanwezig zijn in het Pol III complex slechts één keer voorkomen. Tevens geven onze MS analyses nieuwe informatie over de hetero-oligomere subcomplexen waaruit het Pol III is opgebouwd. Het verkregen model over de complete samenstelling van het RNA Pol III levert een belangrijke bijdrage voor verder biochemisch en structuur analytisch onderzoek van dit belangrijke enzym.

Natieve MS analyses van de RNA polymerases liet tevens zien dat sommige van de eiwitten in het complex waren gemodificeerd. Bijvoorbeeld, de binding van Zn2+ aan C11 en de fosforilatie van Rpb6. De exacte locatie van deze modificaties kan echter niet worden vastgesteld met natieve MS. Hoofdstuk 5 beschrijft de post-translationele modificaties van zowel RNA polymerase II als III. In vivo zijn deze modificaties nodig voor het goed functioneren van de complexen, deze kunnen namelijk de activiteit, lokalisatie en interacties met andere eiwitten regelen. Een vereiste voor het detecteren van deze modificaties is het terugvinden van een zeer hoog percentage van de totale sequentie van de eiwitten. Elk stukje van het eiwit dat gemist wordt kan namelijk ook gemodificeerd zijn zonder dat dit dan wordt gevonden. Hier hebben we gebruik gemaakt van verschillende manieren om de eiwitten in kleinere stukken te krijgen (peptiden) gecombineerd met diverse fragmentatie methoden in de massaspectrometer. Nagenoeg de hele sequentie van alle eiwitten aanwezig in zowel Pol II als III werd teruggevonden, en verschillende nieuwe fosforilatie plekken werden geïdentificeerd.

ACKNOWLEDGEMENTS

At first I would like to thank my supervisor Albert J.R. (!) Heck, without whom this project would not have been possible. His enthusiasm and interest in the field of native mass spectrometry was very addictive not only to me but also to people outside the field, enabling fruitful collaborations. Even though we have not always been of the same opinion he always gave me the freedom to try out new experiments while at the same time carefully keeping me on the right track. Thanks to our collaborators in Munich and Syracuse for their work, the samples and their input. I also want to acknowledge Pim Pijnappel and all the poeple in the Medical Center at the University Utrecht for their help and support. Thanks to Waters for their scientific and financial support.

I would like to acknowledge Kees, whoms knowledge about mass spectrometry is so vast that he always had a solution to any problem concerning the instrumentation and spectra. Without him I would have been stuck for much longer with some problems. I also want to thank him for the fun it was to take some of the instruments apart for cleaning or some fiddling around for improvements. Another of the good souls in the lab is Arjan, whom in his special way was always able to show me that with some humor the solution to some problems almost comes by itself. I would also like to thank all the other "native MS" people: Charlotte for all her input both scientific as well on the social level and of course her contributions to my balcony and music knowledge, Esther for her patience and support, and of course the people whom already moved on: Silvia, Vera. Thanks to Shabaz, whom was starting as a post doc at the same time as I started my PhD, he is now one of the leading people of the group and as Mathias Mann said one of the most promising young researchers in the mass spectrometry field. He made Chapter 5 possible, thanks for all the work and input. I also want to say thanks to Nadia, Sharon, Tieneke, Thin Thin for the fun and always trying out my cooking. I do not want to forget Marco and Hanka the recently arrived German support. Thanks for the fun suggestions for weekend activities, I hope we will be able to enjoy more of those in the near future. Now I know that there can also be not enough wind in Holland, even though I found it hard to believe. Of course I also want to thank all the other current and past members in the lab for their input suggestions and support.
My gratitude also goes to my former supervisors in Kiel, Prof. Dr. Schulz and Dr. Jens Appel. In the group I learned during my diplom how to work scientifically in a creative atmosphere. I will always value their oppinion how scientific work and the scientific community are supposed to be.

