Analysis of endogenous protein complexes by mass spectrometry

Silvia Anna Synowsky

Printed by F&N Boekservice, Amsterdam

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 Pharmaceutical Sciences, Utrecht University)

Analysis of endogenous protein complexes by mass spectrometry

Analyse van endogene eiwit complexen door massa spectrometrie

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof. dr. J.C. Stoof,
ingevolge het besluit van het college voor promoties in het openbaar te verdedigen
op woensdag 14 mei 2008 des ochtends te 10.30 uur

door
Silvia Anna Synowsky
geboren op 8 juli 1976, te Kreuzburg (Polen)

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

The research in this thesis was financially supported by the Dutch Foundation of Scientific Research N.W.O

The printing of this thesis was financially supported by the J.E. Jurriaanse Stichting

Content

Chapter 1

General introduction

A.) Protein complexes

A cell is considered as the most basic building block of a living organism. These organisms are categorized as unicellular, when existing as a single cell, such as bacteria and yeast, and multicellular when comprising an assembly of many diversely differentiated cells such as human and most eucaryotes. Each cell is a complex entity filled with various different macromolecules (proteins, nucleic acids, lipids and glycans) that are organized in different sub-cellular compartments. The interactions between these macromolecules are highly directed and allow a multitude of different functions ranging from regulating cell function to transmitting information to the next generation of cells.

All cells contain the hereditary information in form of DNA which is folded into chromosomes localized in the nucleus. To translate the genetic information to proteins, the DNA is first transcribed into messenger RNA (mRNA), a negative template of the DNA strand with a poly-adenine sequence at the 5'-end. The RNA serves as a disposable temporary storage of the genetic information which is translocated from the nucleus into the cytosol and translated into proteins. Proteins are the most versatile macromolecule in the cell and are nearly involved in every cellular process. They have catalytic, regulatory, structural and mechanical properties. While for example enzymes catalyze many biochemical reactions and are crucial for the cellular metabolism, actin and myosin function as scaffolding proteins to maintain mechanical stability of the cell. Diversity of proteins is achieved prior translation by alternative splicing of genetic transcripts. But more importantly proteins can adapt rapidly to cellular requirements by post-translational modification, conformational changes and modulation of direct protein-protein interactions or a combination thereof.

Hence current functional proteomics studies emphasize the investigation of protein-protein, ligand-protein and substrate-protein interaction and their conformational changes upon binding as well as mapping of post-translational modifications. Many cellular functions are regulated by post-translational modifications such as acetylation [1], methylation [2], sumoylation [3, 4] and phosphorylation [5, 6]. The reversible phosphorylation of a protein is believed to be one of the most important posttranslational modifications in the cell [7-9]. It is believed that about 30% of the mammalian proteome is phosphorylated at a given time. The reversible phosphorylation attaches a phosphate group to one of the side chains in the protein amino acid sequence, in particular to threonine, serine and tyrosine residues. This reaction is catalyzed by protein kinases. Upon phosphorylation of a protein in the cell at neutral pH negative charges are introduced which can disrupt electrostatic interactions in the unmodified protein. The protein can form new ionic bonds thereby changing its conformation. Changes in protein phosphorylation can be involved in ligand binding or activity. Therefore revealing the phosphorylation pattern of protein complexes is useful in understanding activity.

The last decade has seen a growing interest in the characterization of protein complexes as it became clear that many proteins do not act alone to execute cellular functions. Proteins interact with each other to form distinct stable and transient assemblies. Such non-covalent interactions are mediated by van der Waals, electrostatic and hydrophobic forces and hydrogen bonding. The connectivity between several different proteins may give the protein complex a new functional role. Protein assemblies are not static but instead they undergo dynamic changes depending on cellular localization and time. They constantly form and dissociate. A full understanding of any biological system requires the insight into which macromolecules interact with each other.

The investigation of protein complexes and protein-protein interaction networks is therefore of high importance to understand pathways and biological functions in the cell. The interconnectivity of proteins has been explored in several organisms such as *Saccharomyces cerevisiae* [10-15], *Helicobacter pylori* [16], *Drosophila melanogaster* [17], *Caenorhabditis elegans* [18] and *Homo sapiens* [19-21] This research has yielded into complex protein interaction maps (fig. 1).

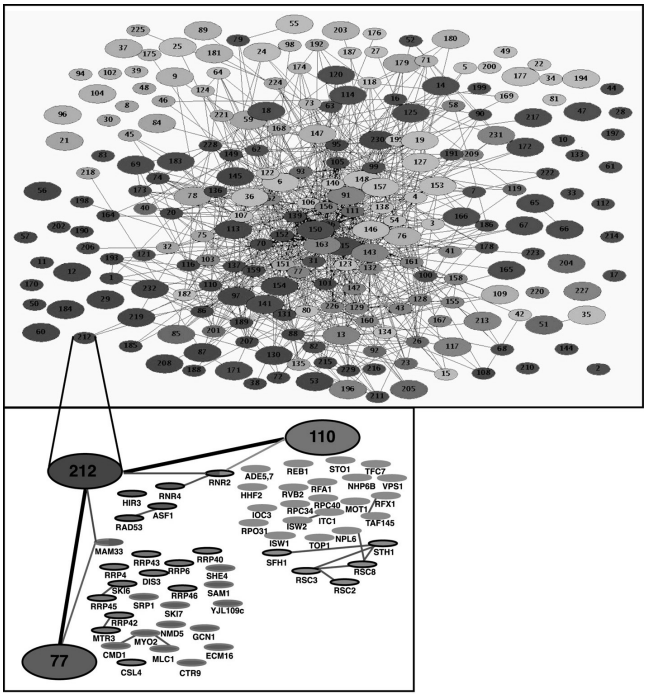


Figure 1

Network of protein interactions in yeast. Each protein complex is represented in modules with high interconnectivity between different protein assemblies (adapted with permission from [10])

Systematic identification of protein-protein interactions was initially introduced by the yeast two hybrid system [22] which was further developed to be applicable to mammalian systems [23, 24]. It utilizes a transcription factor consisting of two modules, a DNA binding domain and a transactivator domain which are only active if they are close in space. If these domains are fused to different proteins their activity depends on protein-protein interaction to reconstitute proximity of the domains and therefore activity. This assay is easy, robust and applicable to high-throughput analyses. Although this system allows insight into "*in vivo*" protein interactions, it also leads to a high number of false positives that hampers the reliability of the method and requires extensive validation. Moreover as only binary protein interactions are probed, this method may not reflect higher order interactions for which multiple proteins are required. A complementary approach is the combination of affinity pulldowns with mass spectrometry. In this approach recombinant proteins are isolated directly from cells and analyzed by proteomics or macromolecular mass spectrometric techniques.

B.) Analysis of protein complexes

The cell is filled with macromolecules and proteins are the most versatile macromolecules of all. They can fulfill enzymatic, regulatory and structural functions. Due to this variety of functions they have very different physiochemical properties and isolation proved to be difficult and tedious. A major breakthrough in the purification of proteins was achieved with the development of affinity tags that can be fused to the target protein. These tags allow the isolation of proteins under native conditions in a fast and efficient way and are generically applicable. A selection of small affinity tags include the Arginine-tag (five to six arginine residues), calmodulin binding peptide (CBP), cellulose binding domain, c-myc epitope tag (EQKLISEEDL), gluthathione S-transferase, FLAG-tag (DYKDDDDK) and Histidine-tag (five to six histidine residues) [25]. These tags are very useful in the purification of recombinant proteins in over-expressed systems. They are commonly used in a single step affinity purification. If the protein is not pure yet classical biochemical techniques like gelfiltration which separates according to hydrodynamic volume or anion/cation exchange for purification depending on charge can be applied. The introduction of genes coding for the target proteins into an over-expression system such as *E.coli* or mammalian expression systems produces large amounts of protein, but also changes the natural environment of the protein and therefore does not always retain essential interactions between proteins in their natural habitat.

The analysis of endogenous protein complexes is interesting as it keeps the proteins in their original cell with close proximity to its interaction partners. It does not inhibit the tendency of proteins to form large molecular machineries and also allows post-translational

modifications. It therefore represents the most authentic situation in the cell. Originally the isolation of endogenously expressed protein complexes was cumbersome. Numerous biochemical procedures had to be subsequently used, and in case of low abundant protein complexes a high amount of starting material was required for a reasonable final yield. Another major disadvantage of this lengthy isolation procedure is the risk of losing transient interaction partners. Since each protein complex shows different characteristics the purification procedure had to be evaluated and optimized for each protein assembly. The introduction of affinity tags revolutionized the isolation of protein complexes [26, 27]. The use of one step affinity purifications in the isolation of endogenous protein complexes leads to several problems in respect to yield and purity. One drawback may be a low affinity towards the tagged protein, which gives major problems in the yield of low to medium abundant endogenous protein complexes. A second major disadvantage of one step affinity purifications is that high abundant proteins in the cell are concomitantly isolated and adulterate the results in regard to protein complex composition. As it is therefore very important to distinguish between specific protein interactions and non-specific contamination the use of a two step affinity approach greatly enhances specificity. It clearly produces purer samples with reduced background proteins. The most popular affinity tag for the isolation of protein complexes is the tandem-affinity-purification (TAP) tag (fig. 2). In the originally developed TAP method the protein of interest is expressed in frame with an epitope tag. The TAP-tag contains a Protein A moiety from *Staphylococcus aureus* and a Calmodulin Binding Peptide (CBP) which are separated by a TEV protease cleavage site. The tag can be fused either to the N- or C-terminus.

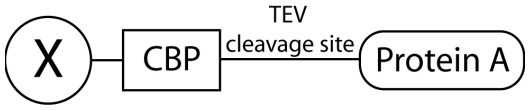


Figure 2

Structure of the TAP tag. The TAP-tag consists of a Protein A moiety followed by a tobacco etch virus (TEV) protease cleavage site and a calmodulin binding peptide (CBP). The circle containing X denotes the tagged protein.

A disadvantage of the TAP technique is the introduction of a large tag to retrieve a protein complex. Due to the tag size it is possible that the protein complex cannot assemble with all its interaction partners and/or that expression levels are lowered. The protein complex isolation could be also affected if the tag is not easily accessible for the purification, hence adulterate the result.

To attach a tag to a protein, standard molecular biology tools can be applied. The first step to a recombinant protein is the amplification of the TAP-tag from a plasmid by PCR. The forward and reverse primers are carefully chosen and contain partially the sequence of the tag coded on the plasmid for successful annealing during the PCR procedure and also a

part of the genetic sequence of the target protein so that proper insertion into the genome can be ensured. After amplification the PCR product can be inserted into host organism. In yeast homologous *in vivo* recombination is highly efficient and the original gene can therefore be replaced by the recombinant gene product. If fused C-terminally the implemented gene is expressed under their own promotor. The constructs described in this thesis were unexceptionally C-terminally tagged. The TAP technique has been also successfully applied to higher eukaryotic protein complexes from human [28].

As the protein assembles with its partners in the cell, this double tag is used to pull out the protein (bait) with its interaction proteins (prey) from the cell. The procedure is schematically illustrated in (fig. 3).

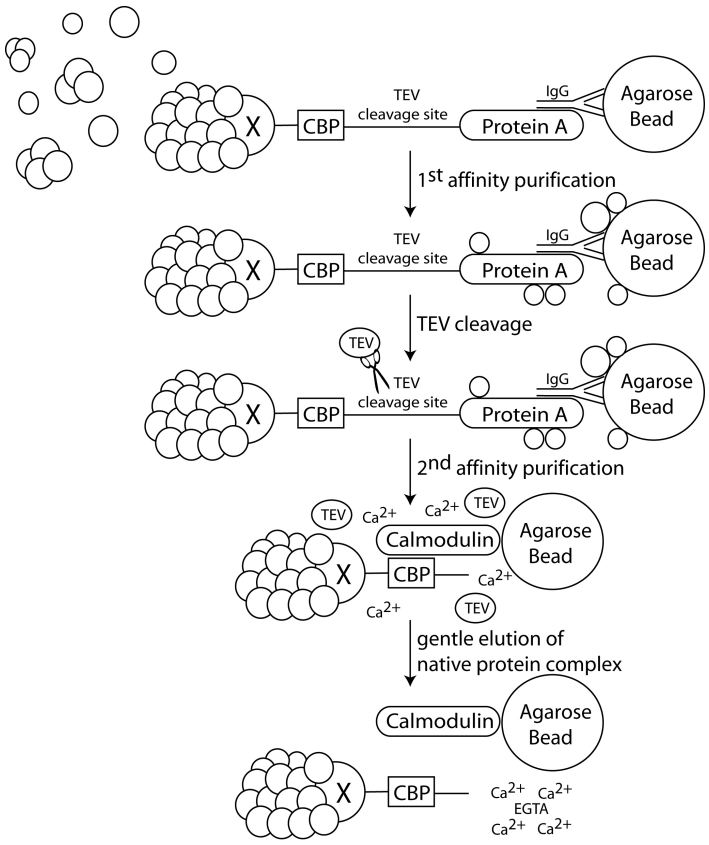


Figure 3

Schematic overview of the isolation of a TAP-tagged protein complex.

A protein complex containing a TAP tagged subunit is retrieved by a subsequent two step affinity purification. The purification is performed via two independent affinity pulldowns on IgG- and calmodulin containing resins. The

immobilized protein complex is eluted from the first step by TEV cleavage and in the second step by chelating calcium ions to weaken the affinity of the protein complex towards calmodulin.

In the first affinity purification the interaction between Protein A and IgG affinity resins is used. As this interaction is very strong the non-bound and aspecifically bound proteins of the total lysate can be removed by washing with a buffer containing a high amount of salt. The salt content should be carefully chosen as to balance between removal of non-specific proteins binding to the IgG beads and to not disrupt more transient interactions within the protein complex. The protein complex is eluted from the IgG-column by proteolytic cleavage using tobacco etch virus (TEV) protease. The TEV protease specifically recognizes a linear amino acid sequence of ENLYFQG, located between Protein A and CBP, and cleaves between Q and G. The eluted protein complex is then applied onto a second affinity column. This column contains calmodulin immobilized agarose beads to which the second affinity tag CBP binds in the presence of calcium ions which are essential for the binding between calmodulin and CBP. The TEV protease and remaining contamination can be removed in washing steps. Finally the protein complex is eluted by chelating the calcium ions with ethylene glycol tetraacetic acid (EGTA). Buffers chosen for both purification steps are compatible with keeping the protein complexes intact as they resemble physiological conditions and give reasonable high yields. The procedure is generic, easy and rapid. With some variations in the standard protocol, for example the substitution and / or the addition of different detergents not only soluble protein complexes but also membrane protein complexes can be purified. The analysis of purified protein complexes is accomplished with mass spectrometry. The proteins are separated and visualized on 1-D-SDS PAGE, proteolytically digested to peptides and analyzed by mass spectrometric techniques (LC-MS/MS or MALDI-MS).

To validate protein-protein interactions reciprocal purification of a complex with any other subunit protein used as bait can be performed. Due to the generic protocol, the TAP procedure allows easy variations in the workflow. Many proteins play an important role in more than one protein complex. If the shared component between two protein complexes is used as bait then it is likely that the second protein complex is co-purified. An alternative TAP approach allows the differentiation between two complexes by substraction of one of these protein assemblies [29]. To selectively enrich for protein complex A with the TAP strategy, a different subunit of complex B is only fused to the Protein A moiety without a protease cleavage site. In the course of purification protein complex B is removed in the first affinity step as it cannot be eluted from the affinity resins.

To reduce high abundant contaminating proteins in the TAP procedure a triple affinity tag may further improve the purity of the pulldown [30]. The third tag should ideally be compatible with the buffer conditions of the two other tags. One disadvantage is a more

lengthy procedure. Whether it will enhance purity at no cost of losing more transient interaction partners still has to be demonstrated.
In addition the introduction of a large tag increases the likeliness that it will influence the properties of the target protein. It may induce changes in complex stability or even at the complex composition. It is therefore important to carefully evaluate the data derived from the TAP technique with known protein-protein interactions in databases. Such a comparison with the MIPS (munich information center for protein sequences) database indicate that the TAP method has a low false negative rate (15%) [31, 32]. Interaction data can be also compared to mRNA co-expression profiles originating from microarray data [33]. In principle interacting proteins of a protein complex are expected to show similar regulation based on mRNA levels. The correlation of TAP datasets to mRNA expression profiles suggests a low false positive rate. The internal reproducibility of high throughput TAP data by reciprocal tagging is 70%, thus only 30% of the interactions should be treated with caution [34]. Alternatively other tandem affinity tags are currently developed by combining a tandem Strep-tag II with a FLAG-tag. These new tags are reduced in size and also allow an even faster purification procedure [35].
The TAP-MS technique allows the systematic analysis of protein-protein- interactions in a cell at physiological relevant levels. The power of a two-step affinity purification procedure under gentle conditions is that it keeps the protein complex intact during the procedure. Even posttranslational modifications are not affected and can therefore be analyzed. The resulting interaction maps provide a good starting point on how proteins are organized within the cell. But the knowledge about the composition of protein complexes in the cell is only the first step in the comprehensive understanding of their function. It does not give information about the internal subunit architecture or the assembling process of the complexes.

C.) Techniques in structural biology

As mentioned above most cellular functions require the coordinated assembly of several proteins into higher-ordered structures for activity. The investigations of 3-dimensional structures and protein-protein-interactions are of central importance for the understanding of biological processes. Established methods in structural biology such as x-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy and electron microscopy (EM) have provided a wealth of structural data on protein complexes. Frequently the biological insight gained from these techniques has been fundamental in the understanding of functionality.
Crystal structures show atomic details about interactions and the architecture of proteins. In principle this technique is not limited to the size of the macromolecule. Nevertheless x-ray crystallography also faces limitations. It is not easy to obtain well-structured crystals with

good diffraction pattern for high-resolution reconstruction. The larger the macromolecule the more difficult it is to get well ordered crystals. Moreover the method suffers from the need of very large quantities of highly pure homogeneous material.
NMR is another prominent tool capable of portraying the three dimensional structure of proteins at atomic resolution. NMR is an appropriate technique to study weak and strong interactions of protein complexes in solution. It also allows studies of conformational changes in time courses. However most NMR structures can be solved of proteins with a molecular weight distribution in the range of 20 kDa and limited to 50 kDa. The NMR spectra for larger biomolecules suffer from a resonance overlap which makes data interpretation difficult. Moreover the technique itself depends on the excitation of molecules. Due to faster relaxation times of large molecules, peaks become broader and less intense and eventually lead to no signal at all.
Electronmicroscopy and image processing is another attractive technique to investigate protein complexes. The most common technique for biomolecules is cryofixation which rapidly freezes the macromolecule and preserves a snapshot of the “in solution” state. It uses highly energetic electrons which are focused onto the biological specimen to create electron-scattering density maps. As the molecules are randomly oriented several two dimensional pictures can be reconstructed into a three dimensional molecule. This technique illuminates the shape of the protein complex. It needs only minute amounts of a protein complex. Unfortunately this method gives mainly rough structures with a low resolution (more than 10 Å) so that modeling of subunits into the overall picture can be questionable.
Mass spectrometry offers multiple techniques in the field of protein complex analyses. As the field is broad, it will be explained in more detail with an emphasis on macromolecular mass spectrometry.

Mass Spectrometry

Mass spectrometry is an indispensable analytical tool to analyze the weight of molecules. It had its starting point in the elucidation of the mass of elements and of molecules in chemistry. Nowadays it is used in various areas to identify compounds ranging from small molecules such as drugs to large molecules such as polymers, peptides, proteins and native protein complexes.
A mass spectrometer is divided into three essential parts: I.) ion source, II.) mass analyzer and III.) ion detector. Ions are produced from neutral molecules that need to be brought into the gas phase. These are generated in the ion source and then accelerated into the mass analyzer by a potential voltage. In the mass analyzer the ions are separated according to their mass over charge (m/z) values and finally detected on the detector. The detector converts the signal to a graphic output with intensity plotted against the m/z value.

Ionization techniques

Early ionization techniques such as electron ionization (EI), chemical ionization (CI), field ionization (FI) and field desorption (FD) were the only tools for analyzing molecules until the 1980's. These ionization techniques were restricted to small volatile compounds and due to high energy levels large molecules were easily fragmented in this process. In the early 1980's fast atom bombardment (FAB) mass spectrometry and plasma desorption (PD) was introduced which permitted the ionization of oligonucleotides, peptides and small proteins. FAB involves the bombardment of a solid analyte in a liquid matrix (a small organic compound for example glycerol or m-nitrobenzyl alcohol) mixture by a fast continuous particle beam of neutral Ar or Xe atoms. The disadvantages were still the upper mass limit, the low ionization efficiency and the high background peaks arising from the matrix [36].

A real technical breakthrough for the application of mass spectrometry in protein chemistry was the development of electrospray ionization (ESI) and matrix-assisted laser desorption / ionization (MALDI) in the late 1980's. These "soft" ionization methods allow the formation of charged ions (such as $[M + nH]^{n+}$ in the positive ion mode and $[M - nH]^{n-}$ in the negative ion mode) and the transfer of biomolecules into the gasphase. This technique does not deposit sufficient energy in the ions to induce fragmentation. More importantly, molecules with a mass above 100 kDa could be analyzed. The development of soft ionization methods was particularly useful to ionize previously involatile macromolecules such as peptides and proteins. ESI ionizes the analyte at atmospheric pressure and then transfers the ions through focusing lenses to higher vacuum stages into the mass spectrometer. A high vacuum inside the mass spectrometer is essential to avoid ion neutralizing collisions. In the process of ionization the analyte solution is forced through a conductive nozzle or capillary with an inner diameter of 100 µm. The applied electric field (1-4 kV) causes the exposed liquid to form an extended structure, also called the Taylor cone that contains a plume of droplets (innerdiameter: 1µm) (fig. 4A). The ionization process was originally described in studies of spraying dilute polymer solutions into an evaporation chamber [37]. The ion formation led to two controversially discussed mechanisms in electrospray ionization: the ion evaporation model (IEM) [38] and the charge residue model (CRM) [37]. The ion evaporation model describes the formation of ions by ejection of one ion from the solution due to a potential difference caused by the high electric field. The IEM model is thought to best describe the formation of small molecules and peptidic ions in the gas phase [39] (fig. 4B). The charge residue model is generally accepted to explain the formation of large ions such as folded globular proteins and protein complexes [40]. The applied potential adds charges to the droplet that are concentrated on the surface. During the evaporation process these charges come closer together until the coulomb repulsion will overcome the surface tension and the droplet will disintegrate into several highly charged smaller droplets. If the analyte is diluted enough this process will lead to a single molecule in a droplet and subsequently an ion in the absence of water (fig. 4C). These ions are typically positively

charged by the attachment of protons, alkali cations or ammonium ions and negatively charged by proton subtraction if the electrospray polarity is reversed.

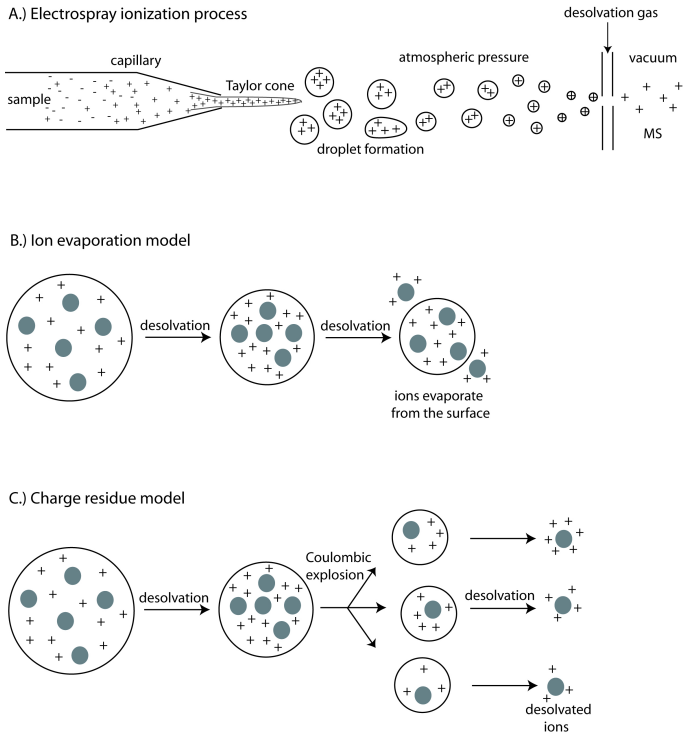


Figure 4

A.) Illustration 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.

B.) Ion evaporation model (IEM)

Small molecular ions (grey circles) 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.

C.) 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 charge residue model is generally accepted to explain the formation of large ions such as folded globular proteins and protein complexes

The charge on the analyte depends on the ion's size. Large molecules carry therefore more charges than small molecules. On average an unfolded protein will carry one charge per kDa depending on the number of basic residues (arginine, lysine and histidin) in the amino acid sequence. The advancement of using ESI for the analyses of biomolecules and the

coupling of liquid chromatography with mass spectrometry via electrospray ionization as an interface took another 15 years [41, 42].
The further development of ESI to a low-flow variant greatly enhanced sensitivity. Due to the decrease in flow rates to approximately 100 nl / min, the ions experience a better desolvation [43-46]. In nanoflow-ESI the sample is sprayed from a very thin capillary with a tip of an innerdiameter of 1-3 µm. This reduces the droplet size 100-1000 fold to a droplet diameter of less than 200 nm with a higher charge to surface ratio [41, 47, 48]. The nanospray ESI source disperses the liquid purely by electrostatic repulsions and the use of nebulizing gas and additional heating becomes obsolete. Furthermore the method benefits from a low sample consumption (picomole to femtomole).
MALDI uses a different ionization principle which is also gentle enough to keep biomolecules intact. The analytes are co-crystallized in a saturated solution of matrix on a metal target plate. Common matrices are α-cyano-4-hydroxycinnamic acid for peptides and small proteins, sinapinic acid for proteins and other large molecules and 2,5 dihydroxy benzoic acid for proteins and glycopeptides. The target plate is inserted into the mass spectrometer and a high vacuum is applied. Upon irradiation with a laser pulse (typically nitrogen laser with 337 nm) the analytes vaporize as ions into the gas phase. The matrix absorbs most of the energy thereby leading to barely any fragmentation of the analyte. The technique is relatively robust towards salts and impurities. This tolerance is explained by the crystallization effect of the analytes with the matrix as it omits impurities from the crystal. Another major advantage is the low consumption of sample (low femtomole to attomoles). However, in the analysis of protein complexes MALDI-MS is not the method of choice. Since it requires the co-crystallization with an acidic matrix, it does not keep the protein complex under physiological conditions. Moreover since the MALDI technique typically produces mainly singly charged ions, it does not give multiple datapoints to calculate the mass of a protein.
John Fenn and Koichi Tanaka were awarded the Nobel Price in 2002 for their contributions in the development of ESI and MALDI.

Mass analyzers

Mass analyzers separate and measure gas phase ions according to their m/z ratio. This is achieved by the application of electric and/or magnetic fields. Previously magnetic sector analyzers were used, but with the development of new ion sources also the mass analyzers had to be adapted to meet new challenges such as the analysis of biomolecules. Mass analyzers have to meet criteria such as accuracy, resolution, mass range, tandem analysis abilities and speed. The most widely used mass analyzers are the time of flight (ToF) and the quadrupole (Q).
The simplest kind of mass analyzer is the linear time of flight (ToF) analyzer. It separates ions based on the time they require to reach the detector. The flight tube is essentially a field

free drift region of known length under a high vacuum. Ions generated in the ion source are accelerated with the same amount of kinetic energy (E_k) into the flight tube. The energy is proportional to the charge (z) and the accelerating potential (V) (equation 1). Ions move at the speed determined by their mass (equation 2). Since all ions have the same energy but can have a different mass they will take different times (t) to hit the detector. The arrival time (t) and tube length (d) can be used to derive the velocity of the ions. Low molecular weight ions with a high velocity will reach the detector earlier than high molecular weight ions with a lower velocity.

The ion motions are controlled by:
equation 1: $E_k = z * V$
equation 2: $E_k = ½ m * v^2$ and $v = d / t$
$=> z * V = ½ m * [d/t]^2$, hence: $m/z = 2 * V * [t/d]^2$
z is the charge, V is the accelerating potential, v is the velocity of the ion, t is the time for the ion to cover the distance, d is the length of the drift tube, m is the mass of the ion. A limitation of a linear ToF analyzer is the poor resolving power, which is mainly due to slight differences in the initial kinetic energy at the starting point.

Nowadays continuous ion sources such as ESI are generally coupled to orthogonal accelerated time of flight (oaToF) analyzers. In the oa-ToF instruments the flight direction of a continuous ion beam from the source is diverted by 90º into the oa-ToF by a pulse from a pusher [49]. This development limits the spread of initial kinetic energy which enhances sensitivity and resolution. Figure 5 schematically represents an ESI-ToF instrument.
The implementation of an electrostatic ion mirror (reflectron) helps in the reduction of the energy spread and improves resolution. The reflectron elongates the drift time of the ions and consequently results in a better separation. In addition minor differences in the initial kinetic energy distribution are corrected as more energetic ions move deeper in the reflectron and are reflected slightly later to the detector. With optimal instrument settings, ions of the same m/z but with slight differences in the initial kinetic energy will still arrive at the detector at the same time. ToF analyzers have virtually an unlimited m/z range and are therefore the method of choice for the analysis of protein complexes.

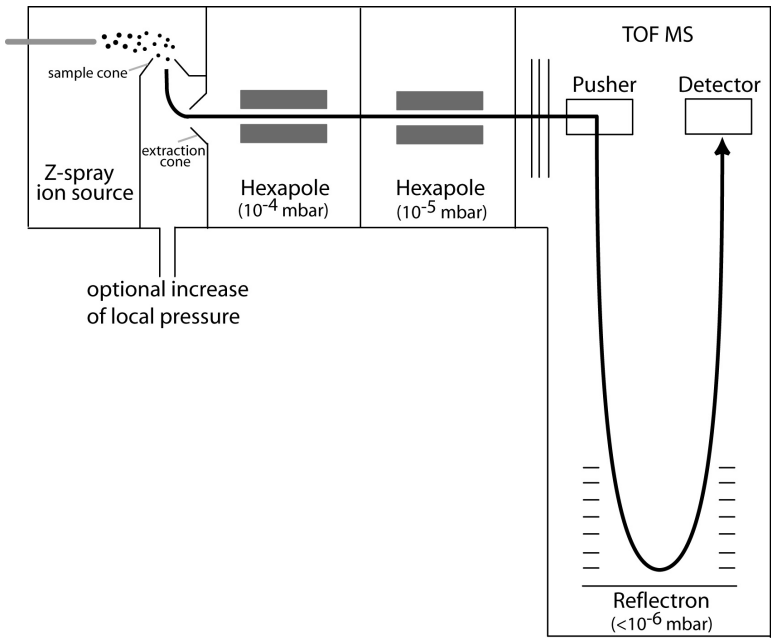


Figure 5

Schematic representation of an ESI-ToF mass spectrometer equipped with a Z-spray source (LCT, Waters). The pressure region between sample cone and extraction cone can be locally increased to facilitate the transfer of non-covalent interaction into the gas-phase.

A quadrupole mass analyzer (Q) separates ions according to selective stable trajectories using a combination of constant (DC) and alternating (AC) voltage. A quadrupole mass analyzer consists of four parallel rods positioned in a radial array with opposing rods carrying the same charge (fig. 6).

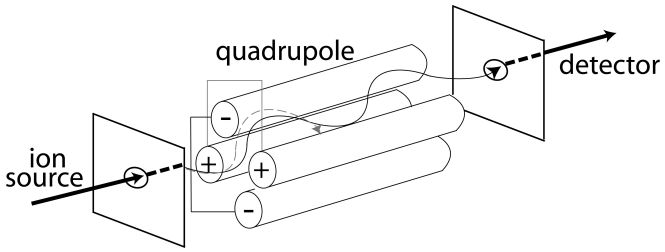


Figure 6

Schematic representation of a quadrupole mass analyzer. Four parallel rods in a radial orientation with opposite rods carrying the same charge. Oscillating an electric field drives the ions in a spiral motion through the quadrupole. Only ions with a stable trajectory will reach the detector, others collide against the rods.

By oscillating this electric field the ion travels through the quadrupole in a spiral motion. The ion is always attracted to the rod with the counter charge. If the potential of the rods changes before the ion is discharged on the rod, it follows a stable trajectory through the mass analyzer. Each specific ion travels through the rod at certain alternating frequency and/or DC potentials. By scanning the voltages, keeping the ratio between the DC and RF constant all ions can be transmitted in a single scan. Quadrupole analyzers are ideal to be coupled to electrospray ion sources since they are robust and do not require high acceleration fields. Conventional quadrupole analyzers are limited in their m/z range which allows them to scan up to 4000 m/z. Due to their architectural design they have a rather low resolution. Another example for scanning mass analyzers is the quadrupole ion trap which captures ions between a ring-electrode with two opposite hyperbolic metal electrodes.

Hybridizing two different mass analyzers with each other creates a tandem mass spectrometer that can be operated in the MS or MS/MS mode. In the MS mode, both mass spectrometers are used equally to transmit the ion produced in the source to the detector. The MS/MS mode allows the selective transmission of a precursor ion through the first mass spectrometer into a collision cell where the ion is fragmented by collisions with gas molecules. The resulting fragments are then analyzed in the second mass spectrometer. It enables the determination of a sequence of peptides and analyses of posttranslational modifications.

Common hybrid mass spectrometers are Q-ToF and ToF-ToF instruments. A new hybrid mass spectrometer has recently been developed that combines ion mobility separation with a Q-ToF mass analyzer. This instrument enables one additional dimension of separation based on the shape of the molecule in the ion mobility cell with the analysis of the mass over charge ratio in the MS or MS/MS mode.

Detectors

After traversing the mass analyzer ions hit the detector, which is the final element of a mass spectrometer. This impact emits secondary electrons or photons from the detector surface that are sensed by light or charge detectors. An electron multiplier is composed of a stack of dynodes from electron emitting material. Upon the initial ejection of secondary electrons an electric potential is applied so that these secondary electrons are accelerated to strike the second dynode emitting more electrons into a continuous cascade until the last dynode is reached (typical gain is 10^6). A photomultiplier functions similar like an electron multiplier. In the initial collision of the ion electrons are freed that hit a phosphorus screen to release a photon from a photocathode that subsequently hits a series of dynodes to liberate more and more electrons towards an anode. The photomultiplier amplifies the initial signal by a factor of 10^8. Multi channel plates (MCP) are the most sensitive detectors. They are designed as a large planar glass plate with thousands of small channels. These channels are not arranged perpendicular to the plate but in a bias angle of several degrees. Ions that enter the channel

release electrons from the coating material. Similarly to multipliers this event releases a cascade of repeatedly liberating more electrons. Due to the geometry of this detector, many ions can be detected at the same time. The sensitivity decreases with increasing mass of the ion as the large ions travel with a lower velocity through the mass analyzer and therefore bounce against the detector with a lower impact thereby releasing less secondary ions. The electrons produce a small current in all detectors. From this current the ion events can be registered and converted into a mass spectrum.
Most common detectors in Q-ToF and ToF instruments are multipliers such as electron multipliers, photo multipliers and multi-channel-plates (MCP).

The role of mass spectrometry in the analysis of protein complexes

Mass Spectrometry is a key player in the analysis of protein complexes. It is used in proteomics and in the study of noncovalent protein interactions.
The constituents of the protein assembly are identified with a combination of protein complex affinity pulldown and proteomics. After the isolation of the protein complex the constituents are separated on a denaturing gel. The one dimensional electrophoretic separation of the proteins is sufficient as the protein complex purification greatly reduced the complexity of the sample compared to a total lysate. The proteins are subjected to proteolytic digestion and the resulting peptides are analyzed with a mass spectrometer in the MS and MS/MS mode. Primary sequence information is deduced from low energy CID (collision induced dissociation) experiments that fragment the protonated peptides. CID splits the peptide into two fragments. One fragment carries the proton charge and the other fragment is neutral. Only the charged fragment is detected. The charge can reside on either of the fragments. A peptide contains three different bonds in its backbone: NH-CH, CH-CO and CO-NH. According to the nomenclature for peptide fragmentation each breakage at any of the bonds gives rise to a different ion depending on which fragment retains the charge. Peptide fragment ions are termed a-, b-, c-ions if the charge is retained on the N-terminus and x-, y- and z-ions if the charge is kept on the C-terminus (fig. 7).
The peptide can be manually sequenced by systematically calculating the mass difference between two adjacent ions of one series that correspond to one amino acid at the position of cleavage. Nowadays computer programs (such as Mascot) facilitate this task. In CID experiments the most common cleavage occurs at the weakest bond between CO-NH, the amide bond, and leads to mainly to b- and y-ions. As leucine and isoleucine have identical masses as well as lysine and glutamine, CID cannot easily distinguish between these amino acids.

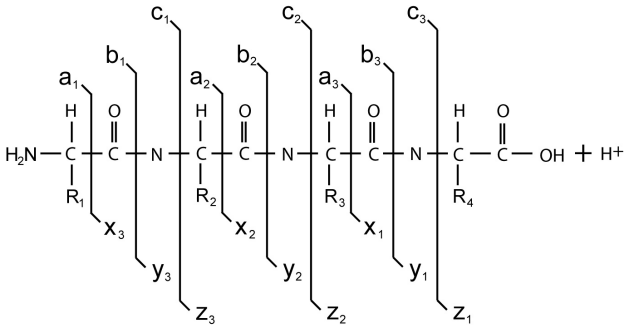


Figure 7

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.

The identification of the constituents of the affinity pulldown experiment enables the categorization of proteins into distinct protein complexes. Additionally mass spectrometry is an indispensable tool for a detailed analysis of protein complexes including post-translational modifications such as protein phosphorylation.

The detection of protein phosphorylation is challenging as they are often substoichiometric and therefore low abundant. The sensitivity and excellent performance of mass spectrometers are very useful in mapping phosphorylation sites in proteins. The most common amino acid residues prone to phosphorylation are serine and threonine and to a lesser extend tyrosine. The mass of phosphorylated amino acids increases by 80 massunits. In CID experiments phosphorylated serines and threonines easily lose their phosphate moiety by gas-phase β-elimination as a neutral molecule H_3PO_4. However the former site of the phosphorylation can be still determined by sequencing the peptide. Previously phosphorylated serine and threonine are now identified as dehydroalanine or dehydroamino-2-butyric acid, respectively with a net loss of 18 mass units compared to a non-modified serine and threonine.

Affinity pulldown mass spectrometry can also be combined with quantitative proteomics to detect relative changes in the protein complex composition [50-52] or to evaluate the specificity of the pulldown [53-55]. In the analysis of protein complexes it is also a useful tool to unravel dynamics of protein assemblies. Several studies have already reported the changes in protein phosphorylation during complex formation [56-59]. The label strategies can be categorized into different classes depending at the time of the incorporation of the quantitative tag to the proteins and peptides. The protein abundance ratio is directly reflected by the ratio of the integrated signal peak of the light and heavy isotopic form of the same peptide. In metabolic labelling the stable isotope nuclei are incorporated into proteins and peptides during cell cultivation in media containing heavy ^{13}C, ^{15}N and/or ^{2}H (stable isotope labeling with amino acids in cell culture, SILAC). The cells are then mixed with cells

that grew on normal medium and do not contain heavy isotopes. The proteins are proteolytically digested and MS analyzed [60-64]. The second approach integrates the label directly to two different pools of proteins prior to mixing and digestion for example by the isotope-coded affinity tag (ICAT) [65]. This tag consists of three parts: i.) a biotin moiety, ii.) an isotopically labeled linker which contains heavy or light isotopes to introduce a mass shift and iii.) a reactive group that specifically alkylates thiol groups in cysteine residues. This labeling technique selectively marks cysteine residues in proteins with the heavy or light version of this tag. These proteins are mixed and digested to peptides. The labeled peptides are selectively enriched by avidin chromatography via the biotin tag and then MS analyzed. In the third category peptides are targeted for labeling before they are mixed. The label can be introduced during digestion with ^{18}O labeled water on the C-terminus [66-69]. Another option is labeling after digestion via chemical reagents such as iTRAQ. These isobaric mass labels are covalently attached to the N-termini and lysines of peptides and produce a strong diagnostic MS/MS signature ions between m/z 114-117 that can be used for relative quantitation [70, 71]. Primary amines can be also labeled via acetylation with acetic anhydride [72] or via reductive amination with light or heavy formaldehyde [73]. Recently also labelfree quantitation methods have been developed [74, 75].

The analysis of protein complexes goes a step further with the investigation of noncovalent interactions of proteins in the assembly. The combination of affinity pulldown mass spectrometry with other mass spectrometric techniques, such as intact mass measurement and native (sub)complex mass spectrometry can determine the supramolecular organization of protein assemblies. This technological progress is owed to the gentle nature of ESI that preserves weaker non-covalent bonds when a large macromolecule is transferred directly from the solution into the gas phase. A major breakthrough was implemented with the development of nano-ESI that facilitates the ionization efficiency due to a decreased flow. Nano-ESI made it possible to analyse proteins from aqueous solutions of appropriate volatile buffers such as ammonium acetate and ammonium bicarbonate at neutral pH, hence under more physiological conditions. Under such conditions the protein complex is kept intact. This revolution in mass spectrometry allowed the investigation of higher ordered structures such as tertiary and even quartenary structure of proteins and protein assemblies. The first example of mass spectrometric detection of a non-covalent protein in the gas phase dates back to 1991. The FKBP receptor with a mass of approximately 12 kDa, complexed with a small organic molecule, was analysed under native condition with mass spectrometry [76]. In the same year the noncovalent interaction between myoglobin and its hemegroup under the influence of the pH was probed [77]. Up to now the detection limit of noncovalent complexes analyzed by mass spectrometry has increased tremendously up to several MDa [78-80]. The successful ionization of macromolecules allowed native mass spectrometry to have a distinct niche in the field of structural biology. The great advantage of this technique is that it requires only minute amounts of analyte (low picomole range).

Heterogeneity in the protein complex composition can be tolerated and symmetry in the complex topology, often an advantage in NMR and x-ray crystallography, is not important. The macromolecules acquire high kinetic energy as they are accelerated in the electric field of the ion source. To sustain the native fold of larger macromolecules (> 100 kDa) it is necessary to withdraw internal energy from the macromolecule before it enters the high vacuum stages in the mass spectrometer. This can be achieved by increasing the pressure at the interface between the ionization at atmospheric pressure and the entrance to the mass spectrometer by reducing the pumping speed in this area. As an alternative an additional gasflow can be introduced [81]. An increased pressure or the additional gasflow facilitates gentle collisions with gas molecules and subducts internal translational energy from the macromolecule and as a result the ions are cooled down [82, 83]. The pressure raise has to be carefully controlled since optimal conditions are m/z dependent and are expressed in relevant changes in the signal intensity in the resulting mass spectra.
Applications of this technique are found in studies of homogeneous [84, 85] and heterogeneous protein-protein [86, 87], protein-DNA [88] and protein-ligand interactions [89-91]. Furthermore conformational changes in the protein fold, measurement of relative dissociation constants and the assembly of proteins could be monitored [92-98].
ESI-MS produces a series of multiply and differentially charged ions which gives rise to a distribution known as charge envelope. Therefore a typical mass spectrum of a protein shows a gaussian shaped ion series reflecting the same protein with different numbers of charges. Generating ions with a high number of charges allows the detection of high masses at relatively low m/z values. The appearance of several peaks corresponding to the same protein helps in defining the molecular weight. The molecular mass can be easily calculated from the m/z values of two adjacent peaks that differ by only one charge that originated from the attachment of the same cation (ideally by a proton). The following equations can be used:

$p_1 * z_1 = M_{protein} + M_{proton} * z_1 = M_{protein} + 1.0079 * z_1$
$p_2 * (z_1 - 1) = M_{protein} + 1.0079 * (z_1 - 1)$
$z_1 = (p_2 – 1.0079) / (p_2 – p_1)$
with p_1 = m/z value of peak 1; p_2 = m/z value peak 2; z_1 = charge of peak 1 ($p_2 > p_1$)

The measured mass of a protein complex is often higher than the expected mass based on the sum of its constituents. This observation is attributed to the incomplete removal of buffer / solvent molecules during the ionization process. The mass determination becomes challenging if the width of the peak increases due to heterogeneity of the non-covalent assembly in form of mixed populations i.e. by post translational modifications.
The choice of the spraying solution has a dramatic impact on the conformation of the protein and the effect can be extrapolated to the charge envelope observed in the mass spectrum.

Organic solvents such as acetonitrile induce unfolding of the protein. The protein loses its conformation and exposes all aminoacids to the solvent with each chargeable residues being charged. The peak distribution shifts to a lower m/z range since more charges are present in the molecule [97, 99]. An alternative is offered by volatile aqueous buffered solutions such as ammonium acetate and ammoniumbicarbonate. They resemble more physiological conditions. If the same protein is sprayed in these buffered solutions at a neutral pH, the protein keeps its folded conformation and only chargeable residues at the outer surface of the complex are exposed to the solvent. The protein acquires less charges and the appearance of the ion signal shifts to higher m/z values. The effect of different spraying solvents is exemplified on the GroEL chaperonin complex. In its native fold in the cell the GroEL chaperonin complex consists of 14 identical subunits with a mass of ~57 kDa each. They form two heptameric back to back stacked rings [100-102]. Together with GroES, a co-chaperonin, it functions in the cell to help the correct folding of many proteins [103, 104]. Aqueous buffered solutions keep the protein assembly intact (here: 50 mM Ammoniumacetate, pH 6.8) (fig. 8).

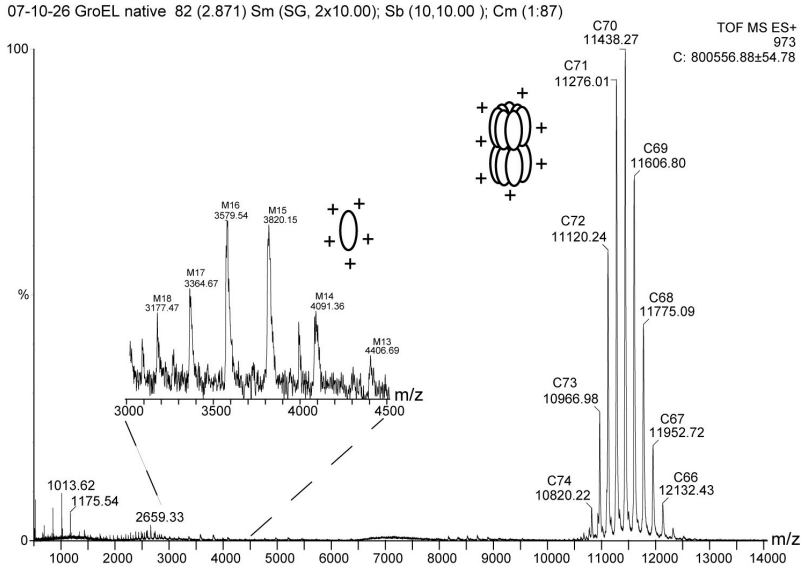


Figure 8

ESI-ToF mass spectrum of 5uM GroEL-chaperonin protein complex electrosprayed from a solution containing 50 mM ammoniumacetat pH 6.8. The main charge envelope at m/z 10,500 to 12,500 with 64 to 74 charges corresponds to the intact 14mer of GroEL. The inset magnifies a minor charge series at m/z 3,000 to 4,500 with 13 to 18 charges representing the folded monomeric protein of GroEL.

The spectrum shows the main charge envelope at m/z 10,500 to 12,500. Basic amino acid residues of the protein complex exposed to NH_4Ac-solution acquire approximately 70

charges in total that allow the mass determination of the protein complex to ~ 801 kDa. Due to the three dimensional structure hydrophobic patches involved in the interaction of all 14 proteins are buried in the interaction interface and not available for charging. The spectrum also shows a minor charge series at m/z 3,000 to 4,500 with about 16 charges. This charge envelope corresponds to a monomeric protein of the protein complex with a determined mass of ~57 kDa. Comparing the number of charges of the overall protein complex to the folded monomeric subunit it is obvious that the monomeric species aquires more charges than the protein complex per monomer. This can be explained with the increased surface of the monomeric protein exposed to the solution relative to monomeric proteins in the protein complex. Under non-physiological solution conditions (here: 50% acetonirile and 0.5% formic acid) the protein complex is not only disrupted into monomeric proteins but these proteins are also completely unfolded and are present in an elongated conformation (fig. 9).

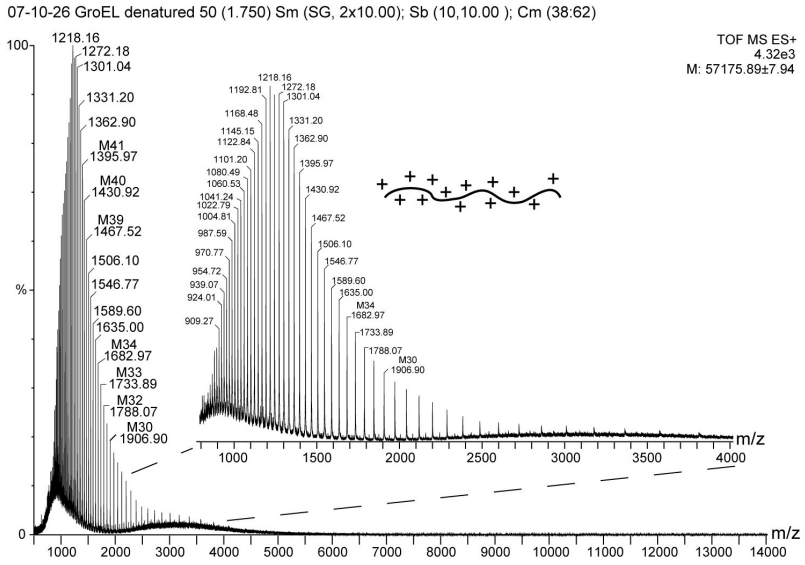


Figure 9

ESI-ToF mass spectrum of 5uM GroEL-chaperonin protein complex electrosprayed from a solution containing 50% acetonitrile, 50% water and 0.5% formic acid. Only one major charge envelope in the low m/z area is present that corresponds to a denatured monomeric GroEL protein.

