Anal. Chem. **2000,** *72,* 267−275

A Generic Strategy To Analyze the Spatial Organization of Multi-Protein Complexes by Cross-Linking and Mass Spectrometry

Juri Rappsilber,[†,‡] Symeon Siniossoglou,[§] Eduard C. Hurt,[§] and Matthias Mann*,[‖]

Protein Interaction Laboratory, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark, European Molecular Biology Laboratory, Meyerhofstrasse 1, D-69012 Heidelberg, Germany, and Biochemie Zentrum Heidelberg, University of Heidelberg, D-69012 Heidelberg, Germany

Most cellular functions are performed by multi-protein complexes. The identity of the members of such complexes can now be determined by mass spectrometry. Here we show that mass spectrometry can also be used in order to define the spatial organization of these complexes. In this approach, components of a protein complex are purified via molecular interactions using an affinity tagged member and the purified complex is then partially cross-linked. The products are separated by gel electrophoresis and their constituent components identified by mass spectrometry yielding nearest-neighbor relationships. In this study, a member of the yeast nuclear pore complex (Nup85p) was tagged and a six-member subcomplex of the pore was cross-linked and analyzed by 1D SDS−PAGE. Cross-linking reactions were optimized for yield and number of products. Analysis by MALDI mass spectrometry resulted in the identification of protein constituents in the cross-linked bands even at a level of a few hundred femtomoles. Based on these results, a model of the spatial organization of the complex was derived that was later supported by biological experiments. This work demonstrates, that the use of mass spectrometry is the method of choice for analyzing cross-linking experiments aiming on nearest neighbor relationships.

Most cellular functions are performed not by individual proteins but by large protein assemblies or multi-protein complexes. Examples of such complexes include the spliceosome, which "edits" premessenger RNA, the ribosome, which translates RNA into proteins, and various complexes involved in vesicle transport and DNA replication and repair. Even transient processes such as those involved in signal transduction from the cell surface to the nucleus are performed by multi-protein complexes. Analysis of these complexes is important for understanding the particular cellular process that they perform. Conversely, proteins that are found to belong to a complex with a defined cellular role are thereby placed into a functional context. We have previously demonstrated that biochemical purification of complexes followed by mass spectrometry is capable of identifying the members of these multi-protein complexes.[1] A large number of complexes have now been identified, including the spindle pole body,[2] the anaphase promoting complex,[3] the yeast nuclear pore complex,[4] and the human spliceosome.[5]

While the identification of members of complexes reveals the genes contributing to a cellular function, it is not the end point of the characterization of a complex. Insight into the spatial organization of a complex would help to integrate detailed functional studies of the individual proteins in order to reconstruct the function of the entire complex. Instances where cross-linking has been used to reveal closest-neighbor relationships in protein assemblies include the 30S subunit[6] and the 50S subunit[7,8] of the ribosome, the F_1-adenosine triphosphatase,[9,10] and the membrane components of several viruses.[11] The three steps in such a procedure usually are purification of the complex followed by a cross-linking reaction and, finally, identification of the cross-linked products. Identification of proteins in cross-linked pairs suggests that these proteins were spatially close in the complex. Despite the simplicity of the concept, each step of the cross-linking experiment has its inherent problems.

Traditionally, the biochemical purification of the protein complex in sufficient amounts and purity was difficult since a different strategy had to be designed for each complex under

* Corresponding author: (e-mail) mann@cebi.sdu.dk.

[†] European Molecular Biology Laboratory.

[‡] Present address: Protein Interaction Laboratory, University of Southern Denmark.

[§] University of Heidelberg.

[‖] University of Southern Denmark.

(1) Neubauer, G.; Gottschalk, A.; Fabrizio, P.; Seraphin, B.; Lührmann, R.; Mann, M. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 385−390.
(2) Wigge, P. A.; Jensen, O. N.; Holmes, S.; Soues, S.; Mann, M.; Kilmartin, J. V. *J. Cell Biol.* **1998**, *141*, 967−977.
(3) Zachariae, W.; Shevchenko, A.; Andrews, P. D.; Ciosk, R.; Galova, M.; Stark, M. J.; Mann, M.; Nasmyth, K. *Science* **1998**, *279*, 1216−1219.
(4) Zhao, Y.; Qin, J.; Chait, B. T. 46th ASMS Conference, Orlando, FL, 1998; pp 767.
(5) Neubauer, G.; King, A.; Rappsilber, J.; Calvio, C.; Watson, M.; Ajuh, P.; Sleeman, J.; Lamond, A.; Mann, M. *Nat. Genet.* **1998**, *20*, 46−50.
(6) Lambert, J. M.; Boileau, G.; Cover, J. A.; Traut, R. R. *Biochemistry* **1983**, *22*, 3913−3920.
(7) Redl, B.; Walleczek, J.; Stöffler-Meilicke, M.; Stöffler, G. *Eur. J. Biochem.* **1989**, *181*, 351−356.
(8) Redl, B.; Walleczek, J.; Stöffler-Meilicke, M.; Stöffler, G. *J. Biol. Chem.* **1989**, *264*, 4231.
(9) Bragg, P. D.; Hou, C. *Arch. Biochem. Biophys.* **1986**, *244*, 361−372.
(10) Bragg, P. D.; Hou, C. *Eur. J. Biochem.* **1980**, *106*, 495−503.
(11) Wong, S. S. *Chemistry of Protein Conjugation and Cross-Linking*; CRC Press: Boca Raton, FL, 1991; pp 237−243.

10.1021/ac991081o CCC: $19.00
Published on Web 12/08/1999

investigation. In contrast to the purification of single proteins, the fragility of many complexes severely limits the repertoire of protein chemical methods that can be used. Modern molecular biological methods now allow transfection of cells with a single complex component that is fused to a sequence coding for a tag at the cDNA level. The tag consists of a complete protein domain or a short amino acid sequence called an epitope tag and is characterized by a strong interaction with a tag-specific solid phase. In this way, the entire complex can be purified by generic means.[12]

In order to obtain cross-linked products, a series of cross-linkers must be tested for each complex. A balance must be kept between obtaining maximal yield of cross-linked products but at the same time not cross-linking more than a few proteins to each other. It can hardly be predicted which cross-linker will give the best results.[13] Currently, no generic methods for complex cross-linking exist and reaction conditions have to be optimized on a case by case basis.

