

Structural characterization by cross-linking reveals the detailed architecture of a coatomer-related heptameric module from the nuclear pore complex

Yi Shi^{1*}, Javier Fernandez-Martinez^{2*}, Elina Tjioe^{3*}, Riccardo Pellarin^{3*}, Seung Joong Kim^{3*}, Rosemary Williams², Dina Schneidman-Duhovny³, Andrej Sali³, Michael P. Rout², and Brian T. Chait¹

¹*Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, New York, NY 10065, USA*

²*Laboratory of Cellular and Structural Biology, The Rockefeller University, New York, NY 10065, USA*

³*Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, California Institute for Quantitative Biosciences, Byers Hall, 1700 4th Street, Suite 503B, University of California, San Francisco, San Francisco, CA 94158, USA*

**These authors contributed equally to this work as co-first authors.*

Corresponding authors

Brian T. Chait, Box 170, Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, New York, NY 10065, USA.

tel: +1 212 327 8849; fax: +1 212 327 7547; chait@rockefeller.edu

Michael P. Rout, Box 213, Laboratory of Cellular and Structural Biology, Rockefeller University, 1230 York Avenue, New York, NY 10021, USA.

tel: +1 212 327 8135; fax: +1 212 327 7193; rout@rockefeller.edu

Andrej Sali, UCSF MC 2552, Byers Hall at Mission Bay, Suite 503B, University of California, San Francisco, 1700 4th Street, San Francisco, CA 94158, USA

tel: +1 415 514 4227; fax: +1 415 514 4231; sali@salilab.org

Running Title: Integrative modeling of yeast Nup84 complex by cross-linking

ABBREVIATIONS

CX-MS: chemical cross-linking with mass spectrometric readout

LC-MS: liquid chromatography–mass spectrometry

DSS: disuccinimidyl suberate

EDC: 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide

NPC: nuclear pore complex

EM: electron microscopy

Nup: nucleoporin

VCC: vesicle coating complexes

dRMSD: distance root-mean-square deviation

SUMMARY

Most cellular processes are orchestrated by macromolecular complexes. However, structural elucidation of these endogenous complexes can be challenging because they frequently contain large numbers of proteins, are compositionally and morphologically heterogeneous, can be dynamic, and are often of low abundance in the cell. Here, we present a strategy for structural characterization of such complexes, which has at its center chemical cross-linking with mass spectrometric readout (CX-MS). In this strategy, we isolate the endogenous complexes using a highly optimized sample preparation protocol and generate a comprehensive, high-quality cross-linking dataset using two complementary cross-linking reagents. We then determine the structure of the complex using a refined integrative method that combines the cross-linking data with information generated from other sources, including electron microscopy, X-ray crystallography, and comparative protein structure modeling. We applied this integrative strategy to determine the structure of the native Nup84 complex – a stable hetero-heptameric assembly (~600 kDa), sixteen copies of which form the outer rings of the 50 MDa nuclear pore complex (NPC) in budding yeast. The unprecedented detail of the Nup84 complex structure reveals previously unseen features in its pentameric structural hub and provides information on the conformational flexibility of the assembly. These additional details further support and augment the protoatomer hypothesis, which proposes an evolutionary relationship between vesicle coating complexes and the NPC, and indicates a conserved mechanism by which the NPC is anchored in the nuclear envelope.

INTRODUCTION

Macromolecular complexes are the building blocks that drive virtually all cellular and biological processes. In each eukaryotic cell, there exist many hundreds such protein complexes (1-3), the majority of which are still poorly understood in terms of their structures, dynamics, and functions. The classical structure determination approaches of nuclear magnetic resonance (NMR), X-ray crystallography, and electron microscopy (EM) remain challenged when determining the high-resolution structures of large, dynamic, and flexible complexes in a living cell (4). Thus, additional robust and rapid methods are needed, ideally working in concert with these classical approaches, to allow the characterization of macromolecular assemblies in the highest structural and functional detail.