The charge envelope dramatically shifts to lower m/z values (m/z 800 to 4000) as all basic amino acid residues are exposed to the solvent. The protein can take up many more charges than folded conformations. Unfolded protein conformations are missing structural information but provide more data points for a precise mass assignment. Ideally each m/z value of a charge envelope should be used for a good mass determination with a reliable standard deviation.

Lowering stepwise the pH of the buffered solutions will gradually unfold the protein and this change in conformation will be reflected in the charge distribution in the mass spectra [105, 106]. At a certain pH value two ion series will appear in the same spectrum; one of which corresponding to the folded conformation and the other to the unfolded conformation. The effects on stability of proteins can be monitored in respect to ionic strength. Ionic strength is a measure of the total concentration of ions in solution. The cell contains above other salts 0.15M KCl. This salt content in particular affects electrostatic interactions between charged amino acids within a protein complex. To mimic physiological condition the ionic strength of the spraying solution is important. Ammonium acetate has the same valency as NaCl and KCl. The effects of different ionic strength were investigated on the capsid protein gp23 of bacteriophage T4 which forms a non-covalent hexameric ring. Mass spectrometric analyses of gp23 in 50 mM ammonium acetate solution were compared to increasing amounts of ammonium acetate (0.2-1M) in the spraying solution. The results showed that increasing salt concentration induced destabilization effects on gp23 [107]. Moreover protein complexes undergo conformational changes at different temperatures. Elevated temperature can also affect complex stability. Wheat small heat shock protein *Ta*Hsp16.9 consists of a dodecamer with a total mass of 200 kDa. Upon heat shock it was shown that this protein complex changes its conformation. Mass spectrometry allowed monitoring the temperature dependant disassembly of the protein complex into oligomers [108].

Gas-phase behaviour of protein complexes versus in solution-phase behaviour

With the achievements in successfully analyzing non-covalent interactions, stoichiometry and conformational changes of proteins and protein complexes by native mass spectrometry, MS has found its niche in structural biology. It is now well established that proteins and protein complexes are maintained in the gas phase. However, it needs to be discussed whether gas phase characteristics of non-covalent proteins and protein complexes resemble in solution properties as in the gas phase the macromolecule is completely devoid of surrounding water molecules. Water molecules play an important role in stabilizing protein structures since it forces hydrophobic groups of the protein to be buried in the core of the folded protein. It was speculated that in the absence of water molecules proteins adopt disordered conformations that do not resemble their native fold anymore.

To correlate and validate mass spectrometry as a tool in structural biology, protein structures derived by native mass spectrometry have been compared to protein structures deduced from other techniques such as small angle x-ray scattering and x-ray crystallography. The analysis of ferredoxin dependent glutamate synthase from *Synechocystis sp.* by small angle x-ray scattering suggests the enzyme to be in a monomeric state in solution whereas the crystal structure shows dimerization. A mass

spectrometric analysis points to an equilibrium between the monomeric and dimeric form. In another example the oligomeric state of 4-oxalocrotonate tautomerase was probed by native mass spectrometry and x-ray crystallography. Both techniques consistently report that this protein forms a hexamer. The urease complex from Helicobacter pylori consisting of 12 $\alpha\beta$ monomers forms four $\alpha_3\beta_3$ subcomplexes. Again these data are in agreement with the structure derived by x-ray crystallography. In addition native mass spectrometry could also reveal that the intact $\alpha_{12}\beta_{12}$ complex dissociates into $\alpha_3\beta_3$ subunits confirming the $(\alpha_3\beta_3)_4$ arrangement. The stoichiometry of protein assemblies in the gas phase analyzed by native mass spectrometry is often in agreement with the stoichiometry in solution phase [94].

Apart from the consistency in the stoichiometry of protein complexes determined by mass spectrometry and x-ray crystallography, it is now investigated if protein complexes also retain their original shape and conformation in the gas-phase. A new player in macromolecular mass spectrometry is an instrument that hybridizes ion-mobility separation with a ToF analyzer (IM-MS). This combination has a great potential to shed light on the architecture of proteins and protein complexes in the gas phase. In ion mobility protein complexes travel through a drift tube filled with inert gas. Depending on their conformation, protein complexes experience a different number of gas collisions that influence the transit time through the ion mobility tube, hence the more volume the macromolecule fills the longer it takes to exit the drift tube. The drift time combined with charge state correction can be used to calculate the collisional cross section of the protein complexes. This information can then be directly used as readout to determine the shape and size of gas-phase protein aggregates. The first ion mobility findings on protein ions indicate that low charged species are rather compact suggesting folded conformations whereas higher charge states show expansion of the molecule hinting to an unfolding of the molecule [109]. Recently the trp RNA binding protein complex consisting of an 11-membered ring was investigated by ion mobility mass spectrometry and revealed that the topology is maintained in the gas phase [110].

Nevertheless structural investigations in the gas phase as it is the nature of mass spectrometry should be still treated with caution in respect to in solution structures. Native mass spectrometry can also lead to unexpected results that do not correlate with other biophysical methods. ESI-MS was used to investigate the interaction between integral membrane proteins and phospholipids [111]. Based on the crystal structure KcsA is a homotetrameric protein complex that forms a K^+ channel in the bacterial membrane. In conjunction with phospholipids the KcsA tetramer is highly stable in lipid bilayers [112]. However the mass spectrometric analyses revealed only monomeric KcsA with no protein-phospholipid interaction since harsh conditions of the mass spectrometer abolished any interactions. When the analyte was mixed with trifluoroethanol (TFE), a membrane mimicking solvent, the non-covalent interactions between the KcsA monomer and the phospholipids could be restored as it allowed more gentle conditions to be used. In addition

SDS-PAGE analysis confirmed that KcsA together with TFE forms only monomeric and not tetrameric conformation. Ideally the combination of several techniques in structural biology will lead to a better understanding of any biological system.

D.) RNA regulating molecular machineries

This thesis focuses on the detailed mass spectrometric analysis of protein complexes. It centers on the protein network concentrated around protein machineries engaged in RNA metabolism.
This paragraph will give an overview on the current knowledge of structural features and function of RNA regulating complexes in the cell.

Identification of the exosome

The exosome was originally identified in yeast in 1997 [113]. Its human homologue was recognized soon after in 1999 [114] and turned out to be a autoantigen in patients suffering from overlap syndrome of polymyositis (PM) and scleroderma (Scl) found already in 1977 [115]. Exosomes can be found in eucaryotes and archaea and besides human and yeast subsequent work identified exosomes in fruit fly (*drosophila melanogaster*) [116-118], protozoan parasites (*trypanosome brucei*) [119], plants (*arabidopsis thaliana*) [120-122] and archaeal bacteria (*sulfolobus sofataricus* [123], *pyrococcus abyssi* and *pyrococcus horikoshii* [124]). It is an evolutionary highly conserved protein complex as several yeast exosome mutant alleles can be complemented and recovered by their human counterpart [113, 114, 120].
The exosome is a multi component complex formed by the non-covalent interaction of several exoribonucleases. Most of these proteins are named ribosomal RNA processing proteins (Rrp) according to their enzymatic function. The assembly of several RNA binding and processing enzymes into one protein complex is still not fully understood but it has been hypothesized that they are essential for distinct RNA substrate recognition and regulation [125].

Function of the exosome

A cell contains numerous RNA types that play a central role in translating genetic information into proteins. For example messenger RNA (mRNA) functions as a carrier for DNA, whereas ribosomal RNA (rRNA) and transfer RNA (tRNA) operate as catalytical molecules that translate the genetic information of such mRNA into proteins. t-RNA does not code for genetic information but is responsible for the transport of nucleotides from the cytoplasm to the ribosome for protein biosynthesis. rRNA is also a short non-coding RNA stretch which acts as a building block for the ribosome. RNAs are also found in the nucleus

such as small nucleolar RNAs (snoRNA) and small nuclear RNA (snRNA) that are involved in the generation and modification of rRNA and mRNA respectively. In recent years a new class of RNA molecules were identified which include microRNA (miRNA) and small interfering RNA (siRNA). These molecules function in association with proteins, attach complementary to DNA and inhibit the transcription of genes. It is important to highly regulate the levels of these types of RNAs. Each step towards gene expression is tightly controlled. The correct processing of RNA, such as quality control, degradation and maturation plays an essential role in many aspects of the cell. One of the crucial key enzymes in the RNA metabolism in the cell is the exosome.

The exosome is a large multi subunit heterogeneous protein complex. It is involved in the precise cropping of the 3'-end of most RNA from precursor RNA to mature RNA [117, 126, 127]. But it also plays an important role in the final degradation of RNA molecules as an important control point for gene expression. The degradation of mRNA is necessary when the transcript is incorrect or when it is sufficiently translated [128, 129]. In eukaryotic cells there exist two independent degradation pathways of mRNA, either in 5'->3' direction or in 3'->5' direction (fig. 10).

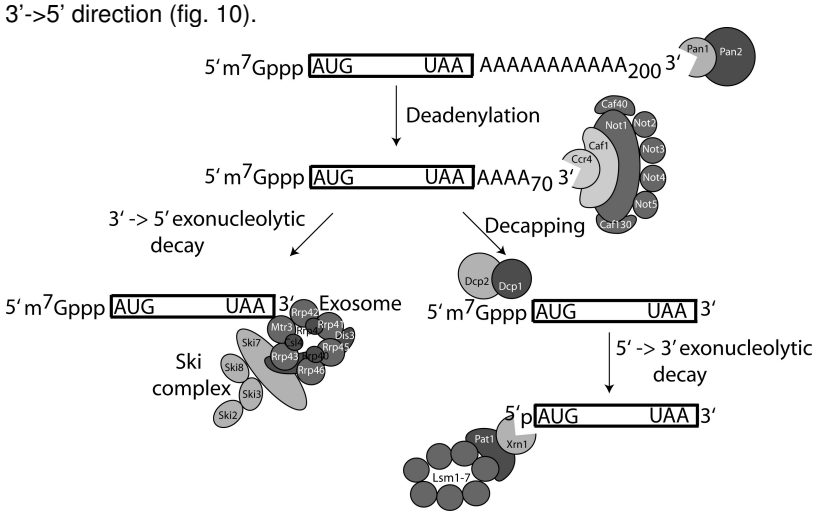


Figure 10

mRNA degradation pathway in yeast in either 5'->3' direction or in 3'->5' direction. Pan2p/Pan3p deadenylase complex removes the polyA tail at the 3' terminus and is then replaced by Ccr/Not-complex. mRNA degradation in 5' -> 3' direction continues with decapping the 5' end by Dcp1p/Dcp2p and complete degradation by exonuclease Xrn1p/Pat1p/Lsm1-7p. 3' -> 5' degradation is mediated by the exosome that forms a supercomplex with the Ski complex.

Both degradation pathways start with removal of the polyA tail at the 3' terminus by two distinct deadenylase complexes, Pan2p/Pan3p and Ccr/Not. Pan2p/Pan3p shortens the polyA tail to approximately 70 nucleotides. Then the Ccr/Not complex, consisting of Ccr4p,

Not1p to Not5p, Caf1p, Caf40p and Caf130p continues degradation of the remaining polyA tail [130-134]. The major pathway in yeast degrades in 5' -> 3' direction and continues with decapping the 5' end by Dcp1p/Dcp2p with subsequent complete degradation of the transcript by the exonuclease Xrn1p. Xrp1p is in complex with regulator proteins Pat1p and Lsm1p to Lsm7p. The second minor pathway degrades in 3' -> 5' direction by the exosome that forms a supercomplex with the Ski complex. In contrast to yeast it is believed that the major degradation pathway of mRNA in mammalian cells proceeds in
3' -> 5' direction and is mediated by the exosome. Moreover the exosome is involved in further mRNA surveillance pathways such as nonsense-mediated decay, nonstop decay and ARE mediated decay [135-140].
In eucaryotic cells the degradation of mRNA takes place in particular section in the cytosol named processing bodies (P-bodies). P-bodies are microscopic structures of distinct foci that co-localize mRNA and many enzymes involved in mRNA turnover which also hold several proteins of the 5' -> 3' degradation machinery i.e. Dcp1/Dcp2, Pat1p, Lsm1–7p and Xrn1p [141-146]. P-bodies need nontranslating mRNA for assembly and respond dynamically in size to changes affecting the degradation process of mRNA in the cell [147, 148].

Structure of the exosome

The exosome is closely related to the bacterial polynucleotide phosphorylase (PNPase) and RNase PH which is the main enzyme in RNA degradation in bacteria [149]. PNPase is composed of two RNase PH domains, a KH domain and a S1 binding domain. KH and S1 binding domains are characteristic for RNA binding. The crystal structure of PNPase of *Streptomyces antibioticus* shows a trimeric RNase PH ringshaped arrangement crowned with the KH and S1 domains of the PNPase proteins [150].
It is now well established that also all exosomes contain a hexameric ring structure with associated proteins. Most of these proteins have a demonstrated or predicted exoribonulease or hydrolase activity (table 1).

The simplest exosome is found in archaeal organisms. The genome codes for two RNase PH proteins (Rrp41 and Rrp42) and three putative S1 RNA binding domain proteins (Rrp4, Csl4, Rrp40). Rrp41, Rrp42 and Rrp4 are located on the same superoperon whereas Csl4 is found in a different operon [151]. The complex is reconstituted of an alternating arrangement of two proteins Rrp41 and Rrp42 to form a hexameric arrangement and three proteins (Rrp4 and/or Csl4) on top. Several atomic resolution crystal structures of archaeal exosomes are now available (fig. 11).

Table 1: Components of the yeast exosome complex

	Molecular weight [kDa]	motifs	function	comments
exosome protein				
Rrp43p	44,011	RNase PH	3'->5' phosphorolytic exoribonuclease	ring protein
Rrp41p	27,560	RNase PH	3'->5' phosphorolytic exoribonuclease	ring protein
Rrp42p	29,055	RNase PH	3'->5' phosphorolytic exoribonuclease	ring protein
Rrp45p	33,962	RNase PH	3'->5' phosphorolytic exoribonuclease	ring protein
Rrp46p	24,407	RNase PH	3'->5' phosphorolytic exoribonuclease	ring protein
Mtr3p	27,577	RNase PH	3'->5' phosphorolytic exoribonuclease	ring protein
Csl4p	31,583	S1	RNA binding	
Rrp4p	39,427	S1 / KH	RNA binding, 3'->5' hydrolytic exoribonuclease	
Rrp40p	26,590	S1 / KH	RNA binding, 3'->5' hydrolytic exoribonuclease	
Rrp44p	113,706	S1 / RNase R (vacB)	RNA binding, 3'->5' hydrolytic exoribonuclease	in human not found complexed with the exosome
Rrp6p	84,038	RNase D	3'->5' hydrolytic exoribonuclease	in yeast only nuclear, in human also cytosolic
Lrp1p	21,045	C1D	nucleotide binding	nuclear, interacts with Rrp6
Ynr024wp	21,142	-	unknown	nuclear
Co-factors				
Ski2p	146,057	DEAD box	ATP dependent RNA helicase	cytosolic, part of Ski complex
Ski3p	163,724	TPR repeats		cytosolic, part of Ski complex
Ski8p	44,231	WD repeats		cytosolic, part of Ski complex
Ski7p	84,778	GTPase		cytosolic, part of Ski complex
Gsp1p	24,810	GTPase	GTPase	cytosolic, interacts with Rrp44
Nip7p	20,381	-	unknown	cytosolic, interacts with Rrp44
Mtr4p	122,054	DEAD box	ATP dependent RNA helicase	nuclear, part of TRAMP complex
Trf4p	66,030	NTP transferase	Poly(A)polymerase	nuclear, part of TRAMP complex
Trf5p	74,179	NTP transferase	Poly(A)polymerase	nuclear, part of TRAMP complex, interacts with Mtr4
Air1p	41,631	RING-finger	RNA binding	nuclear, similarity to Air2p
Air2p	39,343	RING-finger	RNA binding	nuclear, part of TRAMP complex

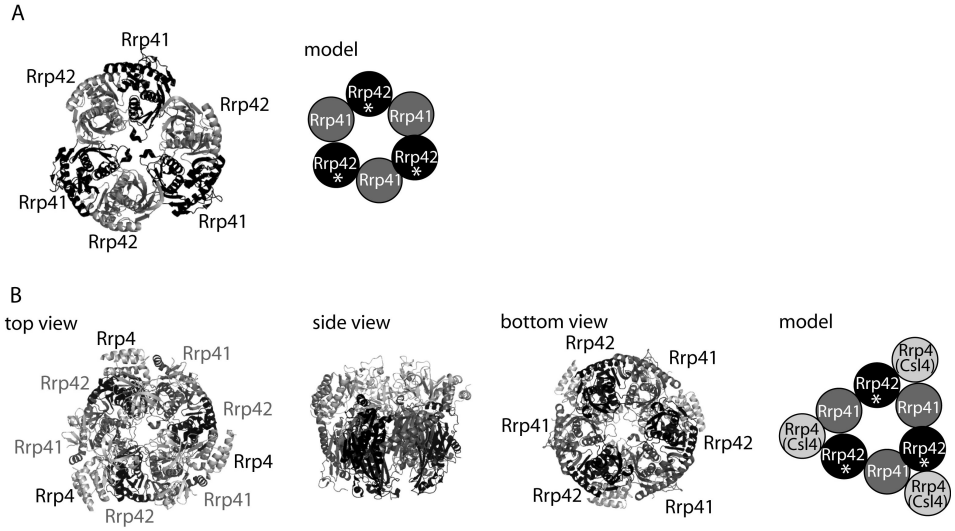


Figure 11

A.) Crystal structure and a schematic model of the archaeal exosome from *Sulfolobus solfataricus*. Two RNase PH proteins (Rrp41 and Rrp42) alternate to form a hexameric ring. Subunits denoted with "*" contain active sites for enzymatic degradation of RNA.

B.) Crystal structure of the nine membered archaeal exosome from *Archaeoglobus fulgidus.* The structures show the top, side and bottom view. On top of the hexameric ring consisting of Rrp41 and Rrp42, three additional proteins (either Rrp4 or Csl4) are bound. The schematic model illustrated active sites of the protein complex denoted with "*".

For example the RNase PH ring structure of *Sulfolobus solfataricus* [152] (fig. 11A) and the nine subunit exosome of the ringstructure in complex with either three molecules of Rrp4 or Csl4 of *Archaeoglobus fulgidus* [153] (fig. 11B). The crystal structure of *Sulfolobus solfataricus* revealed that also in the absence of Rrp4 or Csl4 the ring can bind short RNA oligonucleotides in the catalytic subunit Rrp41. Rrp42 shows no activity but may play an important role in the overall structure of the exosome.

The yeast exosome is a more elaborate protein assembly. The ring comprises six different RNase PH-like subunits (Rrp41, Rrp42, Rrp43, Rrp45, Rrp46, and Mtr3) that can be classified into Rrp41-like proteins (Rrp41, Rrp46 and Mtr3) and Rrp42-like proteins (Rrp42, Rrp43 and Rrp45). Three KH and/or S1 domain-containing RNA binding proteins (Csl4, Rrp4, and Rrp40) are situated on top of the ring structure. Moreover an additional protein named Rrp44 / Dis3, functionally a hydrolase is attached to the complex (for orthologous proteins refer to table 2).

Table 2 yeast exosome and its human, archael and trypanosomal orthologues

Yeast proteins	Yeast gene	MW [kDa]	Human proteins	human gene	MW [kDa]	Archaeal proteins	Trypano-somal proteins
Rrp41p/Ski6	YGR195w	27.6	hRrp41	EXOSC4	26.4	Rrp41	TbRrp41A
Rrp45p	YDR280w	34.0	PM/Scl-75	EXOSC9	49.0	Rrp42	TbRrp45
Rrp46p	YGR095c	24.4	hRrp46	EXOSC5	25.3	Rrp41	TbRrp41B
Rrp43p	YCR035c	44.0	OIP2	EXOSC8	30.0	Rrp42	TbEAP2
Mtr3p	YGR158c	27.6	hMtr3	EXOSC6	28.2	Rrp41	TbEAP4
Rrp42p	YDL111	29.1	hRrp42	EXOSC7	31.8	Rrp42	TbEAP1
Csl4p/Ski4p	YNL232	31.6	hCsl4	EXOSC1	21.5	-	-
Rrp4p	YHR069c	39.4	hRrp4	EXOSC2	32.7	-	-
Rrp40p	YOL142w	26.6	hRrp40	EXOSC3	29.6	-	-
Rrp44p/Dis3p	YOL021c	113.7	(hDis3)	KIAA1008	109.0	-	-
Rrp6p	YOR001w	84.0	PM/Scl-100	EXOSC10	100.8	-	-

The proteins range from 20 to 115 kDa in molecular weight and result in the predicted total exosome complex mass of approximately 400 kDa for the cytoplasmic exosome and in 683 kDa for the nuclear exosome. Exosome proteins reside in the cytosol and in the nucleus to form special exosome assemblies that target different RNA substrates. Immunolocalization studies of the exosome of *Drosophila melanogaster* show a complex distribution pattern of individual exosome proteins in their discrete subcellular compartment [154]. Several, but not all, exosome proteins localize to different cytoplasmic foci suggesting that distinct subcomplexes are formed in the cell. Rrp46p, Rrp4p, Rrp42p and Csl4p are concentrated in distinct foci whereas Mtr3p and Rrp40p are present throughout the cytoplasm and around the nucleus. Rrp41 and Rrp44 are enriched around the nucleus with either Rrp41 cytoplasmic or Rrp44 predominantly nuclear or vice versa. These two proteins may shuttle between the nucleus and the cytoplasm, since Rrp44 is also in direct interaction with Ran, a protein involved in nucleocytoplasmic transport of proteins [155]. The nucleus holds extra proteins that are exclusively associated to the nuclear exosome such as Rrp6, Rrp47 / Lrp1 and Ynr024w. Localization studies illustrate that Rrp6 is enriched in the nucleolus. The direct interaction between Lrp1 and the N-terminal region of Rrp6 has been recently demonstrated where Lrp1 promotes Rrp6 activity [156]. The structure of the yeast exosome was initially investigated by electron microscopy and homology modeling based on bacterial PNPase [157]. All yeast exosomal protein were classified according to their domains and fitted into the overall doughnut like shape of the electron microscopy picture. It suggested a ring arrangement of Rrp43-Rrp41-Mtr3-Rrp45-Rrp46-Rrp42 with Rrp4 on top of Rrp43-Rrp41, Rrp40 crowning Mtr3-Rrp45 and Csl4 on top of Rrp46-Rrp42. Previously yeast two hybrid

analysis detected interactions between Rrp42-Mtr3 [12, 15], Rrp45-Rrp41 [12] and Rrp43-Rrp46 [158]. This information is not in agreement with the suggested structure based on electronmicroscopy but support the human exosome arrangement in a model derived from two hybrid assays. Mammalian two hybrid assay data proposed Rrp41-Rrp42-Mtr3-Rrp43-Rrp46-Rrp45 of the human exosomal ring structure [159]. This model was recently confirmed by the crystal structure of the human exosome. In 2006 the human nine subunit core exosome complex with a total mass of 286 kDa has been reconstituted *in vitro* and crystallized [160] (fig. 12A).

The human exosome contains a homologous protein to yeast Rrp44 but so far it has not been found in complex with the exosome. The arrangement according to the human exosome crystal structure is Rrp41-Rrp45-Rrp46-Rrp43-Mtr3-Rrp42. The subunits Rrp41, Rrp46 and Mtr3 resemble catalytically more its archaeal counterprotein Rrp41. Rrp45, Rrp43 and Rrp42 are alike archaeal protein Rrp42. Hence the ring is formed by three heterodimers (Rrp41-Rrp45, Rrp46-Rrp43, Mtr3-Rrp42) in agreement with the archaeal structure of the exosome. These three dimers are connected to each other by one of three S1 binding domain proteins (Csl4, Rrp4, Rrp40). The interaction point between Rrp45-Rrp46 is stabilized by Rrp40 as well as Rrp43-Mtr3 by Csl4 and Rrp42-Rrp41 by Rrp4.

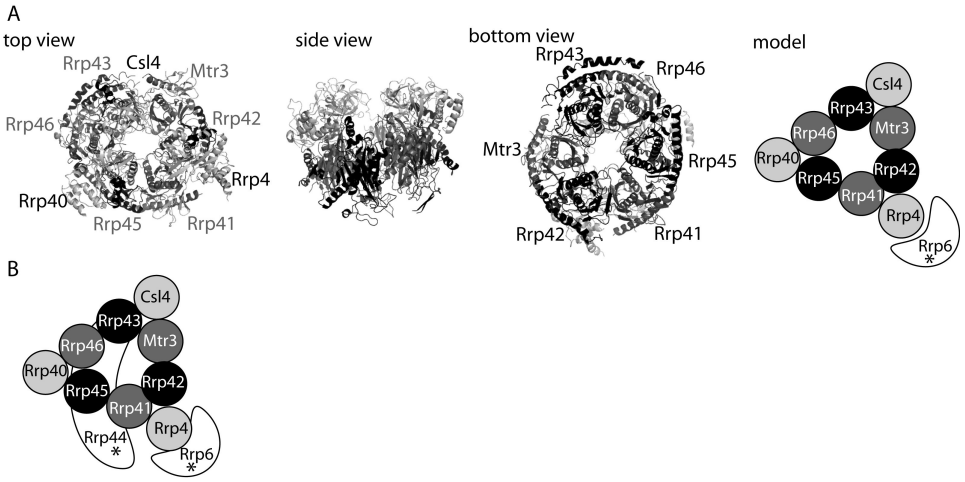


Figure 12

A.) Crystal structure of the nine membered human exosome from the top, side and bottom view. Rrp43, Mtr3, Rrp42, Rrp41, Rrp45 and Rrp46 form a hexameric ring crowned by Csl4, Rrp4 and Rrp40. Csl4 interconnects Rrp43 and Mtr3, Rrp4 binds to Rrp42 and Rrp41 and Rrp40 interacts with Rrp45 and Rrp46. The schematic model of the human exosome additionally places the only functional protein Rrp6 (denoted with "*") into the complex.

B.) Schematic model of the yeast exosome based on electron microscopy image modeling predominantly on the human exosome crystal structure. Additionally Rrp44 could be located in the overall structure and is situated below the hexameric ring. The yeast exosome contains two proteins that show activity, Rrp44 and Rrp6 as denoted with "*".

Trials to crystallize the yeast exosome have not been successful so far. Though in the same study in which the human exosome crystal structure is presented, the in solution reconstruction of the human and yeast exosome was probed. While archaeal exosomes assemble to a hexameric core, the reconstruction of the human and yeast exosome only succeeded when dimeric subcomplexes of the ring were mixed with Rrp4, Csl4 and Rrp40. Human and yeast exosomes are not established by a hexameric core but only form complexes of at least nine subunits. The addition of Rr44 and Rrp6 to the yeast 9 protein exosome led to the formation of 10 protein and 11 protein exosome assembly respectively.
During the time course of the work described in this thesis a new structural model has been proposed for the yeast exosome. The model has been derived from structural investigations based on native mass spectrometry [161]. The yeast complex was isolated directly from cells with the TAP-technique as also performed in this thesis. The protein complex was analyzed in the intact form and was then partially denatured and perturbed by chaotropic reagents (such as DMSO) to generate subcomplexes which were in turn analyzed by MS and MS/MS techniques. Puzzling dimeric and trimeric subcomplex into the overall picture resulted in a proposed arrangement of Rrp41-Rrp45-Rrp46-Rrp43-Mtr3-Rrp42 to form a hexameric ring with the S1 binding domain protein Rrp40 connected to Rrp45-Rrp46, Csl4 to Rrp43-Mtr3 and Rrp4 to Rrp42-Rrp41. This protein assembly corresponds exactly to the subunit configuration of the human exosome in the crystal structure. Additionally the native mass spectrometry data allowed locating Rrp44 in the yeast exosome which is interacting with Rrp45, Rrp41 and Rrp42. Recently single particle electron microscopy imaging revealed the yeast exosome at 23 - 19 Å resolution [162]. The data was reconstructed on the human exosome crystal structure (fig. 12B). Direct comparison between the nine membered exosome (hexameric ring plus Rrp4, Csl4 and Rrp40) and the corresponding exosome including Rrp44 allowed the placement of Rrp44 in the overall topology. According to this structure Rrp44 is a two lobed protein that is situated below the hexameric ring. The RNase II type active site is in direct contact with mainly Rrp43 and Rrp45, whereas the head of Rrp44 interacts with Rrp41. In a yeast-two hybrid screen of the exosome, a weak interaction between Rrp4 and Rrp6 could be established [163].
The exosome from trypanosomes is very similar to the yeast and human exosome. It is also build of six RNase PH-like subunits, TbRrp41A, TbRrp41B, TbRrp45, TbEAP1, TbEAP2, TbEAP4 which are likely homologues to yeast Rrp41, Rrp46, Rrp45, Rrp42, Rrp43 and Mtr3 [119]. And it also contains three S1-binding domain proteins TbRrp4, TbRrp40 and TbCsl4 and the nuclear specific components TbRrp6 and TbEAP3 analogous to yeast and human component Rrp6 and Lrp1 respectively. But in contrast to yeast, Rrp6 in trypanosomes is likely to be present in the cytosol and the nucleus. The topology of the trypanosomal exosome was investigated by yeast two hybrid analyses [164]. A structural arrangement proposes a ring like structure and is consistent with the human exosome. However the yeast two hybrid data could not detect the interaction to close the ring of the trypanosomal

exosome but it is unclear whether this is due to an artifact of the method or indicates a different arrangement of the exosome.

The reoccurring structural feature of one doughnut shaped hexameric RNase PH ring crowned with a second trimeric RNA binding ring prompted the hypothesis that RNA substrates are bound by the RNA binding proteins and consequently channeled to the inner cavity of the protein complex. In the processing chamber of the catalytic ring the RNA is processed and / or degraded [153, 165, 166]. This hypothesis applies to the bacterial PNPase and to the archaeal exosome, where both complexes contain active sites hidden in the processing chamber [150, 152, 167]. The functional mechanism for eukaryotic exosomes must be different since only Rrp44 and Rrp6 show enzymatic activity.

As mentioned earlier, for the efficient mRNA degradation in the cytoplasm a second large protein complex is required that functions as a co-activator complex for the exosome [168]. This complex was originally identified in antiviral response of the cell [169, 170]. It represses dsRNA virus propagation by specifically blocking the translation of viral poly(A) negative mRNA and is therefore termed Superkiller (Ski) complex. The Ski complex has been suggested to be an mRNA decay-specific cofactor for the exosome because mutations in the SKI genes inhibit 3'-to-5' mRNA decay, but have no effect on functions of exosome in nuclear RNA processing [171]. This complex is comprised of three different subunits: Ski3p, Ski2p and Ski8p. It was suggested that the Ski complex is a heterotrimeric protein complex [172]. Ski3p with a mass of 163 kDa contains eleven tetratricopeptide repeats (TPR). This all helical motif is comprised of a repeat of 34 aminoacids and mediates a plethora of cellular processes. The basic function of TPR domains is the assembly of protein complexes [173, 174]. Ski2p contains a DEVH motif and is a RNA helicase [175, 176]. This protein has a molecular weight of 146 kDa. And Ski8p has a molecular weight of 44 kDa and holds several WD-40 repeats. These WD-40 repeats are defined to have a terminal conserved tryptophan and aspartic acid dipeptide and a length of approximately 40 amino acids residues [177]. This motif is found in a wide variety of eukaryotic proteins with multiple functions including adaptor/regulatory modules in signal transduction, RNA splicing, cytoskeleton assembly, gene transcriptional activation and cell cycle control and is rarely found in prokaryotes [178]. The common function is exhibited as a scaffold for the reversible association of binding partners. The crystal structure of Ski8p has been determined and shows a seven bladed β-propeller, which is a common fold of WD-40 repeat proteins [179, 180]. It has also been shown that this structure is a versatile module for the recognition of post-translational modifications. The WD40-repeat domain of β-TRCP and Cdc4 folds into a seven bladed and eight bladed propeller respectively and recognizes phosphorylated serine and threonine containing peptides [181]. At the top surface of the propeller a characteristic hydrophobic region is found which is most likely involved in protein binding of its interaction partners. Ski8p interacts with Ski3p in the Ski complex [180] as proteins containing a TPR motif are often found in association with a WD-repeat protein [178, 182]. Ski8p can be also

connected to Spo11p involved in double-strand break formation during meiotic recombination [183, 184]. Although Ski8p shows only 15.5% sequence identity to the β-subunit of the cell signaling heterotrimeric G-protein complex transducin, both crystal structure superimpose remarkably well with a devitation of only 1.5 Å over 238 C_α-atoms [179, 185, 186]. G-proteins consist of three subunits: α, β and γ. The β- and γ-subunit form a tight heterodimer. The α-subunit contains a GDP/GTP binding domain and hydrolyzes GTP. It binds only to the hydrophobic top region of G_β of the heterodimer in its inactive form when it has bound GDP. Upon GTP binding it dissociates from the protein complex to fulfill diverse functions in the cell. Therefore, the contact surface between G_α and $G_{\beta\gamma}$ has major regulatory importance [187]. These common structural features between G_β and Ski8p suggest that Ski8p plays an important role in bringing Ski3p and Ski2p together to form a protein complex. The Ski complex interacts with the exosome via an adapter G protein named Ski7p [188, 189]. Since both, Ski8p and G_β, have an identical fold and both bind to G-proteins Ski7 and G_α respectively, it is tempting to hypothesize that Ski8p achieved reversible protein binding by using a nucleotide exchange mechanism analogous to G_α and $G_{\beta\gamma}$.
Also the nuclear exosome needs a coactivator complex for activity. The RNA degradation activity in the nucleus is mediated by the TRAMP complex which is the Ski complex equivalent and needed for the specific processing of snRNA and snoRNA [190].

E.) Scope of this thesis

This thesis targets the detailed mass spectrometric analysis of large protein machineries that play a crucial role in the RNA metabolism in the cell. The endogenously expressed protein assemblies were isolated from yeast cells by the TAP technology and subsequently analysed by various mass spectrometric methods. In these analyses emphasis is put on macromolecular mass spectrometry which allows the investigation of non-covalent protein complexes.

Chapter 2 introduces a modified QTOF instrument that enables the analysis of gas phase dissociation behaviour of noncovalent protein complexes. The modifications are set into context and illustrated on a newly identified heterotetrameric dioxygenase (hydroquinonedioxygenase, HQDO) that needs iron(II)-ions for activity and is strongly inhibited by 4-hydroxybenzoate.

The work in **chapter 3** describes the analysis of the endogenously expressed exosome complex from *Saccharomyces cerevisiae* using three different approaches in mass spectrometry. It illustrates an in-depth method to investigate complex constituents, post-

translational modifications, protein complex stoichiometry and relative protein binding in a heterogeneous protein assembly.

Chapter 4 addresses the comparison of the nuclear to the cytoplasmic exosome in the cell. These two protein complexes are closely related and have many proteins in common but are found in two different cellular compartments. Here we set out to relatively quantify the proteins and their phosphorylation status of the nuclear and cytoplasmic exosome by stable isotope labeling and investigate both complexes by macromolecular MS.

The work described in **chapter 5** covers the analysis of the Ski complex, a co-activator complex of the exosome by multiple mass spectrometric methods. It shows the potential of macromolecular mass spectrometry in the determination of protein complex stoichiometry and in the elucidation of the quartenary structure of large protein assemblies. This investigation led to the result that the Ski complex is a heterotetramer and places all Ski proteins into the overall structure.

This thesis ends with a summarizing discussion and concluding remarks in **chapter 6**.

References

1. Yang, X.J. and S. Gregoire, *Metabolism, cytoskeleton and cellular signalling in the grip of protein Nepsilon - and O-acetylation.* EMBO Rep, 2007. 8(6): p. 556-62.
2. Jiang, L., et al., *Global assessment of combinatorial post-translational modification of core histones in yeast using contemporary mass spectrometry. LYS4 trimethylation correlates with degree of acetylation on the same H3 tail.* J Biol Chem, 2007. 282(38): p. 27923-34.
3. Johnson, E.S., *Protein modification by SUMO.* Annu Rev Biochem, 2004. 73: p. 355-82.
4. Wohlschlegel, J.A., et al., *Global analysis of protein sumoylation in Saccharomyces cerevisiae.* J Biol Chem, 2004. 279(44): p. 45662-8.
5. Ptacek, J., et al., *Global analysis of protein phosphorylation in yeast.* Nature, 2005. 438(7068): p. 679-84.
6. Ptacek, J. and M. Snyder, *Charging it up: global analysis of protein phosphorylation.* Trends Genet, 2006. 22(10): p. 545-54.
7. Manning, G., et al., *Evolution of protein kinase signaling from yeast to man.* Trends Biochem Sci, 2002. 27(10): p. 514-20.
8. Ficarro, S.B., et al., *Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae.* Nat Biotechnol, 2002. 20(3): p. 301-5.
9. Cohen, P., *The regulation of protein function by multisite phosphorylation--a 25 year update.* Trends Biochem Sci, 2000. 25(12): p. 596-601.
10. Gavin, A.C., et al., *Functional organization of the yeast proteome by systematic analysis of protein complexes.* Nature, 2002. 415(6868): p. 141-7.
11. Ho, Y., et al., *Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry.* Nature, 2002. 415(6868): p. 180-3.
12. Ito, T., T. Chiba, and M. Yoshida, *Exploring the protein interactome using comprehensive two-hybrid projects.* Trends Biotechnol, 2001. 19(10 Suppl): p. S23-7.
13. Krogan, N.J., et al., *Global landscape of protein complexes in the yeast Saccharomyces cerevisiae.* Nature, 2006. 440(7084): p. 637-43.
14. Miller, J.P., et al., *Large-scale identification of yeast integral membrane protein interactions.* Proc Natl Acad Sci U S A, 2005. 102(34): p. 12123-8.

15. Uetz, P., et al., *A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae.* Nature, 2000. 403(6770): p. 623-7.
16. Rain, J.C., et al., *The protein-protein interaction map of Helicobacter pylori.* Nature, 2001. 409(6817): p. 211-5.
17. Giot, L., et al., *A protein interaction map of Drosophila melanogaster.* Science, 2003. 302(5651): p. 1727-36.
18. Li, S., et al., *A map of the interactome network of the metazoan C. elegans.* Science, 2004. 303(5657): p. 540-3.
19. Gandhi, T.K., et al., *Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets.* Nat Genet, 2006. 38(3): p. 285-93.
20. Rual, J.F., et al., *Towards a proteome-scale map of the human protein-protein interaction network.* Nature, 2005. 437(7062): p. 1173-8.
21. Stelzl, U., et al., *A human protein-protein interaction network: a resource for annotating the proteome.* Cell, 2005. 122(6): p. 957-68.
22. Fields, S. and O. Song, *A novel genetic system to detect protein-protein interactions.* Nature, 1989. 340(6230): p. 245-6.
23. Eyckerman, S., et al., *Design and application of a cytokine-receptor-based interaction trap.* Nat Cell Biol, 2001. 3(12): p. 1114-9.
24. Tavernier, J., et al., *MAPPIT: a cytokine receptor-based two-hybrid method in mammalian cells.* Clin Exp Allergy, 2002. 32(10): p. 1397-404.
25. Terpe, K., *Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems.* Appl Microbiol Biotechnol, 2003. 60(5): p. 523-33.
26. Puig, O., et al., *The tandem affinity purification (TAP) method: a general procedure of protein complex purification.* Methods, 2001. 24(3): p. 218-29.
27. Rigaut, G., et al., *A generic protein purification method for protein complex characterization and proteome exploration.* Nat Biotechnol, 1999. 17(10): p. 1030-2.
28. Westermarck, J., et al., *The DEXD/H-box RNA helicase RHII/Gu is a co-factor for c-Jun-activated transcription.* Embo J, 2002. 21(3): p. 451-60.
29. Bouveret, E., et al., *A Sm-like protein complex that participates in mRNA degradation.* Embo J, 2000. 19(7): p. 1661-71.
30. Honey, S., et al., *A novel multiple affinity purification tag and its use in identification of proteins associated with a cyclin-CDK complex.* Nucleic Acids Res, 2001. 29(4): p. E24.
31. Collins, S.R., et al., *Toward a comprehensive atlas of the physical interactome of Saccharomyces cerevisiae.* Mol Cell Proteomics, 2007. 6(3): p. 439-50.
32. Edwards, A.M., et al., *Bridging structural biology and genomics: assessing protein interaction data with known complexes.* Trends Genet, 2002. 18(10): p. 529-36.
33. Kemmeren, P., et al., *Protein interaction verification and functional annotation by integrated analysis of genome-scale data.* Mol Cell, 2002. 9(5): p. 1133-43.
34. Dziembowski, A. and B. Seraphin, *Recent developments in the analysis of protein complexes.* FEBS Lett, 2004. 556(1-3): p. 1-6.
35. Gloeckner, C.J., et al., *A novel tandem affinity purification strategy for the efficient isolation and characterisation of native protein complexes.* Proteomics, 2007.
36. Williams, D.H., et al., *Fast-atom-bombardment mass spectrometry. A new technique for the determination of molecular weights and amino acid sequences of peptides.* Biochem J, 1982. 201(1): p. 105-17.
37. Dole, M., Mack, L.L., Hines, R.L., Mobley, R.C., Ferguson, L.D., Alice, M.B., *Molecular Beams of Macroions.* The Journal of Chemical Physics, 1968. 49: p. 2240-2249.
38. Iribarne, J.V., Thomson, J., *On the evaporation of small ions from charged droplets.* The Journal of Chemical Physics, 1976. 64: p. 2287-2294.
39. Nguyen, S. and J.B. Fenn, *Gas-phase ions of solute species from charged droplets of solutions.* Proc Natl Acad Sci U S A, 2007. 104(4): p. 1111-7.
40. Kebarle, P., *A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry.* J Mass Spectrom, 2000. 35(7): p. 804-17.
41. Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., Whitehouse, C.M., *Electrospray ionization from mass spectrometry of large biomolecules.* Science, 1989. 246: p. 64-71.
42. Whitehouse, C.M., et al., *Electrospray interface for liquid chromatographs and mass spectrometers.* Anal Chem, 1985. 57(3): p. 675-9.
43. Wilm, M., et al., *Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry.* Nature, 1996. 379(6564): p. 466-9.

44. Ferguson, A.D., et al., *Siderophore-mediated iron transport: crystal structure of FhuA with bound lipopolysaccharide.* Science, 1998. 282(5397): p. 2215-20.
45. Juraschek, R., T. Dulcks, and M. Karas, *Nanoelectrospray--more than just a minimized-flow electrospray ionization source.* J Am Soc Mass Spectrom, 1999. 10(4): p. 300-8.
46. Schmidt, A., M. Karas, and T. Dulcks, *Effect of different solution flow rates on analyte ion signals in nano-ESI MS, or: when does ESI turn into nano-ESI?* J Am Soc Mass Spectrom, 2003. 14(5): p. 492-500.
47. Wilm, M. and M. Mann, *Analytical properties of the nanoelectrospray ion source.* Anal Chem, 1996. 68(1): p. 1-8.
48. Wilm, M.S. and M. Mann, *Electrospray and Taylor-Cone Theory, Doles Beam of Macromolecules at Last.* International Journal of Mass Spectrometry and Ion Processes, 1994. 136(2-3): p. 167-180.
49. Guilhaus, M., D. Selby, and V. Mlynski, *Orthogonal acceleration time-of-flight mass spectrometry.* Mass Spectrom Rev, 2000. 19(2): p. 65-107.
50. Blagoev, B., et al., *A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling.* Nat Biotechnol, 2003. 21(3): p. 315-8.
51. Ranish, J.A., M. Brand, and R. Aebersold, *Using stable isotope tagging and mass spectrometry to characterize protein complexes and to detect changes in their composition.* Methods Mol Biol, 2007. 359: p. 17-35.
52. Corvey, C., et al., *Carbon Source-dependent assembly of the Snf1p kinase complex in Candida albicans.* J Biol Chem, 2005. 280(27): p. 25323-30.
53. Ranish, J.A., et al., *The study of macromolecular complexes by quantitative proteomics.* Nat Genet, 2003. 33(3): p. 349-55.
54. Wang, T., et al., *In vivo dual-tagging proteomic approach in studying signaling pathways in immune response.* J Proteome Res, 2005. 4(3): p. 941-9.
55. Selbach, M. and M. Mann, *Protein interaction screening by quantitative immunoprecipitation combined with knockdown (QUICK).* Nat Methods, 2006. 3(12): p. 981-3.
56. Smolka, M.B., et al., *Dynamic changes in protein-protein interaction and protein phosphorylation probed with amine-reactive isotope tag.* Mol Cell Proteomics, 2005. 4(9): p. 1358-69.
57. Blagoev, B., et al., *Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics.* Nat Biotechnol, 2004. 22(9): p. 1139-45.
58. Hinsby, A.M., J.V. Olsen, and M. Mann, *Tyrosine phosphoproteomics of fibroblast growth factor signaling: a role for insulin receptor substrate-4.* J Biol Chem, 2004. 279(45): p. 46438-47.
59. Pflieger, D., et al., *Quantitative proteomic analysis of protein complexes: Concurrent identification of interactors and their state of phosphorylation.* Mol Cell Proteomics, 2007.
60. Ong, S.E., et al., *Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics.* Mol Cell Proteomics, 2002. 1(5): p. 376-86.
61. Mann, M., *Functional and quantitative proteomics using SILAC.* Nat Rev Mol Cell Biol, 2006. 7(12): p. 952-8.
62. Ong, S.E. and M. Mann, *Stable isotope labeling by amino acids in cell culture for quantitative proteomics.* Methods Mol Biol, 2007. 359: p. 37-52.
63. Romijn, E.P., et al., *Expression clustering reveals detailed co-expression patterns of functionally related proteins during B cell differentiation: a proteomic study using a combination of one-dimensional gel electrophoresis, LC-MS/MS, and stable isotope labeling by amino acids in cell culture (SILAC).* Mol Cell Proteomics, 2005. 4(9): p. 1297-310.
64. Kolkman, A., et al., *Double standards in quantitative proteomics: direct comparative assessment of difference in gel electrophoresis and metabolic stable isotope labeling.* Mol Cell Proteomics, 2005. 4(3): p. 255-66.
65. Gygi, S.P., et al., *Quantitative analysis of complex protein mixtures using isotope-coded affinity tags.* Nat Biotechnol, 1999. 17(10): p. 994-9.
66. Schnolzer, M., P. Jedrzejewski, and W.D. Lehmann, *Protease-catalyzed incorporation of 18O into peptide fragments and its application for protein sequencing by electrospray and matrix-assisted laser desorption/ionization mass spectrometry.* Electrophoresis, 1996. 17(5): p. 945-53.
67. Mirgorodskaya, O.A., et al., *Quantitation of peptides and proteins by matrix-assisted laser desorption/ionization mass spectrometry using (18)O-labeled internal standards.* Rapid Commun Mass Spectrom, 2000. 14(14): p. 1226-32.