In the final step, the proteins covalently attached to each other in the cross-linking reaction and the unreacted proteins have to be separated by gel electrophoresis and subsequently identified. In one approach, this has been achieved using cleavable cross-linkers and additional gel electrophoresis to separate the linked proteins after cleavage of the cross-linker[6,9] and identification of the proteins by their mobility or by specific antibodies. Alternatively, cross-linked partners can be identified using antibodies without prior cleavage.[7,14−16] Both approaches require extensive studies into the investigated complex prior to the topological investigation. For multidimensional electrophoresis, the mobility of the subunits has to be sufficiently characteristic in order to identify the proteins by migration distance. Additionally, each step suffers from sample losses, thus requiring radioactive labeling even if starting amounts are in the picomole range.[9] Although antibodies are able to detect small quantities of proteins, antibody production is both time consuming and labor intensive. Since the ability of the antibody to recognize the unmodified protein is not predictable, it is even more uncertain whether the antibody would recognize the cross-linked and thus modified protein (see also this paper). Traditional methods are thus less suited for rapid and generic analysis of cross-linked proteins.

Over the last years, mass spectrometric tools have emerged as the preferred method for the rapid identification of SDS−PAGE separated proteins.[17] We have previously shown that simple protein mixtures in a gel band are accessible for analysis by MALDI mass fingerprinting.[18]

Based on the progress in the areas of protein complex purification and mass spectrometry we present here a simple generic experimental strategy, devised to rapidly investigate the topology of multi-protein complexes. This strategy consists of the following steps: A complex, purified via a tagged component, is cross-linked under conditions that yield mainly pairwise linked proteins. The products are separated by gel electrophoresis, and the constituents of the bands are identified by MALDI MS mixture analysis. Pairwise or higher interactions as read out by the MALDI protein identifications are used as constraints for the topological models of the multi-protein complex.

The Nup84p complex, a subcomplex of the yeast nuclear pore complex composed of six known members, was chosen as a model (Figure 1). Nup85p is a component of that complex and was tagged with the ProtA tag.[19] The construct also contained a linker, which is cleavable by the tobacco etch virus (TEV) protease. The ProtA tag binds to the heavy chain of IgG antibodies immobilized on beads. The complex is then specifically eluted from the beads by cleavage of the linker by the TEV protease. The purified complex was subsequently cross-linked. The products were separated by 1D SDS−PAGE, digested in situ with trypsin[17] and analyzed by MALDI mass spectrometry, resulting in a model for the organization of the Nup84p complex. After completion of the present work, extensive knockout and tagging experiments produced independent information on the interactions within the complex, which were in full agreement with the model proposed here. We therefore propose that MALDI mass spectrometry can be used to identify the cross-linked products, obviating the need for specific reagents.

EXPERIMENTAL SECTION

Purification of the Nup84p Complex. The Nup85p nucleoporin complex was affinity purified as described[19,20] but using an alternative elution procedure. Briefly, a yeast strain with a deletion in the endogenous NUP85 gene was transformed with a low-copy vector carrying an N-terminally tagged version of NUP85. The tag is composed of two IgG binding domains of the *Staphylococcus aureus* protein A (ProtA tag) followed by a spacer and the TEV protease cleavage site to give a total length of 134 amino acids (~15 kDa). The tagged version of NUP85 was under the control of the NOP1 promoter. Spheroplasts made from yeast cells expressing the fusion protein were lysed in 1% Triton X-100, 150 mM KCl, 20 mM Tris-HCl (pH 8.0), 5 mM $MgCl_2$, supplemented with a cocktail of protease inhibitors. From the whole-cell extract ProtA−Nup85p was affinity purified using IgG-Sepharose (Pharmacia, Uppsala, Sweden). The bound material was specifically eluted from the column under nondenaturing conditions by proteolysis using recombinant TEV protease (Life Technologies, Rockville, MD) in 20 mM potassium phosphate (pH 7.4), 150 mM NaCl.

Cross-Linking Reaction. Judging from the intensity of Coomassie-stained bands after SDS−PAGE, the concentration of Nup84p complex in the cross-linking reactions was estimated to be in the order of 30 nM. Cross-linkers were used at a concentration of 1 μM to 1 mM obtained by dilution from 10 mM stock solution (In DMSO: SMCC, DSS, BMH, BASED. In water: BS^3, DMA[21]). Reaction volume was 35 μL. The cross-linking reaction

(12) Deutscher, M. P., Ed. *Guide to Protein Purification*; Methods in Enzymology 182; Academic Press: San Diego, 1990.

(13) Wong, S. S. CRC Press: *Chemistry of Protein Conjugation and Cross-Linking*; Boca Raton, FL, 1991; pp 223−228.

(14) Walleczek, J.; Redl, B.; Stöffler-Meilicke, M.; Stöffler, G. *J. Biol. Chem.* **1989**, *264*, 4231−4237.

(15) Layh-Schmitt, G.; Herrmann, R. *Infect. Immunol.* **1994**, *62*, 974−979.

(16) Mullins, R. D.; Stafford, W. F.; Pollard, T. D. *J. Cell Biol.* **1997**, *136*, 331−343.

(17) Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M. *Anal. Chem.* **1996**, *68*, 850−858.

(18) Jensen, O. N.; Podtelejnikov, A. V.; Mann, M. *Anal. Chem.* **1997**, *69*, 4741−4750.

(19) Siniossoglou, S. G. P.; Hurt, E. C. In *Cell Biology: A Laboratory Handbook*; Celis, J. E., Ed.; Academic Press: San Diego, 1998; Vol. 2, pp 159−164.

(20) Siniossoglou, S.; Wimmer, C.; Rieger, M.; Doye, V.; Tekotte, H.; Weise, C.; Emig, S.; Segref, A.; Hurt, E. C. *Cell* **1996**, *84*, 265−275.

(21) DMSO, dimethyl sulfoxide (Sigma); SMCC, succinimidyl-4-(*N*-maleimidomethyl)cyclohexane-1-carboxylate; DSS, disuccinimidyl suberate; BMH, bismaleimidohexane; BASED, bis[β-(4-azidosalizylamido)ethyl] disulfide; BS^3, bis(sulfosuccinimidyl) suberate; DMA, dimethyl adipimidate (Pierce).

