Merging Molecular Electron Microscopy and Mass Spectrometry by Carbon Film-assisted Endoproteinase Digestion* S

Florian M. Richter, Bjoern Sander, Monika M. Golas, Holger Stark, and Henning Urlaub

Many fundamental processes in the cell are performed by complex macromolecular assemblies that comprise a large number of proteins. Numerous macromolecular assemblies are structurally rather fragile and may suffer during purification, resulting in the partial dissociation of the complexes. These limitations can be overcome by chemical fixation of the assemblies, and recently introduced protocols such as gradient fixation during ultracentrifugation (GraFix) offer advantages for the analysis of fragile macromolecular assemblies. The irreversible fixation, however, is thought to render macromolecular samples useless for studying their protein composition. We therefore developed a novel approach that possesses the advantages of fixation for structure determination by single particle electron microscopy while still allowing a correlative compositional analysis by mass spectrometry. In this method, we call “electron microscopy carbon film-assisted digestion”, macromolecular assemblies are chemically fixed and then adsorbed onto electron microscopical carbon films. Parallel, identically prepared specimens are then subjected to structural investigation by electron microscopy and proteomics analysis by mass spectrometry of the digested sample. As identical sample preparation protocols are used for electron microscopy and mass spectrometry, the results of both methods can directly be correlated. In addition, we demonstrate improved sensitivity and reproducibility of electron microscopy carbon film-assisted digestion as compared with standard protocols. We show that sample amounts of as low as 50 fmol are sufficient to obtain a comprehensive protein composition of two model complexes. We suggest our approach to be an optimization technique for the compositional analysis of macromolecules by mass spectrometry in general.

An increasing number of proteins are nowadays recognized to fulfill their cellular tasks as part of macromolecular machines (1). Recent advances in the isolation procedures allow purification of many of these assemblies for further biochemical and structural analysis (2–4). Thereby, MS of peptides derived by digestion of the proteins of the assemblies (5, 6) or MS of intact assemblies (7, 8) is used to determine the protein composition, whereas electron microscopy (EM) can be used to provide information about the structure of the assembly using computational averaging and reconstruction techniques (9). Available protocols to combine information derived by MS and EM therefore either analyze the endoproteolytic digestion products (10–13) or the intact assembly (14). However, the sample preparation protocols typically differ between EM and MS. Although for MS the sample is either directly digested in solution (6), subsequent to gel electrophoretic separation of the sample (15–17) or used in an intact form (7, 8), EM typically requires the adsorption of the sample to carbon film (13, 18–20). In the case of sample heterogeneity, the utilization of these different sources of material (i.e. in-solution versus carbon-adsorbed particles) may be associated with the analysis of different particle populations in MS and EM due to for example the individual kinetics of adsorption, thus making a direct correlation of the EM and MS data difficult.

To combine the information obtained by both methods, we thus set out to correlate MS and EM. To do this would require identical material and conditions for the EM and MS analyses to ensure that identical particle populations are measured by both techniques. To meet this need, we developed an approach for digestion of the assemblies directly on EM carbon film followed by an analysis of the generated peptides by MS. Our method differs from established protocols for sample preparation in that identical conditions and input material are used for the analysis of the composition by MS and of the structure by EM.

The abbreviations used are: EM, electron microscopy; APS, average peptide score; ECAD, electron microscopy carbon film-assisted digestion; snRNP, small nuclear ribonucleoprotein; PEEK, polyetheretherketone.
The key to our approach is that disintegration of fragile assemblies can often be reduced by stabilization of the interacting components to reduce sample heterogeneity caused by the sample preparation to a minimum. Chemical fixation with e.g. formaldehyde or glutaraldehyde is known to cross-link fragile and transient assemblies (21) and thus greatly improves the homogeneity of carbon-adsorbed macromolecular samples in EM (3). Optimized sample integrity and a minimum of background (the latter including contaminants and degraded or aggregated components) can be attained by the use of suitable fixation protocols (3). Mild fixation, for example, is achieved by exposure to low concentrations of glutaraldehyde (≤0.1%), which cross-links the accessible primary amines of lysines and protein N termini (21). The improved sample quality seen in the EM led us to hypothesize that the mild fixation used might leave enough protein regions unmodified for them to be cleaved into peptides for sequencing by MS and identification by standard database searches. In that way, the advantage of fixation (improved structural integrity) would be retained, whereas the potential disadvantage of too strong fixation (excessive structural cohesion) would be avoided.

Here we describe this new technique, which we call “electron microscopy carbon film-assisted digestion” (ECAD). It involves mild fixation of the macromolecular assemblies to obtain improved structural integrity of fragile macromolecular assemblies and to allow EM- and MS-based characterization of identical, parallel samples. This adds to the repertoire of approaches for interactome analysis of macromolecular complexes and is especially useful for composition studies of fragile assemblies. ECAD allows a correlation of composition and structure of a macromolecular assembly as discrepancies that may be imposed by differences in the preparation of the respective samples for EM and MS are avoided. We further suggest that the improved integrity of a macromolecule upon ECAD of chemically stabilized macromolecular assemblies can also be an advantage for MS analysis in general independent of the EM analysis.

**EXPERIMENTAL PROCEDURES**

**Supplies**

All chemicals and reagents including the lyophilized mixture of GroEL-GroES and Glu-fibrinogen were purchased from Sigma-Aldrich unless otherwise specified. A 25% EM grade glutaraldehyde solution was obtained from Electron Microscopy Sciences (Hatfield, PA) and was used no more than 1 month after initial puncture of the seal to ensure sufficient activity of the cross-linker. Uranyl formate PA) and was used no more than 1 month after initial puncture of the solution was obtained from Electron Microscopy Sciences (Hatfield, PA) unless otherwise specified. A 25% EM grade glutaraldehyde solution was obtained from Electron Microscopy Sciences (Hatfield, PA).