68. Wang, Y.K., et al., *Inverse 18O labeling mass spectrometry for the rapid identification of marker/target proteins.* Anal Chem, 2001. 73(15): p. 3742-50.
69. Stewart, II, T. Thomson, and D. Figeys, *18O labeling: a tool for proteomics.* Rapid Commun Mass Spectrom, 2001. 15(24): p. 2456-65.
70. Thompson, A., et al., *Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS.* Anal Chem, 2003. 75(8): p. 1895-904.
71. Ross, P.L., et al., *Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents.* Mol Cell Proteomics, 2004. 3(12): p. 1154-69.
72. Che, F.Y. and L.D. Fricker, *Quantitation of neuropeptides in Cpe(fat)/Cpe(fat) mice using differential isotopic tags and mass spectrometry.* Anal Chem, 2002. 74(13): p. 3190-8.
73. Hsu, J.L., et al., *Stable-isotope dimethyl labeling for quantitative proteomics.* Anal Chem, 2003. 75(24): p. 6843-52.
74. Paoletti, A.C., et al., *Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors.* Proc Natl Acad Sci U S A, 2006. 103(50): p. 18928-33.
75. Rinner, O., et al., *An integrated mass spectrometric and computational framework for the analysis of protein interaction networks.* Nat Biotechnol, 2007. 25(3): p. 345-52.
76. Ganem, B., Y.T. Li, and J.D. Henion, *Detection of Noncovalent Receptor Ligand Complexes by Mass-Spectrometry.* Journal of the American Chemical Society, 1991. 113(16): p. 6294-6296.
77. Katta, V. and B. Chait, *Observation of the heme-globin complex in native myoglobin by electrospray-ionization mass spectrometry.* J Am Chem Soc, 1991. 113: p. 8534-8535.
78. Pinkse, M.W., et al., *Macromolecular assembly of Helicobacter pylori urease investigated by mass spectrometry.* J Mass Spectrom, 2003. 38(3): p. 315-20.
79. Sanglier, S., et al., *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, 2003. 14(5): p. 419-29.
80. van Berkel, W.J., et al., *Detection of intact megaDalton protein assemblies of vanillyl-alcohol oxidase by mass spectrometry.* Protein Sci, 2000. 9(3): p. 435-9.
81. Chernushevich, I.V. and B.A. Thomson, *Collisional cooling of large ions in electrospray mass spectrometry.* Anal Chem, 2004. 76(6): p. 1754-60.
82. Tahallah, N., et al., *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, 2001. 15(8): p. 596-601.
83. Ayed, A., et al., *Quantitative evaluation of protein-protein and ligand-protein equilibria of a large allosteric enzyme by electrospray ionization time-of-flight mass spectrometry.* Rapid Commun Mass Spectrom, 1998. 12(7): p. 339-44.
84. van Duijn, E., et al., *Monitoring macromolecular complexes involved in the chaperonin-assisted protein folding cycle by mass spectrometry.* Nat Methods, 2005. 2(5): p. 371-6.
85. Callaghan, A.J., et al., *Quaternary structure and catalytic activity of the Escherichia coli ribonuclease E amino-terminal catalytic domain.* Biochemistry, 2003. 42(47): p. 13848-55.
86. Lorenzen, K., et al., *Structural biology of RNA polymerase III: mass spectrometry elucidates subcomplex architecture.* Structure, 2007. 15(10): p. 1237-45.
87. Ilag, L.L., et al., *Mass spectrometry of Escherichia coli RNA polymerase: interactions of the core enzyme with sigma70 and Rsd protein.* Structure, 2004. 12(2): p. 269-75.
88. Monti, M.C., et al., *Interactions of Kid-Kis toxin-antitoxin complexes with the parD operator-promoter region of plasmid R1 are piloted by the Kis antitoxin and tuned by the stoichiometry of Kid-Kis oligomers.* Nucleic Acids Res, 2007. 35(5): p. 1737-49.
89. van den Heuvel, R.H., et al., *Structural studies on flavin reductase PheA2 reveal binding of NAD in an unusual folded conformation and support novel mechanism of action.* J Biol Chem, 2004. 279(13): p. 12860-7.
90. Tahallah, N., et al., *Cofactor-dependent assembly of the flavoenzyme vanillyl-alcohol oxidase.* J Biol Chem, 2002. 277(39): p. 36425-32.
91. Aquilina, J.A. and C.V. Robinson, *Investigating interactions of the pentraxins serum amyloid P component and C-reactive protein by mass spectrometry.* Biochem J, 2003. 375(Pt 2): p. 323-8.
92. van den Heuvel, R.H., et al., *Coenzyme binding during catalysis is beneficial for the stability of 4-hydroxyacetophenone monooxygenase.* J Biol Chem, 2005. 280(37): p. 32115-21.

93. Sharon, M. and C.V. Robinson, *The role of mass spectrometry in structure elucidation of dynamic protein complexes.* Annu Rev Biochem, 2007. 76: p. 167-93.
94. Heck, A.J. and R.H. Van Den Heuvel, *Investigation of intact protein complexes by mass spectrometry.* Mass Spectrom Rev, 2004. 23(5): p. 368-89.
95. Benesch, J.L. and C.V. Robinson, *Mass spectrometry of macromolecular assemblies: preservation and dissociation.* Curr Opin Struct Biol, 2006. 16(2): p. 245-51.
96. Hanson, C.L. and C.V. Robinson, *Protein-nucleic acid interactions and the expanding role of mass spectrometry.* J Biol Chem, 2004. 279(24): p. 24907-10.
97. McCammon, M.G. and C.V. Robinson, *Structural change in response to ligand binding.* Curr Opin Chem Biol, 2004. 8(1): p. 60-5.
98. Loo, J.A., *Studying noncovalent protein complexes by electrospray ionization mass spectrometry.* Mass Spectrom Rev, 1997. 16(1): p. 1-23.
99. Loo, J.A., et al., *Effect of reducing disulfide-containing proteins on electrospray ionization mass spectra.* Anal Chem, 1990. 62(7): p. 693-8.
100. Braig, K., et al., *The crystal structure of the bacterial chaperonin GroEL at 2.8 A.* Nature, 1994. 371(6498): p. 578-86.
101. Sigler, P.B., et al., *Structure and function in GroEL-mediated protein folding.* Annu Rev Biochem, 1998. 67: p. 581-608.
102. Xu, Z., A.L. Horwich, and P.B. Sigler, *The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex.* Nature, 1997. 388(6644): p. 741-50.
103. Kusukawa, N., et al., *Effects of mutations in heat-shock genes groES and groEL on protein export in Escherichia coli.* Embo J, 1989. 8(11): p. 3517-21.
104. Tilly, K. and C. Georgopoulos, *Evidence that the two Escherichia coli groE morphogenetic gene products interact in vivo.* J Bacteriol, 1982. 149(3): p. 1082-8.
105. Konermann, L., B.A. Collings, and D.J. Douglas, *Cytochrome c folding kinetics studied by time-resolved electrospray ionization mass spectrometry.* Biochemistry, 1997. 36(18): p. 5554-9.
106. Konermann, L., et al., *Acid-induced denaturation of myoglobin studied by time-resolved electrospray ionization mass spectrometry.* Biochemistry, 1997. 36(21): p. 6448-54.
107. Van Duijn, E., *New insight into chapreonin-assisted protein folding revealed by native mass spectrometry*, in *Biomolecular Mass Spectrometry*. 2006, Utrecht: Utrecht. p. 176.
108. Benesch, J.L., F. Sobott, and C.V. Robinson, *Thermal dissociation of multimeric protein complexes by using nanoelectrospray mass spectrometry.* Anal Chem, 2003. 75(10): p. 2208-14.
109. Clemmer, D. and M. Jarrold, *Ion Mobility Measurements and their Applications to Clusters and Biomolecules.* Journal of Mass Spectrometry, 1997. 32: p. 577-592.
110. Ruotolo, B.T., et al., *Evidence for macromolecular protein rings in the absence of bulk water.* Science, 2005. 310(5754): p. 1658-61.
111. Demmers, J.A., et al., *Interaction of the K+ channel KcsA with membrane phospholipids as studied by ESI mass spectrometry.* FEBS Lett, 2003. 541(1-3): p. 28-32.
112. le Coutre, J., et al., *Proteomics on full-length membrane proteins using mass spectrometry.* Biochemistry, 2000. 39(15): p. 4237-42.
113. Mitchell, P., et al., *The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'-->5' exoribonucleases.* Cell, 1997. 91(4): p. 457-66.
114. Allmang, C., et al., *The yeast exosome and human PM-Scl are related complexes of 3' --> 5' exonucleases.* Genes Dev, 1999. 13(16): p. 2148-58.
115. Wolfe, J.F., E. Adelstein, and G.C. Sharp, *Antinuclear antibody with distinct specificity for polymyositis.* J Clin Invest, 1977. 59(1): p. 176-8.
116. Liu, H., et al., *The scavenger mRNA decapping enzyme DcpS is a member of the HIT family of pyrophosphatases.* Embo J, 2002. 21(17): p. 4699-708.
117. Andrulis, E.D., et al., *The RNA processing exosome is linked to elongating RNA polymerase II in Drosophila.* Nature, 2002. 420(6917): p. 837-41.
118. Forler, D., et al., *An efficient protein complex purification method for functional proteomics in higher eukaryotes.* Nat Biotechnol, 2003. 21(1): p. 89-92.
119. Estevez, A.M., T. Kempf, and C. Clayton, *The exosome of Trypanosoma brucei.* Embo J, 2001. 20(14): p. 3831-9.
120. Chekanova, J.A., et al., *Poly(A) tail-dependent exonuclease AtRrp41p from Arabidopsis thaliana rescues 5.8 S rRNA processing and mRNA decay defects of the yeast ski6 mutant*

and is found in an exosome-sized complex in plant and yeast cells. J Biol Chem, 2000. 275(42): p. 33158-66.
121. Chekanova, J.A., et al., *Arabidopsis thaliana exosome subunit AtRrp4p is a hydrolytic 3'-->5' exonuclease containing S1 and KH RNA-binding domains.* Nucleic Acids Res, 2002. 30(3): p. 695-700.
122. Hooker, T.S., et al., *A core subunit of the RNA-processing/degrading exosome specifically influences cuticular wax biosynthesis in Arabidopsis.* Plant Cell, 2007. 19(3): p. 904-13.
123. Evguenieva-Hackenberg, E., et al., *An exosome-like complex in Sulfolobus solfataricus.* EMBO Rep, 2003. 4(9): p. 889-93.
124. Ramos, C.R., et al., *The Pyrococcus exosome complex: structural and functional characterization.* J Biol Chem, 2006. 281(10): p. 6751-9.
125. Mitchell, P. and D. Tollervey, *Musing on the structural organization of the exosome complex.* Nat Struct Biol, 2000. 7(10): p. 843-6.
126. Venema, J. and D. Tollervey, *Processing of pre-ribosomal RNA in Saccharomyces cerevisiae.* Yeast, 1995. 11(16): p. 1629-50.
127. Gatfield, D. and E. Izaurralde, *Nonsense-mediated messenger RNA decay is initiated by endonucleolytic cleavage in Drosophila.* Nature, 2004. 429(6991): p. 575-8.
128. Houseley, J., J. LaCava, and D. Tollervey, *RNA-quality control by the exosome.* Nat Rev Mol Cell Biol, 2006. 7(7): p. 529-39.
129. Hilleren, P., et al., *Quality control of mRNA 3'-end processing is linked to the nuclear exosome.* Nature, 2001. 413(6855): p. 538-42.
130. Chen, J., et al., *Purification and characterization of the 1.0 MDa CCR4-NOT complex identifies two novel components of the complex.* J Mol Biol, 2001. 314(4): p. 683-94.
131. Collart, M.A., *Global control of gene expression in yeast by the Ccr4-Not complex.* Gene, 2003. 313: p. 1-16.
132. Denis, C.L. and J. Chen, *The CCR4-NOT complex plays diverse roles in mRNA metabolism.* Prog Nucleic Acid Res Mol Biol, 2003. 73: p. 221-50.
133. Coller, J. and R. Parker, *Eukaryotic mRNA decapping.* Annu Rev Biochem, 2004. 73: p. 861-90.
134. Daugeron, M.C., F. Mauxion, and B. Seraphin, *The yeast POP2 gene encodes a nuclease involved in mRNA deadenylation.* Nucleic Acids Res, 2001. 29(12): p. 2448-55.
135. Hilleren, P.J. and R. Parker, *Cytoplasmic degradation of splice-defective pre-mRNAs and intermediates.* Mol Cell, 2003. 12(6): p. 1453-65.
136. Chen, C.Y., et al., *AU binding proteins recruit the exosome to degrade ARE-containing mRNAs.* Cell, 2001. 107(4): p. 451-64.
137. Frischmeyer, P.A., et al., *An mRNA surveillance mechanism that eliminates transcripts lacking termination codons.* Science, 2002. 295(5563): p. 2258-61.
138. van Hoof, A., et al., *Exosome-mediated recognition and degradation of mRNAs lacking a termination codon.* Science, 2002. 295(5563): p. 2262-4.
139. Doma, M.K. and R. Parker, *Revenge of the NRD: preferential degradation of nonfunctional eukaryotic rRNA.* Dev Cell, 2006. 11(6): p. 757-8.
140. Doma, M.K. and R. Parker, *Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation.* Nature, 2006. 440(7083): p. 561-4.
141. Bashkirov, V.I., et al., *A mouse cytoplasmic exoribonuclease (mXRN1p) with preference for G4 tetraplex substrates.* J Cell Biol, 1997. 136(4): p. 761-73.
142. Ingelfinger, D., et al., *The human LSm1-7 proteins colocalize with the mRNA-degrading enzymes Dcp1/2 and Xrnl in distinct cytoplasmic foci.* Rna, 2002. 8(12): p. 1489-501.
143. Lykke-Andersen, J., *Identification of a human decapping complex associated with hUpf proteins in nonsense-mediated decay.* Mol Cell Biol, 2002. 22(23): p. 8114-21.
144. van Dijk, E., et al., *Human Dcp2: a catalytically active mRNA decapping enzyme located in specific cytoplasmic structures.* Embo J, 2002. 21(24): p. 6915-24.
145. Parker, R. and U. Sheth, *P bodies and the control of mRNA translation and degradation.* Mol Cell, 2007. 25(5): p. 635-46.
146. Sheth, U. and R. Parker, *Decapping and decay of messenger RNA occur in cytoplasmic processing bodies.* Science, 2003. 300(5620): p. 805-8.
147. Teixeira, D., et al., *Processing bodies require RNA for assembly and contain nontranslating mRNAs.* Rna, 2005. 11(4): p. 371-82.
148. Teixeira, D. and R. Parker, *Analysis of P-body assembly in Saccharomyces cerevisiae.* Mol Biol Cell, 2007. 18(6): p. 2274-87.

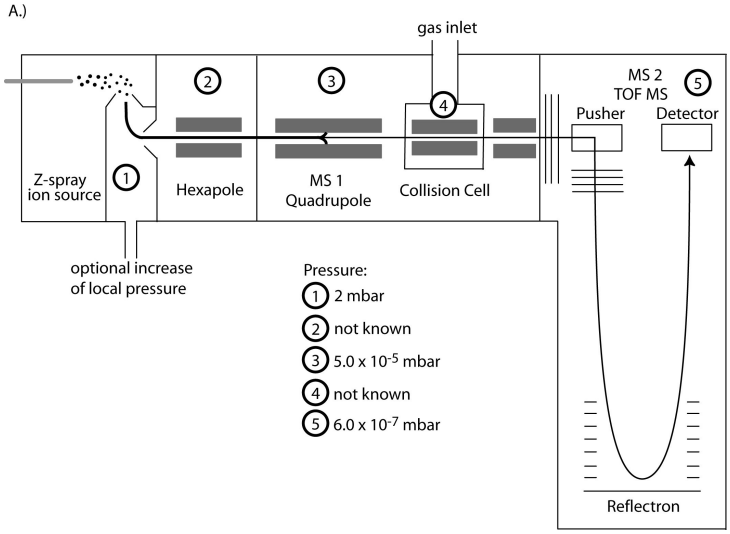
149. Lin-Chao, S., N.T. Chiou, and G. Schuster, *The PNPase, exosome and RNA helicases as the building components of evolutionarily-conserved RNA degradation machines.* J Biomed Sci, 2007. 14(4): p. 523-32.
150. Symmons, M.F., G.H. Jones, and B.F. Luisi, *A duplicated fold is the structural basis for polynucleotide phosphorylase catalytic activity, processivity, and regulation.* Structure, 2000. 8(11): p. 1215-26.
151. Koonin, E.V., Y.I. Wolf, and L. Aravind, *Prediction of the archaeal exosome and its connections with the proteasome and the translation and transcription machineries by a comparative-genomic approach.* Genome Res, 2001. 11(2): p. 240-52.
152. Lorentzen, E., et al., *The archaeal exosome core is a hexameric ring structure with three catalytic subunits.* Nat Struct Mol Biol, 2005. 12(7): p. 575-81.
153. Buttner, K., K. Wenig, and K.P. Hopfner, *Structural framework for the mechanism of archaeal exosomes in RNA processing.* Mol Cell, 2005. 20(3): p. 461-71.
154. Graham, A.C., D.L. Kiss, and E.D. Andrulis, *Differential distribution of exosome subunits at the nuclear lamina and in cytoplasmic foci.* Mol Biol Cell, 2006. 17(3): p. 1399-409.
155. Noguchi, E., et al., *Dis3, implicated in mitotic control, binds directly to Ran and enhances the GEF activity of RCC1.* Embo J, 1996. 15(20): p. 5595-605.
156. Stead, J.A., et al., *The PMC2NT domain of the catalytic exosome subunit Rrp6p provides the interface for binding with its cofactor Rrp47p, a nucleic acid-binding protein.* Nucleic Acids Res, 2007. 35(16): p. 5556-67.
157. Aloy, P., et al., *A complex prediction: three-dimensional model of the yeast exosome.* EMBO Rep, 2002. 3(7): p. 628-35.
158. Oliveira, C.C., F.A. Gonzales, and N.I. Zanchin, *Temperature-sensitive mutants of the exosome subunit Rrp43p show a deficiency in mRNA degradation and no longer interact with the exosome.* Nucleic Acids Res, 2002. 30(19): p. 4186-98.
159. Raijmakers, R., et al., *Protein-protein interactions between human exosome components support the assembly of RNase PH-type subunits into a six-membered PNPase-like ring.* J Mol Biol, 2002. 323(4): p. 653-63.
160. Liu, Q., J.C. Greimann, and C.D. Lima, *Reconstitution, activities, and structure of the eukaryotic RNA exosome.* Cell, 2006. 127(6): p. 1223-37.
161. Hernandez, H., et al., *Subunit architecture of multimeric complexes isolated directly from cells.* EMBO Rep, 2006. 7(6): p. 605-10.
162. Wang, H.W., et al., *Architecture of the yeast Rrp44 exosome complex suggests routes of RNA recruitment for 3' end processing.* Proc Natl Acad Sci U S A, 2007. 104(43): p. 16844-9.
163. Luz, J.S., et al., *Analysis of the Saccharomyces cerevisiae exosome architecture and of the RNA binding activity of Rrp40p.* Biochimie, 2007. 89(5): p. 686-91.
164. Estevez, A.M., et al., *The roles of intersubunit interactions in exosome stability.* J Biol Chem, 2003. 278(37): p. 34943-51.
165. Lorentzen, E. and E. Conti, *The exosome and the proteasome: nano-compartments for degradation.* Cell, 2006. 125(4): p. 651-4.
166. Lorentzen, E., et al., *RNA channelling by the archaeal exosome.* EMBO Rep, 2007. 8(5): p. 470-6.
167. Symmons, M.F., et al., *Running rings around RNA: a superfamily of phosphate-dependent RNases.* Trends Biochem Sci, 2002. 27(1): p. 11-8.
168. Maquat, L.E., *Molecular biology. Skiing toward nonstop mRNA decay.* Science, 2002. 295(5563): p. 2221-2.
169. Toh, E.A., P. Guerry, and R.B. Wickner, *Chromosomal superkiller mutants of Saccharomyces cerevisiae.* J Bacteriol, 1978. 136(3): p. 1002-7.
170. Ridley, S.P., S.S. Sommer, and R.B. Wickner, *Superkiller mutations in Saccharomyces cerevisiae suppress exclusion of M2 double-stranded RNA by L-A-HN and confer cold sensitivity in the presence of M and L-A-HN.* Mol Cell Biol, 1984. 4(4): p. 761-70.
171. Anderson, J.S. and R.P. Parker, *The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI2 DEVH box protein and 3' to 5' exonucleases of the exosome complex.* Embo J, 1998. 17(5): p. 1497-506.
172. Brown, J.T., X. Bai, and A.W. Johnson, *The yeast antiviral proteins Ski2p, Ski3p, and Ski8p exist as a complex in vivo.* Rna, 2000. 6(3): p. 449-57.
173. D'Andrea, L.D. and L. Regan, *TPR proteins: the versatile helix.* Trends Biochem Sci, 2003. 28(12): p. 655-62.

174. Sikorski, R.S., et al., *A repeating amino acid motif in CDC23 defines a family of proteins and a new relationship among genes required for mitosis and RNA synthesis.* Cell, 1990. 60(2): p. 307-17.
175. Dangel, A.W., et al., *Human helicase gene SKI2W in the HLA class III region exhibits striking structural similarities to the yeast antiviral gene SKI2 and to the human gene KIAA0052: emergence of a new gene family.* Nucleic Acids Res, 1995. 23(12): p. 2120-6.
176. Widner, W.R. and R.B. Wickner, *Evidence that the SKI antiviral system of Saccharomyces cerevisiae acts by blocking expression of viral mRNA.* Mol Cell Biol, 1993. 13(7): p. 4331-41.
177. Neer, E.J., et al., *The ancient regulatory-protein family of WD-repeat proteins.* Nature, 1994. 371(6495): p. 297-300.
178. Smith, T.F., et al., *The WD repeat: a common architecture for diverse functions.* Trends Biochem Sci, 1999. 24(5): p. 181-5.
179. Madrona, A.Y. and D.K. Wilson, *The structure of Ski8p, a protein regulating mRNA degradation: Implications for WD protein structure.* Protein Sci, 2004. 13(6): p. 1557-65.
180. Cheng, Z., et al., *Crystal structure of Ski8p, a WD-repeat protein with dual roles in mRNA metabolism and meiotic recombination.* Protein Sci, 2004. 13(10): p. 2673-84.
181. Orlicky, S., et al., *Structural basis for phosphodependent substrate selection and orientation by the SCFCdc4 ubiquitin ligase.* Cell, 2003. 112(2): p. 243-56.
182. van der Voorn, L. and H.L. Ploegh, *The WD-40 repeat.* FEBS Lett, 1992. 307(2): p. 131-4.
183. Tesse, S., et al., *Localization and roles of Ski8p protein in Sordaria meiosis and delineation of three mechanistically distinct steps of meiotic homolog juxtaposition.* Proc Natl Acad Sci U S A, 2003. 100(22): p. 12865-70.
184. Arora, C., et al., *Antiviral protein Ski8 is a direct partner of Spo11 in meiotic DNA break formation, independent of its cytoplasmic role in RNA metabolism.* Mol Cell, 2004. 13(4): p. 549-59.
185. Lambright, D.G., et al., *The 2.0 A crystal structure of a heterotrimeric G protein.* Nature, 1996. 379(6563): p. 311-9.
186. Sondek, J., et al., *Crystal structure of a G-protein beta gamma dimer at 2.1A resolution.* Nature, 1996. 379(6563): p. 369-74.
187. Neer, E.J. and T.F. Smith, *G protein heterodimers: new structures propel new questions.* Cell, 1996. 84(2): p. 175-8.
188. Araki, Y., et al., *Ski7p G protein interacts with the exosome and the Ski complex for 3'-to-5' mRNA decay in yeast.* Embo J, 2001. 20(17): p. 4684-93.
189. Wang, L., M.S. Lewis, and A.W. Johnson, *Domain interactions within the Ski2/3/8 complex and between the Ski complex and Ski7p.* Rna, 2005. 11(8): p. 1291-302.
190. LaCava, J., et al., *RNA degradation by the exosome is promoted by a nuclear polyadenylation complex.* Cell, 2005. 121(5): p. 713-24.

Chapter 2

Novel native tandem mass spectrometry technique: dissection of protein complexes in the gas phase

Multi-protein complexes provide an important level of cell organization and regulate vital cellular functions. The structural investigation of protein complexes is important as it helps in the understanding of biological processes in the cell. However, the analysis of protein complexes is challenging and far from routine analyses. Native mass spectrometry is a new player in studying protein machineries as it allows the investigation of structural features of protein complexes in the gas phase (see chapter 1 and [1-9]). This technique gives valuable information about the total mass of protein complexes and sub-complexes generated in solution. While most native mass spectrometry studies aim at preserving non-covalent interactions, latest developments allow the controlled dissociation of protein assemblies in the gas phase using collision induced dissociation (CID) [10-14]. Native tandem mass spectrometry is particularly useful to dissect heterogeneous multi-protein assemblies to identify building blocks and sub-complexes of protein complexes. This methodology can probe the oligomeric state and overall topology and gives insight into the physical binding strength of proteins in the gas phase. A typical MS/MS experiment is performed by selecting a precursor of a certain m/z value in the first mass analyzer that is subsequently dissociated in the gas phase by collisions with inert rare gas molecule. The resulting fragment ions are then separated in the second mass analyzer prior to detection. In a QToF instrument the first mass analyzer is a quadrupole and the second a time of flight. These two mass analyzers operate consecutively and are only separated by a hexapole collision cell (fig. 1A). Conventional QToF mass spectrometers are used for the analysis of peptides and other small molecules. By fragmenting the peptides, amino acid sequence information can be deduced. In order to make this instrument amenable for the analysis of large non-covalent protein assemblies, several instrument adjustments have to be implemented (fig 1B).



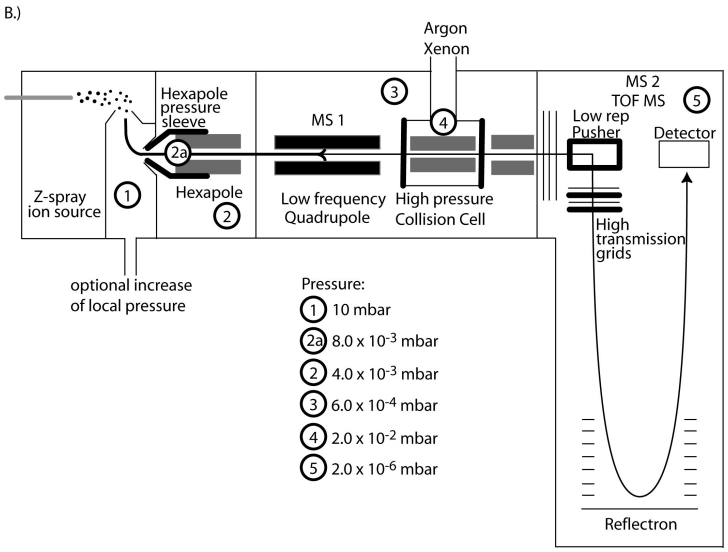


Figure 1

Schematic representation of a Quadrupole-Time of Flight mass spectrometer (QToF). This instrument is comprised of an electrospray ion source, a first hexapole ion guide, a quadrupole analyzer, a hexapole collision cell, a second hexapole ion guide and an orthogonal time of flight (ToF) analyzer. A.) original and B.) modified QTof instrument. Components in bold black are customized for the analysis of large protein assemblies. There are five major adjustments implemented: a pressure sleeve around the first hexapole, a low frequency quadrupole, a high pressure collision cell, a low repetition pusher and high transmission grids. Corresponding to these modifications the pressure in all compartments of the mass spectrometer are optimized to enable better transmission of macromolecular non-covalent protein assemblies.

In general, mass spectrometry requires the generation, transmission, separation and detection of ions in order to investigate biological macromolecules. During the ionization process the analytes have to be transferred from solution to gas phase ions. During this procedure solvent molecules are stripped from the protein assembly. The QToF mass spectrometer is equipped with a Z-flow nano electrospray source. Nano-electrospray-ionization coupled to a QToF is the method of choice for the analysis of protein complexes. It produces multiple charged ions at flow rates between 10-25 nL/min and therefore consumes low amounts of sample. A voltage of 1 to 1.6 kV is applied on the gold-coated capillary. With the formation of smaller initial droplets, nano-ESI is more tolerant towards contaminating buffer salts and small liganding molecules that are disturbing the ionization efficiency but are sometimes needed for protein complex stability. The improved desolvation performance of the ions leads to sharper peaks in the mass spectrum. Due to a smaller diameter of initial droplets, less desolvation energy is required; hence the method has been

shown to be gentler, keeping non-covalent interactions intact and with less non-specific multimeric species formed such as artifactual oligomeric super-complexes of the protein assembly at high analyte concentrations in initial droplets. After ionization the ions have to be transmitted through a small inlet into the different vacuum stages of the mass spectrometer. Since multiply charged ions are produced at atmospheric pressure, the transfer into the mass spectrometer strongly affects the ion kinetics and consequently the non-covalent interactions. The ions are accelerated and obtain a high kinetic energy directly proportional to their charge. Therefore multiple charged protein complexes can initially attain 100 keV of kinetic energy which results in a high velocity of the molecules. For a successful mass detection the speed of the macromolecular ions need to be reduced. This can be achieved by carefully controlling the gas pressures within the mass spectrometer (figure 1b). An increase of pressure from originally 2 mbar up to 8-10 mbar in the region between sample cone and extraction cone tremendously improves high m/z ion transmission into the vacuum stages of the mass spectrometer [1, 15-17]. This phenomenon can be tentatively explained by numerous low-energy collisions between small buffer gas molecules such as air and large analyte ions upon the pressure increase. Small buffer gas ions perform mainly elastic or near-elastic collisions with the analyte that does not lead to fragmentation. Elastic collisions transfer only kinetic energy upon interaction and do not convert into internal energy. As the analyte ion carries a higher energy than the buffer gas molecules the excess kinetic energy from the analyte ion is absorbed by the buffer gas. The collisions alter the momentum of the analyte ion only slightly and are primarily transformed into rotational and vibrational energy. As a result the kinetic energy of the analyte ion is decreased and the ions become decelerated. In a second aspect, which is at least as important, it refocuses the ion beam so that a better transmission of the ions into the higher vacuum stages is ensured. Differences in ion energies are compensated so that all ions contain very similar energies including strongly scattering ions. The gas molecules can be envisioned as a continuous viscous medium that dampens the ion trajectories. The pressure rise in the front part (source) also slightly increases the pressure in the back part (ToF) of the mass spectrometer. This concept of lowering the kinetic energy by elevated pressures was also applied on the subsequent radio frequency (RF) ion guide towards the quadrupole. Ion guides are typically configured and mounted at the beginning and at the end of a vacuum stage. At the entrance orifice to the ion guide neutral background gas molecules collide against the analyte ions absorbing more of their kinetic energy. Thereby additional cooling and focusing of the analyte ions is achieved as they traverse along the ion guide. In the modified QToF an extra pressure sleeve is incorporated around the entrance of the first hexapole to enhance this local effect. These elevated pressures in the first vacuum stages can be accomplished by reducing the pumping speed or adding additional collision gas.

The analyte ion is transferred from the hexapole ion guide into the quadrupole. The quadrupole can function as a wide-band pass filter in the MS mode or scan for selected precursors on the basis of their m/z ratio in the MS/MS mode.
Our conventional quadrupole was operated at a frequency of 832 kHz and had a limited m/z range of up to 4190 mass units. The relation between these parameters and the maximal attainable m/z value for the quadrupole is given by:

$$M_{max} = \frac{7 \times 10^6 \, V_m}{f^2 r_0^2} \qquad (1)$$

with $V_m \cos(2\pi ft)$ as the RF voltage between adjacent rods of the quadrupole ($2V_m$ is the peak to peak amplitude; f is the frequency and t is the time) and r_0 is the radial inner-distance of the arranged rods assembly in meters [18].

This m/z range is not sufficient for most large non-covalent complexes as they often give rise to signals at higher m/z values. This restriction suggested modifying the quadrupole so that the m/z range is extended. In accordance with formula (1) theoretically all three parameters (V_m, f and r_0) could be changed. In practice only the frequency f can be modulated as V_m stabilizes the high voltage between the rods and r_0 defines the transmission of ions through the quadrupole and decrease of the rod inner-distance may lead surface contamination due to ions loss in the quadrupole. However by lowering the frequency of the quadrupole, an extension of the m/z range can be achieved. Therefore the RF power supply of the quadrupole is exchanged to operate at a frequency of 300 kHz which expands the m/z range up to 8fold to m/z values of 32,000. Due to the low frequency a loss in resolution is induced. Since two mass analyzers operate in sequence the loss in resolution does not hamper the overall analysis since the mass of fragment ions are measured in the ToF analyzer which recovers resolution. Moreover when a quadrupole is not in mass resolving mode, more ions can traverse the mass analyzer sometimes even leading to an increase in transmission. Taken together lowering the frequency of the quadrupole has two effects; it allows the selection of precursors with m/z values above 4000 and improves transmission of ions [10]. Following the quadrupole ion selection, the analyte ion is fragmented in the collision cell to yield further structural information.
In the collision cell, the precursor ions must be sufficiently activated prior to collision induced dissociation (CID). Ion activation is achieved by accelerating the analyte ion and multiply colliding it with inert gas molecules. In this process the time period of analyte ion to gas molecule interaction, given by velocity of the precursor ion, and the center-of-mass collisions are crucial in defining the impact towards fragmentation of the precursor ion [19].
It has been previously shown that a collision cell filled with gas at an elevated pressure of 10^{-2} torr improves ion fragmentation and transmission [20]. A higher gas pressure in the

collision cell can be realized by reducing the orifices diameter from a standard of 2.0 mm to 1.5 mm at the entrance and exit plates. Probably two physical processes take place during the collisions. Upon several mainly inelastic collisions with the inert gas molecules, in which the kinetic energy is converted into internal energy, the analyte ion's internal energy increases [21]. Near the exit of the collision cell this internal energy is too high to keep all non-covalent interactions intact, hence the weakest interaction will be disrupted and dissociation of a protein from the complex follows. The maximum increase in internal energy, ΔE_{int}, upon collisions with buffer gas molecules, can be estimated by [22-24]:

$$\Delta E_{int} = zV_a \, (1-[M_i^2 + M_g^2) / (M_i + M_g)^2]^n \qquad (2)$$

with z number of charges on average, V_a voltage acceleration, M_i molecular weight of analyte ion and M_g molecular weight of buffer gas molecule, n number of collisions

It becomes clear that protein dissociation in the gas phase is dependent on the number of collisions with the buffer gas molecules; hence increasing the pressure allows more collisions. Dissociation can be also enhanced by using different buffer gases in the collision cell. The instrument set up allows filling the collision cell with different gases. The choice of different collision gases (e.g. noble gases with increasing masses such as argon, krypton and xenon) has a direct influence on the fragmentation efficiency. The collision efficiency with heavier target gases increases and more energy can be converted into internal energy.
Moreover heavier gas atoms decelerate the analyte ion in the collision cell so that the analyte ion spends more time in the collision cell and can therefore undergo more collisions with the target gas. Hence, the heavier the gas molecules, the higher the impact on the analyte precursor ion and increased fragmentation can be observed [25].
Secondly in the subsequent collisions in the collision cell another effect dominates which refocuses the ions similar to the process at earlier vacuum stages. The ions obtain a mass independent low axial translational energy which enhances the transfer to the second analyzer. Heavier collision gases reduce the overall kinetic energy spread on the analyte ions for better transmission into the ToF analyzer.
Increasing collision energies are defined by elevating the voltage applied to the region between the first hexapole to the entrance of the collision cell. The impact of the increasing collision energy will initially remove attached buffer molecules from the analyte ion that are still present due to insufficient desolvation improving the overall peak shape in the mass spectrum corresponding to the protein complex; they become sharper. When the collision energy becomes too high, fragmentation will occur. Latest implementations in the modified QToF allow the increase of the collision energy from previously 200V to now 400V as maximum.

After successful collisions, the fragment ions are guided by an RF hexapole ion guide and through high transmission grids to the ToF analyzer. The ToF analyzer uses a pulsed ion beam that is ejected into a direction orthogonal to the quadrupole and the collision cell. The fragment ions exiting the collision cell are accelerated by the pusher, separated in the flight tube and finally detected on a microchannel plate detector (MCP). To achieve the maximum flight time, the pusher frequency is lowered from 255 µsec to repetitions of 415 µsec.

Case study

Implementation of tandem mass spectrometry in the characterization of protein complexes exemplified on a novel Dioxygenase

Introduction

Several aerobic microorganisms can metabolize acetophenones as sole carbon and energy source for growth [26-29]. These aromatic compounds are naturally occurring in lignin and other plant components but are also widely found in the environment as degradation products of industrial chemical compounds. The biodegradation of aromatic compounds is very important but so far the enzymes performing these reactions have been poorly characterized. *Pseudomonas fluorescens* ACB is a gram-negative bacterium that possesses several oxidative enzymes for the degradation of 4-hydroxyacetophenone. In initial steps 4-hydroxyacetophenone is converted in the Baeyer-Villiger oxidation to 4-hydroxy-phenyl acetate by 4-hydroxyacetophenone monooxygenase (HAPMO) that is encoded by the gene HapE. This ester is then readily hydrolyzed to hydroquinone by the enzyme encoded by HapD. The ring-cleavage of hydroquinone is the key-reaction in the process of degradation. It is converted to γ -hydroxy-
muconic semialdehyde by the incorporation of two oxygens. The enzyme that is responsible for this reaction is hydroquinone 1,2-dioxygenase (HQDO) [29-31]. γ -hydroxymuconic semialdehyde is further converted to maleylacetate and then to β-ketodiapate by enzymes encoded by HapA and HapB (fig. 2).

HQDO performs a key-reaction in the degradation of 4-hydroxyacetophenone and it is therefore of great interest to characterize this enzyme in more detail. The enzyme is activated by iron(II)-ions “in vitro” and can be stabilized by 4-hydroxybenzoate.

The structural characterization of HQDO isolated from *Pseudomonas fluorescens* ACB is demonstrated using the modified QToF instrument. The data provides insight into the stoichiometry of HQDO and reveals relative gas-phase binding affinities of subunits within the protein complex and suggests binding sites for 4-hydroxybenzoate.

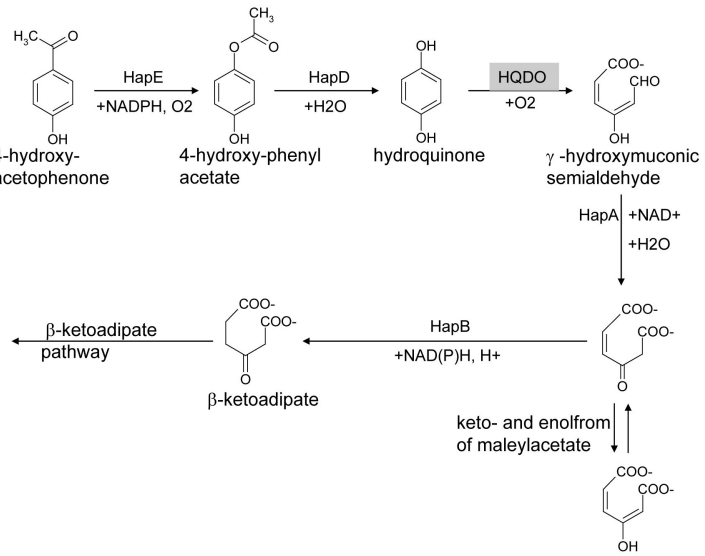


Figure 2

Microbial degradation pathway of 4-hydroxy-acetophenone to β-ketoadipate which is further decomposed via the β-ketoadipate pathway. Hydroquinone, an intermediate product, is converted to γ-hydroxymuconic semialdehyde by Hydroquinone-1,2-Dioxygenase (HQDO) in the presence of oxygen.

P. fluorescens ACB preparation and experimental methods

P. fluorescens ACB [29] was grown on 4-hydroxyacetophenone as described previously [30]. The purification procedure was exclusively performed at 4 °C. 5 g (wet weight) of cells were suspended in a 5 ml buffer containing 20 mM Tris-HCl, pH 7.0, 1 mM 4-hydroxybenzoate, 10 % (v/v) glycerol and 0.1 mM PMSF (buffer A). 1 mM EDTA, 2 mM $MgCl_2$ and 1 mg DNAse were added before disruption of cells in a French press (3 times at 10,000 psi) followed by centrifugation at 27,000 g for 30 min. The supernatent was adjusted to 25 % ammonium sulfate saturation and loaded onto a Phenyl Sepharose column (1.6 x 11 cm) that was equilibrated with buffer A, containing 25 % ammonium sulfate. The column was washed with 3 column-volumes buffer A including 25% ammonium sulfate. HQDO was then eluted in a 100 ml linear gradient of 25 - 0 % ammonium sulfate in buffer A. The eluent was precipitated with 60 % ammonium sulfate, centrifuged at 27,000 g for 30 min and resuspended in 1 ml buffer A. HQDO was further purified using a Superdex 75 PG column (2.6 x 54 cm) and subsequently a Resource 30Q column (1.6 x 5 cm) both equilibrated with buffer A containing 50 mM NaCl. The second column was washed with 3 column-volumes of buffer A plus 50 mM NaCl before HQDO was eluted in 180 ml linear gradient of buffer A containing 50 – 350 mM NaCl. HQDO fractions were combined, concentrated by ultrafiltration (Amicon Ultra-4, 10 kDa membrane) and stored at a concentration of 4.6 mg / ml in buffer containing 20 mM Tris-HCl, pH 7, 10% glycerol and 170 mM NaCl at -20C.

For nanoflow electrospray ionization mass spectrometry of HQDO, the storage buffer was exchanged to 300 mM ammonium acetate, pH 6.7 using a 100 kDa cut-off filter (Millipore). Protein samples were introduced into the nanoflow electrospray ionization source in Z-spray configuration of a Micromass QToF mass spectrometer (Waters), modified for high mass operation and operating in positive ion mode. Mass determinations were performed under conditions of increased pressure to 10 mbar in the source and to 2 x 10^{-2} mbar in the collision cell of the mass spectrometer. HQDO was infused in the mass spectrometer by using in-house pulled and gold-coated borosilicate needles (Kwik-Fil, World Precision Instruments). Borosilicate capillaries were pulled on a P-97 puller (Sutter Instruments) to prepare needles with an orifice of about 5 µM and coated with a thin gold layer (~500 Å) using an Edwards Scancoat six Pirani 501 sputter coater (Edwards High Vacuum International). Electrospray voltages were optimized for transmission of HQDO (capillary voltage 1400 V, sample cone voltage 50 V). In collision induced dissociation experiments HQDO was subjected to collisions with xenon with varied collision voltages between 10 V to 175 V to trigger dissociation of the complex. HQDO was also analyzed in the presence of 4-hydroxybenzoate by nanoflow-electrospray MS(/MS) to determine which subunit binds 4-hydroxybenzoate.

Results and Discussion

HQDO could only be detected with the QToF instrument at elevated pressures of 10 mbar at the entrance of the mass spectrometer and increased pressures to 2 x 10^{-2} mbar in the collision cell. When the QToF is operated in the MS mode the total mass of the complex can be determined. The mass spectra showed a main ion distribution at m/z 4,900 to 5,700 with charges ranging from 20+ to 23+ (fig. 3).

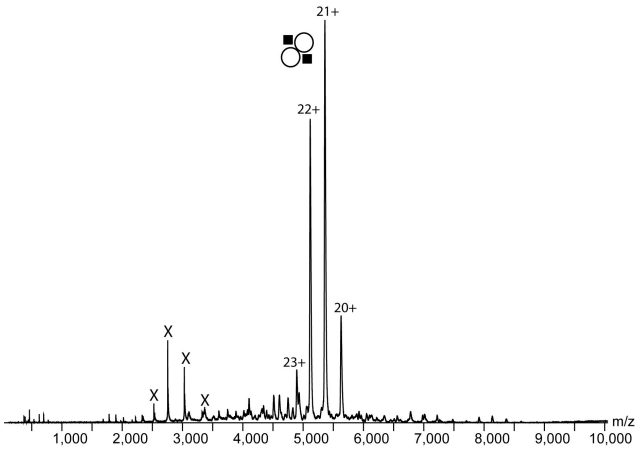


Figure 3

Nano-electrospray MS spectrum of 5 µM HQDO in 300 mM ammonium acetate solution, pH 6.8. The spectrum shows two charge envelopes. The main charge distribution at m/z 4,900 to 5,700 with charges from 20+ to 23+ corresponds to a protein complex of 112 kDa which corresponds to the intact HQDO protein complex. The second charge distribution at m/z 2,500 to 3,400 is likely the translation elongation factor Ts.

Mass calculations revealed a molecular weight of 112 kDa. A second charge envelope at m/z 2,500 to 3,500 denoted with X corresponds to a protein with a mass of 30.3 kDa. Previous N-terminal sequencing revealed this protein to be translation elongation factor Ts and therefore seems to be an impurity in the purification procedure (data not shown). Additionally to the 30.3 kDa protein, previous MS analyses under harsh conditions disrupting the protein complex provided masses for two extra proteins: 17.8 kDa and 38.3 kDa (resuts not shown). To probe the macromolecular arrangement of the protein assembly, the 21+ charge state at m/z 5,350 was isolated in the modified QToF for collision induced dissociation (fig 4). The collision voltage was increased from 10 to 175 V.

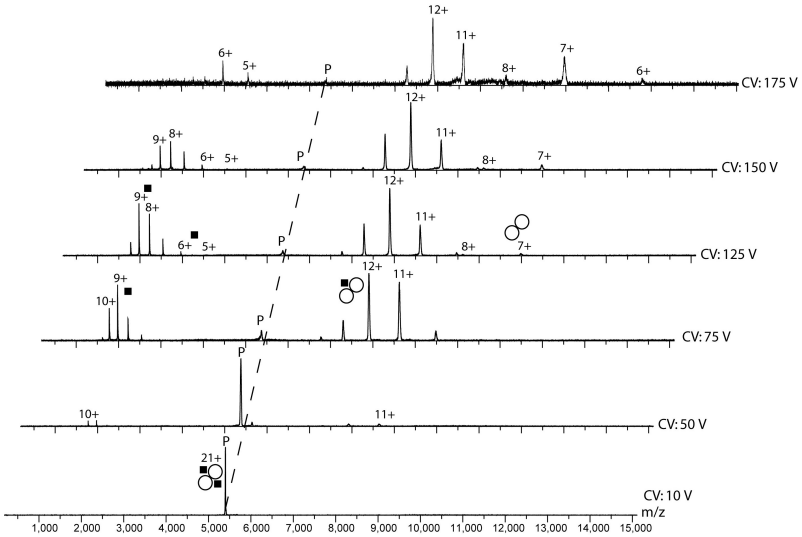


Figure 4

Gas phase dissociation of HQDO (5 µM in 300 mM ammonium acetate solution, pH 6.8) performed at increasing collision voltages (CV) ranging from 10 to 175 V. For this MS/MS experiment, the 21+ charge state of heterotetrameric HQDO was selected as precursor, denoted with (P) and subjected to collision induced dissociation (CID). At a collision voltage of at least 50 V, two fragment ions were produced that were detected at m/z 1,700 to 2,600 with 7+ to 11+ charges and at m/z 6,800 to 9,500 with charges from 10+ to 14+ representing an ejected monomer and a trimeric subcomplex. At a higher collision voltage of 125 V a second fragmentation pair appears at m/z 9,500 to 12,800 with charges 6+ to 8+ and at m/z 2,900 to 3,500 with 5+ and 6+ charges that correspond to a dimeric subcomplex and a second ejected monomer.

At 10 V only the selected heterotetrameric precursor is transmitted without any parent ion dissociation. Fragmentation starts at a collision voltage of 50 V and is enhanced by the subsequent increased collision voltages. While the intensity of the precursor decreases, two charge series appear at m/z 1,700 to 2,600 with charges ranging from 7+ to 11+ and at m/z 6,800 to 9,500 with charges from 10+ to 14 +. The charge envelope in the low m/z region is calculated to 17.8 kDa (α-subunit) and the charge series in the high m/z area corresponds to the counter-ion of 94.5 kDa. The addition of charges from the most intense α-monomer and trimer $\alpha\beta_2$ (12 + and 9+) comprises together the charge of the precursor ion (21+). The successful elimination of an α-subunit protein carries along 9 to 10 of the total 21 charges when it is dissociated. At higher collision voltages (125 to 175 V) another charge envelope in the high m/z area appears. This charge series at m/z 9,500 to 12,800 contains six to eight charges and mass calculation results in 76.8 kDa, which fits nicely to a dimer of the identified protein with a monomeric mass of 38.3 kDa (β-subunit). In the low m/z area this dissociation event gives rise to 5+ and 6+ charge states at m/z 2,950 and 3,560 that correlates to the loss of a second α-subunit. The sum of the charges states from both counter ions (7+ plus 6+ and 8+ plus 5+) identifies the precursor be likely the 13+ charge state of high m/z dissociation fragment with a mass of 94.5 kDa that was produced in the first dissociation event. At a collision voltage of 175 V, peaks of the initially dissociated α-subunit with charges between 7+ and 11+ are strongly diminished. It is possible that at such high collision voltages the higher charged α-monomer is fragmented further into smaller peptides. These dissociation events are summarized in figure 5.

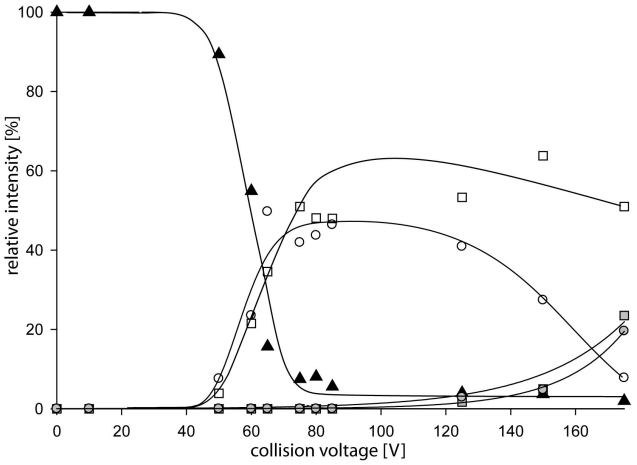


Figure 5

Dissociation curves of HQDO upon increasing collision voltages plotted against the relative intensity of the ions. The dissociation of the tetramer (▲) occurs at low collision voltage giving rise to a trimeric (□) and monomeric (○) fragment. At collision voltages exceeding 100 V, dimeric (■) and a second monomeric (●) fragment ion are observed.

The abundance of the tetramer (▲) decreases concomitant with the emergence of trimeric (□) and monomeric (○) fragment ions. At collision voltages above 100 V, the trimer further dissociates to a dimer (■) and a second monomer (●). The initial ejected monomer likely fragements into smaller peptides at this collision voltage.

The dissociation pathway of HQDO is summarized in figure 6.

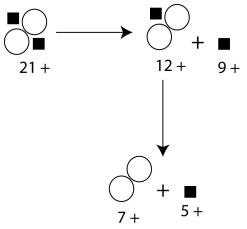


Figure 6

Schematic representation of the gas phase dissociation process of HQDO. Two α-monomers are sequentially ejected from the hetero-tetrameric protein assembly. The dissociation process results in an asymmetric charge partitioning between the produced fragment ions.