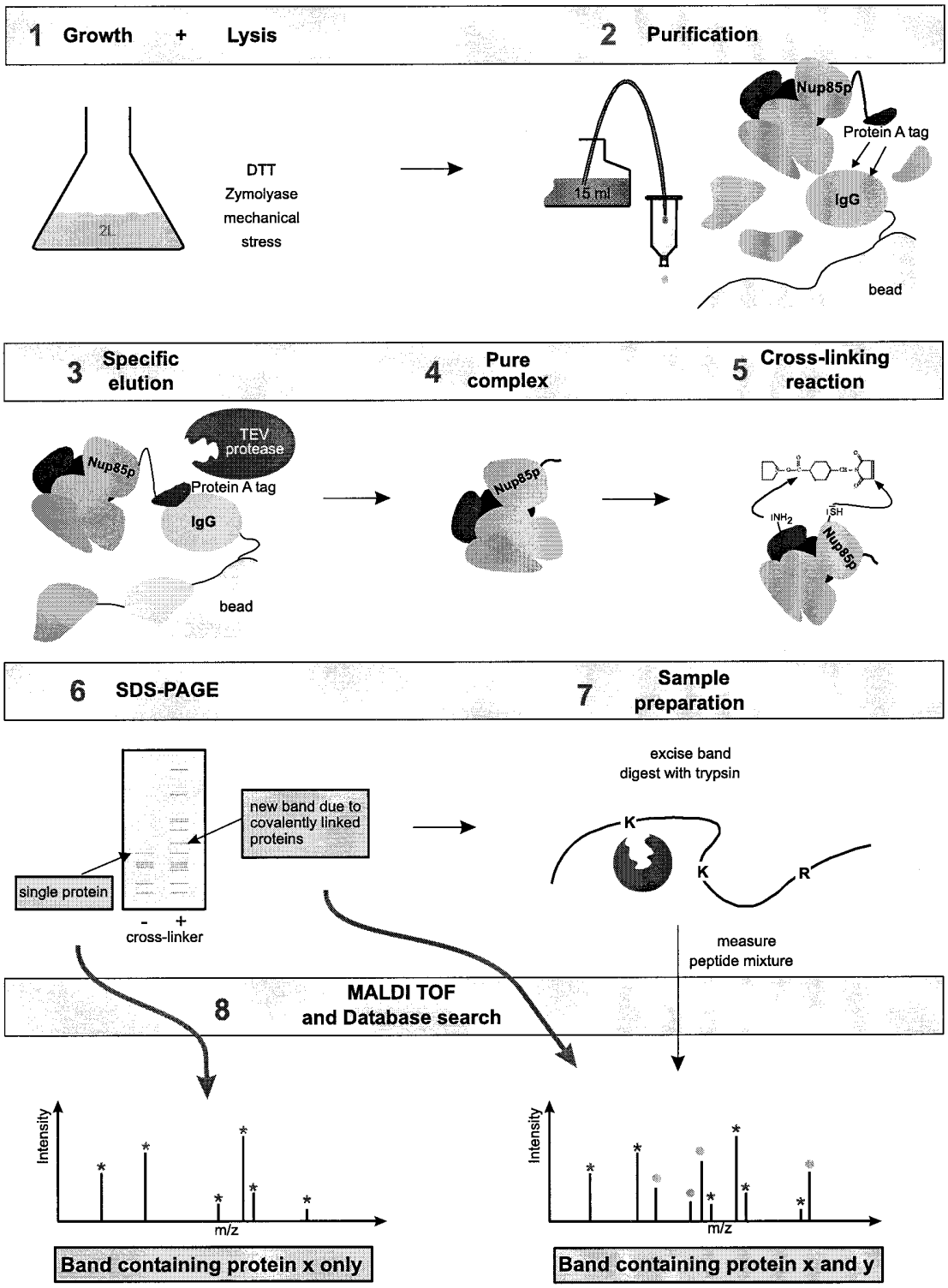

Figure 1. Schematic outline of the approach used to obtain topological information on a protein complex. After the endogenous copy of a complex component has been replaced by a tagged copy, cells are cultured, harvested, and lysed. The complex is purified from the crude lysate by interaction of the ProtA tag with an IgG column. The complex is eluted specifically by proteolytic cleavage of the tag by the TEV protease. A fraction of the purified complex is cross-linked, and the products are separated by 1D SDS−PAGE. In parallel, non-cross-linked complex is subjected to gel electrophoresis in order to recognize the cross-link products. Relevant bands are excised, digested with trypsin, and measured by MALDI MS. The peptide mass maps obtained are searched against the nonredundant database or a complex-specific database. On the basis of the proteins identified in cross-link products, a model of the complex topology can be constructed.

was performed on ice in 20 mM potassium phosphate (pH 7.4), 150 mM NaCl. In the case of BASED, the mixture was irradiated with UV light of 366 nm for 30 min. The reaction mixtures were incubated for 30−60 min. The reaction was terminated by either

switching off the UV radiation (for BASED), addition of Tris-HCl (pH 7.4) (2 M stock solution) and DTT (1 M stock solution) (for SMCC), or addition of Tris-HCl (pH 7.4), in all other cases. The final concentration of the terminating agent was 50 mM.

The protein material obtained after the cross-linking reaction was separated by electrophoresis on 6% SDS−polyacrylamide gels and visualized by silver staining[17] for mass spectrometric analysis or blotted on nitrocellulose membrane for probing with antibodies against components of the complex.

In-Gel Digestion of Proteins. Protein-containing bands were excised from the gel, washed, reduced, and S-alkylated with iodoacetamide, and in-gel digested with an excess of bovine trypsin (sequencing grade, Boehringer Mannheim, Mannheim, Germany) (3 h, 37 °C) as described.[17] A fraction of 0.4 μL of a total of 15−25 μL of digest solution was used for MALDI mass spectrometric analysis.