**Biochemical Methods**

**GroEL-GroES Complexes—**A lyophilized mixture of GroEL-GroES was dissolved in buffer G (50 mM HEPES, pH 8.0, 50 mM KCl, 50 mM MgCl₂) supplemented with 2.5 mM ATP to a concentration of 1 mg/ml. GroEL-GroES (125 μg) was loaded onto a 10–30% glycerol gradient (also containing a gradient from top to bottom of 2.5–0 mM ATP) in buffer G, and ultracentrifugation was performed in a Sorvall TH-660 rotor (Thermo Scientific, Waltham, MA) over 17 h at 30,000 rpm and 4 °C. 25 S U4/U6, U5 Tri-snRNP Particles—Native human 25 S tri-snRNPs were isolated from HeLa nuclear extract by H-20 immuno-affinity chromatography and subsequent glycerol gradient centrifugation as described previously (22). Ultracentrifugation was performed in a Sorvall TH-660 rotor over 17 h at 30,000 rpm and 4 °C. The protein and RNA composition of fractions containing U snRNPs were determined by electrophoresis on denaturing poly-acrylamide gels with SDS (for proteins) or urea (for RNA). Proteins were stained with Coomassie Brilliant Blue, and RNA was stained with silver (23, 24).

**Fractionation and Cross-linking with Glutaraldehyde—**All particles were prepared by using ultracentrifugation in a glycerol gradient as the final purification step before sample preparation for EM and MS. The glycerol gradients were fractionated from the bottom by using a peristaltic pump connected to a fraction collector (3). A 25% glutaraldehyde solution was added to the peak fractions to a final concentration of 0.075% (v/v), and the peak fractions were incubated for 18 h at 4 °C before EM and MS analysis.

**Electron Microscopy Methods**

**Preparation of Carbon Films for EM and ECAD—**Thin (~7–10 nm) carbon films were made in a carbon evaporator (E12E, Boc Edwards, Kirchheim, Germany) by evaporation of spectral carbon (Ringsdorf Werke GmbH, Bonn, Germany) onto freshly cleaved mica (Plano GmbH, Wetzlar, Germany) using a standard indirect coating setup. Before adsorption of particles, the coated mica was freshly cleaved into 2 × 4-mm pieces (except for the analysis of 3.5 fmol of GroEL-GroES where a single 1 × 2-mm piece was used).

**EM Sample Preparation by Negative Staining—**EM grids for negative staining were prepared using the sandwich method (19) except for the freezing step for cryoimaging. The protein concentrations of the samples used for adsorption to the carbon films were in the range suitable for electron microscopy, i.e. 10–200 μg/ml. During preparation, the sample was kept at 4 °C by ice-cooling the sample preparation block. Briefly, the carbon film was floated on the surface of 25 μl of particle solution, allowing adsorption of particles to the carbon film, over a period of 1 min to several hours. Adsorption times were adjusted to yield a total adsorbed particle amount of 50 fmol on a total of up to 10 carbon films as judged by counting the particles on the carbon film in the electron microscope. For negative staining, the carbon film was incubated with an aqueous ~2% uranyl formate solution for 2 min and subsequently attached to a 400 mesh EM copper grid on which a carbon film containing holes of ~1-μm diameter had already been mounted. A second carbon film incubated with uranyl formate solution was then used to form a sandwich, enclosing the specimen in a layer of staining solution. The grids were stored under dry conditions at room temperature until image acquisition.

**Unstained Cryo-EM Preparation by Vitrification—**For cryo-EM in a stain-free buffer, the buffer of the glycerol gradient fractions was exchanged to a glycerol-free buffer using a spin column (Zeba, Pierce). During preparation, the sample was kept at 4 °C. Particles were adsorbed on a piece of carbon film for a defined period of time (1 min to several hours depending on the particle concentration). Subsequently, the carbon film was attached to an EM copper grid covered by a carbon film containing holes. The EM grid was carefully blotted and plunge-frozen in liquid ethane (25). Specimens were stored in liquid nitrogen until imaging in the EM.
Electron Microscopy—The specimens were imaged at room temperature (negative staining) or at cryogenic conditions (unstained cryo-EM) under low dose conditions on a 200-kV transmission electron microscope equipped with a field emission gun (CM200 FEG, Philips/FEI, Eindhoven, The Netherlands) operated at 160 kV. Images were taken on a 4096 × 4096 charge-coupled device camera (TemCam-F415, TVIPS, Gauting, Germany) at a magnification of 122,000× and 2× binning of the charge-coupled device pixels using an electron dose of −20 e−/Å² (26).