The asymmetric separation of mass and charge is frequently observed in gas phase dissociation events of non-covalent protein complexes, even on homodimers [13, 32-34]. It can be rationalized on the basis that the ejected monomer unfolds during the dissociation process and thereby acquires many charges [35, 36]. The currently predominant opinion on the mechanism of collisional induced dissociation of protein complexes is an even charge distribution of the protein assembly before dissociation. As the internal energy of the protein assembly increases upon gas collisions, the weakest bound protein locally starts unfolding. It exposes more surfaces, which are quickly occupied by mobile charge transfer to reduce coulombic repulsion forces on the total protein assembly. This process further destabilizes the partially unfolded protein until complete unfolding is reached and the protein dissociates from the protein complex. Rare examples have been reported that show different dissociation behaviour. For example the homotetramer 2-keto-3-arabinonate dehydratase dissociates mainly into two dimers with a symmetric charge partitioning [11]. Recently collision induced dissociation (CID) and surface induced dissociation (SID), two different activation mechanisms, were carefully examined on a dimer of cytochrome c [37]. While CID gradually disposes low activation energy into the protein complex on a longer time scale, SID deposits high activation in a single fast step. It could be illustrated that CID predominantly results in an asymmetric charge partition and SID forms mainly symmetric fragment ions, at least for this non-natural dimer of cytochrome c.
The collision-induced dissociation experiments performed on HQDO unambiguously define it to be a heterotetramer consisting of two α-subunits with a mass of 17.8 kDa and two β-

subunits with a mass of 38.3 kDa. Consequently the overall mass of HQDO adds up to 112 kDa. Both α-subunits sequentially dissociate from the HQDO complex with increasing collision energy.

Since ligandation with 4-hydroxybenzoate prevents irreversible inactivation of HQDO, native mass spectrometry was used to deduce the stoichiometry and binding site of this ligand in the protein assembly. Therefore HQDO was incubated with an excess of 4-hydroxybenzoate prior to mass spectrometric analysis. At a collision voltage of 50 V the HQDO heterotetramer was fragmented into an α-monomer and a $\alpha\beta_2$-trimer. Upon the addition of 4-hydroxybenzoate to HQDO a peak shift in the trimeric ions of HQDO ($\alpha\beta_2$) corresponding to a mass shift of 270 Da was observed (figure 7).

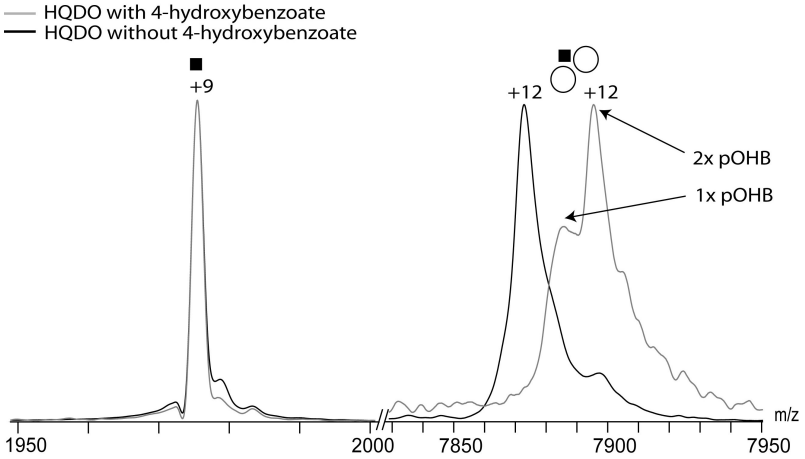


Figure 7

MS analysis of HQDO after the addition of 4-hydroxybenzoate at a collision voltage of 50 V. Overlay and zoom in to the 9+ charge state of the α-subunit and of the 12+ charge state of the trimeric subcomplex. In black the trace of HQDO without 4-hydrozybenzoate is seen whereas the grey trace represents HQDO after the addition of 4-hydrozybenzoate. The 12+ charge state of the trimeric subcomplex shows a considerable mass shift to higher m/z after the addition of 4-hydrozybenzoate whereas the 9+ charge state of the alpha subunit remains unchanged. The mass shift corresponds to a mass of 270 units strongly indicating that two molecules of 4-hydroxybenzoate are bound to the trimeric subcomplex.

This suggests that, under the experimental conditions applied, one 4-hydroxybenzoate molecule (molecular mass 138 Da) binds to one β-subunit. The expelled α-subunit ions stays unchanged upon binding of 4-hydroxybenzoate to the protein complex.

Conclusion

Native mass spectrometry and in particular native tandem mass spectrometry is an ideal tool to determine the stoichiometry of heterogeneous protein complexes. The modified

QToF instrument allows gas phase dissociation experiments of large protein complexes. The high impact of this instrument could be exemplified on a novel hydroquinone 1,2-dioxygenase isolated from *P. fluorescens* ACB. It was unambiguously shown that HQDO is a heterotetramer consisting of two α-subunit with a mass of 17.8 kDa and two β-subunits with a mass of 38.3 kDa comprising to a protein assembly with a total mass of 112.4 kDa. HQDO starts dissociation at a collision voltage of 50 V into α-monomer and the remaining trimeric complex. Further increase of the collision voltage leads to the subsequent release of a second α-monomer from the trimeric sub-complex with a counter ion consisting of a two β-subunits. The β-subunit binds to two molecules 4-hydrozybenzoate that can be easily dissociated at a collision voltage of 50 V.

References

1. van Berkel, W.J., et al., *Detection of intact megaDalton protein assemblies of vanillyl-alcohol oxidase by mass spectrometry*. Protein Sci, 2000. **9**(3): p. 435-9.
2. van den Heuvel, R.H. and A.J. Heck, *Native protein mass spectrometry: from intact oligomers to functional machineries.* Curr Opin Chem Biol, 2004. **8**(5): p. 519-26.
3. Benesch, J.L. and C.V. Robinson, *Mass spectrometry of macromolecular assemblies: preservation and dissociation.* Curr Opin Struct Biol, 2006. **16**(2): p. 245-51.
4. Bothner, B. and G. Siuzdak, *Electrospray ionization of a whole virus: analyzing mass, structure, and viability.* Chembiochem, 2004. **5**(3): p. 258-60.
5. Siuzdak, G., et al., *Mass spectrometry and viral analysis.* Chem Biol, 1996. **3**(1): p. 45-8.
6. Rostom, A.A., et al., *Detection and selective dissociation of intact ribosomes in a mass spectrometer.* Proc Natl Acad Sci U S A, 2000. **97**(10): p. 5185-90.
7. Loo, J.A., et al., *Electrospray ionization mass spectrometry and ion mobility analysis of the 20S proteasome complex.* J Am Soc Mass Spectrom, 2005. **16**(7): p. 998-1008.
8. Heck, A.J. and R.H. Van Den Heuvel, *Investigation of intact protein complexes by mass spectrometry.* Mass Spectrom Rev, 2004. **23**(5): p. 368-89.
9. Kennaway, C.K., et al., *Dodecameric structure of the small heat shock protein Acr1 from Mycobacterium tuberculosis.* J Biol Chem, 2005. **280**(39): p. 33419-25.
10. Sobott, F., et al., *A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies.* Anal Chem, 2002. **74**(6): p. 1402-7.
11. van den Heuvel, R.H., et al., *Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry.* Anal Chem, 2006. **78**(21): p. 7473-83.
12. McCammon, M.G., et al., *Tandem mass spectrometry defines the stoichiometry and quaternary structural arrangement of tryptophan molecules in the multiprotein complex TRAP.* J Am Chem Soc, 2004. **126**(19): p. 5950-1.
13. Benesch, J.L., et al., *Tandem mass spectrometry reveals the quaternary organization of macromolecular assemblies.* Chem Biol, 2006. **13**(6): p. 597-605.
14. Ilag, L.L., et al., *Heptameric (L12)6/L10 rather than canonical pentameric complexes are found by tandem MS of intact ribosomes from thermophilic bacteria.* Proc Natl Acad Sci U S A, 2005. **102**(23): p. 8192-7.
15. Krutchinsky, A., et al., *Collisional Damping Interface for an Electrospray Ionization Time-of-Flight Mass Spectrometer.* J Am Soc Mass Spectrom, 1998. **9**(6): p. 569-579.
16. Tahallah, N., et al., *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, 2001. **15**(8): p. 596-601.
17. Chernushevich, I.V. and B.A. Thomson, *Collisional cooling of large ions in electrospray mass spectrometry.* Anal Chem, 2004. **76**(6): p. 1754-60.
18. Austin, W., A. Holme, and J. Leck, *Quadrupole Mass Spectrometry and its Applications*. American Institute of Physics, 1st edition ed, ed. P. Dawson. 1995, Woodbury/New York.

19. Shukla, A.K. and J.H. Futrell, *Tandem mass spectrometry: dissociation of ions by collisional activation.* J Mass Spectrom, 2000. **35**(9): p. 1069-90.
20. Morris, M., P. Thibault, and R. Boyd, *Characterization of a High-Pressure Quadrupole Collision Cell for Low-Energy Collision-Induced Dissociation.* J Am Soc Mass Spectrom, 1994. **5**: p. 1042-1063.
21. Mauk, M.R., et al., *Tandem mass spectrometry of protein-protein complexes: cytochrome c-cytochrome b5.* J Am Soc Mass Spectrom, 2002. **13**(1): p. 59-71.
22. Benesch, J.L., et al., *Protein complexes in the gas phase: technology for structural genomics and proteomics.* Chem Rev, 2007. **107**(8): p. 3544-67.
23. Douglas, D.J., *Mechanism of the Collision-Induced Dissociation of Polyatomic Ions Studied by Triple Quadrupole Mass Spectrometry.* J Phys Chem, 1982. **86**(2): p. 185-191.
24. Gill, A., et al., *Conformations of bioploymers in the gas phase: a new mass spectrometric method.* International Journal of Mass Spectrometry, 2000. **195**: p. 685-697.
25. Lorenzen, K., et al., *Optimizing macromolecular tandem mass spectrometry of large non-covalent complexes using heavy collision gases.* International Journal of Mass Spectrometry, 2007. **268**(2-3): p. 198-206.
26. Cripps, R.E., *The microbial metabolism of acetophenone. Metabolism of acetophenone and some chloroacetophenones by an Arthrobacter species.* Biochem J, 1975. **152**(2): p. 233-41.
27. Cripps, R.E., P.W. Trudgill, and J.G. Whateley, *The metabolism of 1-phenylethanol and acetophenone by Nocardia T5 and an Arthrobacter species.* Eur J Biochem, 1978. **86**(1): p. 175-86.
28. Havel, J. and W. Reineke, *Microbial degradation of chlorinated acetophenones.* Appl Environ Microbiol, 1993. **59**(8): p. 2706-12.
29. Higson, F.K. and D.D. Focht, *Bacterial Degradation of Ring-Chlorinated Acetophenones.* Appl Environ Microbiol, 1990. **56**(12): p. 3678-3685.
30. Kamerbeek, N.M., et al., *4-Hydroxyacetophenone monooxygenase from Pseudomonas fluorescens ACB. A novel flavoprotein catalyzing Baeyer-Villiger oxidation of aromatic compounds.* Eur J Biochem, 2001. **268**(9): p. 2547-57.
31. Moonen, M.J., I.M. Rietjens, and W.J. van Berkel, *19F NMR study on the biological Baeyer-Villiger oxidation of acetophenones.* J Ind Microbiol Biotechnol, 2001. **26**(1-2): p. 35-42.
32. Versluis, C., et al., *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, 2001. **12**(3): p. 329-36.
33. Kaufman, S.L., et al., *Analysis of a 3.6-MDa hexagonal bilayer hemoglobin from Lumbricus terrestris using a gas-phase electrophoretic mobility molecular analyzer.* Anal Biochem, 1998. **259**(2): p. 195-202.
34. Jurchen, J.C., D.E. Garcia, and E.R. Williams, *Further studies on the origins of asymmetric charge partitioning in protein homodimers.* J Am Soc Mass Spectrom, 2004. **15**(10): p. 1408-15.
35. Felitsyn, N., E.N. Kitova, and J.S. Klassen, *Thermal dissociation of the protein homodimer ecotin in the gas phase.* J Am Soc Mass Spectrom, 2002. **13**(12): p. 1432-42.
36. Jurchen, J.C. and E.R. Williams, *Origin of asymmetric charge partitioning in the dissociation of gas-phase protein homodimers.* J Am Chem Soc, 2003. **125**(9): p. 2817-26.
37. Jones, C.M., et al., *Symmetrical gas-phase dissociation of noncovalent protein complexes via surface collisions.* J Am Chem Soc, 2006. **128**(47): p. 15044-5.

Chapter 3

Probing genuine strong interactions and post-translational modifications in the heterogeneous yeast exosome protein complex

Silvia A. Synowsky [1]
Robert H. H. van den Heuvel [3]
Shabaz Mohammed [1]
W. W. M. Pim Pijnappel[2]
Albert J. R. Heck[1]

[1]*Biomolecular Mass Spectrometry and Proteomics Group, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Sorbonnelaan 16, 3584 CA Utrecht, The Netherlands*

[2]*Department of Physiological Chemistry, University Medical Center Utrecht, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands*

[3]*current address: Organon Biosciences, Postbus 20, 5340 BH Oss, The Netherlands*

Mol Cell Proteomics. 2006 Sep;5(9):1581-92

Abstract

The characterization of heterogeneous multi-component protein complexes, which goes beyond identification of protein subunits, is a challenging task. Here we describe and apply a comprehensive method that combines a mild affinity purification procedure with a multiplexed mass spectrometry approach for the in-depth characterization of the exosome complex from *Saccharomyces cerevisiae* expressed at physiologically relevant levels. The exosome is an ensemble of primarily 3'->5' exoribonucleases and plays a major role in RNA metabolism. The complex has been reported to consist of 11 proteins, in molecular weight ranging from 20 to 120 kDa. By using native macromolecular mass spectrometry we measured accurate masses (around 400 kDa) of several (sub)-exosome complexes. Combination of these data with proteolytic peptide LC tandem mass spectrometry using a LTQ-FT-ICR and intact protein LC mass spectrometry provided us with the identity of the different exosome components and (sub)-complexes, including the subunit stoichiometry. We hypothesize that the observed complexes provide information about strong and weak interacting exosome-associated proteins. In our analysis we also identified for the first time phosphorylation sites in seven different exosome subunits. The phosphorylation site in the Rrp4 subunit is fully conserved in the human homologue of Rrp4, which is the only previously reported phosphorylation site in any of the human exosome proteins. The described multiplexed mass spectrometry-based procedure is generic and thus applicable to many different types of cellular molecular machineries, even if they are expressed at endogenous levels.

Introduction

One of the most intriguing views that have emerged out of large-scale proteome-wide analyses of protein-protein interaction in yeast and other organisms (1-6) is that most proteins do not “act on their own”. Consequently, it has been proposed that a cell may be better described as a network of interlocking assembly lines (3,7,8), each of which is composed of large protein machinery complexes, interacting with DNA, RNA, lipids, carbohydrates and other biomolecules. The components of these assemblies may vary over time as a function of the environment, induced by for instance protein post-translational modifications or signaling molecules. Up to now these large-scale studies of protein-protein interactions have contributed a quite static picture; the dynamic (spatial or temporal) nature has been less reported. It is now accepted that protein complexes are composed of “core” highly co-expressed and tightly interacting components that are decorated by transcriptionally and differentially regulated proteins. The monitoring of the dynamic and temporal assembly/disassembly or the recruitment of specific components of protein complexes requires their isolation from their physiological environment. In that respect, affinity purification coupled to mass spectrometry has proven to be a very powerful approach (9,10); it allows the analysis of protein complexes, which are expressed at physiological levels from endogenous promoters and are assembled *in vivo* (3,6,11).

Besides the great benefits in large-scale analysis of protein networks, the affinity purification coupled to mass spectrometry strategy has also some disadvantages. On the one hand it allows only the pull-down of proteins to the bait-protein that have a reasonable strong physical interaction with the bait, on the other hand the strategy also leads to the non-specific binding of proteins, which for instance bind strongly to the beads or are highly abundant in the cells studied. Additionally, unknowingly and unintentionally, a mixture of different complexes may be isolated at the same time. Finally, the conventional mass spectrometry approach does not provide information about stoichiometry, dynamics, sub-complexes and three-dimensional structure of the protein complex. Therefore, results from affinity purified pull-downs still need to be validated by other methods (3,12). Ideally, one would like to determine the structure of these complexes by high-resolution structural biology approaches, such as electron microscopy (13,14), NMR (15,16) and X-ray crystallography (17,18), which provide supreme detail on molecular structure. Although, these three techniques have become to a different degree amendable to larger proteins and even protein complexes they are still somewhat limited in their applications into very large and/or very heterogeneous protein complexes.

In recent years it has become apparent that the gentle nature of electrospray ionization enables the analysis of intact non-covalent structures, often referred to as native or macromolecular mass spectrometry (macromolecular MS). For this method, biomolecules are directly electrosprayed from aqueous solutions kept at physiological relevant pH

conditions. Coupling of electrospray ionization with time-of-flight mass analysis has greatly increased the mass-to-charge (m/z) range attainable (19), and has thus extended the realm of mass spectrometry also to the field of macromolecular non-covalent complexes such as protein oligomers, chaperone machineries, small viruses and even bacterial ribosome complexes (20-27). With these capabilities mass spectrometry has now been used to analyze quaternary structures and changes therein that occur upon binding of cofactors, metal ions, nucleotides, ligands etc., information which is essential for understanding the cellular functions of protein machineries. Here we report on the use of macromolecular MS to characterize the genuine stronger physical interactions and subunit stoichiometry of the exosome complex from *Saccharomyces cerevisiae*, which was expressed at endogenous levels and purified using the tandem affinity purification (TAP) procedure (9,10). In combination with denaturant gel liquid chromatography tandem mass spectrometry (1D gel LC MS/MS) and intact protein chromatography mass spectrometry (LC MS) our data allowed a comprehensive analysis of the exosome.

The exosome is a conserved multi-protein complex that functions in both processing of 3' extended precursor molecules to mature stable RNAs and complete degradation of other RNAs. The complex was originally discovered in *S. cerevisiae*, but has more recently also been identified in humans, plants, parasites, flies and archaea (28). The archetypal eukaryotic exosome from *S. cerevisiae* consists of nine core components. Six proteins are homologous to bacterial RNase PH (Rrp41, Rrp42, Rrp43, Rrp45, Rrp46 and Mtr3) and three contain a putative S1 RNA-binding domain (Rrp4, Rrp40 and Csl4) (28-31). Additional components include the RNase D homologue Rrp6 (32), which is only found in the nuclear exosome, and the RNase R homologue Dis3 (31). Four of the exosome proteins (Rrp4, Rrp41, Dis3 and Rrp6) have been demonstrated to have 3'->5' exoribonuclease activity *in vitro*. In recent years a number of exosome-associated proteins have been identified (Lrp1, Ski7, Ski2, Ski3, Ski8, Mtr4, Gsp1 and Nip7), which probably participate in the regulation and coordination of the exosome activity in different sub-cellular compartments (28). There is evidence that the six RNase PH-type proteins of the exosome ensemble form a ring-shaped structure with the three S1 RNA-binding domain containing proteins binding on top of this ring (33-36). The recent determination of the X-ray structure of an archaeal exosome, consisting of only Rrp41 and Rrp42 components, confirmed this ring-like arrangement of the RNase PH proteins (37). However, the mode of interaction of the exosome-associated proteins has not been identified yet.

Here we set out to further define and analyze the components, protein stoichiometry, post-translational modifications and relative strength of physical interactions of the exosome complexes using a generic multiplexed mass spectrometry approach. First, we analyzed the exosome-associated proteins by 1D gel LC MS/MS using a linear ion trap fourier transform ion cyclotron resonance mass spectrometer (LTQ-FT-ICR). Second, we analyzed the intact individual protein components protein LC MS. Third, we measured accurate masses of

(sub)-exosome complexes by macromolecular MS. Our analysis also revealed for the first time phosphorylation sites in yeast exosome subunits. We compared our findings with suggested models for the exosome structure based on biochemical essays, homology modelling and electron microscopy.

Experimental Procedures

S. cerevisiae strain, cultivation and protein purification

The *S. cerevisiae* strain MGD35313D, BSY17 containing Csl4 or Rrp42 as the C-terminal tagged entry point was kindly provided by Cellzome AG (Heidelberg, Germany) (3). 2 L of cell culture of *S. cerevisiae* was grown at 30 ºC in yeast extract-peptone-dextrose medium to an optimal density at 600 nm of 3.8. The cell pellets were lysed mechanically with glass beads resulting in 25 mL cell lysate. TAP purifications were performed essentially as described previously (9,10). In the first affinity purification step, 10 mL cell lysate (~200 µg/ml total protein) in 50 mM Tris/hydrochloride buffer, pH 7.5 containing 100 mM NaCl, 1.5 mM $MgCl_2$, 1 mM dithiotreitol and 0.15% (v/v) NP40 was mixed with 200 µL IgG beads (Amersham Biosciences, Sweden) and incubated for 150 min at 4 ºC. The protein complex was eluted from the IgG beads by incubation with 100 U TEV protease (Invitrogen, CA, USA) for 90 min at room temperature. In the second affinity purification step the eluted fraction was incubated with 200 µL calmodulin beads (Stratagene, CA, USA) in the presence of 2 mM calcium chloride for 60 min at 4 ºC. The protein complex was then eluted from the calmodulin beads using phosphate buffered saline buffer in the presence of 5 mM EGTA.

Protein separation, in gel-digestion and LTQ-FT-ICR analysis

Exosome complex concentrations were determined by Bio-Rad DC protein assay (Bio-Rad, Germany). Exosome (5-20 µg; 11.5-46 pmol) components were resolved by one-dimensional polyacrylamide gel electrophoresis using a 12.5% (used entry point Csl4) and 4-12% (used entry point Rrp42) Tris gel. Gels were subsequently stained with Pro-Q Diamond Phosphoprotein gel stain (Invitrogen, The Netherlands) and 0.1% (v/v) Coomassie brilliant blue G250 for 3 min and destained overnight in 10% (v/v) acetic acid and 30% (v/v) methanol. Gel lanes were excised and used to prepare 10 independent samples. Prior to digestion the samples were reduced and alkylated followed by in-gel trypsin (10 ng/µL trypsin) (Roche, The Netherlands) and a combination of trypsin (10 ng/µL) and endoproteinase Glu-C (20 ng/µL) (Roche, The Netherlands) digests for each gel as described previously (38) for 8 hrs at 37 ºC. The digestion was stopped by the addition of 2.5% (v/v) formic acid. LC MS/MS was performed using an LTQ-FT-ICR (Thermoelectron, Bremen, Germany) coupled to an Agilent 1100 Series LC system (vacuum degasser, auto

sampler, and one high-pressure mixing binary pump without static mixer). Peptide mixtures were delivered to a trap column (Reprosil C18AQ; 20 mm x 100 µm, packed in-house) at 5 µl/min 100% eluent A (0.1 M acetic acid). After reducing the flow to 100 nl/min by using a splitter, the peptides were transferred to the analytical column (Reprosil C18AQ 200 mm x 50 µm, packed in-house) with a linear gradient from 0 to 50% eluent B (0.1 M acetic acid in 80% (v/v) acetonitrile) for 60 min. The LTQ-FT-ICR was operated in positive ion mode. With one FT scan, three MS/MS scans were acquired. Peak lists were created using dta files which were combined to mgf files (mgf, Mascot generic format files). Peak intensities below 25 were considered as noise and were therefore removed. The Mascot algorithm (39) (Matrix Science Ltd., version 2.1) was used to interpret the processed raw MS/MS data (mgf files). Initially the data were run against the complete non-redundant proteome database Swiss Prot (search date: 23.08.05) and NCBI (search date: 19.10.05) in the FASTA format of *S. cerevisiae*. The algorithm was set to use trypsin as enzyme, allowing at maximum for two missed cleavages and assuming carbamidomethyl on cysteines as a fixed modification and oxidized methionine, N-terminal acetylation, acetylated lysine and phosphorylated serine/tyrosine/threonine as variable modifications. The peptide tolerance was fixed to 5 ppm and the MS/MS tolerance to 0.9 Da. Individual ion scores of > 25 (Trypsin, NCBInr) and > 22, (Trypsin, Swiss Prot) indicated identity or extensive homology ($p<0.05$). The data was then subjected to a homemade database consisting of exosome proteins previously identified in Swiss Prot and NCBInr (Dis3 (Q08162), Rrp43 (P25359), Rrp4 (P38792), Rrp45 (Q05636), Csl4 (P53859), Rrp42 (Q12277), Mtr3 (P48240), Rrp41 (P46948), Rrp40 (Q08285), Rrp46 (gi 37362654), Rrp6 (Q12149), Lrp1 (gi 6321873), Ski7 (gi 6324650), Ski2 (P35207), Ski3 (P17883), Ski8 (Q02793), Mtr4 (P47047)). The Mascot algorithm was set to use trypsin and endoproteinase Glu-C as enzymes, allowing at maximum for nine missed cleavages and assuming carbamidomethyl, oxidized methionine, N-terminal acetylation, and phosphorylated serine/tyrosine/threonine as variable modifications. The peptide tolerance was fixed to 10 ppm and the MS/MS tolerance to 0.9 Da. Individual ion scores of > 5 indicated identity or extensive homology ($p<0.05$), in reality a minimum score of 20 was used as threshold since the calculated confidence was based on such a small protein database. The Mascot score and sequence coverage were used as an indication for the relative abundance of a protein. Primary sequence alignment of yeast exosome proteins with their human homologues was performed using CLUSTAL W, version 1.82 (EMBL-EBI, UK) (40).

Intact protein LC MS analysis

Before sample injection intact exosome complex (23 pmol) was subjected to 0.5 % (v/v) formic acid. Protein chromatography was performed using an adapted Agilent 1100 Series LC system (vacuum degasser, auto sampler, and one high-pressure mixing binary pump) (Agilent, Palo Alto, CA, USA). Proteins were delivered to a trap column (Poros10 R2; 19 mm

x 150 µm, 10 mm particle size (Applied Biosystems, Framingham, MA); packed in-house) at 5 µl/min 100% eluent A (0.05% (v/v) trifluoro acetic acid. After reducing the flow to 1 µl/min by using a splitter, the proteins were transferred to the analytical column (Vydac TP214 C4 RP; 123 mm x 150 µm, 5 µm particle size; packed in-house) with a linear gradient from 0 to 80% eluent B (0.05% (v/v) trifluoro acetic acid, 80% (v/v) acetonitrile) for 40 min. The column eluent was directly introduced into a modified electrospray ionization time-of-flight instrument (Waters, Micromass LC-T, Manchester, UK) equipped with a Z-spray nanoflow electrospray source. The instrument settings were adjusted for optimal transfer of the ions into the mass spectrometer (capillary voltage 3,000 V, sample cone 80 V, desolvation gas 180 L/hr, desolvation temperature 120 ºC). The mass spectra were externally calibrated with 4 mg/ml cesium iodide and analyzed by MassLynx 4.0 software (Waters).

Macromolecular MS analysis

For the macromolecular MS experiments exosome samples were prepared in 50 mM aqueous ammonium acetate, pH 6.8 by using ultra-filtration units with a cut-off of 100 kDa (Millipore, Bedford, UK). Exosome sample (1 pmol; concentration 0.5 µM) was introduced into the modified electrospray ionization time-of-flight instrument mass spectrometer using nanoflow electrospray glas capillaries. The instrument was modified by introducing a speedivalve between the sample cone and extraction cone. To produce intact ions *in vacuo* from large complexes in solution the ions were cooled by increasing the pressure in the first vacuum stages of the mass spectrometer. In addition efficient desolvation was needed to sharpen the ion signals in order to determine the stoichiometry of the complexes from the mass spectrum. Therefore, source pressure conditions were raised and nanoflow electrospray voltages were optimized for transmission of large complexes (capillary voltage 1400-1600 V, sample cone voltage 100-150 V, source pressure 9.0 mbar, desolvation temperature 80 ºC) (26,41,42). The borosilicate glass capillaries (Kwik-Fil, World Precission Instruments, Sarasota, FL) were pulled on a P-97 puller (Sutter Instruments, Novato, CA) and subsequently coated with a thin layer of gold (Edwards Scancoat, Edwards Laboratories, Milpitas, CA, USA). The mass spectra were externally calibrated with 40 mg/ml cesium iodide and analyzed by MassLynx 4.0 software (Waters).

Results

Multiplexed mass spectrometry approach to characterize the exosome complex

Here we describe a comprehensive method that combines a mild affinity purification procedure with a multiplexed mass spectrometry approach for the in-depth characterization of endogenously expressed exosome complexes from *S. cerevisiae*. The procedure is schematically outlined in Fig. 1.

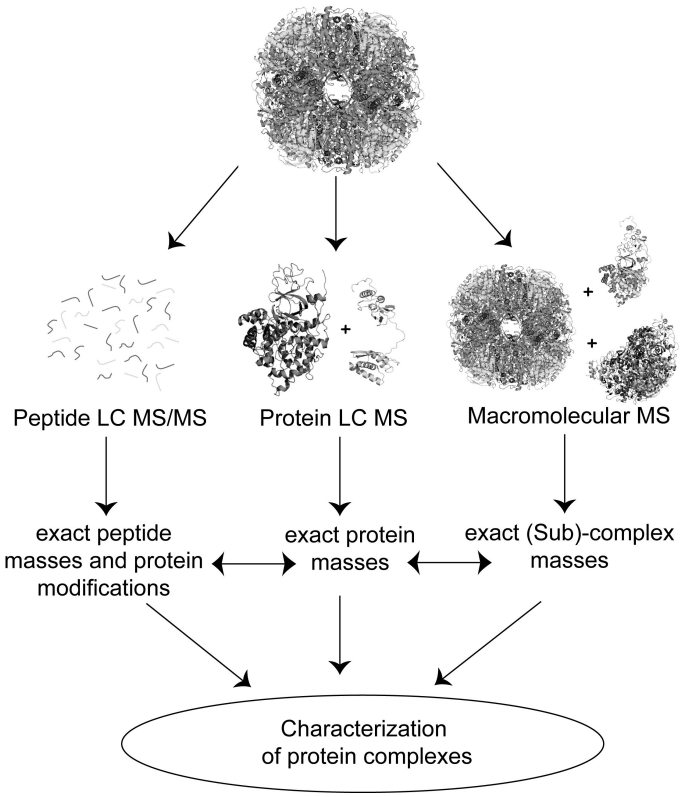


Figure 1

Schematic representation of the multiplexed mass spectrometry approach to characterize the purified exosome complex. The results of the three approaches peptide LC MS/MS, protein LC MS and macromolecular MS were combined. The data yielded i) the identity of the individual proteins, including modifications and phosphorylations, ii) the exact masses of the individual proteins and iii) the identity of the (sub)-exosome complexes including protein stoichiometry and information about strong and weak interacting proteins.

Identification of exosome components by LC MS/MS analysis

For the present study we used both Csl4 and Rrp42 as the tagged entry points to purify the exosome complex with the TAP procedure. Unless stated otherwise the results described in this report are from the measurements with the Csl4 TAP-tagged exosome. We introduced a few modifications to the standard affinity purification procedure to obtain pure exosome complexes with a minimum of non-specifically interacting proteins (see experimental procedures). The amount of exosome purified from 2 L of *S. cerevisiae* culture was about 10-40 µg (11.5-46 pmol) as determined by a protein assay. Half of this purification was then analyzed by denaturant one-dimensional polyacrylamide gel electrophoresis (Fig. 2).

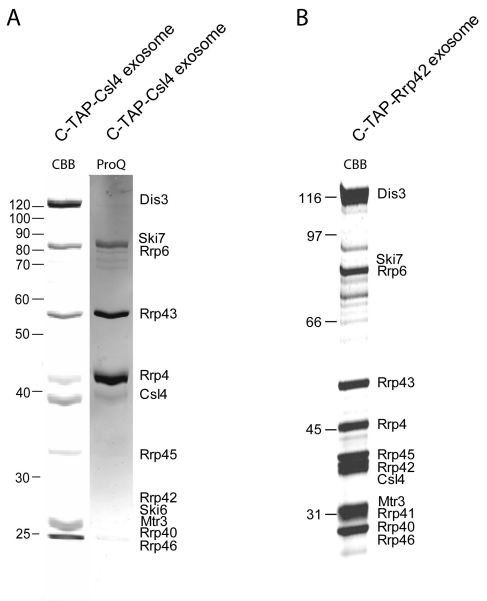


Figure 2

Purification of the *S. cerevisiae* exosome by using a C-terminal TAP tag. Coomassie Brilliant Blue stained denaturant gels representing proteins recovered from purification of *A*, the C-terminal TAP-tagged Csl4 and *B*, the C-terminal TAP-tagged Rrp42 are depicted. Prior to Coomassie staining the gel of the TAP-tagged Csl4 exosome was subjected to the phospho-specific stain ProQ. The left lane represents the molecular weight markers. The gels contained 7 µg and 15 µg of exosome, repectively. The complete gel with TAP-tagged Csl4 was excised and used to prepare 10 independent in-gel trypsin and endoproteinase Glu-C digests for analysis by LC MS/MS.

The gel showed 8 protein bands between 20 and 120 kDa. To assign the protein bands and to obtain information about primary amino acid sequences, modified N- and C-termini and possible post-translational modifications we undertook a LC MS/MS approach using a LTQ-FT-ICR. In-gel trypsin and endoproteinase Glu-C digests were prepared from 10 independent gel slices from one gel for LC MS/MS analyses. Proteins were identified by Mascot database searches using tryptic and endoproteinase Glu-C digests as starting point for matches. The supplementary tables 1 and 2 provide a full overview of the LC MS/MS experimental results and an overview of the exosome protein sequences and the obtained sequence coverage, respectively.

The database searches identified 13 unique proteins, which were either exosome core proteins (Rrp41, Rrp42, Mtr3, Rrp43, Rrp46, Rrp45, Rrp4, Rrp40 and Csl4) or proteins known to be associated with the exosome (Rrp6, Dis3, Lrp1 and Ski7) (28,31). All proteins, apart from Lrp1, had good sequence coverage in between 47 % (Mtr3) and 91 % (Dis3 and Csl4) (supplementary tables 1 and 2). The low sequence coverage of Lrp1 (19%) may indicate that it is only present in sub-stoichiometric amounts within the purified exosome

complex. Lrp1 is known to be only present in the nuclear exosome giving a rational for the sub-stoichiometric presence of this protein (32). Also Rrp46 had relatively low sequence coverage (47%), which may be explained by the relatively low content of lysine and arginine residues in Rrp46. The resulting large tryptic peptides are detected less efficiently in our LC MS/MS procedure. From the previously identified associated proteins we only detected Rrp6, Dis3, Lrp1 and Ski7. We did not detect Ski2, Ski3, Ski8, Mtr4, Gsp1 and Nip7 or any other new exosome-associated proteins. Yeast two-hybrid experiments have shown that the Ski complex (Ski2, Ski3, and Ski8) interacts with the exosome via Ski7 (43). Mtr4 is known to be associated only with the nuclear exosome (29,44) and Gsp1 can interact with Dis3 from the exosome (45). In two-hybrid screens Nip7 has been shown to interact with Rrp43p (46). The low number of exosome-associated proteins is explained by our purification procedure, which leads to relative pure exosome complexes, but may also lead to dissociation of weakly bound proteins.

The phosphorylation-specific staining of the denaturant gel of Csl4-tagged exosome suggested the presence of phosphorylation sites in several exosome proteins (Fig. 2A). By using LC MS/MS, and in agreement with the phoshorylation-specific stain, one or more phosphorylation sites could be identified in Ski7, Rrp6, Rrp43, Rrp4, Csl4, Mtr3 and Rrp46 (table 1, for detailed MS analyses see supplemental data).

Table 1 Overview of identified phosphorylation sites in yeast exosome proteins.

Band in gel	Protein name	Peptide sequences	Amino acid sequences	Phosphorylated residue
1	Ski7	KS**S**NLDPSSSNSFK	305-318	S307
1	Ski7	LSALKK**S**N**S**DLEK	82-94	Ser88, Ser90
1	Rrp6	NTNEEA**T**PIPSSETK	514-528	Thr520
2	Rrp43	VGAC**T**DEEMTISQK	141-154	Thr146
2	Rrp43	I**S**PELSLQR	25-33	Ser26
3	Rrp4	RK**S**ESDELQMR	150-160	Ser152
4	Csl4	SVDA**S**PNDVTR	90-100	Ser94
5	Mtr3	YRDLMI**S**CLMNQET	237-250	Ser243
5	Mtr3	STDL**T**PKGNESEQE	32-45	Thr36
5	Rrp46	IIQDNI**S**PR	245-253	Ser251

For each detected phosphorylated site the phosphorylated and non-phosphorylated peptide could be detected and semi-quantitative analysis revealed that all phosphorylation sites were present in sub-stoichiometric amounts. In all other proteins, Dis3, Rrp40, Rrp42 and Ski6, we did not identify any phosphorylation sites.

On the basis of the sequence data of the N- and C-termini of the identified proteins we could conclude that all proteins except Mtr3 lacked the N-terminal methionine residue (table 2).

Table 2 Comparison of predicted (corrected for protein modifications) and measured masses of exosome-associated proteins. Csl4 was used as C-terminal TAP-tagged entry point.

Database entry	Protein name	Observed protein modifications[a]	Predicted mass[b,c]	Determined mass[c]
			(Da)	(Da)
Q08162	Dis3	- Met, + N-terminal acetyl	113,617.9	113,621 ± 3.5
P25359	Rrp43	- Met, + N-terminal acetyl, + 2 P	43,922.0	-
P38792	Rrp4	- Met, + N-terminal acetyl, + P	39,338.2	-
Q05636	Rrp45	- Met, + N-terminal acetyl	33,872.8	33,846.5 ± 0.6
P53859	Csl4	- Met, + N-terminal acetyl + TAG (5,077 Da), + P	36,571.2	36,575.4 ± 0.8
Q12277	Rrp42	- Met, + N-terminal acetyl	28,966.1	28,970.4 ± 1.9
P48240	Mtr3	+ N-terminal acetyl, + 2 P	27,619.1	27,622 ± 0.3
P46948	Rrp41	- Met, + N-terminal acetyl	27,471.6	27,473.8 ± 2.3
Q08285	Rrp40	- Met, + N-terminal acetyl	26,501.3	26,469.2 ± 0.7
gi 37362654	Rrp46	- Met, + P	24,276.1	24,319.2 ± 0.4
Q12149	Rrp6	- Met, + N-terminal acetyl, + P	83,949.6	-
gi 6321873	Lrp1	- Met	20,914.0	-
gi 6324650	Ski7	- Met, + N-terminal acetyl, + 3 P	84,689.3	-

[a] - Met indicates loss of N-terminal methionine, + TAG (5,077 Da) indicates C-terminal calmodulin binding peptide attached to protein and + P indicates phosphorylation site

[b] The predicted mass was based on the primary amino acid sequence and corrected for observed protein modifications as measured by peptide LC MS/MS.

[c] All phosphorylation sites were present in sub-stoichiometric amounts and therefore they were not taken into account to predict masses.

[d] The determined mass was the protein mass as measured by protein LC MS.

Moreover, all proteins except Rrp46 and Lrp1 contained an N-terminal acetyl moiety. There was no evidence for a modified C-terminus for any of the proteins. Csl4 contained the calmodulin-binding peptide-part of the TAP-tag (5,077 Da) and therefore it was not possible to draw any conclusions about the C-terminal region of this protein. In databases it is reported that Rrp46 exists in two forms: a short (24,407.3 Da) and long (28,331.9 Da) form (29,47). Our LC MS/MS data did not reveal any peptides originating from the long form of Rrp46 suggesting that our purified exosome complexes contain only the short form of Rrp46. Moreover, the exact masses of exosome proteins and (sub)-complexes also pointed

to the exclusive presence of the short form of Rrp46 (see section Analysis and mass measurement of intact proteins of the exosome).

Analysis and mass measurement of intact proteins of the exosome

The LC MS/MS data identified 13 exosome proteins and provided us with information about N- and C-termini and phosphorylation sites on the proteins. These data however did not provide so-called full coverage of the proteins and therefore also did not reveal exact masses of the different proteins. To complete the dataset on the individual proteins we performed intact protein LC MS experiments. The components of the purified exosome complex were first dissociated with 0.5% (v/v) formic acid. The dissociated proteins (13 pmol) were then injected onto a Poros10 trap column after which proteins were transferred to the C4 RP column (Fig. 3A). The subsequent on-line analysis of the eluted proteins using an electrospray ionization time-of-flight mass spectrometer allowed us to identify eight exosome proteins (Dis3, Rrp45, Csl4, Rrp42, Mtr3, Rrp41, Rrp40, Rrp46) with a mass error in between 0.4 and 3.5 Da, i.e. mass error < 0.003% (table 2; Fig 3B).

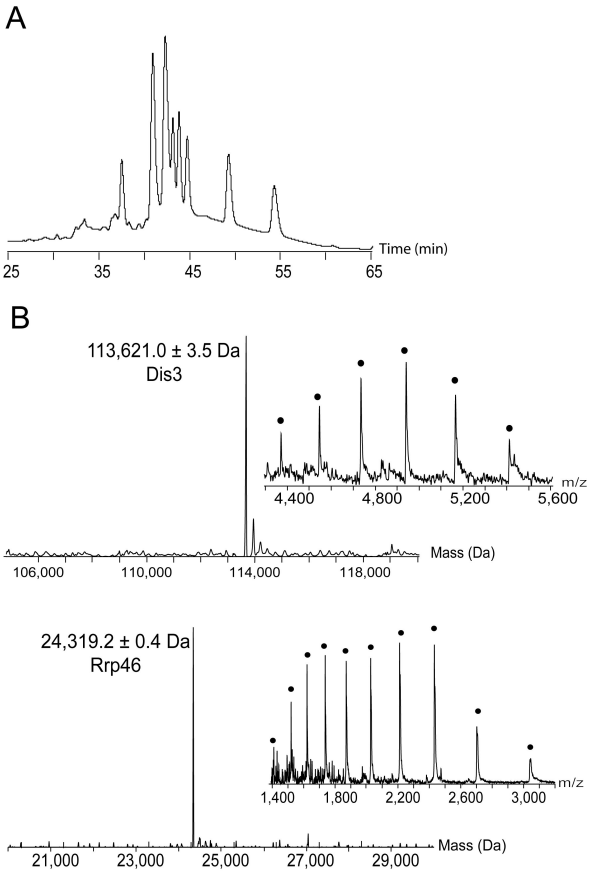


Figure 3

Protein LC MS analysis of purified Csl4 TAP-tagged *S. cerevisiae* exosome complex. The purified exosome (17 µg) was subjected to 0.5% (v/v) formic acid and injected onto a Poros10 trap column. The proteins were transferred to a C4RP column by using a linear gradient of acetonitrile and analyzed on-line by electrospray ionization mass spectrometry. *A*, UV trace of elution of proteins from C4RP column.

B, Deconvoluted mass spectra of two of the eluted proteins (Dis3 and Rrp46). The insets show the raw LC MS data from these two proteins. The spectra were deconvoluted by using a maximum entropy algorithm and yielded exact masses of the proteins (error less than 3.5 Da).

Unfortunately, in this assay we were unable to detect the exosome-associated proteins Rrp6, Ski7 and Lrp1, which was probably due to their sub-stoichiometric presence within the complex. Furthermore, we could not detect Rrp4 and Rrp43, which may be related to their stability in the unfolded form.

The determined molecular masses using this protein LC MS approach differed by 2 to 43 Da as compared to the masses predicted from both the gene sequences and the peptide LC MS/MS data (table 2). We did not take into account the mass increment due to the phosphorylation sites as all these sites were present in sub-stoichiometric amounts. Most predicted and determined masses were with the given uncertainty of maximal 4.3 Da in agreement (Dis3 113,621.0 Da *v.s.*113,617.9 Da; Csl4 36,575.4 Da *v.s.*36,571.2 Da; Rrp42 28,970.4 Da *v.s.*28,966.1 Da; Mtr3 27,622.0 Da *v.s.* 27,619.1 Da and Rrp41 27,473.8 Da *v.s.* 27,471.6 Da, respectively). Only in the case of Rrp45, Rrp40 and Rrp46 the predicted and the determined masses differed up to 43.1 Da (Rrp45 33,846.5 Da *v.s.* 33,872.8 Da; Rrp40 26,469.2 Da *v.s.* 26,501.3 Da and Rrp46 24,319.2 Da *v.s.* 24,276.1 Da respectively). As the measured masses had experimental errors ranging from 0.4 Da (Rrp46) to 3.5 Da (Dis3), the observed differences between predicted and measured masses were likely to be related to other non-identified post-translational modifications or to errors in the primary amino acid sequences. In fact, we identified a peptide (amino acid 359 to 368 (QLTLMGGGAK) by peptide LC MS/MS from Rrp43 that had one amino acid substitution (V->M, position 363) compared to the primary sequence in the database. However, we could not determine a mass of intact Rrp43 and thus we could not validate this data. In line with the peptide LC MS/MS data these protein LC MS data strongly suggested that our purified exosome comprises only the short form of the exoribonuclease Rrp46. We identified a protein with a mass of 24,319.2 Da; 43.1 Da higher than the predicted mass for the short form of Rrp46 (in the absence of the first methionine) and 4,012.7 Da lower than the predicted mass from the long form of Rrp46 (in the absence of the first methionine) (29,47). The measured mass of TAP-tagged Csl4 (36,575.4 Da) was in good agreement with the predicted mass of Csl4 including the calmodulin binding peptide moiety of the affinity tag (36,571.2 Da).

Probing the exosome by macromolecular MS

In the third approach, the most challenging one from an experimental point of view, we investigated exosome (sub)-complexes by macromolecular MS (20-27). Fig. 4A presents a positive ion mass spectrum obtained from an ammonium acetate solution, pH 6.8 of 1 pmol of intact Csl4 TAP-tagged exosome introduced into the nanoflow electrospray source of an electrospray ionization time-of-flight mass spectrometer.

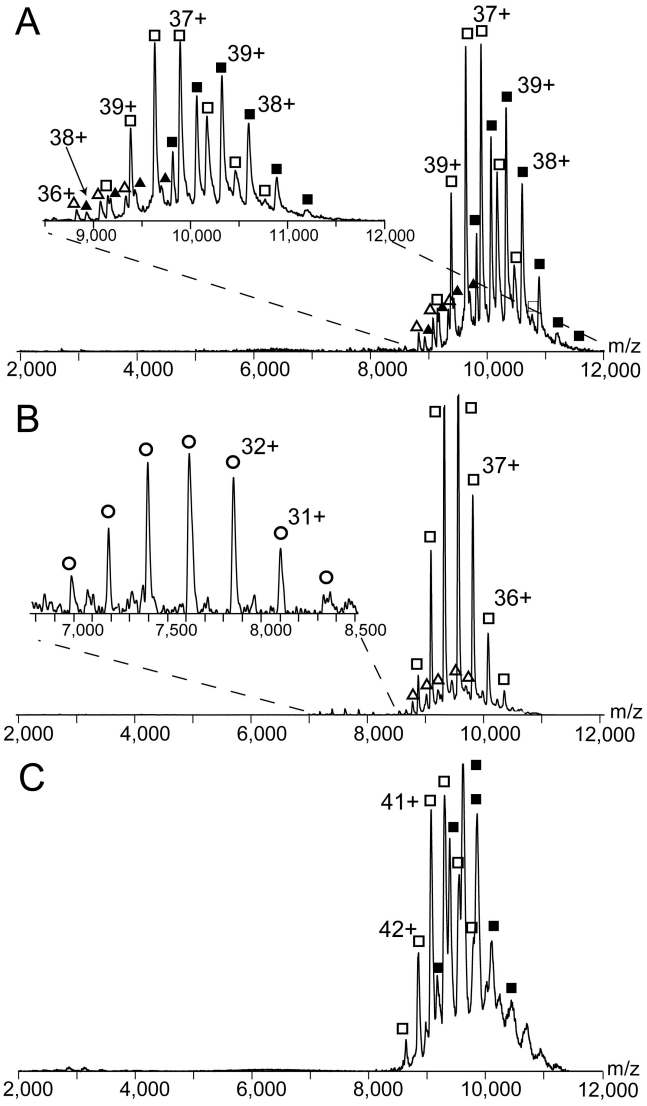


Figure 4

Macromolecular electrospray ionization mass spectra of purified Csl4 or Rrp42 TAP-tagged *S. cerevisiae* exosome complex. Purified exosome (1 pmol) was sprayed from a 50 mM ammonium acetate solution, pH 6.8. *A*, and *B*, represent (sub)-complexes from Csl4-tagged exosome and *C*, represents (sub)-complexes from Rrp42-tagged exosome. Samples represented by *A,* and *B,* were purified from the same cells, but during the analysis *A*, was at 20 ºC whereas *B*, was at 25 ºC. *A*, The four ions series centered around m/z 9,100 (△), 9,400 (▲), 9,800 (□) and 10,400 (■) represent (sub)-exosome complexes having molecular masses of 326,590, 339,404, 365,918 and 402,678 Da, respectively. *B*, The three ions series centered around m/z 7,500 (○), 9,100 (△) and 9,800 (□) represent (sub)-exosome complexes having masses of 251,927, 327,059 and 366,160 Da, respectively. *C*, The two ions series centered around m/z 9,800 (□) and 10,400 (■) represent (sub)-exosome complexes having masses of 371,283 and 403,083 Da, respectively. See also table 3 for a summary.