Sample Preparation for MALDI MS. The fast evaporation method was used to prepare samples for MALDI MS.[22] Briefly, a saturated solution of α-cyano-4-hydroxy-*trans*-cinnamic acid (Sigma, St. Louis, MO) in acetone was mixed with a 10 g/L solution of nitrocellulose (BioRad, Munich, Germany) in acetone/2-propanol (1:1, v/v) in a 4:1 ratio. A 0.3 μL aliquot of the matrix solution was deposited on the target to yield a thin matrix layer after evaporation of the solvent. To prevent dissolution of the matrix, 0.4 μL of digest supernatant was injected into 0.7 μL of 10% formic acid previously deposited onto the matrix layer. The acidified droplet was left to dry at room temperature. Samples were washed twice with 5−10% formic acid in order to remove remaining buffer salts before mass spectrometric analysis.

MALDI MS Peptide Mapping. Peptide mixtures were analyzed on a Bruker Reflex II reflector time-of-flight mass spectrometer equipped with the Scout multiprobe inlet, a 337 nm N_2 laser, and a gridless delayed extraction ion source (Bruker-Franzen, Bremen, Germany). Data acquisition was performed by a Lecroy 9350AM 1 gigasample/s digital storage oscilloscope (LeCroy Corp., Chestnut Ridge, NY). The LaserOne software developed by our group was used to control both the oscilloscope and the mass spectrometer. Mass spectra were obtained as the sum of 200−250 selected measurements, and spectra were internally calibrated using matrix-related signals and trypsin autolysis signals.

Database Searching with Mass Spectrometric Data. Lists of peptide masses were searched against a nonredundant protein sequence database (NRDB) maintained by the European Bioinformatics Institute (EBI, Hinxton, U.K.) and updated regularly at the EMBL. Cysteine was set as *S*-carbamidomethylcysteine, up to two incomplete cleavages per peptide were considered, and the minimum required mass accuracy was 50 ppm. Oxidation of methionine was considered in the second pass search algorithm of PeptideSearch.[23] No constraints on the molecular weight or the organism were applied. Nonmatching peptides from the first round of the search were used in the second round. Nonmatching peptides from the second round of the search were used for the third round to successively identify the multiple components of the protein mixture obtained from a single band in the gel. This iteration was continued until no further statistically meaningful hits were found.[18]

In addition, two further databases were searched. One contained all protein sequences of budding yeast *(Sacharomyces cerevisiae)*, human and sheep keratins, and bovine trypsin. The other database was restricted to the protein sequences of the six components of the Nup84p complex.

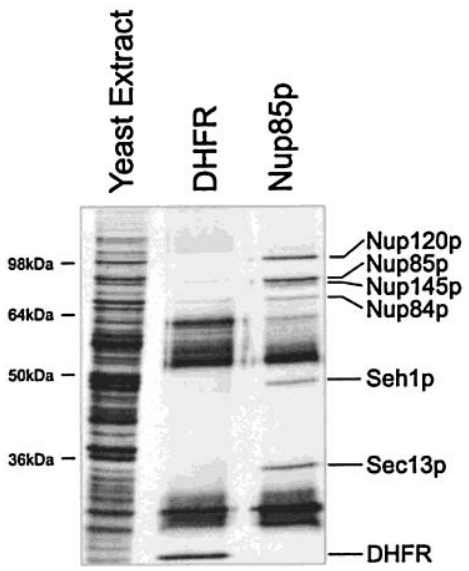

Figure 2. Silver-stained SDS−PAGE gel (10%) of the Nup84p complex (lane 3). As a control, dehydrofolatreductase (DHFR), which has no known binding partners, was purified in parallel to show the background of the purification procedure (lane 2). Total yeast extract is shown in lane 1.

RESULTS AND DISCUSSION

Purification of the Nup84p Complex. The Nup84p complex, a subcomplex of the yeast nuclear pore complex, is composed of six known members, one of which is Nup85p.[20] We have generated a yeast strain in which the endogenous Nup85p is replaced with an N-terminally tagged version of the protein.[24] Expression of the tagged version in this strain does not cause any abnormal phenotype whereas deletions of Nup85p result in severe defects in the nuclear structure and the nuclear membrane. When lysate from the tagged strain is passed over an IgG-Sepharose column, Nup85p is retained owing to the interaction of the ProtA tag with IgG.[19] Five proteins were found to bind to Nup85p, and mass spectrometric protein identification showed that all are known members of the Nup84p complex.[20] The interaction of these proteins with Nup85p is specific as shown by a control purification of ProtA-tagged dehydrofolatreductase (DHFR), a metabolic enzyme with no known binding partners, using the same experimental conditions (Figure 2). To separate the Nup84p complex from proteins bound nonspecifically to the IgG-Sepharose column and to elute the complex in the native state, Nup85p was cleaved from the tag using the TEV protease.[25] Subsequent elution of the entire material by low pH revealed that the vast majority of the

(22) Vorm, O.; Roepstorff, P.; Mann, M. *Anal. Chem.* **1994**, *66*, 3281−3287.
(23) Mann, M.; Højrup, P.; Roepstorff, P. *Biol. Mass. Spectrom.* **1993**, *22*, 338−345.
(24) Siniossoglou, S.; Lutzmann, M.; Santos-Rosa, H.; Fahrenkrog, B.; Pante, N.; Leonard, K.; Aebi, U.; Hurt, E. C., submitted.
(25) Senger, B.; Simos, G.; Bischoff, F. R.; Podtelejnikov, A.; Mann, M.; Hurt, E. *EMBO J.* **1998**, *17*, 2196−2207.

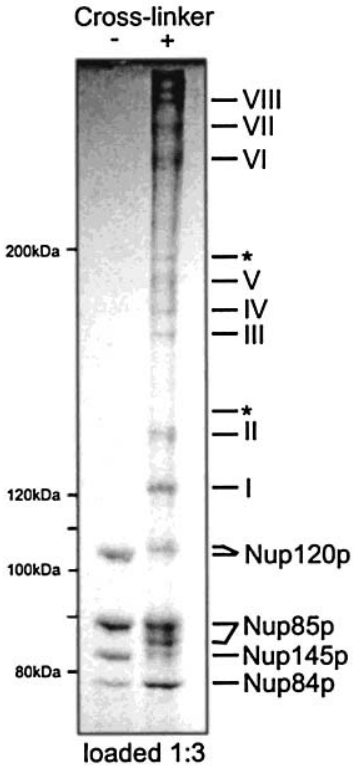

Figure 3. Silver-stained SDS−PAGE gel (6%) of the purified Nup84p complex either non-cross-linked (lane 1) or cross-linked (lane 2) loaded with a ratio of 1:3. At least eight new bands appear after addition of cross-linker (labeled I−IX). Background bands are denoted with an asterisk.

complex was already eluted by treatment with the TEV protease. The purification procedure does not employ any property specific to the Nup84p complex. Instead, the method is based on the tag alone and can therefore be generically used to attempt purification of any complex. This method also results in a consistent protein background, most notably IgG leakage from the column and the heat shock protein HSP70, a protein that is often found bound to proteins outside their in vivo context. This background does not hamper further analysis as it can easily be recognized by comparison to a control purification.