Integrative modeling approaches help address this need, providing powerful tools for determining structures of endogenous protein complexes (5, 6), by relying on the collection of an extensive experimental dataset, preferably coming from diverse sources (both classical and new) and different levels of resolution. These data are translated into spatial restraints that are used to calculate an ensemble of structures by satisfying the restraints, which in turn can be analyzed and assessed to determine its precision and estimate its accuracy (5, 7). A major advantage of this approach is that it readily integrates structural data from different methods and a wide range of resolutions, spanning from a few angstroms to dozens of nanometers. This strategy has been successfully applied to a number of protein complexes (8-16). However, it has proven difficult and time-consuming to generate a sufficient number of accurate spatial

restraints to enable high-resolution structural characterization; thus, the determination of spatial restraints currently presents a major bottleneck for widespread application of this integrative approach. An important step forward therefore is the development of technologies for collecting high-resolution and information-rich spatial restraints in a rapid and efficient manner, ideally from endogenous complexes isolated directly from living cells.

Chemical cross-linking with mass spectrometric readout (CX-MS) (17, 18) has recently emerged as an enabling approach for obtaining residue-specific restraints on the structures of proteins and protein complexes (19-25). In a CX-MS experiment, the purified protein complex is chemically conjugated by a functional group-specific cross-linker, followed by proteolytic digestion and analysis of the resulting peptide mixture by mass spectrometry (MS). However, due to the complexity of the peptide mixtures and low abundance of most of the informative cross-linked species, comprehensive detection of these cross-linked peptides has proven challenging. This challenge increases substantially when studying endogenous complexes of modest to low abundance, which encompasses the great majority of assemblies in any cell (26, 27). In addition, since most cross-linkers used for CX-MS target primary amines, comprehensive detection of cross-links is further limited by the occurrence of lysine, which constitutes only ~6% of protein sequences (<http://prowl.rockefeller.edu/aainfo/struct.htm>), although these lysine residues are generally present on protein surfaces. Use of cross-linkers with different chemistries and reactive groups, especially towards abundant residues, would increase the cross-

linking coverage and could be of great help for downstream structural analysis (28).

The nuclear pore complex (NPC) is one of the largest protein assemblies in the cell and is the sole mediator of macromolecular transport between the nucleus and the cytoplasm. The NPC is formed by multiple copies of ~30 different proteins termed nucleoporins (Nups) that are assembled into discrete subcomplexes (8, 29). These building blocks are arranged into eight symmetrical units called spokes that are radially connected to form several concentric rings. The outer rings of the NPC are mainly formed by the Nup84 complex (a conserved complex, termed the Nup107-Nup160 complex in vertebrates). In budding yeast, the Nup84 complex is an essential, Y-shaped assembly of ~600 kDa that is formed by seven nucleoporins (Nup133, Nup120, Nup145c, Nup85, Nup84, Seh1, and Sec13 in *Saccharomyces cerevisiae*) (30). The Nup84 complex has been shown to have a common evolutionary origin with vesicle coating complexes (VCCs), such as COPII, COPI, and clathrin (31, 32), but the evolutionary relationships between these VCCs have not been fully delineated. The Nup84 complex has been extensively characterized; several of its components have been analyzed by X-ray crystallography (33, 34), its overall shape has been defined by negative-stain electron microscopy (14, 30, 35, 36) and recently efforts were made to define the protein contacts in the Nup84 complex by CX-MS in human (35) and a thermophilic fungus (37). Finally, we recently used an integrative modeling approach, combining domain mapping, negative-stain electron microscopy (38), and publicly available crystal structures to generate a medium resolution map of the native Nup84 complex (14). However, despite all these efforts, the fine features of the complex, and in

particular the intricate domain orientations and contacts within the complex's hub, remain poorly described.