Fascinatingly, the spectrum showed four clear charge state distributions centered around m/z 9,100, 9,400, 9,800 and 10,400, and was dominated by the two distributions around m/z 9,800 and 10,400. The protein mass could easily be determined by using the well-resolved multiple charge states of a protein. The observation of multiple complexes was surprising as our experimental conditions were such that we did not expect disassembly of protein complexes (23,26). Moreover, the number of charges the complexes obtained clearly reveal that all dissociations represent solution-phase events and not gas-phase events.

Mass determination of the ion series with the highest m/z values (m/z 10,400) yielded a molecular mass of 402,678 ± 600 Da (exosome **1** in Fig. 5). Table 3 summarizes the masses measured by macromolecular MS. When we summed the masses of the proteins as determined by intact protein LC MS (Dis3, Rrp45, Csl4, Rrp42, Mtr3, Rrp41, Rrp40 and Rrp46) and the corrected predicted masses of the two proteins that we did not identify by this approach (Rrp43 and Rrp4), assuming a 1:1 stoichiometry for all proteins, we would expect a mass of 402,158 Da for the cytoplasmic exosome (thus core exosome (Rrp45, Csl4, Rrp42, Mtr3, Rrp41, Rrp40, Rrp46, Rrp43 and Rrp4) including Dis3, but excluding Rrp6, Ski7 and Lrp1 (28)).

Table 3 Measured masses of exosome (sub)-complexes by macromolecular MS. Csl4 and Rrp42 were used as the C-terminal tagged entry points.

	Measured mass (Da)				
Exosome[a]	**1**	**2**[b]	**3**	**4**	**5**
Entry point					
Csl4	402,678 ± 600	365,918 ± 400	339,404 ± 400	326,590 ± 400	
		366,160 ± 1,000		327,059 ± 1,000	251,927 ± 1,000
Rrp42	403,083 ± 1,000	371,283 ± 1,000			

[a] Exosome (sub)-complexes 1-5 are schematically presented in Fig. 5

[b] Mass difference of 5,077 Da between the exosome **2** species are due to the use of the two different

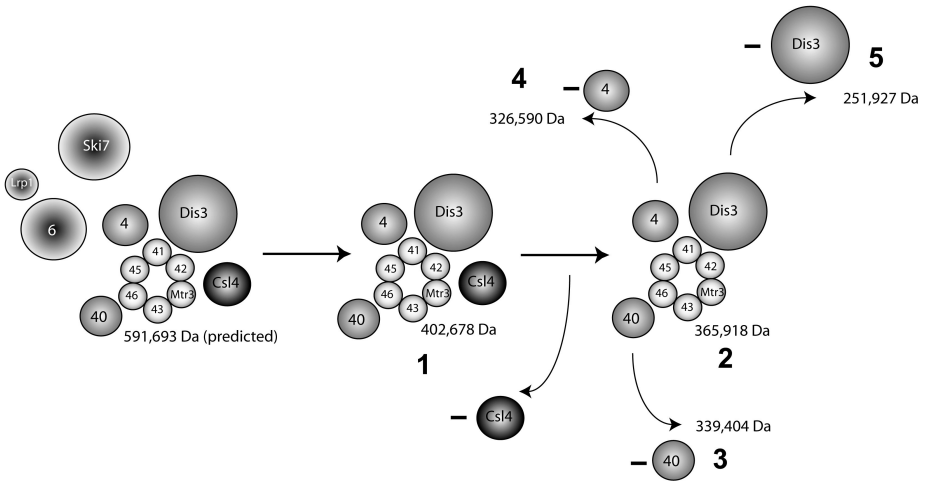


Figure 5

Schematic representation of measured exosome (sub)-complexes. The numbers 1 to 5 refer to the (sub)-complexes measured by macromolecular MS. The inner dark grey (Ski7, Rrp6 and Lrp1), dark grey (Csl4), grey (Rrp40, Rrp4 and Dis3) and light grey (Rrp41, 42, Mtr3, Rrp43, Rrp46, Rrp45) spheres represent proteins that interact very weak, weak, moderately strong and strong, respectively, within the exosome complex. The representation is based on described protein-protein interactions in literature. It should be noted here that Rrp6 and Lrp1 are known to be present only in the nuclear exosome.

This expected mass was only 0.13% lower than the measured mass for this complex (402,678 Da). We also calculated expected masses of exosome complexes having different stoichiometries for the non-ring proteins Csl4, Rrp4, Rrp40 or Dis3 (33), but these calculations did not yield molecular masses close to the measured mass. Therefore, we concluded that all components of the cytoplasmic exosome were present in stoichiometric amounts. The second ion series centered around 9,800 m/z within the mass spectrum had a determined mass of 365,918 ± 400 Da (exosome **2**). The mass difference between the two complexes (402,678 – 365,918 Da) was 36,760 Da, which could only be assigned to the loss of TAP-tagged Csl4 (36,575 Da). Thus, the complex with a mass of 365,918 Da was the cytoplasmic exosome missing the TAP-tagged Csl4 (expected mass 365,582 Da; mass difference 0.09%). These data indicated that under the conditions used (50 mM ammonium acetate, pH 6.8, 20 ºC) Csl4 interacted relatively weak with the exosome complex. The determined masses of the two remaining and low abundant charge state distributions were 339,404 ± 400 Da (exosome **3**) and 326,590 ± 400 Da (exosome **4**). The mass difference

between exosome **2** and **3** was 26,514 Da, which was in good agreement with the measured mass of Rrp40 by protein LC MS (26,469 Da), strongly indicating that exosome **3** was the cytoplasmic exosome missing the TAP-tagged Csl4 and Rrp40. In non-biased calculations the exosome **3** could theoretically also be composed of Dis3, Rrp43, Rrp4, Csl4, Mtr3, Rrp41, Rrp40 and Rrp46 (expected mass 339,341 Da), thus with loss of Rrp42 and Rrp45. It is now well-accepted that *S. cerevisiae* exosome consists of a doughnut-shaped ring structure consisting of Rrp43, Rrp41, Mtr3, Rrp45, Rrp46 and Rrp42 with other proteins associated with one or more of the ring proteins (48). With regard to this ring structure we consider it unlikely that Rrp45 and Rrp42 dissociate from the ring structure resulting in exosome **3** and therefore exosome **3** very likely contains the six ring proteins and Dis3 and Rrp40. The mass difference between exosome **2** and **4** matched almost exactly the mass of Rrp4 (measured mass difference 39,328 Da; expected mass Rrp4 39,338 Da). Thus, the smaller exosome species **4** was assigned to the cytoplasmic exosome missing the TAP-tagged Csl4 and Rrp4. The mass of exosome **4** was in close agreement with its expected mass of 326,244 Da (mass difference 0.10%). Calculations with all possible permutations of the exosome proteins confirmed that only the loss of Csl4 and Rrp4 from exosome **4** could explain the measured mass.

Next, we measured protein complex mass spectra of the Csl4-tagged exosome complexes from a second purification from the same yeast cells (Fig. 4B). The mass spectra of the two purifications were very similar, but also showed some interesting differences, which is likely to be related to a higher temperature (25 ºC) when analyzing the exosome from the second purification (20 *v.s.* 25 ºC). The mass spectrum of the exosome from the second purification revealed three charge state distributions centered at m/z 7,500, 9,100, and 9,800. Mass determination of the ion series around m/z 9,100 and 9,800 yielded masses of 327,059 ± 1,000 Da and 366,160 ± 1,000 Da, respectively. Although the mass accuracy was somewhat lower than in the first experiment (Fig. 4A) we could unambiguously assign the complexes to exosome **2** and **4**, respectively (Fig. 5). The mass difference between the two complexes (366,160 – 327,059 Da) of 39,101 Da was in good agreement with the predicted mass of Rrp4 (39,338 Da). Mass determination of the new ion series around m/z 7,500 yielded a mass of 251,927 ± 1,000 Da (exosome **5**). This molecular species represented the cytoplasmic exosome, composed of Rrp45, Rrp42, Mtr3, Rrp41, Rrp40, Rrp46, Rrp43 and Rrp4 but missing the TAP-tagged Csl4 and Dis3 (expected mass 251,961 Da). Calculations with all possible permutations showed that this protein composition was the only possibility for exosome **5**. In our macromolecular MS approach we never observed smaller complexes or complexes including the components Rrp6, Lrp1 and/or Ski7. Together with the peptide LC MS/MS results we may conclude from this that these proteins were present only in sub-stoichiometric amounts, such that the amounts were too low to be detected by the macromolecular MS approach. However, as an alternative we cannot exclude that the two nuclear proteins and Ski7 disassembled from the exosome during our final preparations.

As one can argue that the TAP-tag to Csl4 may weaken the interaction with the core exosome, we also performed macromolecular MS experiments with the Rrp42 TAP-tagged exosome (Fig. 4C). In comparison with the TAP-tagged Csl4, this TAP-tag resulted in a mass increase of Rrp42 with 5,077 Da to 34,047 Da and a mass decrease of Csl4 with 5,077 Da to 31,498 Da. The mass spectrum showed two charge state distributions around m/z 9,200 and 9,600, respectively. Mass determination of the two species yielded masses of 403,083 ± 1,000 Da and 371,283 ± 1,000 Da. The mass of the first species was in reasonable agreement with the mass of cytoplasmic exosome (exosome **1**), whereas the second species could only be assigned to the cytoplasmic exosome after dissociation of Csl4 (exosome **2**). The mass difference between the two complexes (31,800 Da) matched reasonably well with the predicted mass of Csl4 (31,498 Da). These results were in line with the results of the Csl4 TAP-tagged exosome and confirmed that Csl4 only weakly interacts with the core exosome. Therefore, we concluded that the TAP-tag did not affect the association of Csl4 with the exosome.

Discussion

By using the multiplexed MS approach described here we characterized yeast (sub)-exosome complexes, including protein stoichiometry, protein modifications and strong and weak interacting exosome proteins. The affinity purification of *S. cerevisiae* exosome yielded the core exosome complex (Rrp41, Rrp42, Rrp43, Rrp45, Rrp46, Mtr3, Csl4, Rrp4, and Rrp40) and the known associated proteins Dis3, Rrp6, Ski7 and Lrp1. Our analysis revealed for the first time phosphorylation sites in yeast exosome proteins (table 1, for detailed MS analyses see supplemental data). Intriguingly, the serine phosphorylation at position 152 of Rrp4 is conserved within the human homologue of Rrp4 (Ser124) (49) (Fig. 6), and is the only phosphorylation site previously identified in any of the human exosome components. This serine phosphorylation site is located within the S1 RNA-binding domain of the protein (residues 107-187 in yeast Rrp4).

The *in vivo* relevance of this phosphorylation site remains to be elucidated, but its conservation within the human homologue strongly indicates that it has an important function in the cell.

Ptacek *et al.* (50) have performed a global analysis of protein phosphorylation in yeast and they suggest several protein kinases may have exosome components as substrates. The two kinases Sky1 and Pho85 recognize Dis3 as their substrate, Atg1 and Swe1 recognize Lrp1 and the serine/threonine specific Pkh3 can phosphorylate Csl4 *in vitro*. We could not confirm any phosphorylation sites in Dis3 and Lrp1. Several exosome proteins in which we detected phosphorylation sites (Rrp6, Ski7, Rrp43, Rrp4 and Rrp46) were not tested in the global analysis (50). Our data provide us with the exact phosphorylation sites of several

exosome proteins and may give us an idea of the phosphorylation status of the yeast exosome *in vivo*. Our phosphorylation map of the exosome may not be complete but provides a good starting point to probe the relevance of phosphorylation for the exosome *in vivo*.

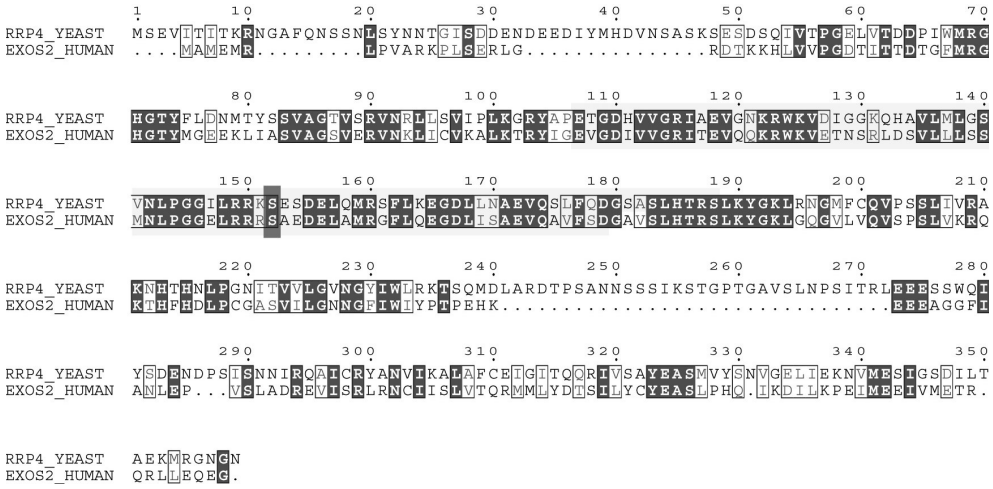


Figure 6

Primary sequence alignment and phosphorylation of yeast and human Rrp4. Conserved residues are shown in boxes and identical residues are shown with a dark-grey background. The predicted S1 RNA-binding domain is indicated with a light-grey background and the phosphorylated Serine (S152) is highlighted. The sequence alignment was performed by using the program Clustal W (40).

By combining the peptide and protein LC MS data with native macromolecular MS we obtained data on intact exosome complexes and sub-complexes thereof. The largest complex that could be observed theoretically would have a mass of 591,710 Da (Fig. 5), assuming that all components were present in a 1:1 stoichiometry. That we have never observed this complex by macromolecular MS is likely due to the sub-stoichiometric amounts of Rrp6, Ski7 and Lrp1 and/or to the weak interactions of these proteins to the core exosome. Rrp6 and Lrp1 are known to be present only in the nuclear exosome (32) and previous studies have indicated that Ski7 interacts only weakly with the exosome core complex (51). The global analysis of protein expression in yeast has provided the abundance of the yeast proteome in the logarithmic growth phase (5). The number of protein molecules per cell may be related to the expected sub-stoichiometric amounts of Rrp6, Ski7 and Lrp1 in the exosome. The core exosome components have in between 3,180 and 10,800 copies per cell, but Dis3 has only a reported copy number of 606 (Supplementary table 1). The copy numbers of Rrp6 and Ski7 are 2,160 and 233, respectively (Lrp1 was not determined in (5)). Thus, the low abundance of Rrp6 and Ski7 is

in line with its sub-stoichiometric presence within (sub)-exosome complexes. The 1D gel also indicates that Rrp6 and Ski7 are only present in sub-stoichiometric amounts as the intensity of the protein band containing both proteins has a somewhat lower intensity than most single protein bands. On the other hand the low abundance of Dis3 was not in agreement with our results. Our macromolecular MS and 1D gel LC MS/MS analyses clearly showed that Dis3 was present in each exosome complex.

The largest complex we observed by macromolecular MS was the core cytoplasmic exosome including the commonly associated protein Dis3 having a measured mass of 402,678 Da. The most abundant complex (exosome **2**) that we detected lacks Csl4. From two-hybrid studies, affinity purifications, RNA interference studies (34-36) and a recent X-ray structure of *Sulfolobus solfataticus* exosome (37) it is now well-accepted that *S. cerevisiae* exosome consists of a doughnut-shaped ring structure consisting of Rrp43, Rrp41, Mtr3, Rrp45, Rrp46 and Rrp42 with other proteins associated with one or more of the ring proteins (48). Our multiplexed mass spectrometry approach provides evidence that only the ring structure is very stable in solution. Electron microscopy data in combinations with structure predictions have suggested that Csl4, Rrp40 and Rrp4 are positioned on top of the doughnut-shaped ring (33). This representation is in agreement with our data, which showed that Csl4 only weakly interacts and that Rrp40 and Rrp4 moderately interact with the ring structure. We also found that Dis3 has a similar affinity to the purified exosome complex as Rrp40 and Rrp4. Our observation that Rrp40, Rrp4 and Dis3 only dissociated in combination with Csl4 may suggest that Csl4 stabilizes the quaternary structure of the exosome.

We conclude that the yeast exosome does not necessarily behave as a single static complex. The exosome complexes are all organized around a stable hexameric ring to which association and dissociation of proteins takes place. Several exosome proteins are sub-stoichiometrically phosphorylated and we found that the phosphorylation site in yeast Rrp4 is conserved within the human homologue of Rrp4. The *in vivo* significance of our data needs to be established, but the data may give an indication about the assembly and disassembly of the exosome *in vitro*. The observation that the macromolecular mass spectra from different purifications and different TAP-tagged target proteins were very similar also revealed the high reproducibility of the current method. The described method is generic and thus applicable to many protein complexes, even when they have been expressed at endogenous (picomol) levels.

References

1. Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., Qureshi-Emili, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadamodar, G., Yang, M., Johnston, M., Fields, S., and Rothberg, J. M. (2000) A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature* **403**, 623-627

2. Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 4569-4574
3. Gavin, A. C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., Cruciat, C. M., Remor, M., Hofert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M. A., Copley, R. R., Edelmann, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G., and Superti-Furga, G. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141-147.
4. Zhao, R., Davey, M., Hsu, Y. C., Kaplanek, P., Tong, A., Parsons, A. B., Krogan, N., Cagney, G., Mai, D., Greenblatt, J., Boone, C., Emili, A., and Houry, W. A. (2005) Navigating the chaperone network: an integrative map of physical and genetic interactions mediated by the hsp90 chaperone. *Cell* **120**, 715-727
5. Ghaemmaghami, S., Huh, W. K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., O'Shea, E. K., and Weissman, J. S. (2003) Global analysis of protein expression in yeast. *Nature* **425**, 737-741
6. Krogan, N. J., Peng, W. T., Cagney, G., Robinson, M. D., Haw, R., Zhong, G., Guo, X., Zhang, X., Canadien, V., Richards, D. P., Beattie, B. K., Lalev, A., Zhang, W., Davierwala, A. P., Mnaimneh, S., Starostine, A., Tikuisis, A. P., Grigull, J., Datta, N., Bray, J. E., Hughes, T. R., Emili, A., and Greenblatt, J. F. (2004) High-definition macromolecular composition of yeast RNA-processing complexes. *Mol. Cell* **13**, 225-239
7. Alberts, B. (1998) The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* **92**, 291-294
8. Gavin, A. C., and Superti-Furga, G. (2003) Protein complexes and proteome organization from yeast to man. *Curr. Opin. Chem. Biol.* **7**, 21-27
9. Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Seraphin, B. (1999) A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* **17**, 1030-1032
10. Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Seraphin, B. (2001) The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* **24**, 218-229
11. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* **422**, 198-207
12. Bouwmeester, T., Bauch, A., Ruffner, H., Angrand, P. O., Bergamini, G., Croughton, K., Cruciat, C., Eberhard, D., Gagneur, J., Ghidelli, S., Hopf, C., Huhse, B., Mangano, R., Michon, A. M., Schirle, M., Schlegl, J., Schwab, M., Stein, M. A., Bauer, A., Casari, G., Drewes, G., Gavin, A. C., Jackson, D. B., Joberty, G., Neubauer, G., Rick, J., Kuster, B., and Superti-Furga, G. (2004) A physical and functional map of the human TNF-alpha/NF-kappa B signal transduction pathway. *Nat. Cell. Biol.* **6**, 97-105
13. 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
14. Sali, A., Glaeser, R., Earnest, T., and Baumeister, W. (2003) From words to literature in structural proteomics. *Nature* **422**, 216-225
15. Fiaux, J., Bertelsen, E. B., Horwich, A. L., and Wuthrich, K. (2002) NMR analysis of a 900K GroEL GroES complex. *Nature* **418**, 207-211
16. Riek, R., Fiaux, J., Bertelsen, E. B., Horwich, A. L., and Wuthrich, K. (2002) Solution NMR techniques for large molecular and supramolecular structures. *J. Am. Chem. Soc.* **124**, 12144-12153
17. Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H., and Noller, H. F. (2001) Crystal structure of the ribosome at 5.5 A resolution. *Science* **292**, 883-896
18. Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science* **289**, 905-920
19. Verentchikov, A. N., Ens, W., and Standing, K. G. (1994) Reflecting time-of-flight mass spectrometer with an electrospray ion source and orthogonal extraction. *Anal. Chem.* **66**, 126-133
20. Loo, J. A. (1997) Studying noncovalent protein complexes by electrospray ionization mass spectrometry. *Mass Spectrom. Rev.* **16**, 1-23
21. Heck, A. J. R., and van den Heuvel, R. H. (2004) Investigation of intact protein complexes by mass spectrometry. *Mass Spectrom. Rev.* **23**, 368-389
22. Robinson, C. V. (2002) Protein complexes take flight. *Nat. Struct. Biol.* **9**, 505-506.

23. 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
24. Hernandez, H., and Robinson, C. V. (2001) Dynamic protein complexes: insights from mass spectrometry. *J. Biol. Chem.* **276**, 46685-46688
25. Videler, H., Ilag, L. L., McKay, A. R., Hanson, C. L., and Robinson, C. V. (2005) Mass spectrometry of intact ribosomes. *FEBS Lett.* **579**, 943-947
26. 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
27. van Berkel, W. J., van den Heuvel, R. H., Versluis, C., and Heck, A. J. (2000) Detection of intact megaDalton protein assemblies of vanillyl-alcohol oxidase by mass spectrometry. *Protein Sci.* **9**, 435-439
28. Raijmakers, R., Schilders, G., and Pruijn, G. J. (2004) The exosome, a molecular machine for controlled RNA degradation in both nucleus and cytoplasm. *Eur. J. Cell. Biol.* **83**, 175-183
29. Allmang, C., Petfalski, E., Podtelejnikov, A., Mann, M., Tollervey, D., and Mitchell, P. (1999) The yeast exosome and human PM-Scl are related complexes of 3' --> 5' exonucleases. *Genes Dev.* **13**, 2148-2158
30. Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M., and Tollervey, D. (1997) The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'-->5' exoribonucleases. *Cell* **91**, 457-466
31. Mitchell, P., and Tollervey, D. (2000) Musing on the structural organization of the exosome complex. *Nat. Struct. Biol.* **7**, 843-846
32. Hieronymus, H., Yu, M. C., and Silver, P. A. (2004) Genome-wide mRNA surveillance is coupled to mRNA export. *Genes Dev.* **18**, 2652-2662
33. Aloy, P., Ciccarelli, F. D., Leutwein, C., Gavin, A. C., Superti-Furga, G., Bork, P., Bottcher, B., and Russell, R. B. (2002) A complex prediction: three-dimensional model of the yeast exosome. *EMBO Rep.* **3**, 628-635
34. Estevez, A. M., Lehner, B., Sanderson, C. M., Ruppert, T., and Clayton, C. (2003) The roles of intersubunit interactions in exosome stability. *J. Biol. Chem.* **278**, 34943-34951
35. Raijmakers, R., Egberts, W. V., van Venrooij, W. J., and Pruijn, G. J. (2002) Protein-protein interactions between human exosome components support the assembly of RNase PH-type subunits into a six-membered PNPase-like ring. *J. Mol. Biol.* **323**, 653-663
36. Lehner, B., and Sanderson, C. M. (2004) A protein interaction framework for human mRNA degradation. *Genome Res.* **14**, 1315-1323
37. Lorentzen, E., Walter, P., Fribourg, S., Evguenieva-Hackenberg, E., Klug, G., and Conti, E. (2005) The archaeal exosome core is a hexameric ring structure with three catalytic subunits. *Nat. Struct. Mol. Biol.* **12**, 575-581
38. Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **68**, 850-858
39. Pappin, D. J., Hojrup, P., and Bleasby, A. J. (1993) Rapid identification of proteins by peptide-mass fingerprinting. *Curr. Biol.* **3**, 327-332
40. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673-4680
41. Krutchinsky, A. N., Chernushevich, I. V., Spicer, V. L., Ens, W., and Standing, K. G. (1998) Studies of noncovalent complexes in an electrospray ionization/time-of-flight mass spectrometer. *J. Am. Soc. Mass Spectrom.* **9**, 569-579
42. 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.
43. Orban, T. I., and Izaurralde, E. (2005) Decay of mRNAs targeted by RISC requires XRN1, the Ski complex, and the exosome. *RNA* **11**, 459-469
44. Liang, S., Hitomi, M., Hu, Y. H., Liu, Y., and Tartakoff, A. M. (1996) A DEAD-box-family protein is required for nucleocytoplasmic transport of yeast mRNA. *Mol. Cell. Biol.* **16**, 5139-5146
45. Noguchi, E., Hayashi, N., Azuma, Y., Seki, T., Nakamura, M., Nakashima, N., Yanagida, M., He, X., Mueller, U., Sazer, S., and Nishimoto, T. (1996) Dis3, implicated in mitotic control, binds directly to Ran and enhances the GEF activity of RCC1. *EMBO J.* **15**, 5595-5605

46. Zanchin, N. I., and Goldfarb, D. S. (1999) Nip7p interacts with Nop8p, an essential nucleolar protein required for 60S ribosome biogenesis, and the exosome subunit Rrp43p. *Mol. Cell. Biol.* **19**, 1518-1525
47. Tettelin, H., Agostoni Carbone, M. L., Albermann, K., Albers, M., Arroyo, J., Backes, U., Barreiros, T., Bertani, I., Bjourson, A. J., Bruckner, M., Bruschi, C. V., Carignani, G., Castagnoli, L., Cerdan, E., Clemente, M. L., Coblenz, A., Coglievina, M., Coissac, E., Defoor, E., Del Bino, S., Delius, H., Delneri, D., de Wergifosse, P., Dujon, B., Kleine, K., and et al. (1997) The nucleotide sequence of Saccharomyces cerevisiae chromosome VII. *Nature* **387**, 81-84
48. Pruijn, G. J. (2005) Doughnuts dealing with RNA. *Nat. Struct. Mol. Biol.* **12**, 562-564
49. Beausoleil, S. A., Jedrychowski, M., Schwartz, D., Elias, J. E., Villen, J., Li, J., Cohn, M. A., Cantley, L. C., and Gygi, S. P. (2004) Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 12130-12135
50. Ptacek, J., Devgan, G., Michaud, G., Zhu, H., Zhu, X., Fasolo, J., Guo, H., Jona, G., Breitkreutz, A., Sopko, R., McCartney, R. R., Schmidt, M. C., Rachidi, N., Lee, S. J., Mah, A. S., Meng, L., Stark, M. J., Stern, D. F., De Virgilio, C., Tyers, M., Andrews, B., Gerstein, M., Schweitzer, B., Predki, P. F., and Snyder, M. (2005) Global analysis of protein phosphorylation in yeast. *Nature* **438**, 679-684
51. Brown, J. T., Bai, X., and Johnson, A. W. (2000) The yeast antiviral proteins Ski2p, Ski3p, and Ski8p exist as a complex in vivo. *RNA* **6**, 449-457

Supplementary table 1 Overview of peptide LC MS/MS and protein LC MS experiments of exosome components. By using peptide LC MS/MS analysis 13 exosome associated proteins were identified.

Database entry	Protein name	SwissProt Mass	Protein abundance[a]	Sequence coverage[b]	Observed protein modifications	N-terminus[c]	C-terminus[c]	Predicted mass[d]		Determined mass[e]
		(Da)		(%)				(Da)		(Da)
								Phosphorylated	Non-phosphorylated	
Q08162	Dis3	113,707.02	606	91	- Met, + N-terminal Acetyl	+	+	-	113,617.9	113,621 ± 3.5
P25359	Rrp43	44,011.12	3,180	79	- Met, + N-terminal Acetyl, + 3 Phosphorylation	+	+	44,162.12	43,922.0	-
P38792	Rrp4	39,427.35	4,840	76	- Met, + N-terminal Acetyl, + 1 Phosphorylation	+	+	39,418.35	39,338.2	-
Q05636	Rrp45	33,962.0	4,800	78	- Met, + N-terminal Acetyl	+	+	-	33,872.8	33,846.5 ± 0.6
P53859	Csl4	31,583.39	5,550	91	- Met, + N-terminal acetyl + TAG (5,077 Da), + 2 Phosphorylation	+	-	36,731.39	36,571.2	36,575.4 ± 0.8
Q12277	Rrp42	29,055.28	7,110	69	- Met, + N-terminal Acetyl	+	+	-	28,966.1	28,970.4 ± 1.9
P48240	Mtr3	27,577.1	7,380	47	+ N-terminal Acetyl + 2 Phosphorylation	+	+	27,779.1	27,619.1	27,622 ± 0.3
P46948	Rrp41	27,560.73	5,150	78	- Met, + N-terminal Acetyl	+	+	-	27,471.6	27,473.8 ± 0.9
Q08285	Rrp40	26,590.50	6,050	74	- Met, + N-terminal Acetyl	+	+	-	26,501.3	26,469.2 ± 0.7
gi 37362654	Rrp46	24,407.31	10,800	54	- Met + 1 Phosphorylation	-	+	24,356.31	24,276.1	24,319.2 ± 0.4
Q12149	Rrp6	84,038.72	2,160	71	- Met, + N-terminal Acetyl + 1 Phosphorylation	+	+	84,029.6	83,949.6	-
gi 6321873	Lrp1	21,045.23	n.d.	19	- Met	-	-	-	20,914.0	-
gi 6324650	Ski7	84,779.09	233	78	- Met, + N-terminal Acetyl + 3 Phosphorylation	+	+	84,929.3	84,689.3	-

[a] The protein abundance is the number of protein copies per cell in exponentially growing yeast (Ghaemmaghami, S., Huh, W. K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., O'Shea, E. K. & Weissman, J. S. (2003) *Nature* **425,** 737-41).

[b] The Mascot score and sequence coverage give an indication of the relative amounts of the individual proteins in the sample.

[c] + and – indicate observed and not observed, respectively.

[d] The predicted mass was based on the primary amino acid sequence and corrected for observed protein modifications as measured by peptide LC MS/MS.

[e] The determined mass was the protein mass as measured by protein LC MS.

Supplementary table 2 Overview of the exosome protein sequences and the obtained sequence coverage.

SwissProt entry	Protein name	Protein Sequence[a,b]
Q08162	Dis3	ac 1 MSVPAIAPRR KRLADGLSVT QKVFVRSRNG GATKIVREHY LRSDIPCLSR 51 SCTKCPQIVV PDAQNELPKF ILSDSPLELS APIGKHYVVL DTNVVLQAID 101 LLENPNCFFD VIVPQIVLDE VRNKSYPVYT RLRTLCRDSD DHKRFIVFHN 151 EFSEHTFVER LPNETINDRN DRAIRKTCQW YSEHLKPYDI NVVLVTNDRL 201 NREAATKEVE SNIITKSLVQ YIELLPNADD IRDSIPQMDS FDKDLERDTF 251 SDFTFPEYYS TARVMGGLKN GVLYQGNIQI SEYNFLEGSV SLPRFSKPVL 301 IVGQKNLNRA FNGDQVIVEL LPQSEWKAPS SIVLDSEHFD VNDNPDIEAG 351 DDDDNNESSS NTTVISDKQR RLLAKDAMIA QRSKKIQPTA KVVYIQRRSW 401 RQYVGQLAPS SVDPQSSSTQ NVFVILMDKC LPKVRIRTRR AAELLDKRIV 451 ISIDSWPTTH KYPLGHFVRD LGTIESAQAE TEALLLEHDV EYRPFSKKVL 501 ECLPAEGHDW KAPTKLDDPE AVSKDPLLTK RKDLRDKLIC SIDPPGCVDI 551 DDALHAKKLP NGNWEVGVHI ADVTHFVKPG TALDAEGAAR GTSVYLVDKR 601 IDMLPMLLGT DLCSLKPYVD RFAFSVIWEL DDSANIVNVN FMKSVIRSRE 651 AFSYEQAQLR IDDKTQNDEL TMGMRALLKL SVKLKQKRLE AGALNLASPE 701 VKVHMDSETS DPNEVEIKKL LATNSLVEEF MLLANISVAR KIYDAFPQTA 751 MLRRHAAPPS TNFEILNEML NTRKNMSISL ESSKALADSL DRCVDPEDPY 801 FNTLVRIMST RCMMAAQYFY SGAYSYPDFR HYGLAVDIYT HFTSPIRRYC 851 DVVAHRQLAG AIGYEPLSLT HRDKNKMDMI CRNINRKHRN AQFAGRASIE 901 YYVGQVMRNN ESTETGYVIK VFNNGIVVLV PKFGVEGLIR LDNLTEDPNS 951 AAFDEVEYKL TFVPTNSDKP RDVYVFDKVE VQVRSVMDPI TSKRKAELLL 1001 K
P25359	Rrp43	ac 1 MAESTTLETI EIHPITFPPE VLARISPELS LQRHLSLGIR PCLRKYEEFR 51 DVAIENNTLS RYADAGNIDT KNNILGSNVL KSGKTIVITS ITGGIIEETS 101 ASIKDLDDFG EEELFEVTKE EDIIANYASV YPVVEVERGR VGACIDEEMT 151 ISQKLHDSIL HSRILPKKAL KVKAGVRSAN EDGTFSVLYP DELEDDTLNE 201 TNLKMKRKWS YVLYAKIVVL SRTGPVFDLC WNSLMYALQS VKLPRAFIDE 251 RASDLRMTIR TRGRSATIRE TYEIICDQTK SVPLMINAKN IAFASNYGIV 301 ELDPECQLQN SDNSEEEEVD IDMDKLNTVL IADLDTEAEE TSIHSTISIL 351 AAPSGNYKQL TLKGGGAKIT PEMIKRSLLL SRVRADDLST RFNI
P38792	Rrp4	ac 1 MSEVITITKR NGAFQNSSNL SYNNTGISDD ENDEEDIYMH DVNSASKSES 51 DSQIVTPGEL VTDDPIWMRG HGTYFLDNMT YSSVAGTVSR VNRLLSVIPL 101 KGRYAPETGD HVVGRIAEVG NKRWKVDIGG KQHAVLMLGS VNLPGGILRR 151 KSESDELQMR SFLKEGDLLN AEVQSLFQDG SASLHTRSLK YGKLRNGMFC 201 QVPSSLIVRA KNHTHNLPGN ITVVLGVNGY IWLRKTSQMD LARDTPSANN 251 SSSIKSTGPT GAVSLNPSIT RLEEESSWQI YSDENDPSIS NNIRQAICRY 301 ANVIKALAFC EIGITQQRIV SAYEASMVYS NVGELIEKNV MESIGSDILT 351 AEKMRGNGN

Q05636	Rrp45	ac 1 MAKDIEISAS ESKFILEALR QNYRLDGRSF DQFRDVEITF GKEFGDVSVK 51 MGNTKVHCRI SCQIAQPYED RPFEGLFVIS TEISPMAGSQ FENGNITGED 101 EVLCSRIIEK SVRRSGALDV EGLCIVAGSK CWAVRADVHF LDCDGGFIDA 151 SCIAVMAGLM HFKKPDITVH GEQIIVHPVN EREPVPLGIL HIPICVTFSF 201 FNPQDTEENI KGETNSEISI IDATLKEELL RDGVLTVTLN KNREVVQVSK 251 AGGLPMDALT LMKCCHEAYS IIEKITDQIL QLLKEDSEKR NKYAAMLTSE 301 NAREI
P53859	Csl4	ac 1 MACNFQFPEI AYPGKLICPQ YGTENKDGED IIFNYVPGPG TKLIQYEHNG 51 RTLEAITATL VGTVRCEEEK KTDQEEEREG TDQSTEEEKS VDASPNDVTR 101 RTVKNILVSV LPGTEKGRKT NKYANNDFAN NLPKEGDIVL TRVTRLSLQR 151 ANVEILAVED KPSPIDSGIG SNGSGIVAAG GGSGAATFSV SQASSDLGET 201 FRGIIRSQDV RSTDRDRVKV IECFKPGDIV RAQVLSLGDG TNYYLTTARN 251 DLGVVFARAA NGAGGLMYAT DWQMMTSPVT GATEKRKCAK PF
Q12277	Rrp42	ac 1 MSLSVAEKSY LYDSLASTPS IRPDGRLPHQ FRPIEIFTDF LPSSNGSSRI 51 IASDGSECIV SIKSKVVDHH VENELLQVDV DIAGQRDDAL VVETITSLLN 101 KVLKSGSGVD SSKLQLTKKY SFKIFVDVLV ISSHSHPVSL ISFAIYSALN 151 STYLPKLISA FDDLEVEELP TFHDYDMVKL DINPPLVFIL AVVGNNMLLD 201 PAANESEVAN NGLIISWSNG KITSPIRSVA LNDSNVKSFK PHLLKQGLAM 251 VEKYAPDVVR SLENL
P48240	Mtr3	ac 1 MNVQDRRRLL GPAAAKPMAF SNTTTHVPEK KSTDLTPKGN ESEQELSLHT 51 GFIENCNGSA LVEARSLGHQ TSLITAVYGP RSIRGSFTSQ GTISIQLKNG 101 LLEKYNTNEL KEVSSFLMGI FNSVVNLSRY PKSGIDIFVY LTYDKDLTNN 151 PQDDDSQSKM MSSQISSLIP HCITSITLAL ADAGIELVDM AGAGEANGTV 201 VSFIKNGEEI VGFWKDDGDD EDLLECLDRC KEQYNRYRDL MISCLMNQET 251
P46948	Rrp41	ac 1 MSRLEIYSPE GLRLDGRRWN ELRRFESSIN THPHAADGSS YMEQGNNKII 51 TLVKGPKEPR LKSQMDTSKA LLNVSVNITK FSKFERSKSS HKNERRVLEI 101 QTSLVRMFEK NVMLNIYPRT VIDIEIHVLE QDGGIMGSLI NGITLALIDA 151 GISMFDYISG ISVGLYDTTP LLDTNSLEEN AMSTVTLGVV GKSEKLSLLL 201 VEDKIPLDRL ENVLAIGIAG AHRVRDLMDE ELRKHAQKRV SNASAR
Q08285	Rrp40	ac 1 MSTFIFPGDS FPVDPTTPVK LGPGIYCDPN TQEIRPVNTG VLHVSAKGKS 51 GVQTAYIDYS SKRYIPSVND FVIGVIIGTF SDSYKVSLQN FSSSVSLSYM 101 AFPNASKKNR PTLQVGDLVY ARVCTAEKEL EAEIECFDST TGRDAGFGIL 151 EDGMIIDVNF NFARQLLFNN DFPLLKVLAA HTKFEVAIGL NGKIWVKCEE 201 LSNTLACYRT IMECCQKNDT AAFKDIAKRQ FKEILTVKEE

P53256	Rrp46	1 MNQLRCIKYL CIYGNNKEPN TKNRLDSAEK KKKMSVQAEI GILDHVDGSS 51 EFVSQDTKVI CSVTGPIEPK ARQELPTQLA LEIIVRPAKG VATTREKVLE 101 DKLRAVLTPL ITRHCYPRQL CQITCQILES GEDEAEFSLR ELSCCINAAF 151 LALVDAGIAL NSMCASIPIA IIKDTSDIIV DPTAEQLKIS LSVHTLALEF 201 VNGGKVVKNV LLLDSNGDFN EDQLFSLLEL GEQKCQELVT NIRRIIQDNI 251 SPRLVV
Q12149	Rrp6	ac 1 MTSENPDVLL SRVINVVRAA SSLASQDVDF YKNLDRGFSK DLKSKADKLA 51 DMANEIILSI DEHHESFELK EEDISDLWNN FGNIMDNLLE MSDHSLDKLN 101 CAINSKSRGS DLQYLGEFSG KNFSPTKRVE KPQLKFKSPI DNSESHPFIP 151 LLKEKPNALK PLSESLRLVD DDENNPSHYP HPYEYEIDHQ EYSPEILQIR 201 EEIPSKSWDD SVPIWVDTST ELESMLEDLK NTKEIAVDLE HHDYRSYYGI 251 VCLMQISTRE RDYLVDTLKL RENLHILNEV FTNPSIVKVF HGAFMDIIWL 301 QRDLGLYVVG LFDTYHASKA IGLPRHSLAY LLENFANFKT SKKYQLADWR 351 IRPLSKPMTA YARADTHFLL NIYDQLRNKL IESNKLAGVL YESRNVAKRR 401 FEYSKYRPLT PSSEVYSPIE KESPWKILMY QYNIPPEREV LVRELYQWRD 451 LIARRDDESP RFVMPNQLLA ALVAYTPTDV IGVVSLTNGV TEHVRQNAKL 501 LANLIRDALR NIKNTNEEAT PIPSSETKAD GILLETISVP QIRDVMERFS 551 VLCNSNISKS RAKPVTNSSI LLGKILPREE HDIAYSKDGL PNKVKTEDIR 601 IRAQNFKSAL ANLEDIIFEI EKPLVVPVKL EEIKTVDPAS APNHSPEIDN 651 LDDLVVLKKK NIQKKQPAKE KGVTEKDAVD YSKIPNILSN KPGQNNRQQK 701 KRRFDPSSSD SNGPRAAKKR RPAAKGKNLS FKR
gi 6321873	Lrp1	1 MEDIEKIKPY VRSFSKALDE LKPEIEKLTS KSLDEQLLLL SDERAKLELI 51 NRYAYVLSSL MFANMKVLGV KDMSPILGEL KRVKSYMDKA KQYDNRITKS 101 NEKSQAEQEK AKNIISNVLD GNKNQFEPSI SRSNFQGKHT KFENDELAES 151 TTTKIIDSTD HIRKASSKKS KRLDKVGKKK GGKK
gi 6324650	Ski7	ac 1 MSLLEQLARK RIEKSKGLLS ADQSHSTSKS ASLLERLHKN RETKDNNAET 51 KRKDLKTLLA KDKVKRSDFT PNQHSVSLSL KLSALKKSNS DLEKQGKSVT 101 LDSKENELPT KRKSPDDKLN LEESWKAIKE MNHYCFLKND PCINQTDDFA 151 FTNFIIKDKK NSLSTSIPLS SQNSSFLSLK KHNNELLGIF VPCNLPKTTR 201 KVAIENFNRP SPDDIIQSAQ LNAFNEKLEN LNIKSVPKAE KKEPINLQTP 251 PTESIDIHSF IATHPLNLTC LFLGDTNAGK STLLGHLLYD LNEISMSSMR 301 ELQKKSSNLD PSSSNSFKVI LDNTKTEREN GFSMFKKVIQ VENDLLPPSS 351 TLTLIDTPGS IKYFNKETLN SILTFDPEVY VLVIDCNYDS WEKSLDGPNN 401 QIYEILKVIS YLNKNSACKK HLIILLNKAD LISWDKHRLE MIQSELNYVL 451 KENFQWTDAE FQFIPCSGLL GSNLNKTENI TKSKYKSEFD SINYVPEWYE 501 GPTFFSQLYL LVEHNMNKIE TTLEEPFVGT ILQSSVLQPI AEINYVSLKV 551 LINSGYIQSG QTIEIHTQYE DFHYYGIVSR MKNSKQILET NTKNNISVGL 601 NPDILEVLVK IHNTEDFTKK QFHIRKGDII IHSRKTNTLS PNLPNTLKLL 651 ALRLIKLSIQ THALSDPVDL GSELLLYHNL THNAVKLVKI LGTNDISINP 701 NQSLIVEVEI IEPDFALNVI DSKYITNNIV LTSIDHKVIA VGRIACQ

[a] Protein sequences are from Swiss Prot and NCBI databases from October 2005.

[b] bold dark-grey indicates matched peptides by LC MS/MS in tryptic and tryptic + GluC digest, bold light-grey indicates peptides only found in tryptic digest, bold black indicates peptides only found in tryptic + GluC digest, black indicates peptides not found in LC MS/MS. Ac indicates acetylated residue. White with dark background denotes changes in primary amino acid sequence detected by LC MS/MS
Highlighted amino acids in dark grey (S and T) show substoichiometric phosphorylation site
The underlined N-terminus in Rrp46 represents additional amino acids for the long form of this protein as found in the database

Chapter 4

Comparative study of the nuclear and cytoplasmic exosome by mass spectrometry

Silvia A. Synowsky
Reinout Raijmakers
Martin van Wijk
Albert J. R. Heck

Biomolecular Mass Spectrometry and Proteomics Group, Bijvoet Center for Biomolecular Research & Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Sorbonnelaan 16, 3584 CA Utrecht, The Netherlands

Manuscript in preparation

Abstract

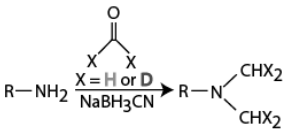
The analysis of endogenously expressed heterogeneous protein complexes is challenging and usually limited to reporting constituents of protein complexes. Different sub-cellular localization and the dynamic nature of closely related protein complexes, which execute different function, are often overlooked. In this report we combine the tandem affinity purification (TAP) approach with quantitation procedures and macromolecular mass spectrometry to gain better insight in different but closely related protein assemblies.

The exosome complex fulfills several different functions in RNA metabolism such as degrading mRNA and processing rRNA, snRNA and snoRNA. To be able to distinguish between these different substrates there exist at least two forms in the eukaryotic cell. These are localized in the cytoplasm and in the nucleus. Both protein complexes have most proteins in common but also differ in their composition by few proteins.

Here we investigate differences between the cytoplasmic and the nuclear form of the exosome. We purify different exosomes by selecting a nuclear specific exosome protein (Rrp6), a cytoplasmic specific protein (Ski7) and a common exosome protein (Csl4) as the TAP tagged bait protein. First we probe both exosome assemblies with a stable isotope labelling technique to detect relative differences among them in their protein complex composition and probe the relative abundance of their post-translational modification status. Then the nuclear exosome is analyzed by macromolecular mass spectrometry to determine the total complex mass and to give insight in subcomplex composition, stoichiometry, relative protein binding strength and the mode of complex dissociation in the gas-phase.

Introduction

The assembly of individual proteins to form protein complexes is very often crucial as these complexes execute most functions in the cell. But the analysis of endogenously expressed protein complexes is challenging due to low abundance of protein complexes in the cell, transient interactions and the dynamic nature of protein complexes. In recent years many single proteins were categorized into protein complexes. A powerful proven strategy is hereby the tandem affinity puritication approach in which endogenously expressed protein complexes are selectively isolated with a high purity [1, 2]. Proteome wide studies have identified many protein complexes of *S. cerevisiae* and mapped them into a protein network [3, 4]. It became clear that many proteins are shared between different protein complexes. However, quantitative differences of closely related protein assemblies that have many proteins in common are usually overlooked. Several quantitation techniques based on stable isotope labelling have been established in proteomic research [5-10]. Such quantitative methods can be utilized in the analysis of related protein complexes to reveal differences in the abundance of individual protein complex constituents [11] but can also give insight into differentially regulated post-translational modifications of protein complex components based on their respective peptides. One of these techniques is based on dimethyl labelling, a chemical labelling which can be applied on peptides after digestion, which is based on reductive amination using formaldehyde and cyanoborohydride. The reaction introduces an additional mass of 28 mass units to each derivatized ε-amino group of lysines and N-termini of peptides (reaction 1). It has been shown that the reaction is fast and efficient with no detectable byproducts formed. An additional four mass units in each labelling event can be achieved when deuterated instead of light formaldehyde is being used, giving a mass difference of 4 Dalton between the heavy and light label [12, 13].



(1)

The differences between related protein complexes can be further investigated by native macromolecular mass spectrometry. This elegant but challenging MS approach allows studying differences in stoichiometry of related protein assemblies. The technique keeps the protein complexes (semi) intact and gives information about the total complex mass and the most stable core (sub-) complex. Relatively new tandem MS techniques on intact protein complexes permit the selective isolation of defined charge states of protein complex ions that can be subsequently exposed to collision induced gas-phase dissociations. The ejection

of proteins from the protein assembly gives information about the relative binding strength of protein complex constituents in the gas-phase.

We exemplify our MS-strategies on the yeast exosome and compare the nuclear with the cytoplasmic variant and the nuclear with a mixed population of the both variants. The exosome is a multisubunit protein assembly consisting of several 3'-> 5' exoribonucleases. Most RNA is processed after transcription by the exosome. It functions in mRNA degradation and snRNA, snoRNA and rRNA processing. Given the functional diversity of the exosome it is suggested that the exosome functions in these different cellular processes by the association of different co-factors in the cell [14-16] and by localization in different cellular compartments [17]. The exosome is a conserved protein complex and is present in Archeael, trypanosomal, yeast and mammalian cells. The composition, overall structure, stoichiometry and relative strength of protein binding of the cytoplasmic yeast exosome have been investigated by macromolecular mass spectrometry [18, 19]. The X-ray structure of the archeael exosome [20] and very recently the X-ray structure of the core complex of the mammalian exosome has been determined [21]. These studies were exclusively performed on the cytoplasmic form of the exosome. The cytoplasmic and nuclear exosome complexes have most proteins in common. The cytoplasmic yeast exosome contains 10 different proteins that range between 25 – 115 kDa in molecular weight. It is well established that the exosome forms a hexameric ring (Rrp41-Rrp42-Mtr3-Rrp43-Rrp46-Rrp45) with three proteins on top of the ring that contain a S1-binding domain (Rrp4, Csl4, Rrp40). Additionally it contains one large protein (Dis3) and is associated with Ski7 in the cytoplasm [14]. The nuclear exosome lacks Ski7 but contains several additional proteins. Rrp6 and Lrp1 are assumed to be present exclusively in the nuclear exosome [22]. Here we set out to investigate differences in abundance of proteins and their phosphorylation status in closely related protein complexes, the cytoplasmic and nuclear exosome, with stable isotope dimethyl labelling. Moreover we perform native mass spectrometry on the heterogeneous nuclear exosome assembly to determine stoichiometry and core components.