Chemical Cross-Linking of the Subunits in the Nup84p Complex. A selection of homo- and heterobifunctional cross-linkers of variable length and chemical specificity were used to optimize the number and the yield of products. The low concentration of the purified Nup84p complex minimized the likelihood of cross-links between complexes.

Five different cross-linkers with reactivity toward sulfhydryl and/or amino groups and one nonspecific linker (BASED) were investigated (data not shown). The cross-linkers SMCC, BS^3, DSS, and BMH resulted in the appearance of different new bands on the 1D polyacrylamide gel at a concentration of 10 μM (~300-fold excess over the complex). BS^3 and DSS both resulted in one major product at about the size of band II (Figure 3) obtained by SMCC. SMCC and BMH had a more equal distribution of yield in the products (data not shown). With BASED, the intensity of nonlinked material on the gel decreased; however, surprisingly, no bands corresponding to cross-linked species were observed. Only one reagent, DMA, did not result in any apparent reaction.

To establish optimal reaction conditions, the parameters agent concentration, temperature, and reaction time were investigated for SMCC and BS^3, the most promising of the cross-linkers that were initially screened. Cross-linking with 100 μM SMCC for 1 h at 2 °C gave the largest diversity of products for this complex, as judged by SDS−PAGE and subsequent silver staining. Consequently, these conditions were chosen for the further analysis of the Nup84p complex (Figure 3).

The ratio of complex to cross-linker is important in maximizing the number of proteins linked as pairs. From the desired distribution of products, it follows that the yield of the cross-link reaction cannot be 100%. Optimal reaction conditions give a mixture of unlinked components, pairwise linked components, ternary complexes, and so on up to a completely cross-linked complex. Too small an amount of cross-linker will not use the protein amount available efficiently. More efficient usage of the unlinked components shifts the product distribution to higher-order cross-link products. As these contain less information, a trade-off between yield and information governs the search for optimal reaction conditions.

The optimal molar ratio of complex to reagent was found to be ~1:3000 (data not shown). A ratio of 1:300 gave low yields of cross-linked products, whereas ratios of 1:10 000 or higher practically linked all the members of Nup84p complex to one another.

Establishing standard conditions for the cross-linking reaction was difficult in that the Nup85p complex could only be purified in relatively small amounts (few hundred-femtomole range), and that quantitation of cross-linking reactions proved difficult as comparison of stained bands only gave an estimate of the complex concentration. The cross-linking reaction is limited by hydrolysis of the reagent[26] and the fact that functional groups on each of the two interacting protein surfaces react with cross-linker over time which then cannot react with each other. Time course experiments using SMCC, BS^3, and BMH revealed that the reaction was essentially complete after 30−60 min (data not shown) presumably reflecting these two effects. Reactions were performed at 2, 24, and 30 °C. It was found that higher temperatures only increased the staining background on the gel but did not increase cross-linking efficiencies and were therefore not used further (data not shown).

Analysis of the Product Composition. The products obtained from chemical cross-linking of the Nup84p complex using SMCC were separated by 1D SDS−AGE and analyzed by two independent methods. The bands were blotted onto nitrocellulose membrane and probed with antibodies specific for components of the Nup84p complex (Western blotting). Alternatively, relevant bands were excised from the gel and digested, the peptide mixture obtained for each product was analyzed by MALDI-MS, and the proteins were identified by database searching.

The amino- and sulfhydryl-specific heterobifunctional cross-linker SMCC yielded at least eight visible bands on silver-stained SDS−PAGE gels which were the direct result of the cross-linking reaction (Figure 3). In addition to the cross-linking products, the gel shows the complex components Nup84p, Nup145p as well as Nup85p (two bands), and Nup120p (two bands). Nup145p mi-

(26) Partis, M. D.; Griffiths, D. G.; Roberts, G. C.; Beechey, R. B. *J. Protein Chem.* **1983**, *2*, 263−277.

grates at ~85 kDa because the N-terminal region is posttranslationally cleaved.[27] The unreacted proteins could be stained by Coomassie while the bands containing the cross-link products were below the detection limit of Coomassie staining. Two small components of the Nup84p complex, Seh1p (40 kDa) and Sec13p (32 kDa), do not appear on the 6% gel as it covers the mass range from 70 kDa to more than 350 kDa. Compared to the unmodified complex, some of the bands representing single components appear to be shifted upward in mass (migrate slower), presumably due to the additional mass acquired from the cross-linker (see Nup145p and Nup120p). However, Nup85p migrates in two bands, one of which has an apparent molecular weight below the mass of the unmodified protein. This may be due to an intramolecular linkage in Nup85p forcing the denatured protein into a more compact structure and thereby increasing its mobility on the gel.

A similar effect may also explain the differences in the band patterns observed by different cross-linkers. Depending on which functional groups they attack and therefore at which positions they link the proteins, products of identical composition may have different structures in the denatured state and therefore different migration properties. In agreement with such a reasoning, BS^3 and DSS, both targeting at amino groups, yielded very similar gel patterns.

Except for Nup120p, specific antibodies against peptides from all members of the complex[20,24,28] were used in Western blots. Without cross-linking, each of the peptide antibodies stained its antigen in the Western blot (data not shown) and the non-cross-linked proteins were weakly recognized after the cross-linking reaction. Surprisingly, none of the cross-link products were recognized by any of these antibodies. A possible explanation is that modification of lysine and cysteine residues at the protein surface hindered recognition by the antibody (this phenomenon is well known for example in antibody staining of fixed cells[29]). Even though antibodies have been employed successfully for the analysis of cross-linked protein complexes (see, for example, refs 1**5** and 16), our results clearly demonstrate that the use of antibodies in the recognition of modified proteins can be problematic.