Mass Spectrometry Methods

In-solution Digestion of Purified Particles—The protein concentration in the glycerol gradient fractions containing GroEL-GroES or tri-snRNP was determined using Bradford assays to c ≈ 0.19 mg/ml and c = 0.2 mg/ml, respectively. The samples were diluted to the following absolute sample amount prior to in-solution digestion: for GroEL-GroES, the sample amounts were 3.5 fmol (3.1 ng), 7 fmol (6.1 ng), and 50 fmol (43.7 ng) (each in a 30-μl volume); for tri-snRNP, we used sample amounts of 10, 20, 50 (85 ng), 100, 150, 250, and 500 fmol (each in a 30-μl volume). Dilution was performed by adding to each sample 7.2 mg of solid urea to obtain a final concentration of 4 M in a 30-μl volume and by adjusting the final volume with 50 mM Tris-HCl, pH 7.5. In the case of glutaraldehyde treatment (see below), samples were incubated with glycine (0.1 M final concentration) for 30 min at room temperature prior to the addition of urea to block residual activity of the cross-linker unless otherwise specified. Samples were incubated for 5 min at room temperature by vigorous shaking. The samples were subsequently diluted with 50 mM Tris-HCl, pH 7.5 to give a final urea concentration of 1 M. The volume of the denatured samples subjected to digestion was thus 120 μl. For tri-snRNP, U4/U6 and U5 small nuclear RNAs were digested for 2 h at 52 °C by adding 0.1 μg of RNase A and 0.16 μg of RNase T1. Samples were chilled briefly on ice, and proteins were digested with 1 μg of trypsin at 37 °C overnight. Digestion was stopped by adding TFA to a final concentration of 0.1% (v/v). Samples were stored at −20 °C until MS analysis. In-solution digests were performed in five technical replicates with the exception of 10 and 20 fmol of tri-snRNPs that were only analyzed twice.

Digestion of Macromolecular Samples Attached to Carbon Film (ECAD)—Fixed macromolecular samples (see above) were adsorbed to the carbon film exactly as performed for the EM analysis, i.e. by putting the mica with the carbon film on top of the particle solution so that the carbon film floats on the solution. After adsorption, the piece of mica is lifted using tweezers and extensively blotted with filter paper (Whatman) without destroying the carbon film. Optionally, the carbon film with absorbed particles can be washed with a washing buffer (e.g., the glycerol-free and glutaraldehyde-free sample buffer). Subsequently, the carbon film is transferred to a reaction tube (Eppendorf) containing 60 μl of a buffer comprising 4 M urea, 50 mM Tris-HCl, pH 7.5, 0.1 M glycine to neutralize the chemical activity of glutaraldehyde. Of note, experiments in which the mica was removed indicated that the MS results are similar irrespective of the presence or absence of the mica. Subsequent to 30-min incubation at room temperature for inactivation of the residual glutaraldehyde activity, samples were sonicated for 15 min at 4 °C to disrupt the carbon film, and the sample volume was adjusted to a concentration of 1 M urea by dilution with 50 mM Tris-HCl, pH 7.5. The volume of the denatured samples subjected to digestion was thus 240 μl. Of note, digestion was performed in the presence of disrupted (i.e. sonicated) carbon film. Hydrolysates of RNA and protein were performed as described above. The samples were centrifuged for 5 min at 13,000 rpm, and the supernatant was collected. Peptides were extracted from the carbon film pellet by adding 120 μl of 60% (v/v) acetonitrile in water containing 0.2% (v/v) formic acid and shaking in a thermomixer for 15 min at 37 °C. The supernatant was dried in a SpeedVac and pooled with the supernatant containing 1 M urea (see above). TFA was added to the pooled supernatants to give a final TFA concentration of 0.1% (v/v). Samples were stored at −20 °C until the MS measurement. Unfixed samples subjected to ECAD were treated likewise except that no glutaraldehyde and no glycine were added to the sample. Different sample amounts for ECAD (e.g., 3.5, 7, and 50 fmol of GroEL-GroES complex and 10–50 fmol of tri-snRNP) were prepared by using a smaller piece of carbon (down to 1 × 2 mm) or, more usually, a greater number of carbon films; in our experience, up to 7–10 carbon films can be processed in a single reaction tube containing 60 μl of the 4 M urea solution. All ECAD experiments using GroEL-GroES and tri-snRNP (50 fmol) were performed five times; the other samples were analyzed only once.

Off-line Nano-LC Analysis—Samples were injected into a nano-LC system (Dual Gradient, Dionex, Idstein, Germany) in six (in-solution) or 12 (ECAD) loading cycles using a 20-μl injection volume for each cycle. The system was equipped with precolumns working in back-flush mode (25 × 0.15 mm packed in house with C18, 5 μm, 300 Å; number 218TPS215, Vydac, Hesperia, CA). Each sample loading cycle was performed for 10 min with a flow rate of 5 μl/min in loading solvent A (3.5% (v/v) ACN containing 0.1% (v/v) TFA in water). After loading of the sample, the precolumn was flushed with loading solvent for 15 min at 5 μl/min between successive loading cycles. Peptides were eluted from the precolumn in back-flush mode and separated on an analytical column packed in house (200 × 0.075 mm, C18, 5 μm, 300 Å; number 218TPS215, Vydac) by a standard gradient from 10% solvent B to 60% solvent B over 3 h (solvent A, 0.1% (v/v)
software v3.5.3 (release date, February 2007). Collision energy in open or the Peak to Mascot tool of the 4000 Series Explorer allowed the 15 highest intensity peptides of every spot to be selected for each precursor; dynamic stop criteria depending on the spectral quality were used. Job-wide interpretation of the MS data was performed on a MALDI-TOF/TOF 4800 analyzer (Applied Biosystems/Sciex MDS, Foster City, CA) equipped with a neodymium-doped yttrium aluminum garnet (Nd:YAG) laser (355-nm wavelength and 200-Hz repetition rate). The eluted peptides were mixed in a T-piece dead volume, Upchurch Scientific Inc., Oak Harbor, WA) with 10 mg/ml α-cyano-4-hydroxycinnamic acid matrix (Sigma-Aldrich) containing 10 fmol/μl Glu-fibrinogen as internal standard in 70% (v/v) ACN, 0.1% (v/v) TFA delivered at a flow rate of 0.9 μl/min. Every 15 s the eluate mixed with matrix was spotted onto a stainless steel MALDI target (Opti-TOFTM LC/MALDI insert, Applied Biosystems) by a Probot Spotter (Dionex). Per gradient, ~600 peptide-containing fractions were collected.