To address these issues, we present here an optimized CX-MS strategy for robust and in-depth structural characterization of endogenous protein complexes. To test the strategy, we generated a comprehensive high-quality CX-MS dataset on the endogenous Nup84 complex using two complementary cross-linkers - disuccinimidyl suberate (DSS) and 1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC). Using the resulting cross-linking restraints together with other sources of information (including electron microscopy, X-ray crystallography, and comparative modeling), we computed a detailed structure of the endogenous Nup84 complex. In addition to providing the overall architecture of the yeast Nup84 complex, the resulting structure reveals the previously unknown architecture of its pentameric structural hub. Our results demonstrate that the present approach provides a robust framework for the standardized generation and use of CX-MS spatial restraints towards the structural characterization of endogenous protein complexes.

EXPERIMENTAL PROCEDURES

Purification and chemical cross-linking of the endogenous Nup84 complex

To purify the native *Saccharomyces cerevisiae* Nup84 complex, we used a procedure that we have optimized over the course of several years (14) (Supplemental Experimental Procedures). The natively eluted Nup84 complex (200 μ L) was cross-linked by adding isotopically labeled Di-Succinimidyl-Suberate (d0:d12= 1:1, Creative Molecules) to yield a final concentration of 1 mM and incubated for 45 minutes at 25°C and 750 rpm of agitation. The reaction was then quenched using a final concentration of 50 mM ammonium bicarbonate.

In the case of cross-linking using 1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC, Pierce), the sample was equilibrated and eluted in EDC cross-linking buffer (20 mM MES pH 6.5, 500 mM NaCl, 2 mM MgCl₂, 0.1% CHAPS, 1 mM DTT); EDC was added to the sample to yield a final concentration of 25 mM and N-hydroxysulfosuccinimide (Sulfo-NHS, Pierce) was added to yield a final concentration of 0.5 mM (*i.e.*, 2% molar ratio with respect to EDC). The sample was incubated for 45 minutes at 25°C and 750 rpm of agitation. After the incubation, Tris-HCl was added to the final concentration of 50 mM and pH 8.0, and β -mercaptoethanol was added to the final concentration of 20 mM, followed by incubation at 25°C for 15 minutes and 750 rpm of agitation to quench the reaction.

The cross-linked samples were either directly processed for in-solution digestion or precipitated using 90% cold methanol and resuspended in SDS-PAGE loading buffer for

in-gel separation and digestion.

Proteolytic digestion of chemically cross-linked Nup84 complex

After cross-linking, the complex was reduced by 10 mM tris-(2-carboxyethyl)-phosphine (TCEP, Invitrogen) at 55°C, cooled to room temperature (RT) and alkylated by 20 mM iodoacetamide for 20 minutes in the dark. The cross-linked complex was digested either in-solution or in-gel with trypsin to generate cross-linked peptides. For in-solution digestion, ~10-20 µg of purified complex was digested using 0.5 µg trypsin (Promega) in 1 M urea and 0.1% Rapigest (Waters). After overnight (12-16 hours) incubation, an additional ~0.3 µg aliquot of trypsin was added to the digest and incubated for a further 4 hours. The resulting proteolytic peptide mixture was acidified, and centrifuged at 13,000 rpm for 5 minutes. The supernatant peptides were collected and desalted using a C18 cartridge (Sep-Pak, Waters), lyophilized in protein LoBind tubes (Eppendorf) and fractionated by peptide size exclusion chromatography (SEC) (below). For in-gel digestion, ~10-20 µg purified complex was heated at 75°C in 1X SDS loading buffer for 10 minutes. The sample was cooled at RT for cysteine alkylation and separated by electrophoresis in a 4-12% Bis-Tris SDS-PAGE gel (Invitrogen). The gel region above 220 kDa was sliced, crushed into small pieces, and digested in-gel by trypsin. After extraction and purification, the resulting proteolytic peptides were dissolved in 20 µL of a solution containing 30% acetonitrile (ACN) and 0.2% formic acid (FA), and fractionated by peptide SEC (Superdex Peptide PC 3.2/30, GE Healthcare) using off-line HPLC separation with an autosampler (Agilent Technologies). Two to four SEC fractions covering the molecular mass range of ~2.5 kD to ~8 kD were collected and analyzed by

LC/MS.