We then investigated whether mass spectrometry, like Western blotting, would be influenced by the modification of proteins by cross-linkers. As mentioned above, the migration positions of the bands for Nup145p and Nup120p were shifted and Nup85p migrates in two bands following cross-linking as a result of chemical modification (Figure 3). To determine whether the modification introduced into the proteins during the reaction with the cross-linking agent still enables digestion with trypsin and subsequent identification of the protein using MALDI MS, the bands of Nup85p and Nup120p were analyzed. The shifts indicated extensive modification of the proteins and therefore provide a good test case for the cross-linked proteins. In both cases, good MALDI spectra were obtained (data not shown), indicating that the modification does not significantly interfere with the digestion. The peptide mass fingerprint clearly identified Nup120p and Nup85p. Comparison to the peptide map of the unmodified Nup85p showed no significant differences in the peptides present, and the sequence coverage was approximately equal in both cases. Therefore, we conclude that the modification does not affect the sequence coverage to a significant extent and that the approach of identifying cross-linked proteins by mass spectrometry is viable.

In principle, peptide peaks in the mass spectrum of the modified protein but not the unmodified one can be used to deduce modification sites and intramolecular linkages.[30] Such an approach could be used for the analysis of intermolecular linkage sites and, thereby, protein contact sites. However, in our experience, the MALDI peptide mass maps are generally not sufficiently quantitative and reproducible to allow assignment of cross-linked peptides by their mass alone. Without further analysis (such as presented in refs 31 and 32), changes are not reliably interpreted in an isolated manner but only in a statistical way with several new peaks pointing toward the same conclusion. In any case, such an investigation of cross-link sites would require more protein material or a way of selectively identifying the modified or cross-linked peptides against the background of the unmodified species in a sensitive way.[33]

While in the present case only combinations of the six known subunits of the Nup84p complex were expected, in other cases—when the cross-linking precedes the purification of the complex—additional proteins could be found in the cross-linked bands. Therefore, we analyzed respective cross-linked bands as if we had no knowledge of their protein composition. Peptide mass fingerprints obtained by MALDI mass spectrometry (Figure 4) were used to search a nonredundant database containing over 300 000 protein sequences. The analysis of cross-linked proteins resembles the identification of proteins in mixtures with the added constraint that the components are expected in a 1:1 stoichiometry. To identify a protein against the background of the signals from the other mixture components, the mass accuracy needs to be very high.[18] Because the mass accuracy in weak peaks is limited by ion statistics, only intense peaks were annotated and used for the search. The quality of the data is shown by the fact that more than 80% of the assigned masses correspond to the top identifications returned by the NRDB search, especially when considering the low amounts (silver-staining levels) of proteins present in the mixtures. Interestingly, more than two protein components could still be identified in MALDI searches, even at these levels. For example, four proteins were identified in band VII of Figure 3. These were Nup85p, Nup145p, Nup84p, and Nup120p, which are all previously known members of the Nup84p complex (Figure 4). Nup120p was identified in band I, Nup85p in band II, and Nup145p in band III.

The two small proteins, Sec13p (32 kDa) and Seh1p (40 kDa), could not be identified unambiguously in one of the cross-link products when the entire NRDB was searched without molecular weight restriction. Both were identified, however, when the search

(27) Teixeira, M. T.; Siniossoglou, S.; Podtelejnikov, S.; Bénichou, J. C.; Mann, M.; Dujon, B.; Hurt, E.; Fabre, E. *EMBO J.* **1997**, *16*, 5086−5097.

(28) Fabre, E.; Boelens, W. C.; Wimmer, C.; Mattaj, I. W.; Hurt, E. C. *Cell* **1994**, *78*, 275−289.

(29) Harlow, E.; Lane, D. *Antibodies: A Laboratory Manual*; CSHL Press: Cold Spring Harbor, NY, 1999; pp 153−219.

(30) Glocker, M. O.; Kalkum, M.; Yamamoto, R.; Schreurs, J. *Biochemistry* **1996**, *35*, 14625−14633.

(31) Kalkum, M.; Przybylski, M.; Glocker, M. O. *Bioconjugate Chem.* **1998**, *9*, 226−235.

(32) Scaloni, A.; Miraglia, N.; Orrù, S.; Amodeo, P.; Motta, A.; Marino, G.; Pucci, P. *J. Mol. Biol.* **1998**, *10*, 4945−4958.

(33) Trester, M. L. H. N.; Muir, T.; Chait, B. T. 47th ASMS Conference, Dallas, TX, 1999.

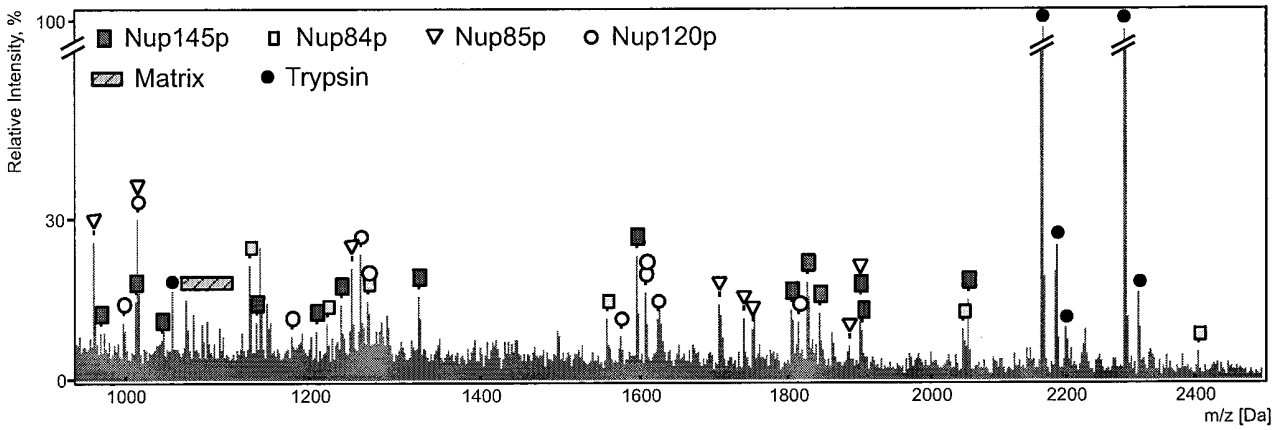

Figure 4. MALDI peptide mass map obtained after in-gel tryptic digestion of band VII. Four proteins were identified by iterative searches in the nonredundant database (NRDB) with greater than 50 ppm mass accuracy. The proteins were identified in the following order: Nup145p, Nup120p, Nup84p, and Nup85p.