Column Packing—Nanoanalytical columns were packed as follows. A 2-mm frit at one end of a fused silica capillary (360-μm outer diameter, 75-μm inner diameter; Polymicro Technologies, Phoenix, AZ) was generated by polymerizing Kvasitl™ (PQ Europe, Amersfoort, The Netherlands). Column material was packed into the capillaries at 50 bars by using a high pressure chamber (Brechbühl) working with helium. The analytical column was connected to the system by using a PEEK polymer tubing and stainless steel nuts and ferrules (PEEK, gray, ½ × 0.615; 400-μm inner diameter; Upchurch Scientific Inc.). Precolumns manufactured in house were packed with fused silica capillaries (375-μm outer diameter, 150-μm inner diameter, 500–750 mm in length; Polymicro Technologies). Precolumns were cut into lengths of 25 mm, and both ends were sealed with fittings (Inline MicroFilters, Upchurch Scientific Inc.) by using MicroTight Sleeves (0.0155 × 0.0255; Upchurch Scientific Inc.) and Inline MicroFilters (PEEK, 0.5 μm; Upchurch Scientific Inc.).

MALDI-MS and MSMS Analysis—MS analysis was performed on a MALDI-TOF/TOF 4800 analyzer (Applied Biosystems/Sciex MDS, Foster City, CA) equipped with a neodymium-doped yttrium aluminum garnet laser (Nd:YAG) and an MaldiTOF/TOF instrument with 200-Hz repetition rate. For MS spectra in positive ion mode, a total of 800 shots were generated, and for MSMS a maximum of 2000 shots were accumulated for each precursor; dynamic stop criteria depending on the spectral quality were used. Job-wide interpretation of the MS data allowed the 15 highest intensity peptides of every spot to be sequenced in MSMS mode. Peak lists were created using “TS2Mascot open” or the “Peak to Mascot” tool of the 4000 Series Explorer software v3.5.3 (release date, February 2007). Collision energy in MSMS mode was set to 1 × 10⁻²⁰ torr with the potential difference between the source II accelerator and the collision cell set at 1 kV.

Protein Identification—Proteins were identified by searching fragment spectra against the databases RefSeq (taxonomy, human; release date, February 29, 2008) with 128,611 and 194,779 sequences, respectively, used for the actual database search or NCBInr (release dates, August 28, 2006 and October 8, 2007) with 70,679 sequences used for the actual database search. Modifications; no fixed modification; MS mass tolerance set to 100 ppm; and average peptide score shown in Fig. 9, A and B. Of note, this particular score for a single “red bold” peptide was chosen as cutoff because our experience with the 4800 MALDI-TOF/TOF instrument revealed that this particular score is still valid for protein identification under the condition that the corresponding product ion spectra (MSMS) are manually validated. Keratins were removed from the protein identification list. Ambiguities caused by the redundancy in the databases as well as by protein isoforms or protein families with shared sequences were corrected on the basis of visual inspection of the Mascot database result. Thereby, the protein entry with the higher Mascot score was kept in the list.

Statistics

Statistical Analysis—p values were computed by paired, two-sided Wilcoxon tests as implemented in the software package R (28). Generation of Color-coded Three-dimensional Models—Peptide location maps in the protein three-dimensional model were generated by combining the data sets from the five independent MALDI-MS (MSMS) analyses. Peptides found in all five independent analyses are shown in red, peptides found four times in the five analyses are shown in orange, peptides found three times are shown in turquoise, peptides found twice are shown in dark blue, and peptides found once are shown in purple. Peptides with their corresponding color code are visualized in the three-dimensional structure of GroEL-GroES in complex with seven molecules of ADP (Protein Data Bank code 1sav (29)). For the color coding of the surface of the GroEL-GroES structure, we
used scripting and visualization tools provided by the PyMOL software (DeLano Scientific LLC, Palo Alto, CA).

**RESULTS**

**ECAD Approach**—Our approach was designed to determine the protein composition of macromolecular complexes using the same sample preparation protocol for both EM and MS analysis (the work flow is summarized in Fig. 1). In particular for EM of low concentrated samples, the assemblies are required to be adsorbed onto carbon film to accumulate sufficient amounts for EM analysis (3, 13). As different assemblies (e.g. broken assemblies, aggregates, or contaminating complexes) may adsorb with individual kinetics, the particle population seen in the electron microscope and in MS can differ. The approach presented herein eliminates such possible differences because the same sample handling is used for EM and MS analyses. To test our approach, we used well characterized macromolecular assemblies with a level of fixation similar to that obtained by using a previously established protocol (3) so that separation of individual proteins by SDS-PAGE was prevented (Fig. 2).