Mass spectrometric analysis of cross-linked peptides

To characterize the composition of our Nup84 complex preparation, ~1 µg of the endogenously purified complex was in-gel digested and 1/50 portion of the purified proteolytic peptides (corresponding to ~20 ng of the purified complex) were loaded onto an EASY-Spray column (15 cm x 75 µm ID, 3µm), followed by an analysis with an Orbitrap Fusion mass spectrometer coupled on-line to the EASY-nLC 1000 nano-LC system (Thermo). A 10 min LC gradient (8% B to 55% B, 0-7 minutes, followed by 55% B to 100% B, 7-10 min, where mobile phase A consisted of 0.1% formic acid and B consisted of 100% acetonitrile in 0.1% formic acid). A top 10 (high-low) method was used where the precursors ($m/z = 300, 1,700$) were measured by the orbitrap, isolated in the quadrupole mass filter (isolation window 1.8 Th), fragmented within the HCD cell (HCD normalized energy 28), and the product ions were analyzed in the low resolution ion trap. Other instrumental parameters include a flow rate of 300 nl/min, spray voltage of 1.7 kv, S lenses (35%), AGC targets of 5×10^5 (orbitrap) and 1×10^4 (ion trap).

The raw data on the immunoprecipitation was converted to an mzXML file (by MM file conversion 3) and searched online by X! Tandem (<http://prowl.rockefeller.edu/tandem>). Database search parameters include mass accuracies of MS1 < 10 ppm and MS2 < 0.4 Da, cysteine carbamidomethylation as a fixed modification, methionine oxidation, N-terminal acetylation, and phosphorylation (at S, T, and Y) as variable modifications. A maximum of one trypsin missed-cleavage sites was allowed. The seven Nup84

components were identified as the top 7 hits by spectral counting. The results are provided in Table S7.

For cross-link identification, the purified peptides were dissolved in the sample loading buffer (5% MeOH, 0.2% FA) and loaded onto a self-packed PicoFrit® column with integrated electrospray ionization emitter tip (360 O.D, 75 I.D with 15 µm tip, New Objective). The column was packed with 8 cm of reverse-phase C18 material (3 µm porous silica, 200 Å pore size, Dr. Maisch GmbH). Mobile phase A consisted of 0.5% acetic acid and mobile phase B of 70 % ACN with 0.5 % acetic acid. The peptides were eluted in a 150-minutes LC gradient (8% B to 46% B, 0-118 minutes, followed by 46% B to 100% B, 118-139 minutes, and equilibrated with 100% A until 150 minutes) using an HPLC system (Agilent), and analyzed with a LTQ Velos Orbitrap Pro mass spectrometer (Thermo Fisher). The flow rate was ~200 nL/minute. Spray voltage was set at 1.9-2.3 kV. The capillary temperature was 275°C and ion transmission on Velos S lenses was set at 35%. The instrument was operated in the data-dependent mode, where the top eight-most abundant ions were fragmented by higher energy collisional dissociation (HCD) (39) (HCD energy 27-29, 0.1 millisecond activation time) and analyzed in the Orbitrap mass analyzer. The target resolution for MS1 was 60,000 and for MS2 it was 7,500. Ions (370-1700 m/z) with charge state of > 3 were selected for fragmentation. A dynamic exclusion of (15 s / 2 / 55 s) was used. Other instrumental parameters include: “lock mass” at 371.1012 Da, monoisotopic mass selection turning off, the mass exclusion window +/-1.5 Th, and the minimal threshold of 5,000 to trigger an MS/MS event. Ion trap accumulation limits (precursors) were 1×10^5 and 1×10^6

respectively for the linear ion trap and Orbitrap. For MS2, the Orbitrap ion accumulation limit was 5×10^5 . The maximal ion injection time for the LTQ and Orbitrap were 100 milliseconds and 500 - 700 milliseconds, respectively.