Table 1. Identification of Cross-Linked Proteins

band	MW_{exp} $\times 10^{-3}$	Sec13p (32 kDa)[a]	Seh1p (40 kDa)[a]	Nup84p (84 kDa)[a]	Nup85p (86 kDa)[a]	Nup145p (85 kDa)[a,b]	Nup120p (121 kDa)[a,c]	linked proteins	MW_{calc} $\times 10^{-3}$
I	122	6/25	6/29	16/25	11/23	14/14	33/35	Nup120p−(Sec13p or Seh1p)	153/161[c]
								Nup84p−(Sec13p or Seh1p)	116/124
								Nup85p−Sec13p	118
								Nup145p−Sec13p	117
II	138	−/−	8/32	−/−	11/21	9/11	2/−	Nup85p−Seh1p	126
								Nup145p−Seh1p	125
III	168	−/−	1/−	−/−	8/11	11/8	−/−	Nup85p−Nup145p	171
IV	175	1/−	−/−	1/−	8/19	6/7	6/9	Nup85p−Nup120p	207
								Nup145p−Nup120p	206
V	185	9/33	−/−	1/−	7/15	6/6	−/−	Nup85p−Nup145p−Sec13p	203
VI	>200	−/−	5/18	2/−	11/15	18/19	11/14	Nup85p−Nup145p−Nup120p−Seh1p	332
VII	>200	1/−	1/−	9/17	8/14	16/15	10/11	Nup85p−Nup145p−Nup120p−Nup84p	376
VIII	>200	8/40	2/−	8/15	9/23	23/24	11/14	Nup85p−Nup145p−Nup120p−Nup84p−Sec13p	408

[a] *x*/*y*%, where *x* represents the number of peptides and *y* the sequence coverage observed for the respective proteins. [b] In mature Nup145p, the N-terminal third of the protein sequence is cleaved so that Nup145p migrates faster than Nup85p on SDS−PAGE. [c] Nup120p runs at an apparent MW of 100 000.

was restricted to yeast proteins and required a protein size of less than 130 kDa. These constraints are reasonable as the complex is purified from yeast and all protein components are smaller than 130 kDa.

The results shown here demonstrate that it is possible to identify proteins from cross-linked products either by searching the entire NRDB or at least by restricting the search to all proteins of the organism under investigation. Even when four proteins with a total mass of over 350 kDa were covalently linked and purified in amounts of only a few hundred femtomoles (silver stain level) (band VII), the analysis by MALDI MS was successful without the need to know the composition a priori.

Analysis of the Product Composition by MALDI Mass Spectrometry. After the initial identification of all the complex members in the cross-linked bands, we next analyzed the individual cross-link bands. The spectra were reanalyzed, assigning also peaks closer to the noise of the spectrum, resulting in improved sequence coverage. In all cases, more than 40% of the assigned peaks matched to peptides of the investigated proteins. For the identification of a single protein, this is a very good result and generally sufficient for identification even in the complete database. For a complex mixture, however, not all components can be identified from the NRDB. As the composition of the complex is known, either from previous information as in this investigation or by initial identification as shown above, it is now possible to use a database that contains only the protein sequences of the complex members.

With this refined analysis, a large number of binary and higher interactions were determined (Table 1). Band III of Figure 3 contains Nup85p and Nup145p. The added masses of Nup85p (86 kDa) and the C-terminal part of Nup145p (~85 kDa) are in agreement with the apparent mass of band III (168 kDa). We therefore conclude that Nup85p and Nup145p are direct neighbors. Band II contained Nup85p, Nup145p, and Seh1p. The apparent molecular weight of the band is 138 000, which agrees within the uncertainty introduced by the unknown migration behavior of cross-linked protein pairs with the protein pairs Nup85p + Seh1p (86 000 + 40 000) and Nup145p + Seh1p (85 000 + 40 000). From these data we conclude that Seh1p is a direct neighbor of both Nup85p and Nup145p. This is further confirmed by the analysis of band V, which migrates at an apparent mass of 185 kDa and contains Seh1p, Nup85p, and Nup145p, which have

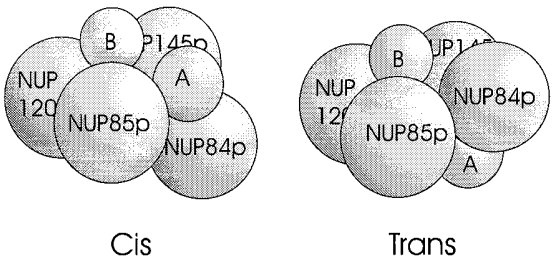

Figure 5. Proposed topology model for the Nup84p complex. The two small complex members may be on the same side of the complex (cis) or on opposing sides (trans). Sec13p may be close to Nup120p and Seh1p close to Nup84p (Sec13p = B and Seh1p = A) or vice versa (Sec13p = A and Seh1p = B).

an added mass of 203 kDa. This band therefore contains all three proteins cross-linked to one another. Similar reasoning on the results of the identifications in band IV shows that Nup120p is a direct neighbor of Nup85p and Nup145p. The higher molecular weight bands VI, VII, and VIII all contain four or more components cross-linked to each other and are therefore less informative. These bands, however, all contain Nup85p and Nup145p, thus supporting a central role for these two proteins. The remaining band (band I) migrates at an apparent mass of 122 kDa and was found to contain all six members of the complex. Considering the apparent mass of the band, each of the four large Nup proteins is potentially linked to either one of the small proteins, Sec13p or Seh1p. Nup85p linked to Seh1p was identified in band II; therefore the combination in band I is likely to be Nup85p + Sec13p. Equally, Nup145p + Seh1p was identified in band II, suggesting the combination of Nup145p + Sec13p in band I.