In our procedure (Fig. 1), we adsorbed the particles onto carbon films of 7–10-nm thickness produced by evaporating carbon onto a freshly cleaved piece of mica. To do this, the films were floated on the particle solution for a defined period (1 min to several hours depending on sample concentration). One of the films was mounted on a standard EM grid and subjected to EM analysis. By single particle EM using negatively stained or vitrified (cryo-EM) specimens (19, 25), the particles on the complete carbon film were counted by extrapolation from several representative areas of the specimen (see Fig. 3A for a negatively stained specimen and Fig. 3B for a native cryospecimen in vitrified buffer). Single particle image processing was performed to determine the structure of the particle (Fig. 3, C and D). The other carbon films were then transferred to a reaction tube containing 60 μM urea and an excess of glycine to block residual activity of glutaraldehyde. After denaturation, digestion of the samples with endoproteinases and ribonucleases was performed directly on the EM carbon film (see Fig. 1). The other carbon films were then transferred to a reaction tube containing 60 μM urea and an excess of glycine to block residual activity of glutaraldehyde. After denaturation, digestion of the samples with endoproteinases and ribonucleases was performed directly on the EM carbon film (see Fig. 1). Of note, an excess of enzymes was used (that was above the recommended enzyme-to-substrate ratio) to ensure that putative adsorption of the enzymes onto the carbon film did not result in too low amounts of enzymes for the digestion of the sample. Digestion products were separated and analyzed by nano-LC off-line MALDI-MSMS or nano-LC ESI-MSMS, and proteins were then identified by searching fragment spectra of unfixed, i.e. non-glutaraldehyde-modified, peptides against a database.
**Required Amount for ECAD**—In general, we noted that an amount of ~2–25 fmol of various macromolecular assemblies on a piece of EM carbon film (2 × 4 mm) showed a particle density suitable for structure determination by single particle EM analysis (Figs. 3 and 4A; Fig. 4B shows an overloaded carbon film that is not suitable for image processing). A correlation of structure and composition using identically prepared samples for both techniques therefore requires that the MS approach also works with low sample quantities, and we show here that our method is indeed applicable for quantities as low as 50 fmol. To establish and validate our approach, we first used a symmetric complex of well known composition, the bacterial GroEL–GroES (containing 14 copies of GroEL and seven copies of GroES (29)). We then examined a more complex assembly, the human U4/U6.U5 tri-snRNP (30) (comprising 29 core proteins of which most occur singly and a few occur as double copies).

**Bacterial GroEL–GroES as Simple Model System**—GroEL–GroES contains 14 copies of GroEL and seven copies of GroES (29) and forms two stacked homoheptameric GroEL subcomplexes that are covered by a homoheptameric GroES lid (Fig. 3). We compared the ECAD performance of fixed material with standard in-solution digest of unfixed material. Examples of MSMS spectra are given in Fig. 5. For the GroEL-GroES assembly, we used 50 fmol of glutaraldehyde-treated sample for ECAD (left) and 50 fmol of non-fixed material for in-solution digestion (right). Both peptides displayed here were identified five times in five technical replicates. A, the MSMS-generated sequence information of the GroEL peptide AAVEEGVAGGGVALIR (residues 405–421) is compared. The absolute intensity (and the signal-to-noise ratio) is higher by a factor of ~6.5. Sequencing of the ECAD-generated peptide derives an almost complete coverage of the y-type ion series. B, similar results were observed for the GroES-specific peptide VGDIVFNDGYGVK (residues 61–74); the absolute ion intensity was higher by a factor of 1.8.
peptide AAVEEGVVAGGGVALIR (residues 405–421), the absolute intensity and thus signal-to-noise ratio was comparable or even higher using ECAD as compared with the in-solution digest (Fig. 5A). ECAD showed an almost complete coverage of the y-type ion series in MSMS. Likewise, the GroES peptide VGDIVIFNDGYGVK (residues 61–74) showed a comparable or even higher absolute ion intensity (Fig. 5B). Together these results indeed suggest an improved quality of the spectra using the ECAD approach.

For GroEL-GroES, we analyzed 3.5–7 fmol of particles to account for the multiple occurrence of the two constituent proteins (see also supplemental Table S1 for a comprehensive list of identified peptides in the respective analyses and supplemental Table S2 for a summary of the data). With such low amounts, the sensitivity and reproducibility of MS might become an issue. Therefore, we performed five independent ECAD experiments using adsorbed, glutaraldehyde-fixed GroEL-GroES, taking the frequency of identification of a given non-glutaraldehyde-modified peptide in these experiments as a measure of reproducibility (see also supplemental Table S1), whereas in a parallel set of experiments, we digested unfixed samples in solution following a standard protocol. In a range of 3.5–50 fmol of adsorbed, fixed GroEL-GroES complexes, the detected peptides still covered a large portion of the protein sequence (Fig. 6; see also supplemental Table S1), and the reproducibility (five times in five replicates indicated by red colored peptides in Fig. 6; see also supplemental Table S1) was surprisingly good even at 3.5 fmol. Overall, the sequence coverage of ECAD using 3.5 fmol of particles was comparable with that obtained from 50 fmol in in-solution digests (Fig. 6).

Sensitivity and Reproducibility Using Complex Macromolecular Assembly—We next examined a more complex macromolecular assembly, the human U4/U6.U5 tri-snRNP that comprises three small RNAs and 29 core proteins ranging in size from 8.5 to 274 kDa (30). Most of its proteins occur singly, and a few (the Sm proteins) occur as double copies (see also Fig. 2B). We therefore classified the proteins into a group of particle-specific proteins larger than 40 kDa, a group of particle-specific proteins smaller than 40 kDa, and into two groups of the single copy LSm and the double copy Sm proteins that both form a rigid seven-membered ring (31, 32). In the EM, human tri-snRNP adopted a typical triangular shape (Fig. 4) like that observed in our previous study (33).