The raw data were transformed to MGF (mascot generic format) by pXtract 2.0 and searched by pLink (version 1.16) (40) using a target-decoy search strategy with a concatenated FASTA protein sequence database containing the seven subunits of the Nup84 complex (accession numbers are YGL092W, YDL116W, YJR042W, YKL057C, YLR208W, and YGL100W) and bovine serum albumin (BSA). We also included the BSA sequence for target-decoy database search to quickly pre-filter the false positive identifications containing the BSA sequence. An initial MS1 search window of 5 Da was allowed to cover all isotopic peaks of the cross-linked peptides. The data were automatically filtered using a mass accuracy of MS1 ≤ 10 ppm (parts per million) and MS2 ≤ 20 ppm of the theoretical monoisotopic (A0) and other isotopic masses (A+1, A+2, A+3, and A+4) as specified in the software. Other search parameters include cysteine carbamidomethylation as a fixed modification, methionine oxidation, and protein N-terminal methionine cleavage as a variable modification. A maximum of two trypsin missed-cleavage sites were allowed. The initial search results were obtained using a 5% false discovery rate (FDR) - a default parameter estimated by the pLink software (40). We then treated the 5% expected FDR as an initial (permissive) filter of the raw data. We then manually applied additional stringent filters to remove potential false positive identifications from our dataset. For positive identifications, both peptide chains should contain at least five amino acid residues. For both peptide chains, the

major MS/MS fragmentation peaks must be assigned and follow a pattern that contains a continuous stretch of fragments. The appearance of dominant fragment ions N-terminal to proline and C-terminal to aspartic acid and glutamic acid for arginine-containing peptides was generally expected (41, 42). The precise cross-linking site cannot be determined for ~20% EDC cross-links due to the appearance of consecutive aspartic and glutamic acid in the tryptic peptide sequences (Supplemental Figure S5C and Supplemental Table S2). Moreover, because cross-linking by EDC (hydrolysis) resembles peptide bond formation, intra-molecular cross-link candidates (directly identified by the software) that are composed by two adjacent tryptic peptide sequences of the same protein are likely to be a single, missed cleavage peptide (Supplemental Figure S5D). Such ambiguous EDC cross-linking identifications were eliminated from further consideration. In essence, these additional filters represent a standard procedure for eliminating false positives at the cost of removing some true positives. The net result is a more conservative list of cross-linking restraints that are less likely to manifest as artefacts in the resulting structural models. The cross-link maps (Figure 2, A and C) were generated using AUTOCAD (Autodesk, Inc., educational version).

Molecular architecture of the endogenous Nup84 complex revealed by integrative modeling

Our integrative structure modeling proceeds through four stages (7, 13, 14, 43) (Figure 4): (1) gathering of data, (2) representation of subunits and translation of the data into spatial restraints, (3) configurational sampling to produce an ensemble of models that satisfies the restraints, and (4) analysis and assessment of the ensemble. The modeling

protocol (*i.e.*, stages 2, 3, and 4) was scripted using the *Python Modeling Interface* (<https://github.com/salilab/pmi>), version *be72c15*, a library to model macromolecular complexes based on our open source *Integrative Modeling Platform* (IMP) package (<http://salilab.org/imp/>), version *829c3f0* (44). Files containing the input data, scripts, and output models are available at <http://salilab.org/nup84>.

Stage 1: Gathering of data

163 unique DSS and 104 EDC cross-linking peptides were identified by MS (Supplemental Tables S1 and S2; Supplemental Figures S9 and S10; 20% of the EDC cross-links were not determined unambiguously, resulting in 123 possible EDC cross-links). The atomic structures for some of the yeast Nup84 complex components and their close homologs have been previously determined by X-ray crystallography (Supplemental Figure S3) (31, 45-54). In addition, structurally defined remote homologs (PDB codes 2QX5 (chain B) and 4LCT (chain A)) (55, 56) were detected for the C-terminal domain in Nup85 by HHpred (57, 58) (Supplemental Figure S3). Secondary structure and disordered regions were predicted by PSIPRED (59, 60) and DISOPRED (61), respectively (Supplemental Figure S3). Domain mapping data (14), an EM class average (14), and a density map from single particle EM reconstruction of the Nup84 complex (36) were also considered.