The lack of observation of certain protein linkages can also suggest topological information: Sec13p was not found linked to Seh1p (the corresponding mass range is not shown on the gel in Figure 3), and Nup84p was not found directly linked to Nup120p. This suggests that Sec13p is not directly adjacent to Seh1p and that Nup84p does not have direct contact to Nup120p. In addition, for each of the two larger proteins, Nup84p and Nup120p, there is only one band, indicating a link to one of the smaller proteins, Sec13p and Seh1p (band I). This is consistent with Nup120p being directly adjacent to either one of the small proteins and Nup84p to the other one, with both pairs being separate from each other.

Taking all these data together, a model for the Nup84p complex can be proposed in which Nup85p and Nup145p form a central pair with two opposing pairs at the side, one composed of Nup84p with either Sec13p or Seh1p and the other being Nup120p with either Seh1p or Sec13p. At present, the orientation of the two opposing pairs, cis or trans, remains unclear (Figure 5). This model satisfies all the constraints derived from the data in Table 1.

After this work was completed, biochemical analysis of the Nup84p complex was performed which corroborates the model given above: Components of the Nup84p complex were ProtA tagged and other components deleted or mutated, followed by purification of the complex by the ProtA tag.[24] Briefly, the outcome was that when one of either Nup85p, Nup145p, or Nup120p is absent, the complex cannot form. Nup84p, Seh1p, or Sec13p deletion/mutation, however, does not seem to affect the assembly of the other five components, which still bind with 1:1 stoichiometry. Additionally, in the same analysis, Nup85p was found to interact directly with Seh1p. These data suggest that Nup85p, Nup145p, and Nup120p are in the core of the complex, whereas Nup84p, Seh1p, and Sec13p are on the periphery. This resembles closely our model (Figure 5) based only on the data from cross-linking experiments.

CONCLUSION

In this study we were able to show that three established techniques, gene tagging to allow generic protein complex purification, chemical cross-linking of a purified complex, and mass spectrometric identification of proteins in mixtures, can be combined into a powerful approach for the topological investigation of protein complexes. Analyses were possible in the subpicomole range, opening a wide range of biological areas for analysis. As an example, the Nup84p subcomplex of the nuclear pore complex in yeast was purified and analyzed. Up to five cross-linked components could be identified at the femtomole range in the same band by MALDI mass spectrometry. On the basis of the constraints obtained from our analysis, we have proposed a topological model for the Nup84p complex. This kind of analysis is generic and can in principle be performed rapidly on any multiprotein complex.

Exhaustive cross-linking of a complex in conjunction with the mass spectrometric determination of the resulting mass can give insight into the multimeric form of the complex.[34−36] For the analysis of nearest-neighbor relationships, however, the cross-linking reaction has to be incomplete. The starting material is distributed over several products and is not entirely utilized. As a result, an individual product can contain as little as 5% of the educt. Therefore, 3−5 pmol of the complex is required as starting material for the cross-link reaction. For large complexes, the band pattern on the gel will become so complex that more and more products will overlap on the 1D gel. Complexes, like the ribosome, which are composed largely of small proteins, can be investigated using 2D electrophoretic techniques. For most complexes, however, the cross-linked products will be above 100 kDa, precluding the use of 2D gel electrophoresis. Therefore, such an approach is currently limited to complexes with up to 10−15 members.

The next step in the characterization of protein interactions in a complex would be the identification of contact surfaces. Recent work[32,37−41] suggests that the combination of cross-linking and mass spectrometry can also be utilized to address this question. Cross-linked peptides are isolated that define the interacting surfaces in the two proteins if both peptides can be identified.

(34) Farmer, T. B.; Caprioli, R. M. *Biol. Mass. Spectrom.* **1991**, *20*, 796−800.
(35) Farmer, T. B.; Caprioli, R. M. *J. Mass. Spectrom.* **1998**, *33*, 697−704.
(36) Helin, J.; Caldentey, J.; Kalkkinen, N.; Bamford, D. H. *Rapid Commun. Mass. Spectrom.* **1999**, *13*, 185−190.
(37) Hegyi, G.; Mák, M.; Kim, E.; Elzinga, M.; Muhlrad, A.; Reisler, E. *Biochemistry* **1998**, *37*, 17784−17792.
(38) Mills, J. S.; Miettinen, H. M.; Barnidge, D.; Vlases, M. J.; Wimer-Mackin, S.; Dratz, E. A.; Sunner, J.; Jesaitis, A. J. *J. Biol. Chem.* **1998**, *273*, 10428−10435.
(39) Vater, J.; Heinze, K.; Friedrich, B.; Kalbitz, B.; Blokesch, A.; Irrgang, K.-D.; Thiede, B.; Salnikow, J. *Ber. Bunsen-Ges. Phys. Chem.* **1996**, *100*, 2107−2111.
(40) Yang, T. H. D.; Mahan, K. J.; Zaluzec, E. J.; Watson, T. J.; Gage, D. A. *Anal. Biochem.* **1996**, *242*, 55−63.
(41) Yu, Z.; Friso, G.; Miranda, J. J.; Patel, M. J.; Lo-Tseng, T.; Moore, E. G.; Burlingame, A. L. *Protein Sci.* **1997**, *6*, 2568−2577.

Very large amounts of material are still required for this task. If the cross-linker contains a tag that enables its identification at low levels[42] or allows its purification,[33] more biologically relevant levels may be reached by this approach.

Purification of multi-protein complexes followed by mass spectrometric identification of the constituents is currently one of the most established and successful analytical steps in proteomics. Here we show that it can be followed by a generic step to obtain information on the topology of the complex. In the future, it may become possible to define the molecular interactions down to interacting surfaces through the use of mass spectrometry.

(42) Bennett, K. L.; Kussmann, M.; Björk, P.; Godtfredsen, M.; Godzwon, M.; Sørensen, P.; Roepstorff, P., submitted.

ACKNOWLEDGMENT

We thank Bernhard Küster, Akhilesh Pandey, and Keiryn Bennett for critically reading the manuscript. J.R. acknowledges the assistance received from Berthold Rutz and other members of the group of Bertrand Seraphin at the EMBL in establishing the yeast work in M.M.'s laboratory. J.R. was supported by a Kecule scholarship of the Fonds der Chemischen Industrie. Work in M.M.'s laboratory is funded by a grant from the Danish National Research Foundation.

Received for review September 20, 1999. Accepted November 11, 1999.

AC991081O