We first performed MS analyses of 10, 20, and 50 fmol of peptides derived from an in-solution digest of tri-snRNPs to confirm that instrument sensitivity of the nano-LC off-line MALDI-TOF/TOF system used is not a limiting factor for sequencing of 20 fmol of peptides (data not shown). Next, we compared the UV chromatograms and MS and MSMS spectra of peptides derived from the ECAD procedure with those derived from in-solution digested samples (50 fmol of tri-snRNPs each). We observed a similar and surprisingly strong UV absorbance for both ECAD and in-solution digestion procedures (Fig. 7, A and B). The UV chromatograms indicate that equal amounts of sample were separated by LC before spotting with MALDI matrix (Fig. 7C). The visible peaks originated from the sample itself as well as from partial digestion of the added RNases (during in-solution digestion) or complete digestion of these RNases (during ECAD), contaminating keratins (data not shown), and unspecific background contamination as revealed by “mock” experiments without protein sample (i.e. tri-snRNP). The latter clearly showed that both procedures gave rise to species that showed an absorption of UV light of up to 20 milliabsorbance units (data not shown).
shown). These experiments revealed that the added RNases were only (if at all) partially digested during in-solution digests and eluted from the LC in three strong peaks (at ~120, ~138, and ~160 min; see also Fig. 7B). Of note, RNases are surprisingly stable against endoproteolytic digestion (34, 35). However, the RNases were not present in the elution profile of the ECAD samples (Fig. 7A and mock experiments (data not shown)), suggesting that the RNases were hydrolyzed by trypsin or were adsorbed by the carbon films. Of note, we always used a large excess of enzyme (higher than the usual enzyme-to-substrate ratio; see “Experimental Procedures”) to ensure that a sufficient amount of enzymes is present in the sample and not completely adsorbed by the carbon film.

We then evaluated the performance of ECAD in comparison with in-solution digests with respect to the minimum sample amount for the MS analysis required to obtain a comprehensive proteome (Fig. 8, A and B; compare also supplemental Table S3). Based on these results, we conclude that a sample amount of 50 fmol is sufficient for MS analysis. This is shown by accidentally applying incomparable amounts of the sample and not completely adsorbed by the carbon films. As this advantage of ECAD was apparently not effected by the enzyme-to-substrate ratio; see “Experimental Procedures”) to ensure that a sufficient amount of enzymes is present in the sample and not completely adsorbed by the carbon film.

**Effect of Different Cross-linking Levels on MS Analysis**—
Subsequently, we were interested in the effect of different cross-linking levels on the performance of the approach and tested this by selecting different glutaraldehyde concentrations (Fig. 8C). Glutaraldehyde concentrations in the range of about 0.075–0.15% are typically used in the molecular EM field. At these concentrations, a separation of proteins by SDS-PAGE was not possible (compare Fig. 2A). A concentration of 0.075% (v/v) of fixative as, for example, used in a typical GraFix sample was compatible with the ECAD digest, whereas higher concentrations resulted in a decrease of the detection of peptides (compare +Gly conditions in Fig. 8C).

**Effect of Glycine**—The activity of glutaraldehyde can efficiently be inactivated by incubation with an excess of glycine. We thus analyzed the necessity to block the residual activity of glutaraldehyde by adding or omitting glycine (0.1 m glycine, pH 7.9) to the ECAD and in-solution samples. Accordingly, prior inactivation of the fixative by addition of glycine markedly improved the detection and sequencing of peptides in ECAD (compare +Gly and −Gly conditions in Fig. 8C) and in-solution digest (supplemental Fig. S1). However, the presence of glycine in the fixed in-solution digested sample was not sufficient to show an effect comparable with ECAD (supplemental Fig. S1).

**Peptide Scores**—The improved performance of ECAD is finally supported by the peptide scores of in-solution digests of unfixed specimens and ECAD digests of fixed material. For the U4/U6.U5 tri-snRNP, we observed that higher Mascot protein scores were generated from non-redundant peptides by ECAD as compared with the in-solution digests (Fig. 9A). For example, a Mascot protein score of $15 \times 10^3$ was achieved from ~280 non-redundant peptides in ECAD, whereas ~320 non-redundant peptides derived from in-solution digest would be required to obtain this particular Mascot protein score.

We therefore selected the average peptide score (36) (APS; defined as the ratio of the protein score to the number of non-redundant peptides) using Mascot as a measure for the confidence of identification as the peptide score indicates the spectrum quality. In the range of 10–50 fmol, the APS for the tri-snRNP proteins varied only marginally between 45.3 and 58.9 for ECAD, whereas it was substantially lower for the in-solution digest (Fig. 9B; see also supplemental Table S3). By interpolation of the values, it can be estimated that ~400 fmol of sample would be required for the in-solution digest to achieve an average APS comparable with 50 fmol using ECAD. Additionally, the scattering of the average APS values was substantially higher for in-solution digests compared with ECAD, indicating a better reproducibility of ECAD. Moreover,