Material and Methods

Yeast strain, cultivation and protein purification

The *S. cerevisiae* strain MGD35313D, BSY17 containing Rrp6, Ski7 or Csl4 as the C-terminal tagged entry point was purchased at Euroscarf (Frankfurt, Germany) [3]. S. cerevisiae cells with different entry points were grown separately in 20 L yeast peptone dextrose at 30 ºC to an optical density at 600 nm of 3.5. The cells were harvested and mechanically lysed with zirkonia beads in a bead beater (Biospec Products, LabServices BV, The Netherlands) [23] in 350 mM Hepes pH 8.0 containing 350 mM NaCl, 1.5 mM $MgCl_2$, 0.15% (v/v) Igepal + 0.5 mM DTT, 1/1000 (v/v) protease inhibitor (leupeptin,

aprotinin, benzamidine), 1 mM pepstatin, 1 mM PMSF resulting in 150 ml lysate. The purification was performed in parallel using Rrp6, Ski7 and Csl4 as entry points (EP). Please note that the purification protocol was scaled up 3 fold using EP Rrp6 due to the lower abundance of the nuclear exosome in the cell. In the first affinity purification step 10 ml of lysate with EP Csl4 and Ski7 and 30 ml of lysate with the EP Rrp6 (~200 µg/ml total protein) were incubated with 200 µl IgG-Sepharose-beads (Amersham) for EP Csl4 and Ski7 and 500 µl IgG-Sepharose beads for EP Rrp6 for 150 min at 4 ºC. Subsequently the beads were thoroughly washed with 40 ml of a buffer A containing 20 mM Tris, pH 8.0, 350 mM NaCl, 1.5 mM $MgCl_2$, 0.15% (v/v) Igepal, 0.5 mM DTT and the protein complex was eluted from the IgG beads by incubation with 100 units of tobacco etch virus protease (Invitrogen) for 90 min at room temperature. In the second affinity purification step the protein complex was incubated with 200 µl calmodulin beads for EP Csl4 and Ski7 and 600 µl calmodulin beads (Stratagene) for EP Rrp6 in the presence of 2 mM $CaCl_2$ for 60min at 4 ºC. Prior to elution the beads were washed with buffer A plus 2 mM $CaCl_2$ and the protein complex was eluted using buffer A containing 5 mM EGTA.

Protein separation, in gel-digestion, in solution digestion, labelling protocol and LTQ-FT-ICR analysis

The different exosome proteins, retrieved with each EP, were separated on a 10% gel by one-dimensional-polyacrylamid gel electrophoresis and stained using 0.1% (v/v) coomassie brilliant blue G250 and destained overnight in 12% (v/v) acetic acid and 50% (v/v) methanol. Protein bands were excised, reduced and alkylated and tryptically digested (10 ng/mL trypsin) (Roche, The Netherlands) according to Shevchenko et al [24] for 8h at 37 ºC. Peptides were acidified using 2.5% formic acid (v/v).

Identification of the proteins was achieved by LC MS/MS analysis using an LTQ-FT mass spectrometer (Thermoelectron, Bremen, Germany) coupled to an Agilent 1100 series LC system (vacuum degasser, auto sampler, and one high-pressure mixing binary pump without static mixer). Peptide were trapped on a pre-column (Reprosil C18AQ; 20 mm x 100 µm, packed in-house) at 5 µl/min 100% eluent A (0.1 M acetic acid). By using a splitter, the peptides were eluted and separated at a rate of 100 nl/min to the analytical column (Reprosil C18AQ 200 mm x 50 µm, packed in-house) with a linear gradient from 0 to 50% eluent B (0.1 M acetic acid in 80% (v/v) acetonitrile) for 45 min. The LTQ-FT was operated in positive ion mode. MS/MS scans from the eight most intense peaks in the MS mode were acquired. Peak lists were created using dta files which were combined to mgf. files. Only peak intensities above 25 were accepted. The Mascot algorithm [25] (Matrix Science Ltd., version 2.1) was used to interpret the raw MS/MS data. The data were run against the complete non-redundant proteome database Swiss Prot (searchdate: 23.06.06) of *S. cerevisiae*. The algorithm was set to use trypsin as enzyme, allowing at maximum for two missed cleavages and assuming carbamidomethyl on cysteines as a fixed modification and oxidized

methionine, N-terminal acetylation, acetylated lysine and phosphorylated serine/tyrosine/threonine as variable modifications. The peptide tolerance was fixed to 10 ppm and the MS/MS tolerance to 0.9 Da.

For dimethyl labelling the exosome samples isolated with EP Rrp6 and EP Csl4 were concentrated and the buffer was exchanged to 100 µl 100 mM triethylammonium bicarbonate pH 8.0 and an in-solution digest was performed. Proteins were reduced and alkylated and tryptically digested (10 ng/mL trypsin) (Roche, The Netherlands) and incubated for 8 h at 37 ºC. In the dimethyl labelling intended to compare the EP Rrp6- and EP Ski7-exosome, a gel based approach for the tryptic digest was used instead of the in-solution digest. Subsequently peptides from EP Rrp6 and EP Csl4 exosome were labelled with 0.16% (v/v) heavy and light formaldehyde respectively and mixed with 22.5 mM sodium cyanoborohydride by incubation at room temperature for 45 min. The reaction was quenched by the addition of 0.15% ammonia. Finally the peptides were acidified with 2.5% formic acid.

LC-MS/(MS)analysis of phosphorylated amino acid residues

The labelled tryptic peptides of the in-gel digested exosome of EP Rrp6 (heavy labelled) with EP Ski7 (light labelled) and of the in-solution digested EP Rrp6 (heavy labelled) with EP Csl4 (light labelled) were mixed in equal amounts. A quantitative analysis of the peptide mixture was achieved by a two-dimensional chromatographic technique using an Agilent 1100 series LC system for the selective enrichment of phosphopeptides directly coupled to an LTQ-FT-ICR-MS (Thermoelectron, Bremen, Germany) essentially as described in [26]. The trap column consisted of three subsequent separate pre-column in a sandwich alignment (30 mm × 100 µm Aqua C18 precolumn, followed by a 5 mm × 100 µm TiO_2 precolumn, followed by another 30 mm × 100 µm Aqua C18 precolumn). Peptides were trapped at a flow rate of 3 µl/min 100 % eluent A (0.1 M acetic acid) on the first 30 mm Aqua C18 pre-column. Elution is achieved in a 70 min gradient from 0 to 40% solvent B (80% acetonitrile, 0.1 M acetic acid) at a flow rate of ~100 nl/min. While phosphopeptides and acidic peptides bind to the second pre-column consisting of TiO_2, all other peptides were separated on the analytical column (200 mm × 50 µm ReproSil-Pur C18-AQ) and subsequently detected. The elution of peptides bound to TiO_2 was achieved by successive injection of 30 µl 250 mM ammonium bicarbonate pH 9 and 20 µl of 5% formic acid followed by another 70 min gradient from 0 – 40% solvent B. LTQ-FT-ICR-instrument settings were similar as described above. Peak lists using an intensity cut-off of 10 were created. The data was initially searched in a non-redundant database NCBI (searchdate 04.01.08) using a maximum of 3 missed cleavages, carbamidomethyl (C), as a fixed modification and acetylation (N-term), oxidized methionine, dimethylation (K, N-term), dimethyl-D2 (K, N-term) as variable modification as Mascot parameters and then in an in-house database containing only exosome proteins. For this Mascot search the following parameters were

used: trypsin as enzyme, allowing for four missed cleavages and assuming carbamidomethyl (C), as a fixed modification and actylation (N-term), deamidation (NQ), oxidized methionine, dimethylation (K, N-term), dimethyl-D2 (K, N-term) and phosphorylation (STY) as variable modifications. The peptide tolerance was each time fixed to 10 ppm and the MS/MS tolerance to 0.9 Da. Proteins were quantified and manually examined using MS Quant (version 1.4.2).

Macromolecular mass spectrometry

The TAP purified exosomes using either Rrp6, Ski7 or Csl4 as entry point were concentrated and subjected to buffer exchange to 150 mM ammonium acetate solution, pH 6.8 using ultracentrifugation filter devices with a cut-off of 100 kDa (Millipore). The protein solution was sprayed from in-house gold coated nano-electrospray needles (Kwik-Fil, World Precision Instruments, Sarasota, FL). Borosilicate capillaries were pulled (P-97 puller, Sutter Instruments, Novato , CA) and subsequently coated with gold (Edwards Laboratories, Milpitas, CA). Electrospray ionization mass spectra were recorded using a commercially available orthogonal time of flight mass spectrometer (LCT, Waters, UK) equipped with a nano-electrospray Z-source. To allow a transfer of large ions to the mass spectrometer, the pressure between the sample cone and extraction cone was raised to 9 mbar. Prior to the analysis the mass spectrometer was calibrated with a 20 mg/ml CsI solution. Typical spray conditions were a capillary voltage of 1200 V and a sample cone voltage of 150 V. Gas phase dissociation experiments were performed on a QTOF1 mass spectrometer (Waters, UK) [27] modified for high mass operation. These modifications allow the increase of the pressure between sample cone and extraction cone and in the first hexapole that is surrounded by a flow restriction sleeve. Moreover a low frequency RF generator for the quadrupole enables the analysis of ions up to 30,000 m/z. The collision cell was exchanged to be compatible with higher gas pressure. High transmission grids and a low frequency pusher were implemented to facilitate the transfer of ions to the detector. For gas phase dissociation experiments a precursor is selected in the quadrupole and fragmented in the xenon-filled collision cell. These fragments are separated in the time of flight analyzer and subsequently detected. Typical values are 1400 V for the capillary voltage and 150 V for the cone voltage. To observe fragmentation the collision voltage was increased from 50 V to 175 V. All mass spectra were analyzed by mass lynx software (version 4.0).

Results

The composition, structure and stoichiometry of the cytoplasmic exosome have been documented but so far little is known about changes in exosome complexes that are present in different sub-cellular compartments. Here, the exosome protein complex was selectively

retrieved by TAP using three different entry points (Fig.1). Ski7p and Rrp6p are unique components for the cytoplasmic and nuclear exosome respectively and therefore when used as entry point yields only one of these exosome complexes. However, Csl4p is a common protein between the cytoplasmic and nuclear exosome and using that entry point allows the purification of a mixture of both sub-cellular exosome protein complexes. Differences between exosomes using these three proteins as TAP tagged entry points were investigated by (quantitative) proteomics and macromolecular mass spectrometry.

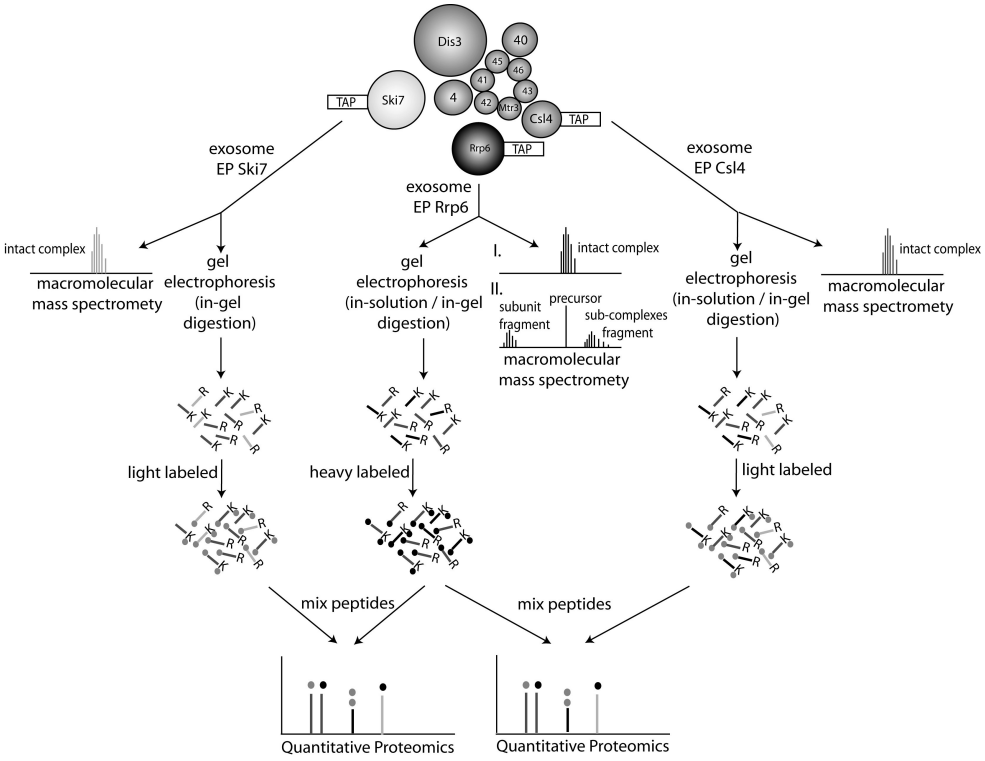


Figure 1

Schematic representation of the experimental procedure to investigate relative changes of different exosome protein complexes

The exosome protein complex is isolated from S. cerevisiae using three different entry points: Ski7p, Rrp6p and Csl4p. Differences between these exosomes are investigated by a 2-dimensional mass spectrometry approach. In the first MS-technique the exosome proteins are separated by gel-electrophoresis. The proteins are excised and digested into peptides by trypsin. After identification of the proteins by mass spectrometry, the peptides are chemically labelled (Ski7-exosome and Csl4-exosome with light dimethyl-labelling and Rrp6-exosome with heavy dimethyl-labelling) and mixed in equal amounts. Information about potential relative differences in the abundance of proteins as well as differential post-translational modifications is obtained by a quantitative proteomics analysis. In the second MS-

technique the non-covalent protein complexes are investigated by macromolecular (tandem) mass spectrometry. In this approach the overall mass of the intact protein complex and relative protein binding affinities in the protein complex are determined.

In the first mass spectrometric approach, the exosome proteins were separated on a 1-D-SDS gel. The excised proteins were then tryptically digested and identified by nano-LC-MS/MS. The remaining peptides were chemically derivatized using stable isotope dimethyl labelling. Peptides of the exosome isolated with entry point Ski7p and Csl4p were labelled with light stable isotopes (using CH_2O) whereas the peptides originating from Rrp6p as entry point were modified with heavy stable isotopes (using CD_2O). Heavy and light labelled peptides were then mixed in equal amounts and peptide pairs were analyzed and identified using nano-ESI-LC-MS/MS. This analysis reveals relative differences on the peptide level of the exosome. The proteins can be quantified relative to each other and differences in post-translational modifications, in particular of phosphorylations can be obtained. In the second mass spectrometric analysis, the different exosomes were investigated using macromolecular mass spectrometry. In this approach the non-covalent interactions are kept intact and therefore permit the determination of the total mass of the intact protein complex. Additional information is obtained by tandem mass spectrometry of protein complexes and allows conclusions about protein stoichiometry and relative protein binding affinities in the gas-phase.

Identification of protein components of different exosomes using nano-LC-MS/MS

Initially exosomes using Rrp6p, Ski7p and Csl4p as entry points were analyzed by 1-D-gel electrophoresis (Fig. 2). The gel shows clear differences between the purifications. A mass spectrometric analysis identified the proteins in the pulldown using EP Rrp6 as the known components of the nuclear exosome (Rrp41, Rrp42, Mtr3, Rrp43, Rrp46, Rrp45, Rrp40, Csl4, Rrp4, Dis3, Rrp6 and Lrp1).

With similar gel band intensity to the exosome components, Srp1 and Kap95 were found which together form the importin α/β heterodimer. Additionally, the hypothetical protein Ynr024w was identified. The protein sequence of Ynr024w shows little but significant amino acid sequence similarities with M-phase phosphoprotein 6 (MPP6) from several organism such as human, drosophila, bovine and xenopus. MPP6 has been previously found be in complex specifically with the nuclear human exosome [28]. It functions as a co-factor in the maturation in the 5.8S rRNA. The amino acid sequence of Ynr024w exhibits basic stretches of amino acid residues in the middle of the sequence and at the C-terminus, which are predicted to bind to RNA [29, 30]. The absolute abundance of this protein is determined to 1,350 molecules per cell [31]. As most of the core exosome components are present in 3,180 to 10,800 molecules per cell, except Dis3 with a reported copy number of 606 molecules per cell, it is likely that Ynr024w is only bound to the nuclear fraction of the

exosome. Furthermore subcellular localization studies using GFP-fusion proteins reveal Ynr024w to be in the nucleus [32].

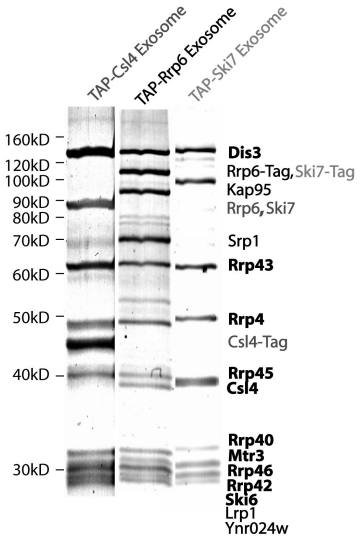


Figure 2

Purification of the S.cerevisiae exosome by using Rrp6, Csl4 and Ski7 as the C-terminal entry point.

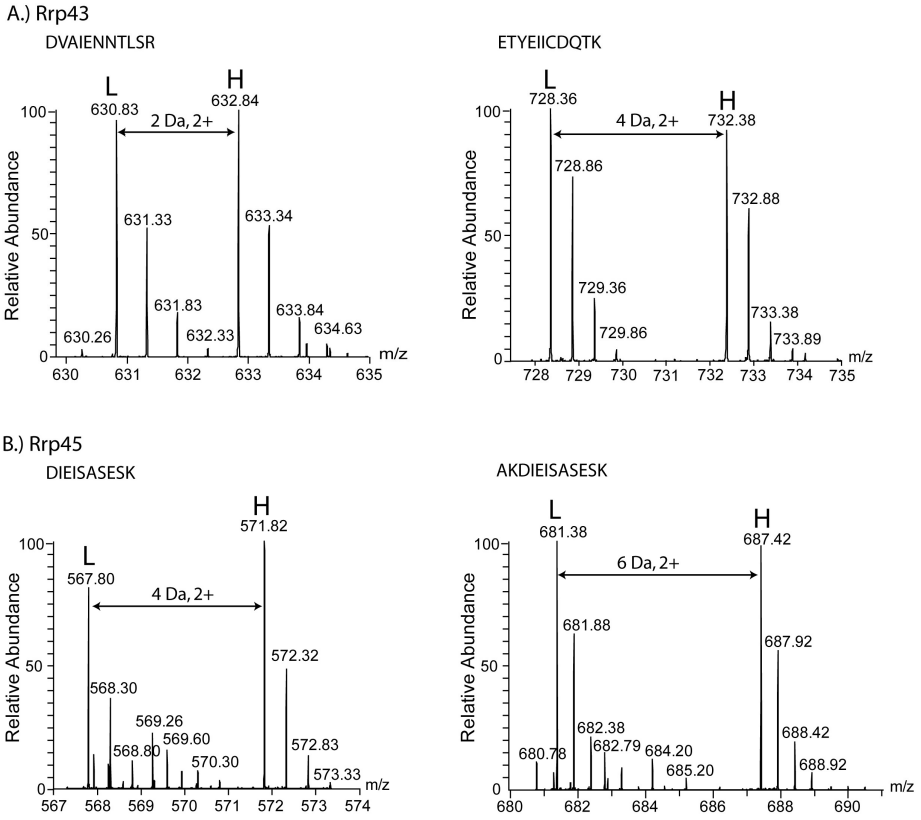
Coomassie stained 1-dimensional SDS gel of separated proteins from the tandem affinity purification of the exosome protein complex. In the left lane the proteins of the exosome using Csl4 as TAP-entry point are visualized. Next in the middle lane the protein complex retrieved using Rrp6 as entry point is seen. And the right lane shows the exosome complex isolated by Ski7 as entry point. At the left side the molecular weight marker is denoted. The proteins were identified by nano-ESI-LC-MS/MS. Protein common in all purifications are annotated in bold black, specific proteins for Csl4, Rrp6 and Ski7 are marked with dark-grey, black and light-grey respectively.

When Csl4 was used as an entry point, mass spectrometric analysis detected a mixture of proteins originating from the nuclear and cytoplasmic exosome. Common exosomal proteins between nuclear and cytoplasmic variant (Rrp41, Rrp42, Mtr3, Rrp43, Rrp46, Rrp45, Rrp40, Csl4, Rrp4 and Dis3) and cytoplasmic and nuclear specific proteins, Ski7and Rrp6, were found. The total exosome distribution in the cell is believed to be split into 1/5 of nuclear and 4/5 of cytoplasmic exosome [22]. Therefore using EP Csl4 mainly isolates the cytoplasmic exosome and in low abundance the nuclear exosome. This finding is supported with the detection of only a few peptides of the nuclear component Rrp6.

The isolation of the exosome using EP Ski7 results in a pure pool of cytoplasmic exosome, as only common exosomal and cytoplasmic specific components (Rrp41, Rrp42, Mtr3, Rrp43, Rrp46, Rrp45, Rrp40, Csl4, Rrp4, Dis3 and Ski7) were identified. Interestingly, also importin α/β was not detected in the latter purification.

Relative quantitation of exosome complexes by stable isotope dimethyl labelling

To determine and compare relative abundance of each protein in the cytoplasmic and nuclear exosome we utilized stable isotope dimethyl labelling to quantitate all proteins of the exosome. The dimethyl labelling yielded 95 % of all peptides to be either heavy or light labelled (data not shown). Peptides originating from the exosome using EP Rrp6 were labelled with heavy formaldehyde whereas peptides using EP Ski7 and Csl4 were labelled with light formaldehyde. By mixing heavy labelled exosomal peptides from EP Rrp6 with light labelled exosomal peptides from either EP Ski7 or EP Csl4 in equal ratio and analyzing peptide pairs by quantitative mass spectrometry the consistency and differences between these exosome complexes become apparent. The correct mixing proportions was first inspected on several peptide pairs originating from several proteins that form the stable hexameric ring and are likely not to change between different exosomes (Fig. 3).



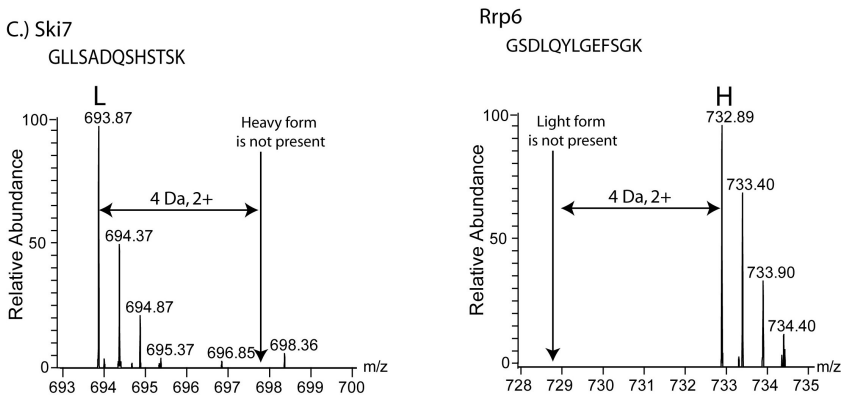


Figure 3

Examination of ratios of labelled peptides from exosome proteins

Doubly charged heavy (H) and light (L) labelled tryptic peptides originating from the ring component Rrp43 (A.) and Rrp45 (B.), which show a ratio of 1. In contrast, the cytoplasmic and nuclear specific proteins Ski7 and Rrp6 in C.) do not show an equal ratio. The peptides are derivatized with dimethyl groups adding 28 mass units to the original peptide for the light label and 32 mass units for the heavy label.

An overview of the quantitative analysis between Rrp6 and Ski7 and Rrp6 and Csl4 is given in table 1.

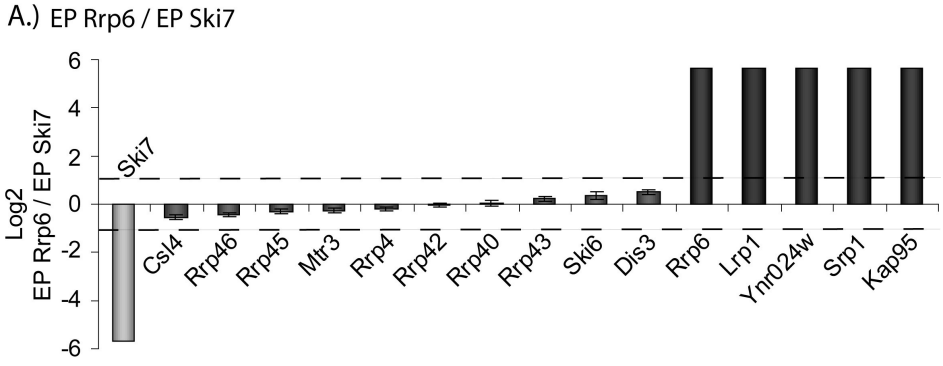
Table1 Overview of quantitative proteomics dataset used in figure 4

All peptide pairs used for ratio calculations were manually examined.

EP Rrp6 / EP Ski7					
protein	total number of peptides	no of peptides used for quantitation	ratio	stdev	log ratio
Ski7	16	1	0.02	-	-5.64
Csl4	16	16	0.7	0.15	-0.54
Rrp46	9	9	0.76	0.16	-0.44
Rrp45	13	11	0.83	0.16	-0.3
Mtr3	5	3	0.85	0.21	-0.27
Rrp4	15	13	0.88	0.17	-0.2
Rrp42	7	7	0.98	0.18	-0.04
Rrp40	9	6	1.06	0.25	0.04
Rrp43	14	14	1.19	0.3	0.22
Ski6	6	4	1.34	0.51	0.36
Dis3	36	30	1.44	0.31	0.5
Rrp6	29	0	50	-	5.64
Lrp1	11	0	50	-	5.64

Ynr024w	5	0	50	-	5.64
Srp1	9	0	50	-	5.64
Kap95	7	0	50	-	5.64
EP Rrp6 / EP Csl4					
protein	total number of peptides	no of peptides used for quantitation	ratio	stdev	log ratio
Ski7	17	1	0.02	-	-5.64
Rrp40	16	2	0.54	0.35	-0.89
Rrp42	8	4	0.76	0.09	-0.40
Mtr3	12	3	0.78	0.89	-0.36
Rrp46	11	4	1.05	0.69	0.07
Rrp43	21	11	1.13	0.72	0.18
Csl4	16	9	1.14	0.5	0.19
Ski6	16	5	1.27	1.07	0.34
Rrp4	24	6	1.43	1.49	0.52
Dis3	63	19	1.64	1.22	0.71
Rrp45	21	8	2.73	1.66	1.45
Rrp6	26	2	20.89	15.37	4.38
Lrp1	7	0	50	-	5.64
Kap95	10	0	50	-	5.64
Srp1	16	0	50	-	5.64
Ynr024w	10	0	50	-	5.64

The comparison of the exosome purified with Rrp6 and Ski7 should give the most drastic changes since these two entry points retrieve selectively the nuclear and cytoplasmic form of the exosome (Fig. 4A).



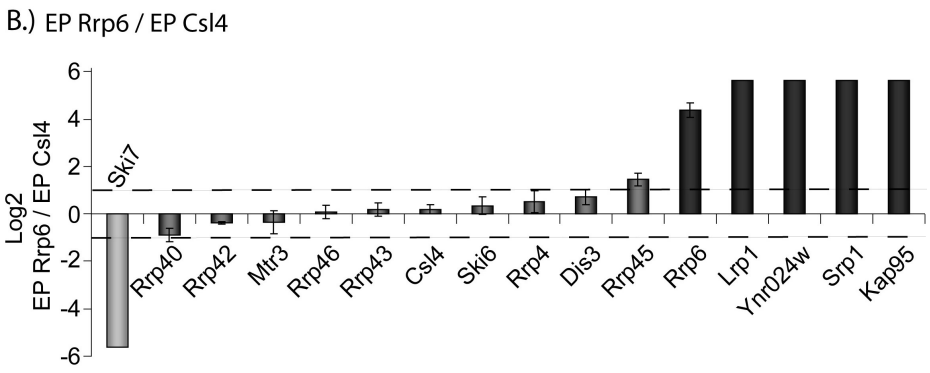


Figure 4

Comparative analysis of different exosome assemblies in the cell

Differences between the exosome purified with Rrp6/Ski7 and with Rrp6/Csl4 are illustrated in 3A and B respectively. Using Rrp6 and Ski7 shows only distinct changes in the nuclear and cytoplasmic specific proteins while the common exosome proteins stay unchanged (A). In the comparison of Rrp6 with Csl4, all common exosome proteins are unchanged except Rrp45, which seems to be higher abundant in the nuclear preparation. Some peptides of nuclear Rrp6 were detected in the Csl4-purified sample, confirming an exosome mixture was isolated (B). Only differences in the proteins of > 1 are considered to be significant.

The data reveals no differences between the common exosome proteins (Rrp41, Rrp42, Mtr3, Rrp43, Rrp46, Rrp45, Rrp4, Rrp40, Csl4 and Dis3). Clear differences between both exosomes were observed in the presence or absence of Rrp6, Lrp1, Ynr024w, Srp1, Kap95 and Ski7 (Fig. 4 A). In this experiment only a shift of > 1 is considered to be significant.

A change in the level of Rrp6 could be clearly observed in figure 4B. The high abundance of Rrp45 in the nuclear exosome was very surprising. It suggests that a subset of the exosome contains more than one copy of Rrp45.

The same approach can also be used to investigate differential post-translational modifications. The exosome assembly has been shown to contain several phosphorylation sites that are distributed over at least seven different proteins. Whether or not phosphorylation patterns change between different exosomes is still unknown. For this purpose we have selected the pure nuclear heavy labelled exosome isolated using EP Rrp6 and the light labelled strictly cytoplasmic exosome purified by Ski7. Nano-LC-MS/MS analysis with specific enrichment of phosphopeptides using TiO_2 allowed the identification of five different phosphorylation sites in four different exosome proteins (table 2). Four out of these six phosphorylation sites were previously reported [18] and two are novel.

Table 2 Overview over identified phophorylation sites in the yeast exosome

protein	peptide sequence		phosphorylated residue	previously reported
Csl4	EGTDQSTEEEKSVDASPNDVTR	aa 79 - 100	Ser90	-
Csl4	SVDASPNDVTR	aa 90 - 100	Ser94	X
Rrp46	IIQDNISPR	aa 244 - 253	Ser251	X
Rrp6	RFDPS***SS***DSNGPR	aa 703 - 715	Ser708 or Ser709	-
Ski7	KSNSDLEKQGKSVTLDSK	aa 87 - 104	Ser88 and Ser90	X

Additionally we could now relatively quantify all identified phosphorylation sites. The phosphorylation on Serine 94 is present in the nuclear and cytoplasmic Csl4. It appears to be approximately twice as abundant in Csl4 originating from the nuclear preparation compared to the cytoplasmic Csl4 (Fig 5 A, left panel). To determine significance of this ratio, the ratio of the phosphopeptides was related to all not phosphorylated peptides originating from Csl4. Normalized Csl4 peptides against all exosome ring proteins show a log ratio of -0.46 ± 0.15. The normalized log ratio of the phosphopeptide was 1.17 ± 0.10 (averaged over two charge states of the peptide and one charge state of a miscleaved peptide containing the same modification). The phosphorylation level is significantly changed between the nuclear and cytoplasmic Csl4 protein, with a P-value of > 0.001, as determined by a one-sided, two tailed T-test. The functional implications of this differentially regulated phosphorylation site remains to be investigated.

At the first glance, the phosphorylation site of Rrp46 seems to be inversely upregulated (Fig. 5 B, left panel). By determining an overall log ratio of peptides of Rrp46 with a standard deviation of -0.36 ± 0.15 it became clear that the log ratio of the potentially differential phosphorylation of -0.25 (one doubly charged phosphopeptdide) falls within one standard deviation. It is therefore likely that the phosphorylation status of this particular phosphopeptide stays constant between the nuclear and cytoplasmic exosome.

Two additional phosphopeptides were found that are attributed to either the cytoplasmic exosome or the nuclear exosome (Fig. 5 C). The corresponding peptide pair is missing as the phosphopeptides are found in exculsively cytoplasmic and nuclear specific exosome proteins; Ski7 and Rrp6 respectively.

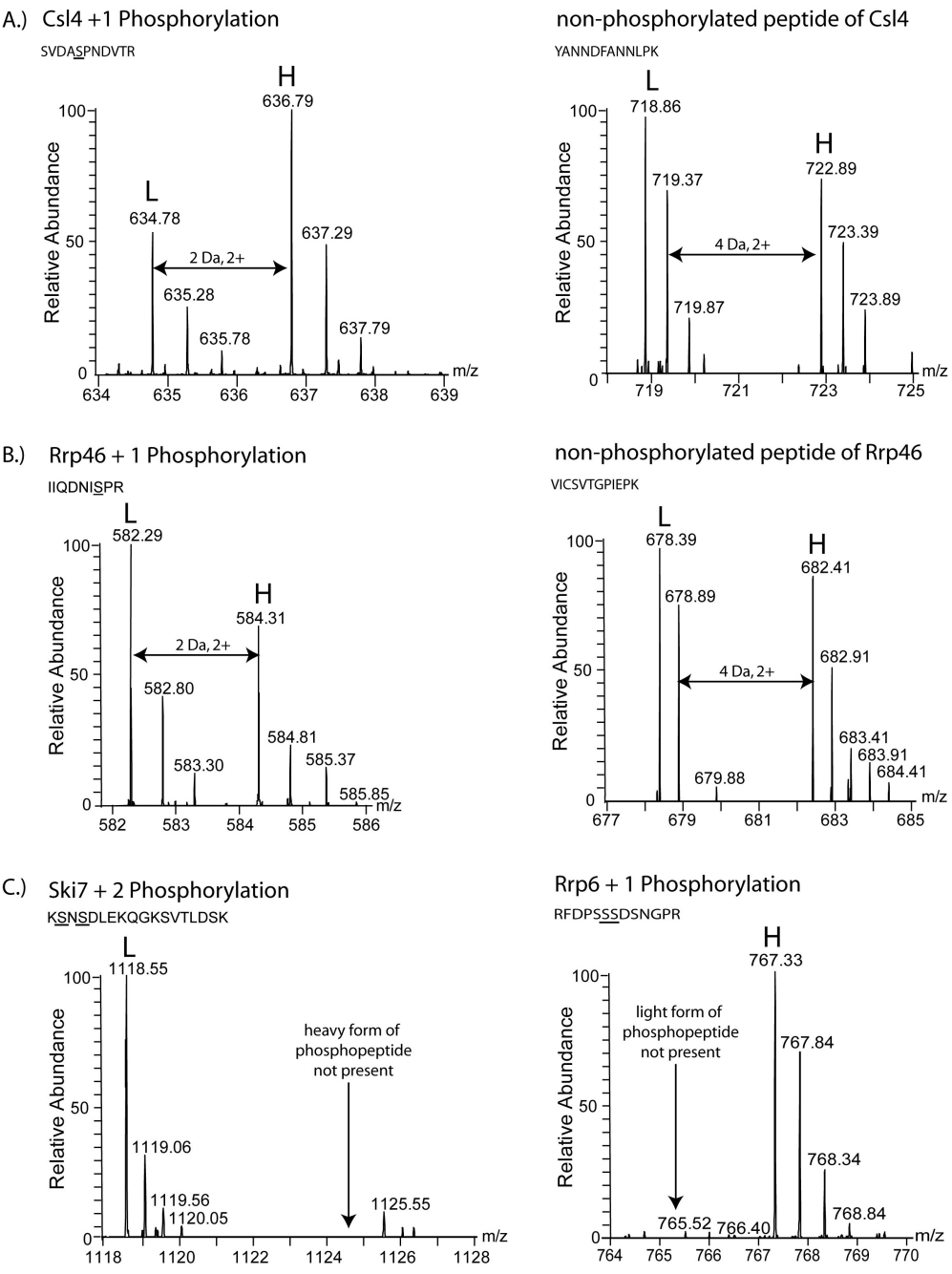


Figure 5

Relative quantitation of four different phosphorylated exosome peptides by stable isotope dimethyl labelling (A – C).

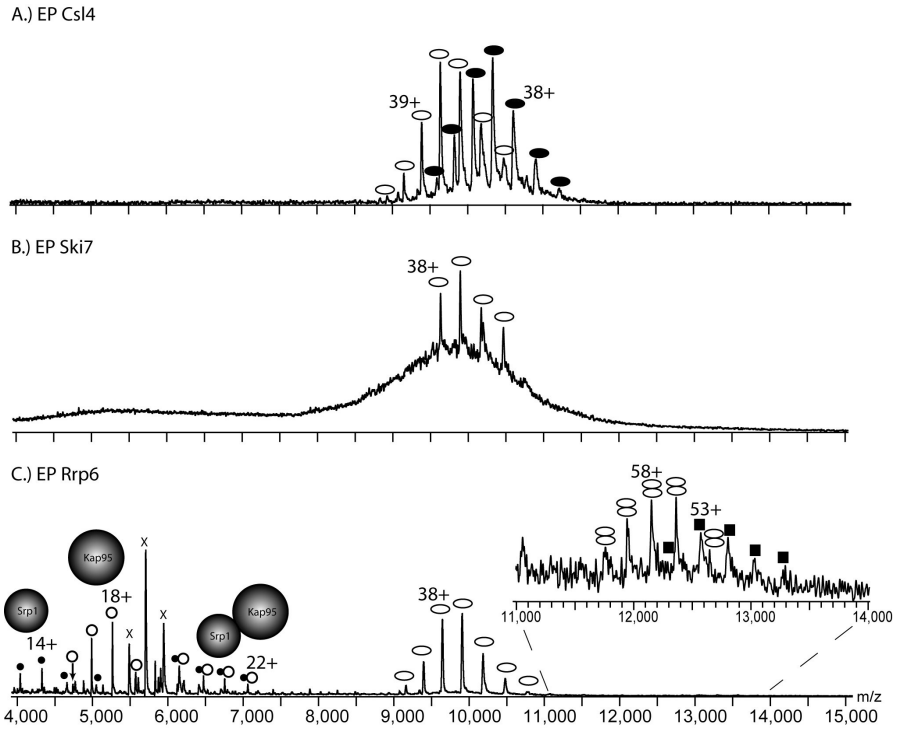
A.) The ratio Csl4-phosphopeptide suggests approximately a two-fold higher abundance of this phosphorylation site in the heavy labelled exosome, hence in the nucleus. In contrast the non-phosphorylated peptide in Csl4 shows an inverse ratio, indicating the difference in phosphorylation of Csl4 to be significant.

B.) The ratio of the phosphorylated peptide is in line with a non-phosphorylated peptide of Rrp46 suggesting no changes between the cytoplasmic and nuclear exosome.

C.) The light and heavy labelled peptides originate from the pure cytoplasmic and nuclear exosome respectively. The two phosphopeptides that were assigned to either Ski7 or Rrp6 could only be found in one of the exosome preparations and are therefore clearly specific for their sub-cellular localization.

Investigation of different exosomes by macromolecular mass spectrometry

To validate our quantitation results and investigate differences of the exosome while non-covalent interactions are maintained, macromolecular mass spectrometry was performed. This technique focuses on the analysis of the intact exosome assembly and its sub-complexes. Even when only minor amounts of endogenously expressed protein complex are isolated, this MS-approach is ideal to probe absolute stoichiometries, topology and gas-phase binding affinities of exosome components.



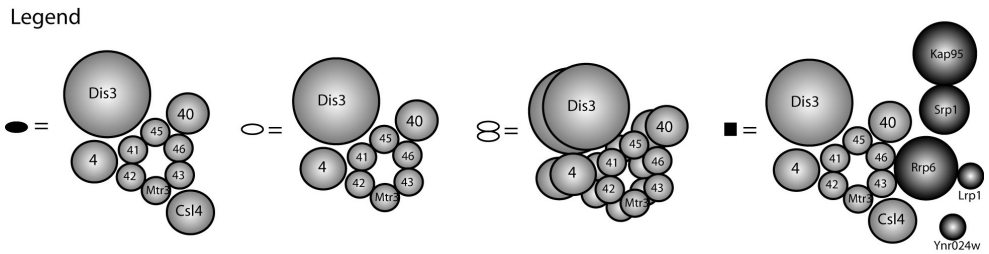


Figure 6

Macromolecular mass spectra of the TAP-tagged exosome from *S.cerevisiae* using three different proteins as entry points

A.) The exosome assembly purified via Csl4. This mass spectrum shows two main charge envelopes. The first charge series (⬭) corresponds to an exosome complex with a mass of 366, 700 kDa consisting of ring proteins (Rrp42, Rrp41, Rrp45, Rrp46, Rrp43, Mtr3), Rrp4, Rrp40 and Dis3. The second charge distribution (⬬) with a mass of 403,100 kDa relates to an exosome with the same protein composition plus Csl4.

B.) The exosome assembly retrieved by using Ski7 as EP gives rise to only one charge envelope (⬭) with a mass of ~367,100 kDa and has identical composition as seen in A.).

C.) The exosome assembly isolated by Rrp6 as EP. Several different charge distributions are observed. The charge envelope denoted with (⬭) corresponds to 366,500 kDa as observed in A.) and B.). At lower m/z values proteins with a molecular weight of 60,600 kDa (⬮), Srp1, and 94,700 kDa (◯), Kap95, and their heterodimeric form, Srp1-Kap95 with a mass of 155,300 kDa (⬮◯) are present. Additionally two very low abundant charge series are observed at high m/z values. They represent the total nuclear exosome containing the ring, Rrp4, Rrp40, Csl4, Dis3, Rrp6, Lrp1, Ynr024w, Srp1 and Kap95 with a molecular weight of ~ 691 kDa (■) and a dimeric form of the ~366 kDa exosome species with a molecular weight of ~ 729 kDa (⬭⬭). The charge series denoted with (X) is a common impurity that we frequently observe in the macromolecular analysis of TAP-isolated protein complexes.

At first exosomes originating from different subcellular localization are investigated by intact macromolecular mass spectrometry. In Fig 6, ESI mass spectra of the exosome isolated with three different entry points are shown. Using EP Csl4 two main charge distributions centered at m/z 10,000 are seen as reported earlier [18]. One of these charge series corresponds to the exosome consisting of the ring proteins (Rrp42, Rrp41, Rrp45, Rrp46, Rrp43 and Mtr3), three proteins with a RNA binding motif (Rrp4, Rrp40 and Csl4) and Dis3 with a molecular weight of 403,100 kDa as denoted with (⬬). The second charge envelope (⬭) can be determined to a molecular weight of 366,700 kDa and consists of identical proteins but is missing Csl4 (Fig 6 A). As described earlier using EP Csl4 is present in exosomes originating from the cytoplasm and the nucleus and therefore a mixed population of exosome is expected to be purified. The detected non-covalent exosome complexes neither include nuclear specific proteins (Rrp6, Lrp1, Ynr024w, Srp1 and Kap95) nor cytoplasmic specific proteins (Ski7).

When Ski7 is used as entry point, a pure preparation of the cytoplasmic exosome is expected. In the native mass spectrum only one low abundant charge series (⬭) corresponding to a mass of 367,100 kDa is identified (Fig. 6 B). The charge envelope is identical to the previously observed charge series when EP Csl4 is used, indicating that this detected exosome is again lacking Csl4 and Ski7, though in this experiment Ski7 is used as the entry point. The third entry point to isolate the exosome is Rrp6. The native mass spectrometric analyses of Rrp6 as entry point revealed several more charge envelopes than the previous entry points. In comparison to exosomes isolated with Csl4 and Ski7 it also shows on the protein complex level interesting similarities and differences. The spectrum is dominated by charge state distributions in the m/z range of 4,000 to 7,000 and between 9,000 and 11,000 of which the masses were determined. The charge distribution at m/z 9,000-11,000 (⬭) is identical to the exosome complex observed by EP Csl4 and Ski7 centered at this m/z value. The determined molecular weight is 366,500 kDa. This finding indicates that the most stable core exosome regardless of cytoplasmic or nuclear origin contains the hexameric ring plus Rrp40, Rrp4 and Dis3. Csl4 and all nuclear specific proteins such as Rrp6, Srp1, Kap95, Lrp1, Ynr024w and cytoplasmic specific protein Ski7 have dissociated from the exosome during the purification procedure. Additionally the mass spectrum of the nuclear exosome reveals several charge envelopes in the m/z range of 4,000 to 7,000 which correspond to Srp1 (importin α) with a mass of 60,600kDa (m/z 4,000 – 5,100, (●)), Kap95 (importin β) with a mass of 94,700kDa (m/z 4,700- 5,600, (◯)) and to the importin α/β dimer with a mass of 155,300kDa (●◯). The intensity of the peaks corresponding to importin suggests that it is highly abundant in the pulldown as has been also previously indicated by the 1-D-SDS gel. Two minor charge distributions can be observed between m/z 11,500-13,500. Due to the low intensity of the peaks, a precise mass determination of these species was difficult. The charge envelope from m/z 11,700 to 12,700 (⬭) led to a mass of ~ 729 kDa, which corresponds well to a dimeric form of the core exosome lacking Csl4 with a theoretical mass of 732,800 kDa. The second charge distribution at m/z 12,500 – 13,300 (■) reveals a mass of 691 kDa. We are tempted to believe this to be the nuclear exosome. The total nuclear exosome including the TAP tagged Rrp6, importin α/β, Lrp1 and Ynr024w has an expected molecular mass of 683,300 kDa. Our determined mass deviates only 1.2% to the expected mass.

The dissociated nuclear exosome specific proteins such as monomeric TAP tagged Rrp6, Lrp1, Ynr024w and cytoplasmic specific Ski7 as well as the common exosome protein Csl4 could not be found in any of the native mass spectra. These proteins may have dissociated either during the purification procedure or during the sample concentration and the buffer exchange. Another explanation is the precipitation of the proteins in the electrospray process.

As all exosomes independent from the used entry point exhibit a common abundant stable core protein sub-complex with a mass of ~366 kDa it prompted us to investigate the gas

phase behaviour of this particular exosome species. In this experiment macromolecular tandem mass spectrometry was employed to selectively dissociate the exosome sub-complex isolated by EP Rrp6 with a mass of 366,500 kDa as precursor carrying 40+ charges and follow the dissociation pattern of proteins in the gas-phase upon increasing collision voltages from 50 V to 175 V (Fig. 7).

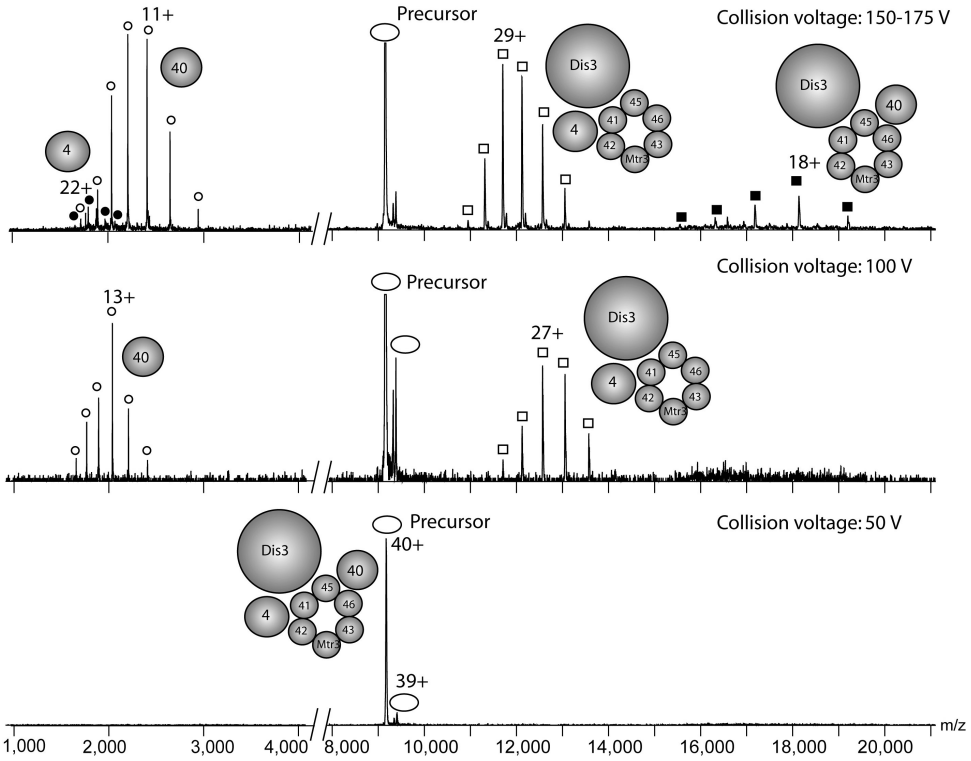


Figure 7

Tandem mass spectra of the most stable common sub-complex (core proteins without Csl4) of the exosome isolated using EP Rrp6. The abundant exosomal sub-complex consisting nine exosome proteins (ring proteins, Rrp4, Rrp40 and Dis3) containing 40+ charges (⬭) was selected as a precursor in native MS/MS experiments. Upon increasing collision voltage from 50 V to 175 V, the dissociation of the exosome sub-complex can be followed. At a collision voltage of 50 V no fragmentation can be observed (bottom panel). When the collision voltage is increased to 100 V, two additional charge distributions emerge; (○) corresponding to a mass of 26,471 kDa (Rrp40) and (□) with a mass of 339, 300 kDa representing the remaining exosome. At higher collision voltages of 150 to 175 V a second fragmentation parallel to the first dissociation event occurs. It gives rise to two more charge envelopes; (●) with a mass of 39,400 kDa and (■) with a mass of 326,400 kDa. In this fragmentation exosome sub-complex precursor ejects Rrp4 and not Rrp40.

