PROBING GENUINE STRONG INTERACTIONS AND POST-TRANSITIONAL MODIFICATIONS IN THE HETEROGENEOUS YEAST EXOSOME PROTEIN COMPLEX

Silvia A. Synowsky¶, Robert H. H. van den Heuvel¶, Shabaz Mohammed¶, W. W. M. Pim Pijnappel‡, and Albert J. R. Heck¶§

¶Department of Biomolecular Mass Spectrometry, Bijvoet Center for Biomolecular Research & Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Sorbonnelaan 16, 3584 CA Utrecht, The Netherlands and ‡Department of Physiological Chemistry, Division of Biomedical Genetics, University Medical Center Utrecht, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands

§To whom correspondence should be addressed: Albert J. R. Heck, Department of Biomolecular Mass Spectrometry, Bijvoet Center for Biomolecular Research & Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Sorbonnelaan 16, 3584 CA Utrecht, The Netherlands, E-mail: a.j.r.heck@chem.uu.nl, Telephone: (+31) 30 253 6797, Fax: (+31) 30 251 8219

Running title: Yeast exosome (sub)-complexes

1 The abbreviations used are: TAP, tandem affinity purification; LTQ-FT-ICR, linear ion trap coupled to fourier transform ion cyclotron resonance mass spectrometer; 1D gel LC MS/MS, denaturant gel liquid chromatography tandem mass spectrometry; protein LC MS, intact protein liquid chromatography mass spectrometry; macromolecular MS, native macromolecular mass spectrometry
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 \textit{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 subcellular 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 °C 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 MgCl₂, 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 °C. 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 °C. 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 °C. 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 \textit{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 °C). 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 °C) (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.

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). 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 I and II 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 I and II). 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 phosphorylation-specific stain, one or more phosphorylation sites could be identified in Ski7, Rrp6, Rrp43, Rrp4, Csl4, Mtr3 and Rrp46 (Table I, for detailed MS analyses see supplemental data). 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 II). 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 II; Fig 3B). 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 II). 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. 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 III 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)). 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 spectr a 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 °C) when analyzing the exosome from the second purification (20 v.s. 25 °C). The mass spectrum of the exosome from the second purification revealed three charge state distributions centered around 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 I, 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*.

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 solfataricus* 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.
ACKNOWLEDGEMENTS

This work was supported by the Netherlands Organization for Scientific Research (NWO) by a VENI grant (700.54.402) to R.H.H.vdH. We thank Cees Versluis for excellent mass spectrometry assistance and the Netherlands Proteomics Centre for financial support.
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., Maimneh, 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., Mangan, 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., Gläser, 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' exonuclease. *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. Pappin, D. J., Hojrup, P., and Bleasby, A. J. (1993) Rapid identification of proteins by peptide-mass fingerprinting. *Curr. Biol.* 3, 327-332
39. 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
40. 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
41. 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.
42. 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
43. 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
FIGURE LEGENDS

Fig. 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.

Fig. 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, respectively. 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.

Fig. 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).
**Fig. 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 III for a summary.

**Fig. 5.** Schematic representation of measured exosome (sub)-complexes. The numbers 1 to 5 refer to the (sub)-complexes measured by macromolecular MS. The green, purple, blue and orange 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.

**Fig. 6.** Primary sequence alignment and phosphorylation of yeast and human Rrp4. Conserved residues are shown in blue boxes and identical residues are shown with a red background. The predicted S1 RNA-binding domain is indicated with a yellow background and the phosphorylated Serine is indicated with a turquoise background. The sequence alignment was performed by using the program Clustal W (40).
### Table I
Overview of identified phosphorylation sites in yeast exosome proteins.

| Band in gel | Protein name | Peptide sequences | Amino acid sequences | Phosphorylated residue |
|-------------|--------------|-------------------|----------------------|------------------------|
| 1           | Ski7         | KSSNLDPSSSNSFK    | 305-318              | S307                   |
| 1           | Ski7         | LSALKKSNSDLEK     | 82-94                | Ser88, Ser90           |
| 1           | Rrp6         | NTNEEATPIPSSETK   | 514-528              | Thr520                 |
| 2           | Rrp43        | VGSACTDEEMTISQK   | 141-154              | Thr146                 |
| 2           | Rrp43        | ISPESLQQR         | 25-33                | Ser26                  |
| 3           | Rrp4         | RKSESDELQMR       | 150-160              | Ser152                 |
| 4           | Csl4         | SVDASPNDVTR       | 90-100               | Ser94                  |
| 5           | Mtr3         | YRDLMISCLMNQET    | 237-250              | Ser243                 |
| 5           | Mtr3         | STDLTPKGNESEQE    | 32-45                | Thr36                  |
| 5           | Rrp46        | IIQDNISPR         | 245-253              | Ser251                 |
Table II  
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 | Predicted mass<sup>b,c</sup> (Da) | Determined mass<sup>d</sup> (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                         | -                                |

<sup>a</sup>  - 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

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

<sup>c</sup>  All phosphorylation sites were present in sub-stoichiometric amounts and therefore they were not taken into account to predict masses.

<sup>d</sup>  The determined mass was the protein mass as measured by protein LC MS.
Table III  Measured masses of exosome (sub)-complexes by macromolecular MS. Csl4 and Rrp42 were used as the C-terminal tagged entry points.

| Entry point | Exosome<sup>a</sup> | 1            | 2<sup>b</sup> | 3            | 4            | 5            |
|-------------|---------------------|--------------|--------------|--------------|--------------|--------------|
| 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|             |              |              |

<sup>a</sup> Exosome (sub)-complexes 1-5 are schematically presented in Fig. 5

<sup>b</sup> Mass difference of 5,077 Da between the exosome 2 species are due to the use of the two different TAP-tagged proteins (Csl4 and Rrp42).
Fig. 3

A

B

113,621.0 ± 3.5 Da
Dis3

24,319.2 ± 0.4 Da
Rrp46