Of course there are always the people outside work, without whom I would have been able to get to the point where I stand now. My honest gratitude goes foremost to my father. Du warst immer für mich da in deiner unvoreingenommenen und nicht beeinflussenden Weise. Ich würde mir wüschen, dass ich (falls ich mal welche haben sollte) die Kraft besitze meine Kinder auch so großzuziehen wie du es mit uns getan hast. Ich sollte an dieser Stelle auch meiner Mutter danken, in ihrer eigenen Weise hat sie mich auch zu dem gemacht was ich heute bin. Vielen Dank auch an meine Geschwister. Ela, du hast mich immer und zu jeder Zeit in deine inzwischen eigene Familie aufgenommen, Stephan, der immer wenn ich wieder mal nach Hause zu Besuch kam Aufopferungsvoll zurückgestanden hat und meine jüngste Schwester Felica die mich immer wieder mit ihrer pragmatisch fröhlichen Art in ihren Bann zieht. Ich möchte hier auch nicht Anne, Leonie und Ismena vergessen die unsere Familie sehr bereichert haben

Thanks also to the dutch people I came to know outside work. Maarten, without you I would probably not have ended up here in Holland. You taught me about self-confidence and personality, maybe not on the easiest way but at the end I think I learned some lessons there. Lianne and Bart, thank you for your support and acceptance. Bartho and Heather, I hope you will have a long happy time together! Lex, never give up your happy and enthusiastic way of life.

Christian, wahrscheinlich wirst du dies nie lesen, aber ich möchte dir danken, dass du einen so langen Weg mit mir gegangen bist. Auch durch dich bin ich erst zu dem Menschen geworden der ich jetzt bin, im guten wie im schlechten. Ich wäre gerne stärker gewesen manchmal.

Andreas, danke dass du mich immer wieder aufgefangen hast und für mich da warst, wo andere schon längst das Handtuch geworfen hätten. Ich möchte dir noch für viel mehr danken, aber das bleibt zwischen uns.

And last but not least: DD was here! Thanks a bunch guys (and girls) for all the distraction and fun I had with you lot in the past years. Thanks Negy for introducing me to them. Thanks Traight for keeping us together through some troublesome times, but also Anti, Obs (thanks for offering help with the introduction), Puddy, Ekos, Cho, Ippy, Bubbs, Ever, Phee, Til, Raggy and Heth whom I came to know IRL and Z, Tolg and all the others… I hope I will see some of you at my promotion.

name	Kristina Lorenzen
place and date of birth	31.01.1978 in Braunschweig, Germany

education

08/1984 - 07/1988	Grund- und Hauptschule Glücksburg
08/1988 - 06/1998	Förde-Gymnasium Flensburg - Abitur

studies

10/1998	Study of biology at the Christian Albrecht Universität Kiel
10/2000	Intermediate diploma biology
10/2002	Associated member in the Graduiertenkolleg 820
07/2003	Diploma (cell biology, botany, toxicology)
09/2003 – 01/2005	Continuing work on diploma topic: Optimizing the bio hydro gen production of cyanobacteria at the Christian Albrecht Universität Kiel
02/2005 12/2008	Ph.D. study: Validation and structure of protein assemblies
12/2008	Post Doctoral postition at the Universtiy Utrecht, Biomolecular Mass Spectrometry and Proteomics Group

van den Heuvel, R. H. H., van Duijn, E., Mazon, H., Synowsky, S. A., **Lorenzen, K**., Versluis, C., Brouns, S. J. J., Langridge, D., van der Oost, J., Hoyes, J., and Heck, A. J. R. (2006) Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. Analytical Chemistry 78, 7473-7483.

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

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

Lorenzen, K., Olia, A. S., Uetrecht, C., Cingolani, G., and 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.

Mohammed, S., **Lorenzen, K**., Kerkhoven, R., van Breukelen, B., Vannini, A., Cramer, P., and Heck, A. J. (2008) Multiplexed proteomics mapping of yeast RNA polymerase II and III allows near-complete sequence coverage and reveals several novel phosphorylation sites. Anal Chem 80, 3584-3592.