At a collision voltage of 50 V the selected 40+ charge state (⬭) of the exosome sub-complex consisting of the ring proteins, Rrp40, Rrp4 and Dis3 does not exhibit any fragmentation yet as seen in the bottom panel of Fig. 7. At increasing collision voltage to 100V the selected precursor starts to dissociate into two fragments (middle panel Fig. 7).
Two charge envelopes can be observed at m/z 11,000 – 13,500 with charges between 25+ and 31+ (□) and at m/z 1,700 – 3,000 with charges of 9+ to 15+ (○). These fragments correspond to masses of 339,357 kDa and 26,471 kDa. The fragments clearly show that the stable core exosome lacking Csl4 with a mass of 365,800 kDa loses initially Rrp40 with a theoretical mass of 26,469 kDa at intermediate collision voltages. The remaining exosome containing the hexameric ring, Rrp4 and Dis3 represent the counter-ion of 339,357 Da. With increasing collision voltage two additional charge state distributions at m/z 15,000 – 19,500 with charges of 17+ to 21+ (■) and at m/z 1,700 – 2,100 with charges of 20+ to 23+ appear (●) (top panel, Fig. 7). Mass determination identified the charge envelope in the high m/z area to be 326,381 kDa and the charge distribution at the low m/z area to be its counter-ion with a mass of 39,438 kDa. It demonstrates that at higher collision voltages a second dissociation pathway can occur. It leads to the fragmentation of the stable core exosome lacking Csl4 with a mass of 365,800 kDa as a precursor into the exosome containing the hexameric ring plus Dis3 that also still holds Rrp40 and loses Rrp4 instead.
A second preparation of the exosome using Rrp6 as entry point resulted in a similar mass spectrum as shown in Fig. 6. Repeatedly the MS spectrum revealed a charge distribution at m/z 10,000 but this time the peaks were much broader in width than previously observed. It indicated that possibly two charge envelopes may be overlapping in this m/z area and that potentially a moderate to low abundant second exosome species was present. Therefore a second precursor was selected and MS/MS experiments determined it to be an exosome species with a mass of ~ 398 kDa that consisted of the ring proteins, Dis3, Rrp4, Rrp40 and additionally Csl4 and carried 42+ charges (⬬). This precursor was subjected to collisional induced dissociation by increasing the collision voltage from 50 V to 175 V (Fig 8).
As seen before the first dissociation is observed at a collision voltage of 100 V. The ion series generated are very similar to the previous tandem MS analysis of the previous exosome sub-complex. The proteins mass for the charge series in the low m/z area with 11+ charges (○) was determined to be 26,470 kDa corresponding to Rrp40 and the counter-ion in the high m/z region (□) to be 371,000 kDa which was missing Rrp40. Upon the increase of the collision voltage, two additional charge series in the high m/z region appeared. The corresponding counter-ions at low m/z values were not present. The first charge envelope at m/z 16,000 to 20,000 with charges around 20+ (■) was calculated to a mass of 357,900 kDa and was therefore identified to be the precursor lacking Rrp4. This fragmentation was also observed in the exosome lacking Csl4. The final charge series at m/z 13,500 to 15,200 with 25+ charges (★) had a mass of 365,800 kDa and unambiguously determined Csl4 to be missing in the fragment ion. Even though the corresponding ejected protein is not

present for the second and third dissociation event, the charges these would attain can be calculated from the charge of the high m/z fragment that is to be substracted from the precursor charge. Therefore it is predicted that dissociated Rrp4 would carry 22+ charges as seen before and Csl4 was ejected with 17+ charges.

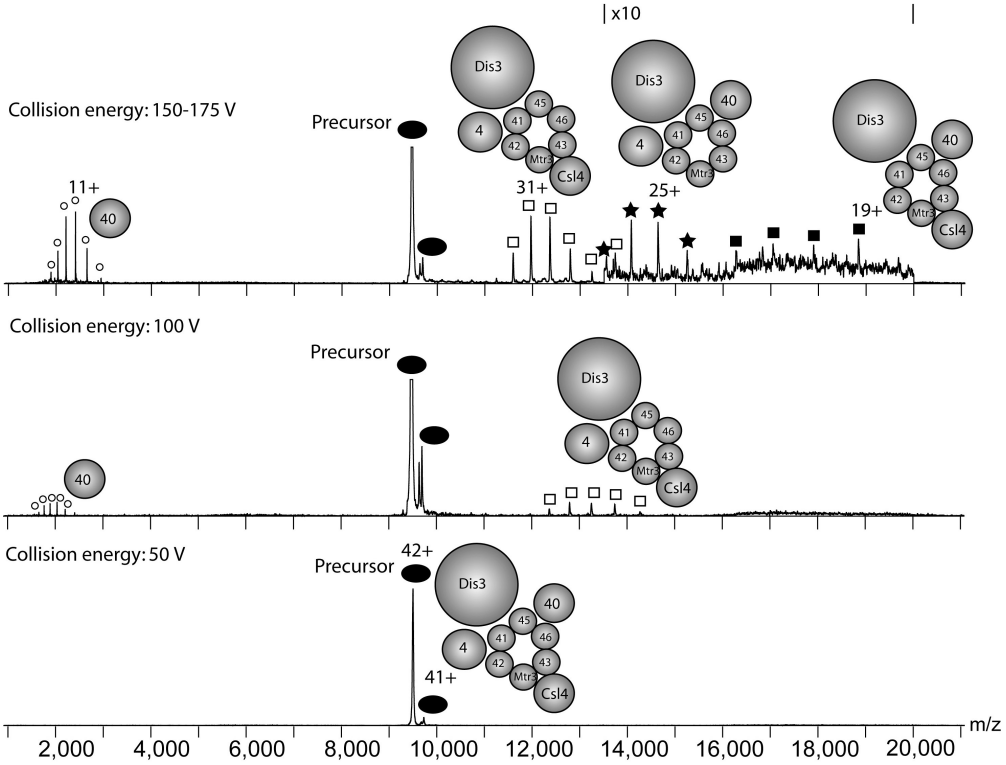


Figure 8

MS/MS spectra of the exosome (core proteins including Csl4) isolated using EP Rrp6

The exosome (ring proteins, Rrp4, Rrp40, Csl4 and Dis3) with a mass of ~398 kDa containing 42+ charges (●) was selected as a precursor in native MS/MS experiments. The collision voltage was increased from 50 V to 175 V. At a collision voltage of 50 V no fragmentation can be observed (bottom panel). When the collision voltage is increased to 100 V, two additional charge distributions emerge. The identical dissociation pattern has been observed in figure 8. The low m/z charge series Da (○) with a mass of 26,470 corresponds to Rrp40. The remaining exosome sub-complex (□) with a mass of 371, 000 kDa can be observed in the high m/z region. At higher collision voltages (150 to 175 V), a second and third fragmentation additional to the first dissociation event occurs. The charge series of the second fragmentation (■) can be calculated to a molecular mass of 357,900 kDa. In this event Rrp4 was ejected from the precursor. The third dissociation (★) reflects a molecular weight of 365,800 kDa and is therefore the fragment ion that has lost Csl4 with a mass of 31,557 kDa.

The corresponding fragment to be found in the low m/z region is not present.

Surprising is the different charge separation of the precursor on the corresponding fragment ions in both dissociation events. While Rrp40 is ejected from the exosome sub-complex with only 11+ charges, Csl4 acquires 17+ and Rrp4 22+ charges when dissociated. The remaining charges of the precursor remain with the counter-ion fragment.

The process can be summarized as followed:

1. $(\text{exosome sub-complex excl. Csl})^{40+} \longrightarrow (\text{Rrp40})^{11+} + (\text{exosome sub-complex} - \text{Rrp40})^{29+}$
2. $(\text{exosome sub-complex excl.Csl4})^{40+} \longrightarrow (\text{Rrp4})^{22+} + (\text{exosome sub-complex} - \text{Rrp4})^{18+}$

1. $(\text{exosome complex incl. Csl4})^{42+} \longrightarrow (\text{Rrp40})^{11+} + (\text{exosome sub-complex} - \text{Rrp40})^{29+}$
2. $(\text{exosome complex incl. Csl4})^{42+} \longrightarrow (\text{Rrp4})^{22+} + (\text{exosome sub-complex} - \text{Rrp4})^{20+}$
3. $(\text{exosome complex incl. Csl4})^{42+} \longrightarrow (\text{Csl4})^{17+} + (\text{exosome sub-complex} - \text{Csl4})^{25+}$

Especially the very low charge of Rrp40 (11+) is remarkable. It is generally believed that the dissociation process favors the local unfolding of the ejected protein. Thereby it can acquire many charges. For Rrp40 this theory is not correct.
Even though Csl4 and Rrp4 have a mass of 31,6 kDa and 39,4 kDa respectively and therefore a higher molecular weight than Rrp40 with a mass of 26,5 kDa, this mass difference cannot account for such a dramatic change in the charge partitioning upon gas-phase dissociation. The number of charges that a protein obtains during dissociation can be directly related to the changes in the native conformation of the protein in this process. If the protein is more folded it can take up fewer charges than proteins that start unfolding. As only 11+ charges for Rrp40 are observed it proposes a near native fold when ejected from the exosome. Csl4 and Rrp4 dissociate with17+ and 22+ charges and must therefore be partially unfolded during this process.
The gas-phase dissociation pathway of exosomal proteins from the (sub-) complex is summarized in figure 9.

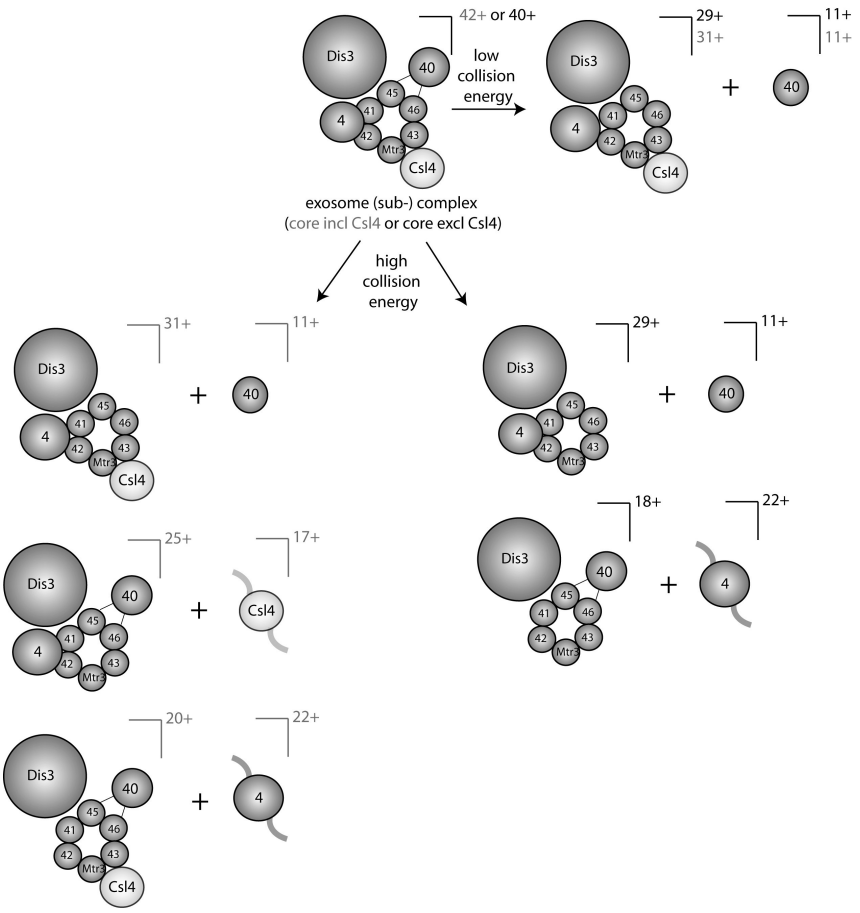


Figure 9

Summary of gas-phase protein dissociation pathway from two different exosome (sub-) complex precursors

Two exosome precursors were selected to perform gas-phase dissociation experiments at low and high collision voltages. The exosome sub-complex consisting of the hexameric ring, Dis3, Rrp4 and Rrp40 but missing Csl4 with 40+ charges (denoted with charges in black) and the 42+ precursor of the same sub-complex that included Csl4 were fragmented (denoted with charges in grey). The fragmentation pattern at lower collision voltages (100 V) yielded only in Rrp40 and the corresponding fragment ion. At higher collision voltages (150 – 175 V) a second fragmentation into Rrp4 and the counter-ion from the selected precursor occurred. These pathways were identical when both exosome (sub-) complexes was used as a precursor. When the exosome including Csl4 was fragmented a third dissociation event took place that ejected Csl4 from the precursor concomitant with the dissociation of Rrp4. Since Rrp40 already dissociates at lower collision voltages than Csl4 and Rrp4 and obtains relatively few charges, it can be hypothesized that its affinity towards the exosome complex is lower than of Csl4 and Rrp4. Csl4 and Rrp4 acquire relative to Rrp40 many more charges which indicate that these two proteins are partially unfolded in the dissociation process.

Discussion

Based on stable isotope dimethyl labelling when EP Rrp6 and Ski7 is used to retrieve the exosome complexes we could establish that all ring components (Rrp41, Rrp42, Mtr3, Rrp43, Rrp46, Rrp45), all proteins containing an S1-binding domain (Rrp40, Rrp4 and Csl4) and Dis3 are present in equal abundance in the nuclear and cytoplasmic exosome. This is expected considering that these proteins form the core of the exosome. Both exosomes differ only in nucleus and cytoplasmic specific proteins. Several known nuclear specific proteins could be identified such as Rrp6, Lrp1, Ynr024w and rather surprisingly also Srp1 and Kap95. The identified cytoplasmic specific protein is Ski7. When exosome isolated by Rrp6 and Csl4 were quantified, most of the common exosome proteins again showed no significant changes except Rrp45. Rrp45 is an exosome core ring component and therefore tightly associated in the assembly. Potential dynamic changes of any of the ring proteins would be very novel and would require further evidence. So far exosomes that are present in other organisms such as archaea have been shown to contain multiple copies of exosome proteins in their assembly. The ring consists of three copies of Rrp41 and Rrp42. Three copies of Rrp4 or Csl4 are crowned on top of this ring. A similar situation likely exists in *Caenorhabditis.elegans*. Here three genes are identified to code for ring proteins (B0564.1, F37C12.13, and C14A4.5) and one gene (F56C6.4) for an S1 domain containing protein [33]. Since the overall structure of the exosome is conserved it is expected that each ring protein is present in two copies with possibly three S1 domain containing protein on top. Rrp6 is found primarily in the purification using Rrp6 as EP but some peptides could be detected in the purification using Csl4 as EP. As Csl4 is a member of all exosomes, some peptides originating from Rrp6 are to be expected. Also in this analysis four additional proteins, Lrp1, Ynr024w, Srp1 and Kap95 were found exclusively in the nuclear exosome. Lrp1, which is exclusively present in the nucleus in substoichiometric amounts, was previously found to be a cofactor of the nuclear exosome. Associated to Rrp6 it facilitates binding of RNA substrates to the nuclear exosome [34].

Ynr024w is a hypothetical protein but it shows significant sequence homology to human M-phase phosphoprotein 6 (MPP6). This protein is associated with the human nuclear protein and is needed for the maturation of rRNA. Strikingly different between the cytoplasmic and the nuclear exosome is the high abundance of Srp1 and Kap95. These two proteins belong to the class of importins. Importins are essential in transporting proteins actively across the nuclear membrane via the nuclear pore complex. Srp1 (importin α) and Kap95 (importin β) exist as a heterodimer. Srp1 binds to target protein that contains a classical nuclear localization signal (cNLS). Such cNLS generally consists of short stretches of basic amino acid residues [35, 36]. Subsequently Kap95 binds the complex to the nuclear pore complex and facilitates its translocation into the nucleus in an energy dependent manner. Rrp6 contains such a cNLS recognition site C-terminally (aa700-722,

KKRRFDPSSSDSNGPRAAKKRRP). Based on gel band intensity we could determine that Rrp6 and Srp1 are present in stoichiometric amounts in the nuclear exosome, suggesting that each nuclear exosome complex could be bound to one protein of Srp1. Kap95 in turn binds to Srp1 and is also equal abundant to Srp1 and Rrp6. It is known that the Srp1-Kap95 heterodimer complex is tightly bound to the target protein in the cytoplasm. When the protein assembly is translocated into the nucleus Srp1-Kap95 dissociates from the target protein. This mechanism is facilitated by Gsp1 [37, 38]. Several exosome proteins from different organisms also exhibit a cNLS site such as human Rrp41, PM/Scl-75 (homologue of yeast Rrp45), PM/Scl-100 (homologue of yeast Rrp6). Up to date the cNLS of PM/Scl-75 has been validated in in vitro assays and a model for nuclear import involving the human exosome is proposed [39]. In this model the exosomal core consisting fo the hexameric ring with the S1 proteins attached is formed in the cytoplasm and via a NLS-sequence in hRrp41 translocated into the nucleus. Once in the nucleus the exosome can further translocate into the nucleolus possibly in association with PM/Scl-100. Nuclear localization signals could be also found in exosome proteins dDis3, dRrp6, and dRrp4 in Drosophila [17]. Still the import and assembly of the yeast nuclear exosome may be less complicated as only Rrp6 contains a C-terminal cNLS motif as was validated in vitro with mutants [40].
A structural model for the nuclear and cytoplasmic exosome is proposed in Fig 10.

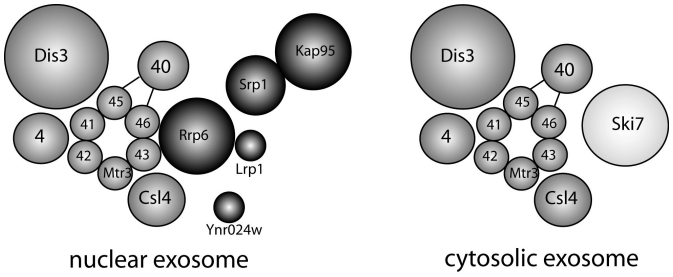


Figure 10

Structural model of the nuclear and cytoplasmic exosome

The common proteins between both exosome variants are represented in grey. Nuclear and cytoplasmic specific exosome proteins are coloured in black and light-grey respectively.

Our study confirmed four known and revealed one novel phosphorylation site. These phosphorylation sites could be relatively quantitated between the nuclear and cytoplasmic exosome. Of particular interest is the phosphorylaton site in Csl4. Csl4 phosphorylated serine 94 is significantly higher abundant in the nuclear than in the cytoplasmic exosome. The phosphorylation is located in the RNA binding domain and could therefore have important functions that are predominantly assigned to the nuclear exosome.
Differences between exosomes on the organisational level of protein-complexes were probed by native (tandem) mass spectrometry. The native mass spectrum exhibited

differences but also similarities between different exosome complexes. In particular Rrp6 revealed an unexpected mass spectrum. Only a very low fraction of the total nuclear exosome was associated including all nuclear specific components. Main differences compared to other exosomes, i.e purified with Ski7 and Csl4, were high signals for importins that have dissociated from the exosome. The interaction between Srp1 and Kap95, which is known to be tight, was still present as it formed heterodimers but also the respective monomers were observed. When the exosome was selected as precursor for MS/MS studies, it was interesting to observe that at low collision voltage Rrp40 with only 11+ charges dissociated, while at higher collision voltage also Rrp4 with 22+ charges and Csl4 with 17+ charges were ejected not in subsequent dissociation events but instead from the selected precursor. It was obvious that Rrp40 dissociated with a very low amount of charges, indicating a more compact and folded conformation while Rrp4 and Csl4 acquired more charges. Several structural studies have shown that Csl4, Rrp40 and Rrp4 are situated on top of the hexameric ring of the exosome. All of them probably interact weaker with other exosomal proteins in the complex compared to the ring-proteins. Therefore it is likely that these proteins dissociate more easily from the total exosome assembly and keep the stable hexameric ring intact. Dis3 is also peripheral attached to the exosome assembly and therefore prone to dissociation. Since it is the largest protein in the total assembly with a molecular weight of ~113,7 kDa, it requires much more energy to be dissociated compared to smaller proteins such as Rrp40, Csl4 and Rrp4. Rrp40 is the easiest protein to be ejected. These observations cannot be explained by the increase in molecular weight and thereby in size. It is conceivable that Rrp40 is the protein with the lowest affinity to the exosome when it is in the gas-phase. This is in contrast to the stability of the exosome protein in solution phase. As previously observed the main protein that was missing from the exosome has been repeatedly Csl4 [18].

Conclusion

We have shown that the combination of TAP, quantitative proteomics and native mass spectrometry has the potential of providing more detailed information about closely related protein complexes. We quantitated all exosomal proteins in relative abundance between the nuclear exosome and the cytoplasmic and between the nuclear and a mixture of exosomes in the cell. The common proteins between exosomes have been shown to be present in similar amounts in all exosome complexes. Absolute differences are observed in the sub-cellular specific exosome proteins Rrp6, Lrp1, Ynr024w, Srp1 and Kap95 for the nucleus and Ski7 for the cytoplasm. Our technique also allows the investigation of differential phosphorylation and we have confirmed and quantified several known exosome phosphorylations, one of which was shown to be upregulated in the nuclear exosome. The

protein complex stoichiometry and relative gas-phase binding strength of the exosome was probed using native (tandem) mass spectrometry. Exosomes have a common stable core complex that consists of common exosome subunits. The complete nuclear exosome is rather instable as only a minor fraction of the intact nuclear exosome survived the sample preparation. Most of the nuclear specific components were dissociated from the common exosome matrix. In tandem mass spectrometry studies the stable core exosome was subjected to collision induced dissociation. At high enough energy several gas-phase dissociation events can take place at the same time from the precursor. In particular the exosomal sub-unit Rrp40 attains very few charges in this process, so that it can be hypothesized that it dissociates in near native fold and also has the lowest affinity towards the exosome.

References

1. Rigaut, G., et al., *A generic protein purification method for protein complex characterization and proteome exploration.* Nat Biotechnol, 1999. **17**(10): p. 1030-2.
2. Puig, O., et al., *The tandem affinity purification (TAP) method: a general procedure of protein complex purification.* Methods, 2001. **24**(3): p. 218-29.
3. Gavin, A.C., et al., *Functional organization of the yeast proteome by systematic analysis of protein complexes.* Nature, 2002. **415**(6868): p. 141-7.
4. Krogan, N.J., et al., *Global landscape of protein complexes in the yeast Saccharomyces cerevisiae.* Nature, 2006. **440**(7084): p. 637-43.
5. Heck, A.J. and J. Krijgsveld, *Mass spectrometry-based quantitative proteomics.* Expert Rev Proteomics, 2004. **1**(3): p. 317-26.
6. Krijgsveld, J., et al., *Metabolic labeling of C. elegans and D. melanogaster for quantitative proteomics.* Nat Biotechnol, 2003. **21**(8): p. 927-31.
7. Gygi, S.P., et al., *Quantitative analysis of complex protein mixtures using isotope-coded affinity tags.* Nat Biotechnol, 1999. **17**(10): p. 994-9.
8. Ong, S.E., et al., *Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics.* Mol Cell Proteomics, 2002. **1**(5): p. 376-86.
9. Yao, X., et al., *Proteolytic 18O labeling for comparative proteomics: model studies with two serotypes of adenovirus.* Anal Chem, 2001. **73**(13): p. 2836-42.
10. Stewart, II, T. Thomson, and D. Figeys, *18O labeling: a tool for proteomics.* Rapid Commun Mass Spectrom, 2001. **15**(24): p. 2456-65.
11. Mousson, F., et al., *Quantitative proteomics reveals regulation of dynamic components within TATA-binding protein (TBP) transcription complexes.* Mol Cell Proteomics, 2007.
12. Hsu, J.L., et al., *Stable-isotope dimethyl labeling for quantitative proteomics.* Anal Chem, 2003. **75**(24): p. 6843-52.
13. Hsu, J.L., S.Y. Huang, and S.H. Chen, *Dimethyl multiplexed labeling combined with microcolumn separation and MS analysis for time course study in proteomics.* Electrophoresis, 2006. **27**(18): p. 3652-60.
14. Araki, Y., et al., *Ski7p G protein interacts with the exosome and the Ski complex for 3'-to-5' mRNA decay in yeast.* Embo J, 2001. **20**(17): p. 4684-93.
15. de la Cruz, J., et al., *Dob1p (Mtr4p) is a putative ATP-dependent RNA helicase required for the 3' end formation of 5.8S rRNA in Saccharomyces cerevisiae.* Embo J, 1998. **17**(4): p. 1128-40.
16. LaCava, J., et al., *RNA degradation by the exosome is promoted by a nuclear polyadenylation complex.* Cell, 2005. **121**(5): p. 713-24.
17. Graham, A.C., D.L. Kiss, and E.D. Andrulis, *Differential distribution of exosome subunits at the nuclear lamina and in cytoplasmic foci.* Mol Biol Cell, 2006. **17**(3): p. 1399-409.

18. Synowsky, S.A., et al., *Probing genuine strong interactions and post-translational modifications in the heterogeneous yeast exosome protein complex.* Mol Cell Proteomics, 2006. **5**(9): p. 1581-92.
19. Hernandez, H., et al., *Subunit architecture of multimeric complexes isolated directly from cells.* EMBO Rep, 2006. **7**(6): p. 605-10.
20. Lorentzen, E., et al., *The archaeal exosome core is a hexameric ring structure with three catalytic subunits.* Nat Struct Mol Biol, 2005. **12**(7): p. 575-81.
21. Liu, Q., J.C. Greimann, and C.D. Lima, *Reconstitution, activities, and structure of the eukaryotic RNA exosome.* Cell, 2006. **127**(6): p. 1223-37.
22. Allmang, C., et al., *The yeast exosome and human PM-Scl are related complexes of 3' --> 5' exonucleases.* Genes Dev, 1999. **13**(16): p. 2148-58.
23. Logie, C. and C.L. Peterson, *Purification and biochemical properties of yeast SWI/SNF complex.* Methods Enzymol, 1999. **304**: p. 726-41.
24. Shevchenko, A., et al., *Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels.* Anal Chem, 1996. **68**(5): p. 850-8.
25. Pappin, D.J., P. Hojrup, and A.J. Bleasby, *Rapid identification of proteins by peptide-mass fingerprinting.* Curr Biol, 1993. **3**(6): p. 327-32.
26. Pinkse, M.W., et al., *Highly Robust, Automated, and Sensitive Online TiO2-Based Phosphoproteomics Applied To Study Endogenous Phosphorylation in Drosophila melanogaster.* J Proteome Res, 2007.
27. van den Heuvel, R.H., et al., *Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry.* Anal Chem, 2006. **78**(21): p. 7473-83.
28. Schilders, G., et al., *MPP6 is an exosome-associated RNA-binding protein involved in 5.8S rRNA maturation.* Nucleic Acids Res, 2005. **33**(21): p. 6795-804.
29. Terribilini, M., et al., *Prediction of RNA binding sites in proteins from amino acid sequence.* Rna, 2006. **12**(8): p. 1450-62.
30. Terribilini, M., et al., *RNABindR: a server for analyzing and predicting RNA-binding sites in proteins.* Nucleic Acids Res, 2007. **35**(Web Server issue): p. W578-84.
31. Ghaemmaghami, S., et al., *Global analysis of protein expression in yeast.* Nature, 2003. **425**(6959): p. 737-41.
32. Huh, W.K., et al., *Global analysis of protein localization in budding yeast.* Nature, 2003. **425**(6959): p. 686-91.
33. van Hoof, A. and R. Parker, *The exosome: a proteasome for RNA?* Cell, 1999. **99**(4): p. 347-50.
34. Mitchell, P., et al., *Rrp47p is an exosome-associated protein required for the 3' processing of stable RNAs.* Mol Cell Biol, 2003. **23**(19): p. 6982-92.
35. Robbins, J., et al., *Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence.* Cell, 1991. **64**(3): p. 615-23.
36. Dingwall, C. and R.A. Laskey, *Nuclear targeting sequences--a consensus?* Trends Biochem Sci, 1991. **16**(12): p. 478-81.
37. Melchior, F., et al., *Inhibition of nuclear protein import by nonhydrolyzable analogues of GTP and identification of the small GTPase Ran/TC4 as an essential transport factor.* J Cell Biol, 1993. **123**(6 Pt 2): p. 1649-59.
38. Goldfarb, D.S., et al., *Importin alpha: a multipurpose nuclear-transport receptor.* Trends Cell Biol, 2004. **14**(9): p. 505-14.
39. Raijmakers, R., G. Schilders, and G.J. Pruijn, *The exosome, a molecular machine for controlled RNA degradation in both nucleus and cytoplasm.* Eur J Cell Biol, 2004. **83**(5): p. 175-83.
40. Phillips, S. and J.S. Butler, *Contribution of domain structure to the RNA 3' end processing and degradation functions of the nuclear exosome subunit Rrp6p.* Rna, 2003. **9**(9): p. 1098-107.

Chapter 5

The yeast Ski complex is a hetero-tetramer

Silvia A. Synowsky
Albert J. R. Heck

Biomolecular Mass Spectrometry and Proteomics Group, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Sorbonnelaan 16, 3584 CA Utrecht, The Netherlands

Protein Sci. 2008 Jan;17(1):119-25

Chapter 5

Abstract

The yeast Ski complex assists the exosome in the degradation of mRNA. The Ski complex consists of three components; Ski2, Ski3 and Ski8, believed to be present in a 1:1:1 stoichiometry. Measuring the mass of intact isolated endogenously expressed Ski complexes by native mass spectrometry we unambiguously demonstrate that the Ski complex has a hetero-tetrameric stoichiometry consisting of one copy of Ski2 and Ski3 and two copies of Ski8. To validate the stoichiometry of the Ski complex, we performed tandem mass spectrometry. In these experiments one Ski8 subunit was ejected concomitant with the formation of a Ski2/Ski3/Ski8 fragment, confirming the proposed stoichiometry. To probe the topology of the Ski complex we disrupted the complex and mass analyzed the thus formed sub-complexes, detecting Ski8-Ski8, Ski2-Ski3, Ski8-Ski2 and Ski8-Ski8-Ski2. Combining all data we construct an improved structural model of the Ski complex.

Introduction

The recently introduced combination of tandem-affinity-purification (TAP) [1] and peptide mass spectrometry has added substantial more information to large databases of protein-protein interaction networks in particular in yeast [2-4]. Although very powerful approaches, unfortunately detailed information on each of the identified protein complexes, such as complex stoichiometry, topology of protein-protein interactions and overall structure, is not easily addressed. Advances in mass spectrometry have enabled the investigation of non-covalent protein interactions of heterogeneous large macromolecular complexes, providing a new complementary tool in structural biology [5-8]. The potential of macromolecular mass spectrometry to also probe endogenously expressed heterogeneous protein complexes has been recently demonstrated [9, 10] whereby the endogenously expressed TAP purified yeast exosome was probed for complex stoichiometry, overall topology of the protein complex and individual protein post-translational modifications. In this work we focused on the exosome related Ski complex endogenously expressed and purified from yeast using the TAP tagging technique.

The Ski complex is involved in exosome mediated 3' -> 5' mRNA degradation. It is hypothesized that the Ski complex recruits the exosome to its substrate, where it is subsequently degraded [11]. Next to its involvement in exosome directed degradation of RNA, the Ski complex has been shown to protect the cell from viral replication by blocking synthesis of extrinsic mRNA transcripts [12].

The Ski complex consists of three subunits. The largest subunit Ski3 has a mass of 163 kDa, and contains a tetratricopeptide (TPR) motif [13]. Next, Ski2 has a mass of 146 kDa, and is a putative RNA Helicase [12]. Finally, Ski8 is a protein with a mass of 44 kDa, containing a WD-repeat domain [14]. The crystal structure of Ski8 shows a seven-bladed β propeller and may therefore function as a scaffolding protein in the Ski complex [15, 16]. Ski3 and Ski8 were found to co-purify by using Myc-tagged Ski2, and by using [^{35}S]-methionine labelled cells and semi-quantitative Western Blotting a 1:1:1 stoichiometry for Ski2, Ski3 and Ski8 in the Ski complex has been proposed. Ski deletion strains have suggested association of Ski2 to Ski3, and Ski2 to Ski8 in the presence of Ski8 and Ski3, respectively. Additionally, a weak interaction between Ski3 and Ski8 is proposed [17]. Based on suitable structural templates the structure of the intact Ski complex was recently modeled using computational structure predictions [18]. This model predicted a non linear Ski8-Ski2-Ski3 conformation of the complex for direct interactions between Ski8-Ski2 and between Ski2-Ski3.

Here we analyze the endogenously expressed Ski complex from *S.cerevisiae* using macromolecular mass spectrometry. We retrieve the Ski complex using the tandem affinity purification technique, taking each of the three proteins subsequently as entry-point. Our data allows us to construct a new structural model for the yeast Ski complex, and

unambiguously shows for the first time that the stoichiometry of the Ski complex is such that it contains one copy of Ski2 and Ski3, but two copies of Ski8.

Material and Methods

Yeast strain, cultivation, protein purification

The *S.cerevisiae* strain MGD35313D, BSY17 containing Ski2, Ski3 or Ski8 as the C-terminal tagged entry point was purchased at Euroscarf (Frankfurt, Germany). Cell cultivation and protein purification with Ski2, Ski3 and Ski8 were essentially performed as described previously [10]. All purifications of the Ski complex, using separately Ski3, Ski2 and Ski8 as bait, were performed using a salt concentration of 300mM NaCl in the lysis buffer.

Proteomic analysis of the Ski complex

The purified Ski proteins were separated on a 10% SDS-gel and stained using 0.1% (v/v) Coomassie Brilliant Blue G250. Peptide preparation and LC-MSMS analyses using an LTQ Mass Spectrometer (Thermoelectron, Germany) was perfomed as described previously [10]. The peptide tolerance was fixed to 0.5 Da and the MS/MS tolerance to 0.9 Da.

Analysis of the Ski complex by macromolecular mass spectrometry

For the mass spectrometric analysis of the intact protein complexes, the elution buffer of the TAP purified Ski complex was exchanged to 150 mM NH_4Ac-buffer, pH 6.8. The mass spectrometer was externally calibrated using aqueous CsI (100mg/ml) solutions. Analysis of the intact protein complex and sub-complexes thereof was performed using a LCT mass spectromter (Waters, UK). Typical spraying conditions were capillary voltage 1200 V, sample cone voltage 150 V. To investigate the topology of the Ski complex it was disrupted by adding 0.5% formic acid (v/v) to the sample solution prior to mass spectrometric analysis. Tandem mass spectrometry experiments were performed essentially as described previously [6]. The collision energy of the ions was gradually increased from 50 Volt to 200 Volt using a xenon gas pressure of approximately $2x10^{-2}$ mbar.

Results

In Figure 1 we have summarized the experimental approach to investigate the stoichiometry and topology of the Ski complex by standard proteomics methods and macromolecular mass spectrometry. The Ski complex was endogenously expressed in *S.cerevisiae*, and purified using the tandem TAP-TAG affinity approach. Ski2, Ski3 and Ski8 as the known constituents of the Ski complex were independently used as tagged entry points.

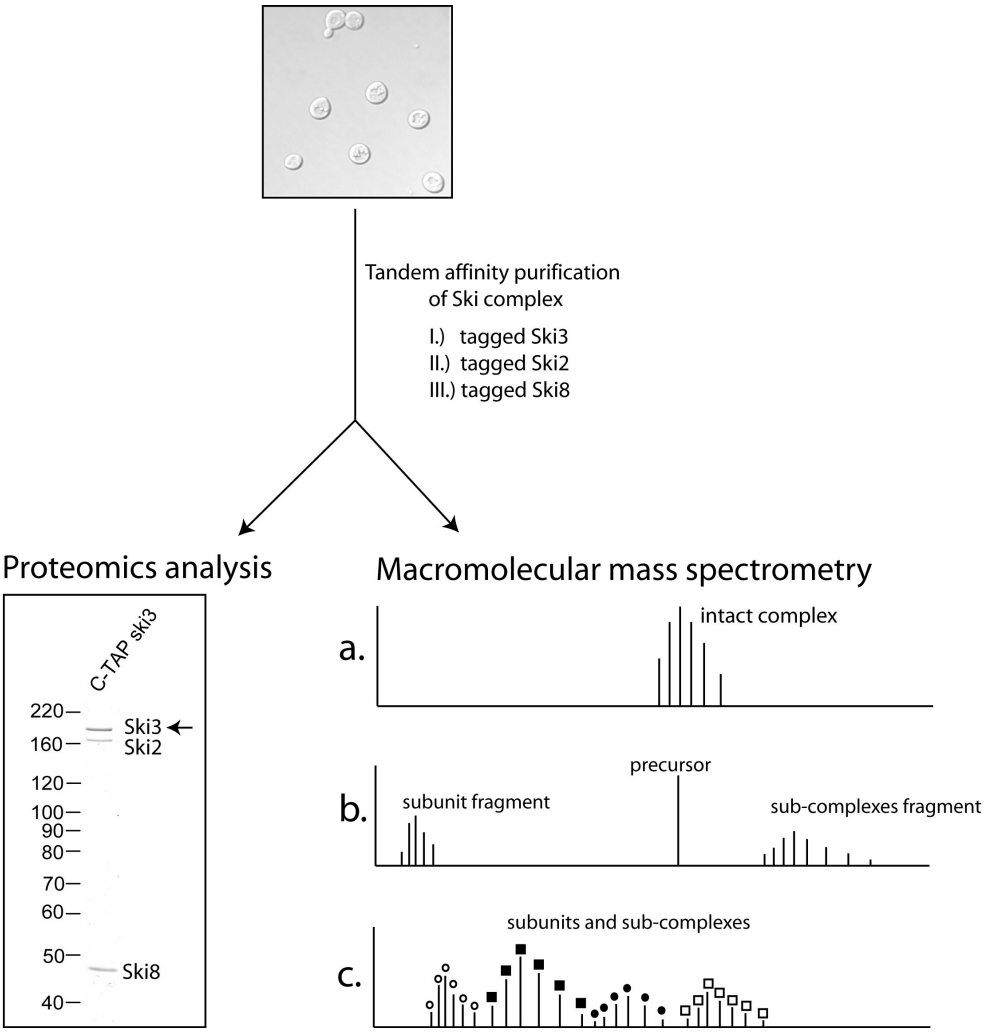


Figure1

Schematic overview of the experimental approach. The Ski complex was tandem affinity purified from yeast cell using either Ski3, Ski2 or Ski8 as bait protein. Proteins of these purified complexes were first separated by 1D gel SDS PAGE and identified by standard proteomics analyses. The 1D gel shown was obtained using Ski-3 as bait (as indicated with the arrow), and revealed only three bands, identified as indicated. Then, the purified protein complexes were buffer-exchanged into a mass spectrometry amenable buffer. Next, the masses of the intact Ski complexes were measured by macromolecular mass spectrometry (a) and tandem mass spectrometry was used to confirm the stoichiometry of the complexes (b). To probe the subunit topology of the Ski complex, the buffer solution was acidified, disrupting the Ski complex leading to the formation and mass detection of Ski subunits and Ski sub-complexes (c).

Proteomics analysis of the yeast Ski complex

To evaluate the quality of the Ski complex purification we analyzed the samples on a 1D gel. In Figure 1 the gel for the complex with Ski3 as entry point is shown. MS analysis clearly showed that the three bands originated from Ski2, Ski3 and Ski8, with no other significant hits. Similar results were obtained from purifications with the Ski2 and Ski8 entry points (data not shown). In a high-throughput proteomic analysis, it was reported that with any tagged Ski protein several other proteins co-purified [19]. In contrast, our purifications were very pure, which may be surprising as at least Ski8 plays a role in another protein complex. Ski8 relocalizes from the cytoplasm to the nucleus where it is involved in double-strand breaks (DSBs) formation catalyzed by Spo11 [20]. However, we did not detect Spo11 when Ski8 was used as bait. Since our purifications were performed on whole cell lysates nuclear proteins may be of relatively low abundance. The isolation of solely the Ski complex may be due to our more stringent and careful washing and purification procedures but generally point at the lower specificity of TAP
pulldowns in high throughput experiments.

Macromolecular mass spectrometry on intact Ski complexes.

In Figure 2A the mass spectrum is shown of the purified Ski complex, using Ski3 as the entry point.

The spectrum reveals two charge distributions, originating from two species. The ion series at m/z values between 8,500 and 11,000 (■) corresponds to a mass of 403,527 Da ± 500 Da. By adding the theoretical masses of tagged-Ski3 (168,803 Da), Ski2 (146,059 Da) and Ski8 (44,232 Da) assuming a 1:1:1 stoichiometry [17], the expected mass should be 359,094 Da . The discrepancy between the expected and the actual mass of the Ski complex is 44,433 Da, close to the mass of Ski8, suggesting an unexpected stoichiometry of 1:1:2 with Ski8 present in two copies.

Another charge distribution is present at values between 3,000 and 4,000 m/z originating of species with a mass of 44,178 Da ± 25 Da, which is in agreement with the expected mass of Ski8, indicating that some free Ski8 is present in our purified samples. Importantly, we detect no other ion signals, such as for a possible 1:1:1 Ski complex, indicating that the observed 1:1:2 stoichiometry is the exclusive stoichiometry present in the yeast Ski complex. The obtained mass spectra using Ski2 as entry point looked very similar to those presented in Figure 2A (data not shown).

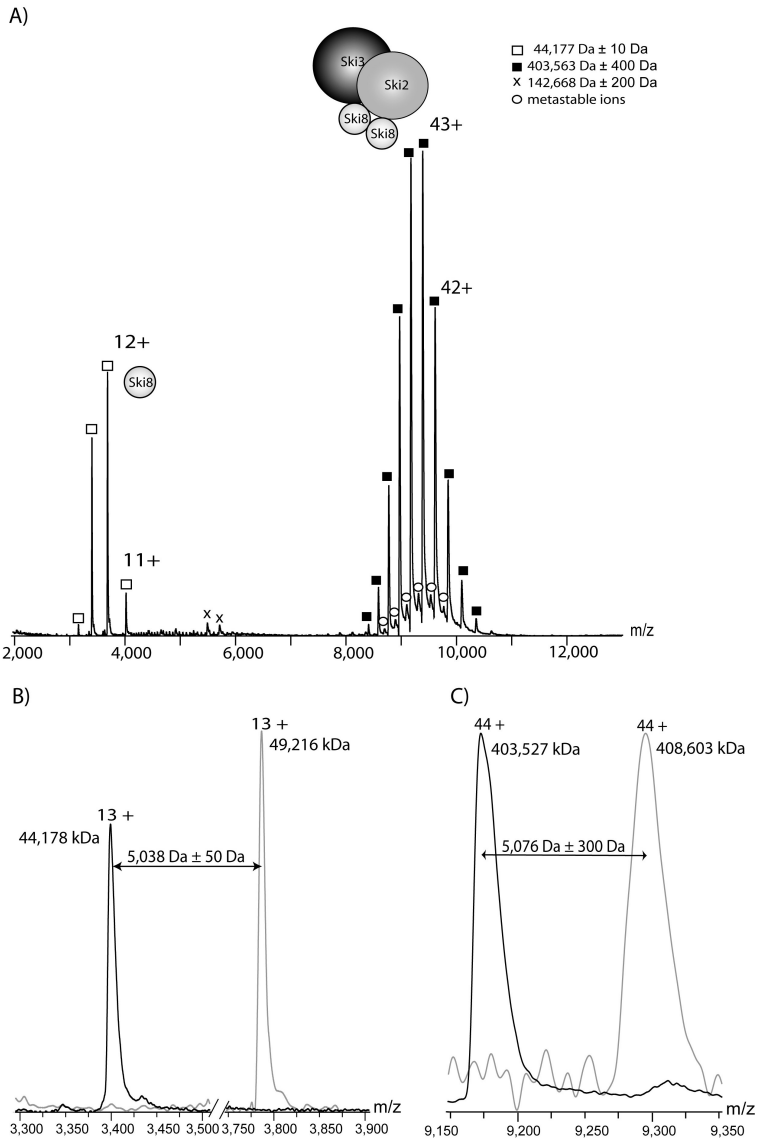


Figure 2

Macromolecular mass spectrometry on intact Ski complexes. (A) Mass spectrum of the purified Ski complex using Ski3 as entry point. The spectrum shows two main ion series (■ and □) originating from species with a mass of 403,527 kDa and 44,178 kDa, representing the full intact Ski complex and monomeric Ski8, respectively. The lower abundant charge distributions are due to metastable ion formation (○) and a protein contamination (x) always observed in our purifications. (B) and (C) Zoomed-in and overlaid parts of the mass spectra of the intact Ski complex purifications using

Ski3 and Ski8 as TAP tagged proteins, respectively. The black trace shows the ion signal of Ski3 as TAP tagged protein and the grey trace of TAP tagged Ski8. In (B) the 13+ ions of Ski8 are depicted, whereby the ion signal of monomeric Ski8 is shifted to higher m/z values, in the tagged Ski8 purification, caused by the presence of the TAP tag of 5,038 Da. In (C) the 44+ ions of the intact Ski complex are depicted. The Ski complex purified with Ski8 shows a shift of 5,076 Da to higher m/z values, clearly indicating that a second TAP tag was present in this complex.

The obtained mass of the major species detected in this purification was 403,561 Da ± 1000 (see table 1), thus again hinting at the 1:1:2 stoichiometry.

Table 1 Masses of the intact yeast Ski complex using different entry points.

TAP TAG bait protein	Measured mass of Ski complex (Da) [#]	Complex assignment
Ski2	403,561 ± 1000	(CBP-Ski2)-Ski3-Ski8-Ski8
Ski3	403,527 ± 500	Ski2-(CBP-Ski3)-Ski8-Ski8
Ski8	408,603 ± 400	Ski2-Ski3-(CBP-Ski8-CBP-Ski8)

[#] The indicated error range was estimated from the standard deviations in the measured masses as measured by mass spectrometry following several different purifications.

If our suggested 1:1:2 stoichiometry is correct, we should, with the mass accuracy of our method, be able to distinguish the expected mass difference between the Ski complex purified from either the Ski2/Ski3 entry points and the Ski8 entry point. Using the tagged-Ski8 for the Ski complex purification we introduce two TAP-TAGs into the complex, if there are indeed two copies of Ski8 present. After purification this extra TAG leads to mass increase of 5,077 Da corresponding to the second remaining calmodulin binding peptide (CBP) of the Tag of the total complex compared to Ski2 and Ski3 as entry points. The mass spectrum using Ski8 as entry point shows a species of 408,603 Da ± 400 (see table 1), again confirming the 1:1:2 stoichiometry. We also detected minor signals for free Ski8, now having a mass of 49,216 Da ± 15 Da. For illustrative purposes, we overlaid in Figure 2B the detected 13+ charged ions of Ski8 from the Ski complex isolated with TAP tagged Ski3 and TAP tagged Ski8, directly showing the expected mass shift of 5,038 ± 50 Da originating from the tagged Ski8. Similarly, we overlaid the ion signals of the 44+ ions of the Ski complex, isolated with TAP tagged Ski3 and TAP tagged Ski8, as depicted in Figure 2C. The ion signals are shifted, corresponding to shift in mass of 5,077 ± 300 Da. This unambiguously proves that two Ski8 molecules are present in the yeast Ski complex. Our results are in contradiction to a 1:1:1 stoichiometry for Ski2, Ski3 and Ski8 in the Ski complex on the basis of semi quantitative Western Blotting [17].

Tandem mass spectrometry on intact Ski complexes.

To further probe the identity and stoichiometry of the purified protein complexes we explored tandem mass spectrometry on the intact Ski complexes. In such tandem mass spectrometry experiments precursor ions are mass selected and subjected to collision induced dissociation [6, 7, 21, 22]. Figure 3 shows a typical tandem mass spectrum whereby the Ski complex precursor ions having 41+ charges were mass-selected (■).

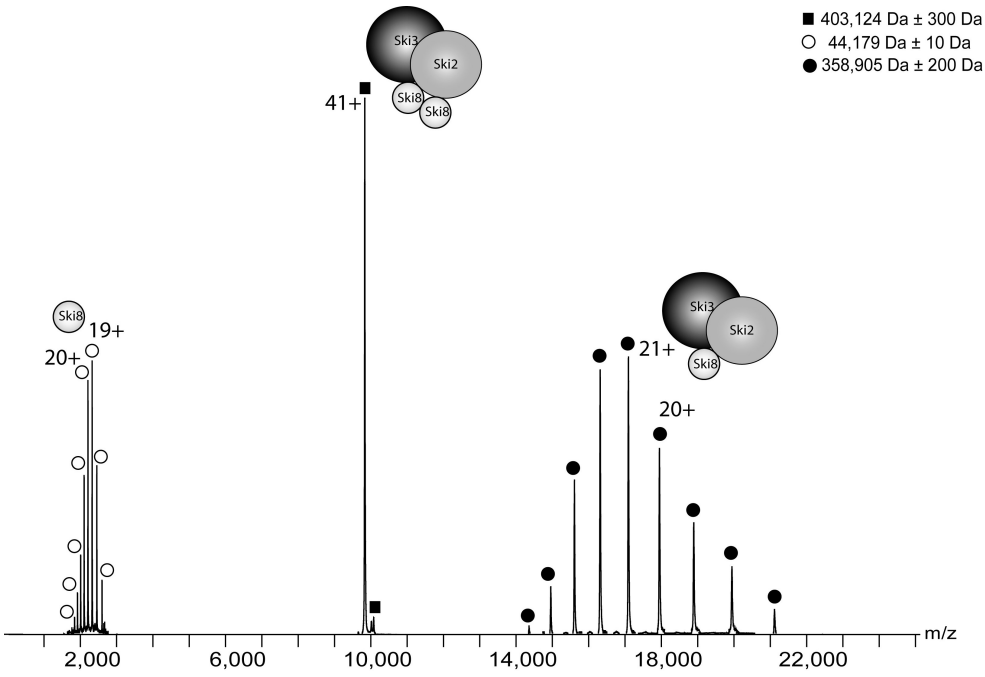


Figure 3

Tandem mass spectrometry on intact Ski complexes. Tandem mass spectrum of the Ski3 TAP-tagged Ski complex. The 41+ charge state of the total Ski complex (■) was subjected to collision induced dissociation using a collision voltage of 175 V. The CID process resulted in the formation of two concomitant fragment ion series, corresponding to a species consisting of Ski3-Ski2-Ski8 (●) with a mass of 358,905 Da and Ski8 (○) with a mass of 44,179 Da, respectively.