**Fig. 7. Assessment of digested sample amounts loaded onto LC system for ECAD and in-solution digests.** A and B, UV chromatograms of the off-line LC analysis of 50 fmol of tri-snRNP particles treated by ECAD (A) and by in-solution digestion (B). To analyze the area below the curve at 220 nm, a peptide region (indicated as “peptide area”) is defined where mainly small digestion products suitable for MS analysis are expected. This transition occurs after an elution time of ~160–180 min as judged visually and by inspection of the corresponding MS spectra. Furthermore, the total integral (indicated as “total area”) from 0 to 226 min is calculated for the five independent runs. For ECAD (A), a “hill” formed by many individual peaks can be seen in the peptide area, which indicates a broad separation of species and is usually favorable for the MS analysis; few undigested products appear at the washout of the LC run. For in-solution digests (B), however, more non-digested material is observed at the washout of the run. C, box-and-whisker plots of the areas under the five 220 nm curves as described above. There is no significant difference in the distribution of the total areas as judged by unpaired, two-sided Wilcoxon tests (left panel; $p = 0.42$). Also, there is no significant difference in the distribution of the peptide areas (right panel; $p = 0.22$), indicating that similar amounts of material were loaded for ECAD and for in-solution digests, supporting the conclusion of good comparability between these two experimental series. D and E, technical comparison of the ECAD and in-solution MS analysis based on representative MS spectra (for the corresponding peak in the LC analysis indicated by the letters “d” and “e,” compare A and B). IS denotes the Glu-fibrinogen peptide contained within the MALDI matrix at a concentration of 10 fmol/µl for internal standardization to ensure constant quality of the MS spectra. Furthermore, a representative ribonuclease T1 peak shows that the proteolytic digestion conditions and the detection sensitivity of equal amounts of enzyme added in the later step of the digestion protocol were similar during the ECAD and in-solution measurements. Thus, the UV chromatograms and MS and MSMS spectra do not argue for any significant difference between ECAD-treated and in-solution digested samples in terms of the input amount, so that the amounts that were initially used for both methods of hydrolysis are well comparable. mAU, milliabsorbance units.
when a list of proteins was generated by a Mascot search for proteins with an APS of ≥20, more proteins showed a high APS of 40–70 when ECAD was applied; in contrast to an exponential decrease in the APS histogram in the case of the in-solution digest, a distinct plateau ranging from an APS of 30 to 60 was seen for ECAD (Fig. 9C).

**DISCUSSION**

The objective of the work presented herein was to develop an approach that enables us to determine the protein composition of exactly the particle population imaged in the electron microscope and thus used for further structure determination. Such an approach allows correlation of both the results of MS and EM and thus overcomes uncertainties that arise from different specimens being analyzed by MS and EM. Our ECAD approach addresses this issue by using the same sample preparation protocol for both MS and EM and by digesting the proteins directly from the EM carbon film. We demonstrated that peptides can be generated by tryptic digestion of proteins in a quality and quantity sufficient to identify the protein components of macromolecular complexes. Furthermore, our approach also showed an improved sensitivity and reproducibility of detection of peptides as compared with standard in-solution digestion.

We developed ECAD for the compositional analysis of mildly chemically fixed samples as they are obtained e.g. by using the recently introduced GraFix protocol (3). GraFix combines mild chemical fixation of macromolecular assemblies with high pressure conditions that occur during gradient ultracentrifugation and results in improvements of the structural integrity among other favorable effects. GraFix has been used to improve the sample quality for single particle EM of a
A growing number of different assemblies including small nuclear ribonucleoproteins (18, 37) and spliceosomes (3, 38), a box C/D small ribonucleoprotein (39), RNA editing complexes (13), survival of motor neuron complexes (40), the pre-mRNA 3’ processing complex (41), the RNA-induced silencing complex (RISC)-loading complex (42), the RNA polymerase (10), the DNA gyrase (10), the DNA topoisomerase (10), a multifunctional replication protein (43), basal transcription factors (44), and p53 (45).

So far, an analysis of the composition of these samples could only be performed indirectly either on aliquots of the eluate prior to fixation (i.e. an aliquot of the eluate was used for a compositional analysis by MS, whereas another aliquot was loaded on the GraFix gradient for EM) or by running a parallel gradient without the addition of glutaraldehyde and subjecting these to MS analysis and/or SDS-PAGE/Western blotting. Such approaches, however, do not ensure that identical particle populations are analyzed in the GraFix-treated and the untreated sample. Moreover, some of these complexes such as the RNA editing machinery (13) isolated from trypanosomes suffer from ultracentrifugation in non-cross-linking gradients. Thus, insufficient amounts of particles are obtained for a compositional analysis of the unfixed material that excludes the application of standard protocols for analysis of their composition. Likewise, we showed here that a gel electrophoretic analysis of the fixed assemblies is not possible as it does not result in a separation of the cross-linked proteins. Finally, also the eluate can represent a mixture of different assembly stages, thus limiting the value of a compositional analysis of the eluate. Although the protein composition of intact macromolecular assemblies can also be determined (46–48), such an approach has not been applied to carbon-adsorbed macromolecular assemblies and might also impose significant challenges in the assignment of peaks to particle populations due to the complex nature of the chemical reaction profile of glutaraldehyde in aqueous solution (21).

In contrast, the protocol presented herein offers a unique possibility to determine the protein composition of fixed macromolecular assemblies. We observed an increase in sensitivity and reproducibility by using the ECAD approach as compared with standard processing of fixed (and also unfixed; see below) material. Thus, our approach overcompensates for a possible decrease in the sensitivity that might be caused by the cross-linking of lysine residues and the resulting inhibition of the enzymatic cleavage by the lysine- and

![Peptide scores.](image)

**Fig. 9.** Peptide scores. A, plot of the Mascot protein score versus the number of non-redundant peptides for ECAD (light grey) and the in-solution digest (dark grey). Overall, ECAD achieves a higher Mascot score based on fewer non-redundant peptides, i.e. ECAD detects selected peptides with a better score. A Mascot protein score of 15,000 is obtained from ∼280 non-redundant peptides in ECAD, whereas ∼320 non-redundant peptides would be required to achieve this particular Mascot protein score in the in-solution digest (dashed lines). B, semilogarithmic plot of the APS for tri-snRNP proteins as a function of sample amount (light grey, ECAD; dark grey, in-solution digest). C, histogram of the number of database protein hits from digests according to APS determined from five repeated experiments using 50 fmol of sample for the in-solution digest (dark grey) and ECAD (light grey).
arginine-specific endoproteinase trypsin. In this respect, it should be noted that we searched and detected non-glutaraldehyde-modified peptides, indicating that despite cross-linking a sufficient quality and quantities of non-cross-linked peptides are still generated by endoproteinase digestion.