Stage 2: Representation of subunits and translation of the data into spatial restraints

The domains of the Nup84 complex subunits were represented by beads of varying

sizes, arranged into either a rigid-body or a flexible string, based on the available crystallographic structures and comparative models (Supplemental Figure S3). In a rigid-body, the beads have their relative distances constrained during configurational sampling, while in a flexible string the beads are restrained by the sequence connectivity, as described later in this section.

To maximize computational efficiency while avoiding using too coarse a representation, we represented the Nup84 complex in a multi-scale fashion, as follows.

First, the crystallographic structures of each Nup84 complex domain were coarse-grained using two categories of resolution, where beads represent either individual residues or up to 10-residue segments. For the 1-residue bead representation, the coordinates of a bead are those of the corresponding C_{α} atoms. For the 10-residue bead representation, the coordinates of a bead are the center of mass of all atoms in the corresponding consecutive residues (each residue is in one bead only). The crystallographic structures covered 54% of the residues in the Nup84 complex.

Second, for predicted non-disordered domains of the remaining sequences, comparative models were built with MODELLER 9.13 (62) based on the closest known structure detected by HHPred (57, 58) and the literature (Supplemental Figure S3). Similarly to the X-ray structures, the modeled regions were also coarse-grained using two categories of resolution, resulting in the 1-residue and 10-residue bead representations. The comparative models covered 30% of the residues in the Nup84

complex.

Finally, the remaining regions without a crystallographic structure or a comparative model (*i.e.*, regions predicted to be disordered or structured without a known homolog) were represented by a flexible string of beads corresponding to up to 20 residues each. The residues in these beads correspond to 16 % of the Nup84 complex.

To improve the accuracy and precision of the model ensemble obtained by satisfaction of spatial restraints (below), we also imposed crystallographic interface constraints based on 3 crystallographically defined interfaces, including Nup85¹²³⁻⁴⁶⁰ - Seh1 (31, 53), Nup145c¹⁴⁵⁻¹⁸¹ - Sec13 (49, 50), and Nup145c¹⁴⁵⁻¹⁸¹ - Nup84¹⁻⁴⁸⁸ (49, 50). The 3 constrained dimers were simply represented as rigid bodies. These constraints are justified because their absence only decreases the precision of the model ensemble, but does not significantly change the average model (Supplemental Figure S7).

With this representation in hand, we next encoded the spatial restraints based on the information gathered in Stage 1, as follows.

First, the collected DSS and EDC cross-links were used to construct a Bayesian scoring function that restrained the distances spanned by the crosslinked residues (63), while taking into account the ambiguity of some cross-linked residue identifications (Supplemental Experimental Procedures). The cross-link restraints were applied to the 1-residue bead representation for the X-ray structures and comparative models as well

as to the 20-residue bead representation for the remaining regions.

Second, the excluded volume restraints (7) were applied to the 10-residue bead representation for X-ray structures and comparative models as well as to the 20-residue bead representation for the remaining regions. The excluded volume of each bead was defined using the statistical relationship between the volume and the number of residues that it covered (7, 64).

Third, we applied the sequence connectivity restraint, using a harmonic upper-bound function of the distance between consecutive beads in a subunit, with a threshold distance equal to four times the sum of the radii of the two connected beads. The bead radius was calculated from the excluded volume of the corresponding bead, assuming standard protein density (7, 14, 64).

Finally, the EM 2D restraint (65) was imposed on the highest resolution representation of each subunit, using the cross-correlation coefficient between the EM class average (14) and the best-matching projection of the model as the score (Supplemental Experimental Procedures).