The intact Ski complex fragmented into two products. One of these products, appearing between m/z 1,700 and 2,700 (○) originates from a Ski8 with a mass of 44,179 ± 25 Da. The second concomitant product is detected between m/z 14,500 and 21,200 (●) and corresponds to CBP-Ski3-Ski2-Ski8 with a mass of 358,905 ± 200 Da. Indeed, summing up the detected masses of the two fragment ions, leads to a precursor mass of 403,124 ± 500 Da, which is in agreement with the expected mass for the intact complex (see table 1). Tandem mass spectrometry experiments on different charge states of the Ski3-tagged Ski

complex revealed identical fragmentation patterns, as did similar experiments on the complexes obtained from the two different entry points (data not shown). These tandem mass spectrometry results therefore confirm our hypothesis that the Ski complex is a hetero-tetramer consisting of a single copy of Ski2 and Ski3, and two copies of Ski8.

Topology of the Ski complex

Here we probe the overall topology of the Ski complex by a gentle disruption of the complex induced by the addition of 0.5 % formic acid (v/v), prior to mass spectrometric analysis. The spectrum shown in Figure 4 reveals a multitude of charge state distributions, from which the masses of several species could be derived.

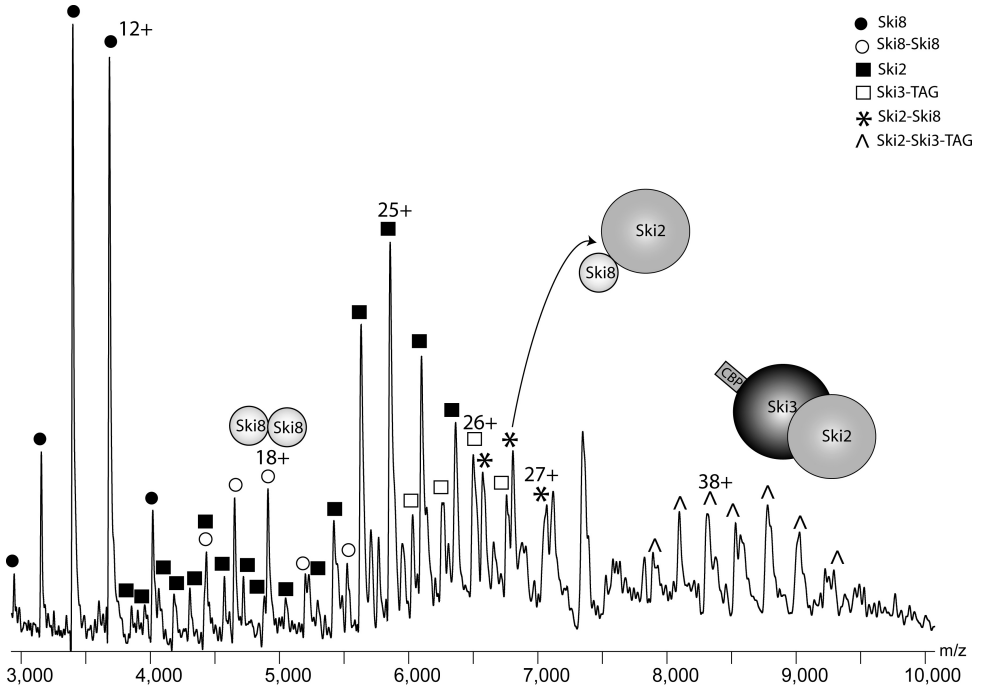


Figure 4

Topology of the Ski complex. Mass spectra acquired of the Ski3 TAP-tagged Ski complex following acid induced disruption. The mass spectrum reveals ion signals of several subunits and sub-complexes, corresponding to Ski8 (●), Ski2 (■) and Ski3 (□) and Ski8-Ski8 (○), Ski2-Ski8 (*) and Ski3-Ski2 (∧)

First of all, no charge distribution was observed to correspond to the intact Ski complex, indicating that the complex was completely dissociated. Between m/z 3,000 and 4,000 a charge envelope from 11+ to 15+ as denoted with (●) matches to a mass of 44,165 ± 25 Da or Ski8. More interestingly, a second charge distribution between m/z 4,300 and 5,600 with charges from 16+ to 20+ is present (○) originating from species with a mass of 88,388 ± 56

Da, matching perfectly to a dimer of Ski8. Furthermore, we detected monomeric Ski2 (between m/z 3,700 to 6,400 with charge states from 23+ to 38+ (■)) with a measured mass of 146,214 ± 150 Da, and monomeric tagged-Ski3 with a mass of 168,909 ± 150 Da (the charge envelope observed between m/z 6,000 to 6,600 with charges from 25+ to 28+ (□)). Two additional subcomplexes were identified. As denoted with (∗) at m/z 6,500 to 7,100 covering charges from 27+ to 29+ and corresponding to a mass of 190,484 ± 100 Da a dimeric subspecies is detected. Taken the experimental mass as determined for monomeric Ski8 and Ski2, a mass of 190,392 Da was expected, clearly assigning this subspecies to Ski2-Ski8. Moreover another charge envelope at m/z 7,700 to 9,300 with charge states of 34+ to 40+ (∧) results in a mass of 315,730 ± 610 Da suggesting a dimeric subcomplex consisting of TAP-Ski3-Ski2. The expected mass of TAP-Ski3-Ski2 is 315,123 Da. Due to the complexity of the spectrum it is not possible to further annotate more charge distributions with a high confidence. However, in an analysis of another Ski complex purification with Ski3 as bait we detected clear signals of a species corresponding to a mass of 234,499 ± 75 Da, suggesting a trimeric subcomplex of Ski2-Ski8-Ski8. A theoretical calculation based on experimental masses suggests 234,570 Da. The experimental measured masses of monomeric proteins and subcomplexes are summarized in table 2 and table 3.

Co-immunoprecipitation experiments in yeast deletion strains have already shown that Ski2-Ski3, Ski2-Ski8 and Ski3-Ski8 are associated with each other [17]. Computational modeling based on suitable structural templates of the individual Ski proteins described the interactions between Ski8-Ski2 and Ski2-Ski3 [18].

Table 2 Masses of the Ski subunits compared with predicted masses as extracted from the gene sequences. CBP with a mass of 5077 Da indicates that the TAP-tag was linked to that subunit.

Ski-Proteins	Measured mass (Da) [#]	Theoretical mass (Da)
Ski8	44,178 ± 25	44,232
Ski2	146,214 ± 150	146,059
Ski3	163,827 ± 30	163,726
Ski8-CBP	49,216 ± 15	49,309
Ski2-CBP	-	151,136
Ski3-CBP	168,909 ± 150	168,803

[#] The indicated error range was estimated from the standard deviations in the measured masses as measured by mass spectrometry following several different purifications.

Table 3 Measured masses of detected sub-complexes compared to the mass of these sub-complexes obtained by summing up the mass of each subunit present. CBP indicates the TAP-tag on the protein.

Sub-complex	Measured mass (Da) #	Theoretical mass (Da)
Ski8-Ski8	88,388 ± 56	88,356
Ski8-CBP-Ski8-CBP	98,404 ± 85	98,432
Ski2-Ski8	190,484 ± 100	190,392
Ski2-Ski3-CBP	315,730 ± 610	315,123
Ski8-Ski3-CBP	-	213,087
Ski2-Ski8-Ski8	234,499 ± 75	234,570

The indicated error range was estimated from the standard deviations in the measured masses as measured by mass spectrometry following several different purifications.

Our mass spectrometric analyses of the resultant sub-complexes confirm most previously identified interactions, but provide even more data to propose a new overall topology. By using each protein of the Ski complex as bait and isolating the Ski complex under stringent conditions we were able to consistently confirm the hetero-tetrameric stoichiometry of the Ski complex. Our data reveal that two copies of Ski8, forming a dimer, are present in the Ski complex. This Ski8 dimer could be further placed into the Ski complex by detecting a trimer consisting of Ski2, and the dimer of Ski8, indicating that at least one of the Ski8 protein is in contact with Ski2. Our data also revealed that Ski2 and Ski3 are connected to each other, which was expected. We did not detect any sub-complexes confirming directly the existence of a direct link between Ski3 and Ski8, but we can not exclude that such a connection exists. Taking all our data together we propose a new structural model for the yeast Ski complex, which is depicted in Figure 5.

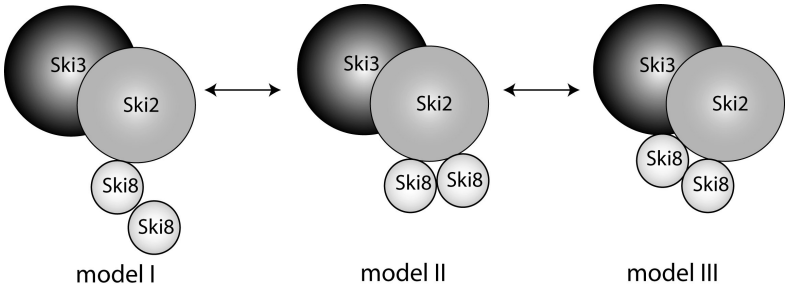


Figure 5

Structural model of the Ski complex. Ski3 is in direct contact with Ski2. Ski2 is connected to at least one Ski8. The second Ski8 is bound to the first Ski8. The direct interaction of the second Ski8 with Ski2 and Ski3 could not be unambiguously established, leading to three similar, albeit slightly different predicted overall topologies.

In Figure 5 three alike albeit different structural topologies are depicted, based on the fact that the two Ski8's are interconnected, Ski2 and Ski3 are linked and at least one Ski8 binds to Ski2. Our data does not allow us to distinguish whether just one (model I) or two Ski8 (model II and III) subunits interact with Ski2, individually, and could therefore be symmetric in the overall structure. Interestingly, in Model I and III both Ski8 proteins are not identical, and may also fulfill different functions in the Ski complex, but this requires further experimental evidence. Model III allows the previously proposed interaction between Ski8 and Ski3.

Discussion

Structure determination on endogenously expressed heterogeneous protein complexes is still a challenging task. Due to minute amounts of purified protein complex, well established techniques such as x-ray crystallography and NMR often fail. In the approach presented here we use the powerful combination of tandem affinity purification combined with standard proteomics methods and macromolecular mass spectrometry to structurally investigate the yeast Ski complex in detail.

We propose a new stoichiometry and topology for the yeast Ski complex. The Ski complex is not a hetero-trimer with 1:1:1 stoichiometry but a hetero-tetramer with a 1:1:2 stoichiometry, whereby one copy of Ski3 and Ski2 and two proteins of Ski8 are assembled per Ski complex. Our mass spectrometric data corroborate the proposed direct interaction between Ski3 and Ski2 as well as between Ski2 and Ski8. The second molecule of Ski8 forms a dimeric structure with the other Ski8 subunit but its exact topology is difficult to be pin-pointed in the overall assembly. Although, our data leave some remaining questions about the Ski complex topology (Figure 5), it will provide a useful template for further more in-depth structural and functional biology approaches into the structure of the Ski complex.

References

1. Rigaut, G., et al., *A generic protein purification method for protein complex characterization and proteome exploration.* Nat Biotechnol, 1999. **17**(10): p. 1030-2.
2. Gavin, A.C., et al., *Functional organization of the yeast proteome by systematic analysis of protein complexes.* Nature, 2002. **415**(6868): p. 141-7.
3. Krogan, N.J., et al., *High-definition macromolecular composition of yeast RNA-processing complexes.* Mol Cell, 2004. **13**(2): p. 225-39.
4. Collins, S.R., et al., *Toward a comprehensive atlas of the physical interactome of Saccharomyces cerevisiae.* Mol Cell Proteomics, 2007. **6**(3): p. 439-50.
5. Sharon, M., et al., *Mass spectrometry reveals the missing links in the assembly pathway of the bacterial 20S proteasome.* J Biol Chem, 2007.
6. van den Heuvel, R.H., et al., *Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry.* Anal Chem, 2006. **78**(21): p. 7473-83.

7. van Duijn, E., et al., *Tandem mass spectrometry of intact GroEL-substrate complexes reveals substrate-specific conformational changes in the trans ring.* J Am Chem Soc, 2006. **128**(14): p. 4694-702.
8. Ilag, L.L., et al., *Mass spectrometry of Escherichia coli RNA polymerase: interactions of the core enzyme with sigma70 and Rsd protein.* Structure, 2004. **12**(2): p. 269-75.
9. Hernandez, H., et al., *Subunit architecture of multimeric complexes isolated directly from cells.* EMBO Rep, 2006. **7**(6): p. 605-10.
10. Synowsky, S.A., et al., *Probing genuine strong interactions and post-translational modifications in the heterogeneous yeast exosome protein complex.* Mol Cell Proteomics, 2006. **5**(9): p. 1581-92.
11. Anderson, J.S. and R.P. Parker, *The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI2 DEVH box protein and 3' to 5' exonucleases of the exosome complex.* Embo J, 1998. **17**(5): p. 1497-506.
12. Widner, W.R. and R.B. Wickner, *Evidence that the SKI antiviral system of Saccharomyces cerevisiae acts by blocking expression of viral mRNA.* Mol Cell Biol, 1993. **13**(7): p. 4331-41.
13. Rhee, S.K., T. Icho, and R.B. Wickner, *Structure and nuclear localization signal of the SKI3 antiviral protein of Saccharomyces cerevisiae.* Yeast, 1989. **5**(3): p. 149-58.
14. Matsumoto, Y., et al., *A yeast antiviral protein, SKI8, shares a repeated amino acid sequence pattern with beta-subunits of G proteins and several other proteins.* Yeast, 1993. **9**(1): p. 43-51.
15. Cheng, Z., et al., *Crystal structure of Ski8p, a WD-repeat protein with dual roles in mRNA metabolism and meiotic recombination.* Protein Sci, 2004. **13**(10): p. 2673-84.
16. Madrona, A.Y. and D.K. Wilson, *The structure of Ski8p, a protein regulating mRNA degradation: Implications for WD protein structure.* Protein Sci, 2004. **13**(6): p. 1557-65.
17. Brown, J.T., X. Bai, and A.W. Johnson, *The yeast antiviral proteins Ski2p, Ski3p, and Ski8p exist as a complex in vivo.* Rna, 2000. **6**(3): p. 449-57.
18. Aloy, P., et al., *Structure-based assembly of protein complexes in yeast.* Science, 2004. **303**(5666): p. 2026-9.
19. Krogan, N.J., et al., *Global landscape of protein complexes in the yeast Saccharomyces cerevisiae.* Nature, 2006. **440**(7084): p. 637-43.
20. Arora, C., et al., *Antiviral protein Ski8 is a direct partner of Spo11 in meiotic DNA break formation, independent of its cytoplasmic role in RNA metabolism.* Mol Cell, 2004. **13**(4): p. 549-59.
21. Sharon, M. and C.V. Robinson, *The Role of Mass Spectrometry in Structure Elucidation of Dynamic Protein Complexes.* Annu Rev Biochem, 2007.
22. Sobott, F., et al., *A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies.* Anal Chem, 2002. **74**(6): p. 1402-7.

Chapter 6

Summarizing discussion and concluding remarks

Chapter 6

In the cell many genes are transcribed and translated simultaneously and give rise to numerous different proteins. The expression levels of these different proteins vary considerably; some proteins are present in only a few copies and others in thousands per cell. These proteins do not exist as single entities but they form large, heterogeneous non-covalent protein assemblies. The physical interaction between proteins is very often crucial for functionality in the cell. This functionality of protein complexes is in turn frequently regulated by post-translational modifications such as phosphorylation, acetylation etc. The interrelation between protein-protein interactions and protein post-translational modifications determines to a large extent the functional organization of a cell. The analysis of endogenously expressed protein assemblies is therefore especially relevant as it represents the most authentic situation. Proteins are expressed in their natural environment, assemble with their natural interactions partners and are modulated according to a genuine situation.

With the development of the tandem affinity purification, the large scale isolation of endogenous protein complexes from cells has become possible. The procedure consists of two subsequent affinity purification steps which keep the protein interactions intact. The technique is relatively fast, reproducible and yields pure protein assemblies. By proteomics approaches large protein-protein interaction networks of *S.cerevisiae* have been unravelled, which indicated that proteins are organized into modules of protein complexes. These proteomics experiments provided very useful information about the constituents of different protein complexes. The next step towards a comprehensive analysis of protein assemblies is to characterize these protein assemblies in detail. It is important to determine the direct interaction between proteins, the overall topology of protein complexes, strong and weak interacting proteins and how these interactions are modulated.

Native mass spectrometry may fill this niche as it offers the unique possibility to analyze different aspects of heterogeneous endogenously expressed protein complexes. The requirement of just a low amount of protein complex (few µL of a minimal concentration of 0.5 µM) for a comprehensive mass spectrometry analysis makes this technique ideal for the analysis of endogenous protein assemblies. This area of mass spectrometry investigates proteins in their highest state of protein organization by keeping the non-covalent interactions intact.

In this thesis latest developments and accomplishments in the analysis of large heterogeneous protein assemblies are described. Combining mass spectrometry of intact protein (-complexes) and proteomics approaches with affinity purifications techniques highlights the fascinating developments in the detailed characterization of large endogenously expressed protein assemblies. Each single mass spectrometric technique (peptide-, protein- and protein complex analysis) provides complementary information, which contributes to elucidate the complex puzzle of protein assemblies. Proteins of the purified protein assemblies and possible post-translational modifications are initially identified by proteomics approaches. Individual protein masses and the total protein

complex mass are then subsequently determined by native mass spectrometry and related to stoichiometry, direct protein-protein interactions, topology and relative gas-phase binding affinities.

The work described in this thesis particularly focuses on the in-depth characterization of protein complexes isolated from *S.cerevisiae* that play an important role in the RNA metabolism in the cell.
A tandem mass spectrometer customized for the analysis of large protein assemblies is described in chapter 2. The modified first generation QToF allows tandem MS analyses on non-covalent protein complexes. Several modifications in the QToF instrument transform it now into a powerful tool in the structural analysis of heterogeneous protein assemblies. It enables not only the determination of the total protein complex mass and thereby the stoichiometry, but also offers the controlled dissection of protein complexes in the gas-phase. At this moment the majority of gas phase dissociations mostly yield in the asymmetric charge partitioning with regards to the mass. Typically the smallest protein is ejected with many charges from the remaining low charged protein complex. This dissociation pathway is often very different to in-solution behaviour. In the literature only few examples are documented with symmetric charge distribution upon gas phase protein complex dissociation in which tetrameric protein complexes dissociate into dimeric sub-complexes. A future challenge is the selective manipulation of gas-phase dissociation to extrapolate the position of the proteins within the protein complex.
In chapter 3, three different mass spectrometric techniques are applied and demonstrated on the endogenously expressed exosome protein complex. At first proteomics techniques identified all known constituents of the exosome with high amino acid sequence coverages up to 75%. This analysis also revealed a first generation phosphorylation map of the yeast exosome. Seven different exosome proteins contain at least one phosphorylation site at either serine or threonine residues. The phosphorylation site at serine152 in exosome component Rrp4 is identical to the phosphorylation site in the human counterpart. Serine 152 is situated in the S1 RNA binding domain, a structural motif that contains several basic residues and is involved in the binding of RNA molecules. When the crystal structure of the human exosome was finally released at the end of the year 2006, this particular phosphorylation site could be located in the three dimensional structure of the human exosome. As the structure of the core exosome consists of a hexameric ring crowned with three RNA-binding proteins, it was interesting to observe that this particular phosphorylation site faces the inner cavity. Further functional studies of the yeast exosome will reveal the exact mechanism in the processing and degradation of RNA. In the second mass spectrometry approach the exosome complex was disrupted into monomeric proteins. By reversed phase chromatography the protein mixture was separated and mass analyzed. These masses were used to predict the total mass of the exosome complex and sub-

complexes. In the third approach the successful ionization and detection of a heterogeneous endogenously expressed protein assembly was demonstrated by native mass spectrometry for the first time. The native mass spectrum of the exosome yielded in several different charge envelopes corresponding to the total exosome, as envisioned in the cytoplasm but not associated with Ski7, and four different sub-complexes. The missing proteins between the total exosome and its sub-complexes allowed establishing relative binding affinities within the protein complex. All sub-complexes shared the characteristic that one protein containing an S1-motif (Csl4) was not present. Two additional proteins containing S1 RNA-binding domains (Rrp4, Rrp40) located on top of the hexameric ring as well as a peripheral protein of the exosome (Dis3) were selectively dissociated from the exosome assembly. The hexameric ring proteins (Rrp41, Rrp42, Rrp43, Rrp45, Rrp46, and Mtr3) were always stably associated with each other.

Since protein complexes are very dynamic in the cell, they can readily associate with different proteins. The cell contains at least two distinct forms of exosome protein complex that are localized in different cellular compartment: the cytoplasm and the nucleus. Both share a pool of common proteins (Rrp41, Rrp42, Rrp43, Rrp45, Rrp46, Mtr3, Csl4, Rrp4, Rrp40 and Dis3), but differ in several decorating proteins. Chapter 4 summarizes relative differences between the nuclear and primarily the cytoplasmic exosome. Using stable isotope labeling all common exosome proteins were present in equal stoichiometry. However several distinct different proteins were characteristic for each exosome complex. The nuclear exosome is associated with Rrp6, Lrp1, Ynr024w, importin α and β whereas the cytoplasmic exosome contains Ski7. Sequence alignments suggest the nuclear hypothetical protein Ynr024w be the homolog of human exosome co-factor M-phase phosphoprotein 6 (MPP6), which is involved in the processing of rRNA. Interesting is the high abundance of importin α and β associated to the nuclear exosome. Both importins are generally thought to help in translocating proteins from the cytoplasm into the nucleus. The specific association with the nuclear exosome suggests that the importins may play an important role in the translocation from the cytoplasm into the nucleus. Our labeling strategy could confirm several previously identified exosomal phosphorylation sites (S94 in Csl4, S251 in Rrp46, S88 and S90 in Ski7) and detected one novel phosphorylation site (S708/709 in Rrp6). Some of these phosphorylated peptides were found in the nuclear and cytoplasmic exosome and could be relative to each other quantified. In particular S94 in Csl4 was significantly higher abundant in the nuclear exosome.

Both exosomes were also studied by native (tandem) mass spectrometry using the modified Q-ToF. The native mass spectra of the nuclear exosome displayed as main species an exosome assembly that consisted of all shared exosomal protein lacking Csl4. This sub-complex is identical to the main charge envelope originating from the exosome isolated with entry point Ski7 and Csl4. It suggests that the most stable sub-complex of exosome species consists of the hexameric ring plus Rrp4, Rrp40 and Dis3. The comparison of the native

mass spectra of both exosome species also pointed to clear differences. A low abundant charge series at high m/z values in the analysis of the nuclear exosome corresponds to the total nuclear exosome including all shared components, Rrp6, Lrp1, Ynr024w and importin α/β with an overall mass of 691 kDa. Moreover strong signal for monomeric importin α and β as well as for the hetero-dimeric form were observed. MS/MS experiments with the selection of the stable sub-complex with EP Rrp6 as precursor illustrated different dissociation pathways. At lower energy Rrp40 with a low number of positive charges was ejected whereas at higher energy Rrp40 or Rrp4 dissociate from the precursor. In a second lower abundant exosome complex the precursor still contains Csl4. Here the same dissociation pattern is observed as before but also Csl4, which is attaining a high number of charges, can be ejected from the precursor at high collision energy. The highly asymmetric charge distribution upon dissociation implies that Rrp40 is dissociating from the exosome in a near native fold and suggests that it has lower binding affinities towards the exosome in the gas-phase.

Protein complexes that are involved in RNA metabolism were further explored by investigating the Ski complex as described in chapter 5. The Ski complex is localized in the protein complex network environment of the exosome. Both protein assemblies are essential for the RNA degradation in the cytoplasm and share a common linker protein (Ski7p). The mass spectrometric analyses of the Ski complex revealed a hetero-tetrameric protein assembly, a newly defined stoichiometry comprised of Ski2, Ski3 and two copies of Ski8. Gas-phase dissociation experiments that were performed on the modified Q-ToF instrument selectively ejected one copy of Ski8 from the total protein complex at high collision energies and thereby confirmed our proposed stoichiometry. In-solution disruption of the Ski complex gave insight in the direct protein-protein interaction by the mass spectrometric detection of dimeric and trimeric sub-complexes (Ski8-Ski8, Ski2-Ski3, Ski8-Ski2 and Ski8-Ski8-Ski2). This data allowed placing the second Ski8 protein within the Ski complex structure. Finally an improved model for the Ski complex was reconstructed.

Challenges

While this thesis clearly demonstrates the power of mass spectrometric techniques in a detailed analysis of protein complexes, so far the routine semi-automated analysis of endogenous protein complexes using native mass spectrometry has not been accomplished. With the development of mass spectrometers that feature high sensitivity in the analysis of non-covalent protein complexes, the main “bottleneck” still persists in the interface coupling biochemistry to mass spectrometry. The sample preparation is crucial for a valuable mass spectrometric analysis. It has to meet several MS compatible requirements. During purification procedures of protein complexes the use of nonvolatile buffers and solubilizing agents such as glycerol, DTT, EDTA and in particular detergents ensures the conformational integrity of the complex. Since ESI mass spectrometry is sensitive to non-

volatile salts and is incompatible with detergents, these additives must be removed prior to MS analysis.

The standard TAP protein complex purification protocol has been successfully applied to generate a comprehensive map of the interactome in yeast [1, 2]. In such studies it is not required to keep the protein complex intact after the purification since proteins are digested and peptides are subjected to nano LC-MS/MS in these “bottom up” proteomics techniques. In this procedure the proteins are precipitated and electrophoretically separated. The proteins are then proteolytically digested and the resulting peptides are analyzed by mass spectrometry. SDS-gel based electrophoresis easily removes low molecular weight contaminations that would be disturbing the MS analyses.

For native mass spectrometry, it is indispensable to preserve the non-covalent interactions of the protein complex while preparing a mass spectrometry friendly sample. One attractive and simple technique is the use of spin columns with cut-off filters. These filters are used for removing molecules below the utilized cut-off such as small molecules and detergent, as they rush through the membrane. In this procedure the purification buffer can be exchanged to any volatile mass spectrometry buffer of any desired ion strength and pH. Moreover highly diluted protein complex solutions can be concentrated to amounts detectable by MS.

These filter devices pose several implications in the analysis of TAP-purified endogenously expressed protein complexes. Though these cut off filters are generally claimed to have a low affinity towards protein binding, they do bind certain amounts to saturate their membrane. Since purifications of endogenous protein complexes yield only several micrograms of material, if they are moderately expressed, they are frequently absorbed on the spin column membrane. This problem cannot be easily solved. Efforts in coating the membrane with an abundant carrier protein prior to exposure to the endogenous protein assembly proved unsuccessful. Traces of small molecular weight carrier protein are likely to contaminate the final protein complex preparation and may ionize more efficiently in the native mass spectrometric analysis. Another option to overcome this restriction is to scale up the starting material, in terms of yeast cells, number of purifications and final yield of intact protein assembly for the mass spectrometric clean up procedure. This approach emerged to be effective and in several cases successful. Unfortunately several variables were difficult to be controlled such as reproducible purity and yield of numerous individual protein complex purifications, which are finally combined for the native MS analysis. The parallel processing of several individual purifications considerably increased not only the length of the general purification procedure but also the time-course during the usage of the filter devices prior to the native MS analysis. Protein complexes feature strong but also transient interactions. These transient interactions may be easily lost during time-consuming isolation procedures. During the purification protocol mass spectrometry incompatible detergents are used. Problems arise in the reduced ionization efficiency of the macromolecule in the presence of only very low amounts of detergent. Ideally these should

be removed through the cut-off membrane of these filter devices. However detergents form micelles in aqueous solution due to their amphiphilic properties. These micelles can generate globular structures with varied diameters and may become too large to pass the membrane. Low amounts of detergent in the final mass spectrometry sample may inhibit the ionization efficiency of the protein complex. Increasing the pore size of the cut-off membrane (5 kDa versus 100 kDa) allows removal of large micelles. Protein complexes are dynamic entities in solution they constantly form and dissociate. If the cut-off filter exceeds the size of the smaller proteins of the protein complex, these are prone to also flush through the filter membrane and are subsequently lost in the final native MS analysis.

The identity of proteins comprising to a protein complex only originate from proteomics experiments as native mass spectrometry does not give any information about the amino acid sequence of proteins. Native mass spectrometry in turn reveals the masses of proteins and higher ordered structures of protein complexes. To facilitate the interpretation of the native mass spectrometry results, the protein complex preparation needs to yield in a highly pure sample. The purification procedures have to be optimized to discriminate against high abundant and sticky proteins. Relatively high salt washes during the tandem affinity purification procedure may be necessary to remove non-specific proteins. Again it cannot be excluded that during these procedures transient interactors of the protein complex are also lost.

Future Outlook and concluding remarks

Contemporary proteomics studies on protein complexes are often limited to reporting constituents of protein assemblies in different organisms under defined conditions. With current advances in native mass spectrometry a powerful method has been developed to address structural aspects of intact multi-component protein assemblies. As protein complexes have a high molecular weight and frequently a heterogeneous protein composition, established structural techniques such as NMR and x-ray crystallography are often not applicable. Native mass spectrometry is not limited by these characteristics of protein complexes and has therefore found a complementary niche in the field of structural biology. This technique gives valuable information about the mass of intact protein assemblies and sub-complexes in a single analysis. Combined with information about the individual mass of proteins, native mass spectrometry can solve the complex organizational level of protein assemblies. It can probe stoichiometry and topology of large protein machineries. Latest developments in this structural field allow the targeted dissection of protein assemblies in the gas-phase by tandem mass spectrometry. This technique can be used to reconfirm stoichiometry and to probe gas-phase stabilities of protein complexes. An appealing advantage of this structural technique is the fast analysis and the low amount of sample consumption. Therefore this technique is particularly important in the structural investigation of endogenously expressed protein assemblies which yield only minute

amounts of sample when isolated from cells. Increasing sensitivity of mass spectrometric instrumentation together with further improvements in the isolation of endogenous protein complexes will strengthen the role of native (tandem) mass spectrometry in the analysis of protein complexes that are physiologically relevant in the cell. It will then become possible to routinely investigate endogenous protein assemblies and functional assays, site-specific mutational studies and changes in protein complex characteristics upon external stimuli will become increasingly interesting.

References

1. Gavin, A.C., et al., *Functional organization of the yeast proteome by systematic analysis of protein complexes.* Nature, 2002. **415**(6868): p. 141-7.
2. Krogan, N.J., et al., *Global landscape of protein complexes in the yeast Saccharomyces cerevisiae.* Nature, 2006. **440**(7084): p. 637-43.

Samenvatting in het Nederlands

Voor veel genen vindt gelijktijdig transcriptie en translatie plaats met behulp van een groot aantal eiwitten in de cel. In de laatste jaren werd het duidelijk dat eiwitten vaak niet alleen functioneren, maar dat ze zijn georganiseerd in grote heterogene eiwit-assemblages, die een extensief netwerk in de cel vormen. Eiwitten zijn de meest veelzijdige macromoleculen en hun directe interactie met elkaar is vaak onmisbaar voor het functioneren van de cel. De activiteit van deze eiwit-assemblages kan door post-translationale modificaties, zoals fosforylering geregeld worden. Eiwitcomplexen zijn vereist in verschillende cellulaire processen en vervullen katalytische, regulatorische, structurele en mechanische functies. Het bestuderen van endogene eiwit-assemblages is in het bijzonder interessant en belangrijk omdat deze macromoleculen zich in hun originele omgeving bevinden, en daarmee de werkelijke situatie in de cel weerspiegelen. Met de ontwikkeling van tandem affiniteit zuivering (TAP) werd de isolatie van endogene eiwitcomplexen op grote schaal mogelijk. Deze procedure bestaat uit twee opeenvolgende affiniteits zuivering stappen, die het eiwit complex intact houden. De techniek is relatief snel, herhaalbaar en levert het zuivere eiwit-complex van interesse .

Het grote eiwit-eiwit netwerk van *S.cerevisiae* (gist) werd met behulp van proteomics bekend en wijst op de organisatie van eiwitten in modulen van eiwitcomplexen. Deze experimenten hebben nuttige informatie over de bouwstenen van eiwitcomplexen geleverd. De volgende stap naar een gedetailleerde analyse van deze eiwit-assemblages is de karakterisering van de individuele eiwitcomplexen in detail. Massa spectrometrie biedt unieke mogelijkheden om verschillende aspecten van heterogene endogene eiwitcomplexen te analyseren. Zoals eerder genoemd, proteomics is de keuze om de identiteit van eiwitten vast te leggen. Het eiwit wordt gedigesteerd en de sequentie van de peptiden wordt door LC-MS/MS bepaald. In combinatie met de verrijking van peptiden, die gefosforyleerd zijn, is het mogelijk de fosforylerings-site te bepalen. Deze data geeft informatie over de post-translationale status van het eiwitcomplex en hoe de interacties geregeld worden.

Verder is het ook belangrijk de stoichiometrie, de directe interacties tussen eiwitten, de algemene topologie van eiwitcomplexen en de affinitiet van de interacterende eiwitten te bepalen. Deze informatie kan worden verkregen door gebruik te maken van natieve (tandem) massa spectrometrie.

Deze techniek wordt gebruikt, om de intacte massa van de complexen te bepalen. Tevens kunnen de eiwitcomplexen in de gas-fase gecontroleerd worden gedissociëerd, waardoor extra informatie over de opbouw van het complex kan worden verkregen. De techniek onderzoekt de eiwit-assemblages in hun functionele vorm omdat ook de niet-covalente bindingen tussen de eiwitten onderling behouden blijven. Het minimale verbruik van monster (enkele µl's met een minimale concentratie van 0.5µM) maakt de techniek ideaal voor de analyse van endogene eiwitcomplexen.

Dit proefschrift behandelt de laatste ontwikkelingen en toepassingen van de analyse van grote heterogene eiwit-assemblages. De combinatie van diverse massa-spectrometrische methoden met tandem affiniteit zuivering onderstreept de fascinerende ontwikkelingen in de gedetailleerde analyse van endogene eiwitcomplexen. Elke individuele massa-spectrometrische methode, zoals peptide-, eiwit- en eiwit assemblage analyses levert complementaire informatie op, die een bijdrage levert om de complexe puzzel van intacte functionele eiwit-complexen op te lossen. In het hier beschreven onderzoek staat de karakterisering van eiwit-assemblages, die een belangrijke rol in het RNA metabolisme in *S.cerevisiae* spelen, in het middelpunt. Hoofdstuk 2 beschrijft een tandem massa-spectrometer (QToF1 instrument), die voor de structurele analyse van grote niet-covalente eiwitcomplexen aangepast werd. Dit instrument staat niet alleen de massa bepaling van het intacte eiwit-assemblages toe, maar maakt het ook mogelijk het eiwitcomplex in de gas-fase te dissociëren. Met deze informatie kunnen conclusies over de stoichiometrie, affiniteiten van individuele eiwitten aan het complex in de gas-fase en eiwit interacties met kleine moleculen (zoals bijvoorbeeld remmers) getrokken worden. De mogelijkheden van deze massa-spectrometer worden in hoofdstuk 2 met behulp van de heterogene tetrameer Hydroquinone 1,2-dioxygenase geïllustreerd.

In hoofdstuk 3 worden drie verschillende massa-spectrometrische technieken gebruikt om het endogene, heterogene exosoom eiwit complex nader te onderzoeken. Het cytoplasmatische exosoom bestaat uit tien eiwitten, die een hexameer ring vormen (Rrp43-Mtr3-Rrp42-Rrp41-Rrp45-Rrp46) met nog drie eiwitten aan de bovenkant van de ring (Rrp4, Csl4, Rrp40) en Rrp44 onder de ring. Het exosoom werd met behulp van de TAP techniek gezuiverd en dit resulteerde in een zuiver intact eiwit-assemblage. Allereerst werd het eiwit complex op een 1-D gel gescheiden en proteolytisch digesteerd. Met peptide LC-MS/MS methoden werden alle bekende exosoom onderdelen geïdentificeerd, waarvan tot 75% van de sequentie werd teruggevonden. Deze analyse resulteerde ook in de eerste overzichtskaart van fosforyleerings-sites in het gist exosoom. Zeven verschillende exosoom eiwitten bevatten minstens een fosforylerings-site op een serine of threonine residu. De fosforyleerings-site serine 152 in Rrp4 is identiek met het homologe eiwit in de humane exosoom. Serine 152 bevindt zich in het S1 RNA bindings domein, wat een structureel motief is met veel basische aminozuren en bekend is om RNA te binden. Met de opheldering van de kristal-structuur van het humane exosoom in 2006, werd deze specifieke fosforyleerings-site gelokaliseerd aan binnenzijde van Rrp4.

Nadat met de eerste techniek het eiwit complex op het peptide niveau geanalyseerd werd, focusseert de tweede techniek op individuele eiwitten. Door het toevoegen van zuur viel het intacte eiwit-assemblage uit elkaar. Het mengsel van eiwitten werd op een LC-system gescheiden en de exacte massa's van de eiwit monomeren werden middels eiwit massa-spectrometrie bepaald. Deze massa's werden vervolgens gebruikt, om de totale te

verwachten massa van het exosoom te voorspellen. Met de derde analytische techniek werden de massa's van het intacte complex en van sub- complexen gemeten door natieve massa-spectrometrie. Het natieve spectrum leverde meerdere lading series op, die overeen stemden met het totale intacte exosoom en exosomen die ofwel Csl4, Csl4 + Rrp4, Csl4 + Rrp40 of Csl4 + Rrp44 verloren hadden. De hexameer-ring was altijd stabiel met elkaar verbonden. Met deze data kan informatie over sterk en zwak bindende eiwitten binnen het exosoom worden verkregen.

De cel bevat tenminste twee verschillende exosoom complexen, waarvan één gelokaliseerd is in het cytoplasma en de andere in de kern. Hoofdstuk 4 beschrijft de relatieve verschillen tussen deze twee exosomen. Met gebruik van stabiele isotoop-labeling werd vastgesteld, dat de basis-opbouw van beide exosomen niet van elkaar verschilden (Rrp43, Mtr3, Rrp42, Rrp41, Rrp45, Rrp46, Csl4, Rrp40, Rrp4, Rrp44), maar alleen onderscheid aanwezig was in de decoratieve eiwitten, die zwakker met het exosoom interacteren. Het cytoplasmatische exosoom bevat Ski7, dat niet aanwezig is in het nucleaire exosoom. Daarentegen associeert het nucleaire exosoom met Rrp6, Lrp1, Ynr024w en importin α/β, die niet aanwezig zijn in het cytoplasmatische exosoom. Ynr024w is een hypothetisch eiwit en is wellicht een homoloog van het humane exosoom co-factor M-phase fosfo-eiwit 6 (MPP6), die bij het produceren van rRNA een belangrijke rol speelt. Verder was de hoge abundantie van importin α/β verrassend. Importins staan er om bekend macromoleculen tussen het cytoplasma en de kern te vervoeren. Met de labellings-strategie konden ook bekende en onbekende fosforyleerings-sites gedetecteerd en gekwantificeerd worden. De fosforylering op serine 94 in Csl4 komt vaker voor op het exosoom dat aanwezig is in de kern dan op het cytoplasmatische exosoom terwijl de fosforylering op serine 251 in Rrp46 gelijk is voor beide complexen. Beide exosomen werden ook geanalyseerd door natieve massa-spectrometrie, waarbij de basis van het complex van beide exosomen zonder Csl4 (en in een kleine lading series met Csl4) als het meest stabiele subcomplex geïdentificeerd werd. Daarnaast was er in het spectrum van het nucleaire exosoom nog signaal aanwezig dat overeenkwam met het totale intacte nucleaire exosoom alsmede voor de importins. Tandem massa-spectrometrie aan het stabiele basis complex van het exosoom zonder Csl4 (en met Csl4) liet Rrp40 met een laag aantal aan ladingen en Rrp4 (en Csl4) met een hoog aantal aan ladingen dissociëren. De lage lading wijst op een gevouwen structuur op het moment van dissociatie, dat een zwakkere binding van Rrp40 in het exosoom in de gas-fase impliceert.

Een ander eiwit complex, dat een belangrijke rol bij RNA metabolisme speelt, is het Ski complex. Structureel onderzoek van dit eiwit complex met natieve (tandem) massa-spectrometrie wordt beschreven in hoofdstuk 5. Het Ski complex bevindt zich in het cytoplasma en functioneert als co-factor van het exosoom bij de correcte afbraak van boodschapper RNA. Een nieuwe stoichiometrie van het Ski complex werd met behulp van natieve massa-spectrometrie ontdekt. Het Ski complex is een hetero-tetrameer eiwit complex dat bestaat uit Ski2, Ski3 en twee kopieën van Ski8. Deze stoichiometrie werd

bevestigd met gebruik van tandem massa-spectrometrie, uitgevoerd met een gemodificeerd QToF1 instrument zoals eerder beschreven in hoofdstuk 2. Het toevoegen van weinig zuur aan het eiwit complex induceert dissociatie van het eiwit complex en dimeer en trimeer subcomplexen van Ski8-Ski8, Ski2-Ski3, Ski8-Ski2 en Ski8-Ski8-Ski2 werden geanalyseerd. Deze data geeft informatie over directe eiwit-eiwit interacties en positioneert het tweede molecuul Ski8 in de totale Ski complex structuur. Hierdoor kon uiteindelijk een verbeterd model van het Ski complex worden gereconstrueerd.

Curriculum vitae

Silvia Synowsky was born on 08th July, 1976 in Kreuzburg, Poland. After secondary school in Remscheid, Germany, she received her qualifications for university entrance in 1996. Thereafter, she started her university education in 1997 with an undergraduate degree (equivalent) in biology at Georg-August University in Goettingen, Germany. After spending a year as an exchange student at University of Birmingham, UK, she continued her master degree (diploma equivalent) at Ruprecht-Karls Univerisity of Heidelberg, Germany, specializing in biochemistry where she graduated in 2003. During her diploma thesis she unravelled direct protein protein interaction of the yeast exosome by using cross-linking, mass spectrometry and western blotting. This work was performed at Cellzome AG, Germany under supervision of Bernhard Kuester.

In January 2004 she set off to Utrecht, The Netherlands to pursue her Ph. D. degree at Biololecular Mass Spectrometry and Proteomics Group under supervision of Prof. Dr. Albert Heck. As described in this thesis she focused on the detailed analysis of heterogeneous protein complexes by mass spectrometry and will be defending her Ph.D. thesis in May 2008.

List of Publications

- **Synowsky SA**, van den Heuvel RH, Mohammed S, Pijnappel PW, Heck AJ. Probing genuine strong interactions and post-translational modifications in the heterogeneous yeast exosome protein complex.
  Mol Cell Proteomics. **2006**;5(9):1581-92.

- **Synowsky SA**, Heck AJ. The yeast Ski complex is a hetero-tetramer.
  Protein Sci. **2008**;17(1):119-25.

- **Synowsky SA**, Raijmakers R, van Wijk M, Heck AJ. Comparative study of the nuclear and cytoplasmic exosome by mass spectrometry. *Manuscript in preparation*

- Moonen MJH*, **Synowsky SA***, van den Berg WAM, Westphal AH, Heck AJR, van den Heuvel RHH, Fraaije MW, van Berkel WJH. Hydroquinone Dioxygenase from Pseudomonas fluorescens ACB:
  A Novel Member of the Family of Non Heme Iron(II)-Dependent Dioxygenases. *submitted to J Bacteriol.* *These authors contributed equally to this work

- van Baal I, Malda H, **Synowsky SA**, van Dongen JL, Hackeng TM, Merkx M, Meijer EW. Multivalent peptide and protein dendrimers using native chemical ligation. Angew Chem Int Ed Engl. **2005**; 12;44(32):5052-7.

- van den Heuvel RH, van Duijn E, Mazon H, **Synowsky SA**, Lorenzen K, Versluis C, Brouns SJ, Langridge D, van der Oost J, Hoyes J, Heck AJ. Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. Anal Chem. **2006**; 1;78(21):7473-83.

Acknowledgements

When I started my Ph.D, in native mass spectrometry of protein complexes at the beginning of 2004 in Utrecht I was excited to learn more about this rather exotic MS-method and was looking very much forward to living in The Netherlands. I very much enjoyed and appreciated the advice and discussions I had with Albert. I met him as an enthusiastic, reasonable, uncomplicated person who gave me a lot of freedom to do research in his lab. His special interest in mass spectrometry of non-covalent protein complexes often made him sneak into the lab, if time allowed.
And what would I have done in my first year without Kees? He took such great care of me and taught me most of my knowledge in native mass spec. I very much appreciated all his practical advice and his help in- and outside of the lab. He definitely knows how to make a foreigner feel welcome!
At the beginning of my Ph.D. the group was rather small and most of my colleagues were dutch. It took me some time to get used to the group atmosphere and the dutch humor. And it was nice to also spend some time with Pim at the Medical Centre to perform the biological side of the project and I appreciated his input on how to best purify protein complexes from cells. During the coming years our group expanded tremendously and in that respect the proportion of foreigners in the group became larger. With Kristina, Andreas and Wilma (and later Charlotte (my german bread supplier) and Nikolai) more germans joined the group. Kristina and I started up our own little lunch group in the coffee room with nice warm dishes (mostly prepared by Kristina). The food was delicious and such a relief from the fried krokets (and whatever they offer) from the cantine. At the beginning Kristina's and my project resembled each other and it was good to have someone to share knowledge and frustration with. I also remember so many nice evenings with colleagues, BBQ's and food at her place.
In general I had great officemates: Silvie (excellent in communication), Anne (very nice and quiet), Marco (best with computers) and Wilma (most hard-working colleague I ever met). I was so lucky that I could share an office with Esther in my last year. Who could have been a better match to me than her with all her native MS expertise? She is so clever, efficient, easily approachable, always willing to share knowledge and is still modest about it. I am sure there is a great future ahead of her! I also very much appreciated her correction of any dutch letter I had to write. It somehow always turned out much better than my own versions ☺.
And I am so proud to know one of only 400 000 maltese people in the world: Sharon. There is no one who is as funny and talkative as her and her mediterranean character definitely spices up the group. I very much liked our „moments" at occasional coffeebreaks with kinder bueno. It was great to have scientific exchange visits with the MS group from Odense, Denmark. At this occasion I met Shabaz and Nadia who joined our group later on. Shabaz knowledge in proteomics is remarkable and I value his opinions regarding science a lot. His future oriented thinking and support helped me so much when it was time to find a new position for the coming years. Nadia's sparkling positive energy is infectious. She is such a kind and warm hearted person and knows how to bond with people very well. I remember so many great moments with her: exploring NL, saving a sheep and a bird, dancing tunak-tunak, watching movies and most importantly her work about lysN. And Reinout, my exosome expert. He is such a gain to the whole group and has definetely put the group onto a higher level in science. His knowledge, support, efficiency and scientific creativity make him such a great collegue. And then there is Thin Thin, an affectionate and simply a lovely person. I so much enjoyed spending time with her eating Danish hot dogs and visiting Madurodam of which we took at least one hundred pictures and they all look the same. When Jeffrey, my favourite caveman, joined the group I was immediately fascinated by his spirit for adventure and his outdoor appearance. We had such a great time together and I loved hearing stories about the Himalaya, Greenland, Sweden and all his unconventional

friends! And there were so many colleagues, Martina (lovely irish accent when speaking dutch), Maarten, Sharon, Nadia, Reinout, Soenita, Mirjam and Shabaz with loads of patience to involve me in maintaining mass spectrometers dedicated to proteomics. I hope my thousands of questions were not too demanding. There are probably many, many more good memories to be mentioned here, please forgive me that I stop. When it comes down to what I have learned within these years in the Netherlands, I would say a great deal of mass spectrometry and I met colleagues and friends that I will remember for a long time.

At the end I would like to thank Albert, Kees, Pim, Nadia, Sharon, Esther, Thin Thin, Jeffrey, Maarten, Tineke, Wilma, Anne, Marco, Silvie, Kristina, Andreas, Charlotte, Nikolai, Paul, Simone, Jeroen, Arjen, Joost, Javier, Reinout, Shabaz, Jeroen, Alessandro, Poupak, Martina, Mirjam, Bas, Henk, Robert, Madalina, Bobby, Natasja, Arjan, Renske, Onno, Gideon, Bas, Soenita, Anca, Monique, Corine, Werner and Nasrin for your scientific and social impact on my work and life here in The Netherlands.

All the best and take care!