Experiments using unfixed material for ECAD showed likewise a reproducible detection of peptides spread over the entire sequence including peptides that were not detected in the MS analysis of fixed sample, indicating that ECAD is also applicable for unfixed macromolecular samples. However, an EM analysis and correlation of MS and EM might be challenging in the case that the sample integrity suffers from the adsorption to carbon film without prior stabilization of the complexes by mild chemical cross-linking (3). In these cases, already the identification of intact particles versus disintegration products and thus counting of particles on the carbon film using EM images might become an issue.

Together, our results suggest that the MS analysis of macromolecular complexes benefits from the usage of carbon film. Possible explanations for the improvement of the MS analysis include an increased activity of the endoprotease in the presence of carbon as reported previously for other enzymes (49) and/or the prevention of aggregation of adsorbed, partly digested proteins in which the more hydrophobic protein core has been exposed upon endoproteolytic digestion. In a solution, such partly digested particles may be prone to form aggregates due to the (partial) uncovering of hydrophobic residues. This may in turn interfere with the further digestion of the sample. In contrast, the presence of glycerine in the fixed in-solution digested sample is not sufficient to show an effect comparable with ECAD.

In conclusion, hydrolysis of macromolecular complexes on EM carbon films and the subsequent analysis of the proteins by MS confer several advantages upon structure-related proteomics studies. The ECAD approach closes a gap between current state-of-the-art single particle EM and MS analysis. The MS analysis of macromolecular assemblies benefits from mild chemical fixation that is in turn required for high quality imaging of fragile assemblies by single particle EM. Fixation of assemblies with formaldehyde, which can be reversed by heating before MS, has previously been used to trap loosely aggregated (51). However, in these studies, the modification is reversed prior to proteomics analysis, and thus, none of these studies were able to address the question of how to process samples in a concentration and amount suitable for a correlative MS/EM approach. With the ECAD approach, we have a tool at hand to study the protein composition of fixed and also of unfixed macromolecular samples with improved sensitivity in general.

Acknowledgments—We thank M. Raabe and U. Plessmann for expert technical assistance in MS; T. Conrad, P. Kempkes, and H. Kohansal for expert technical help with cell cultures; and N. Fischer for help with mock experiments.

* This work was supported in part by a Young Investigator Program grant of the European Alternative Splicing Network of Excellence (EURASNET) of the 6th European Union framework (to H. U.) and the Bundesministerium für Bildung und Forschung and the 6th European Union framework program "3D Repertoire" (to H. S.).

This article contains supplemental Fig. S1 and Tables S1–S3.

These authors contributed equally to this work.

Present address: Max Planck Inst. of Immunobiology, Stübeweg 51, 79108 Freiburg, Germany.

Supported by grants from the Danish Center for Scientific Computing (DCSC); the Agnes and Poul Friis’ Foundation, Denmark; the Novo Nordisk Foundation; the A. P. Møller Foundation for the Advancement of Medical Sciences; and the Fru Astrid Thaysens Foundation.

Supported by grants from the DCSC; the Agnes and Poul Friis’ Foundation, Denmark; the Danish Council for Independent Research, Medical Sciences; and the Danish National Research Foundation.

To whom correspondence may be addressed. Tel.: 49-551-201-1305; Fax: 49-551-201-1197; E-mail: hstark1@gwdg.de.

To whom correspondence may be addressed. Tel.: 49-551-201-1060; Fax: 49-551-201-1197; E-mail: henning.urlaub@mpi-bpc.mpg.de.

REFERENCES

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

2. Forler, D., Kocher, T., Rode, M., Gentzel, M., Izaurralde, E., and Wilm, M. (2003) An efficient protein complex purification method for functional proteomics in higher eukaryotes. Nat. Biotechnol. 21, 89–92

3. Kastner, B., Fischer, N., Golas, M. M., Sander, B., Hartmut, K., Deckert, J., Fauer, H., Wolf, E., Uchtenhagen, H., Urlaub, H., Herzog, F., Peters, J. M., Poerschke, D., Lührmann, R., and Stark, H. (2008) GraFix: sample preparation for single-particle electron cryomicroscopy. Nat. Methods 5, 53–55

4. Oeffinger, M., Wei, K. E., Rogers, R., DeGrasse, J. A., Chait, B. T., Aitchison, J. D., and Rout, M. P. (2007) Comprehensive analysis of diverse ribonucleoprotein complexes. Nat. Methods 4, 951–956

5. Yang, W., Stein, H., and Freeman, M. R. (2008) Proteomic approaches to the analysis of multiprotein signaling complexes. Proteomics 8, 832–851

6. Link, A. J., Eng, J., Schietz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, B. M., and Yates, J. R., 3rd (1999) Direct analysis of protein complexes using mass spectrometry. Nat. Biotechnol. 17, 676–682

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

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

9. van Heel, M., Gowen, B., Matadeen, R., Orlova, E. V., Finn, R., Pape, T., Cohen, D., Stark, H., Schmidt, R., Schatz, M., and Patwardhan, A. (2000) Single-particle electron cryo-microscopy: towards atomic resolution. Q. Rev. Biophys. 33, 307–369

10. Kühner, S., van Noort, V., Betts, M. J., Lasker, K., Mann, M., and Baumeister, W. (2009) Insights into the molecular architecture of the 26S proteasome. Proc. Natl. Acad. Sci. U.S.A. 106, 11943–11947

2 M. M. Golas, F. M. Richter, B. Sander, H. Urlaub, and H. Stark, submitted manuscript.