In summary, all information gathered in Stage 1 is encoded into a Bayesian scoring function (66) (Supplemental Experimental Procedures). The likelihood function reflects the cross-linking data, while the prior depends on the sequence connectivity, excluded volume, and EM 2D restraint. Most of the remaining information (crystallographic

structures of the subunits, their homologs, and the 3 dimeric interfaces) is included in the representation, while the domain deletion data and the density map from single particle EM reconstruction were used only for validating our final model. See the scripts at <http://salilab.org/nup84> for details and parameter values.

Stage 3. Sampling the configurations

Structural models of the Nup84 complex were computed using Replica Exchange Gibbs sampling, based on the Metropolis Monte Carlo algorithm (66). The Monte Carlo moves included random translation and rotation of rigid bodies (up to 2 Å and 0.04 radians, respectively), and random translation of individual beads in the flexible segments (up to 2 Å). 64 replicas were used, with temperatures ranging between 1.0 and 2.5. 20 independent sampling calculations were performed, each one starting with a random initial configuration. A model was saved every 10 Gibbs sampling steps, each consisting of a cycle of Monte Carlo steps that moved every rigid body and flexible bead once. The sampling produced a total of 15,000 models from the 20 runs. Models that have a score lower than 300 were included in an ensemble of 6,520 solutions for subsequent analysis. The entire sampling procedure took approximately 2 weeks on a cluster of 1,280 CPUs.

Stage 4. Analysis and assessment of the ensemble

First, the thoroughness of configurational sampling was assessed by comparing a subset of 3,413 solutions from runs 1-10 to another subset of 3,107 solutions from runs 11-20. Each subset of solutions was converted into a density map that specifies how

often are grid points of the map occupied by a given protein (the 'localization density map'), using VMD (67) (Supplemental Figure S6). The localization density map of a subunit was contoured at the threshold that results in 2.5 times its volume estimated from sequence (Supplemental Table S6). Importantly, the two localization density maps were similar to each other, demonstrating that the Monte Carlo algorithm likely sampled well all solutions that satisfy the input restraints. The final localization density maps of the Nup84 subunits and the whole complex were computed from the complete ensemble of the 6,520 solutions (Figure 6, A and B).

Second, the ensemble of solutions was assessed in terms of how well they satisfy the data from which they were computed, including the cross-links as well as the excluded volume, sequence connectivity, and the EM 2D restraints. We validated the ensemble of solutions against each of 163 DSS and 123 EDC cross-links (Supplemental Tables S1 and S2; Supplemental Figures S9 and S10); a cross-link restraint was considered to be satisfied by the ensemble if the median distance between the surfaces of the corresponding beads was smaller than a distance threshold of 35 and 25 Å for the DSS and the EDC cross-links, respectively. The shape implied by the EM class average (14) was satisfied by the ensemble if the average EM 2D cross-correlation was higher than 0.8. The excluded volume and sequence connectivity were considered to be satisfied by an individual solution if their combined score was less than 20.

Third, to quantify the precision, the solutions were grouped by C_{α} root-mean-square deviation (RMSD) quality-threshold clustering (68), based on either the hub subunits

(Nup145c, Nup85, Seh1, Sec13, and the C-termini of Nup120) or all subunits, using the C_{α} RMSD threshold of 40 Å. The precision of each cluster was calculated as the average C_{α} distance root-mean-square deviation (dRMSD) (69) between the individual solutions and the centroid solution, defined as the solution with the minimal sum of the dRMSD's to the other solutions in the cluster (Supplemental Tables S4 and S5); distance difference terms involving a distance larger than 60 Å were omitted from the dRMSD calculation. The localization density maps of the clusters were computed as described above (Figure 6C).

Finally, the proximities of any two residues in each cluster were measured by their relative 'contact frequency', which is defined by how often the two residues contact each other in the cluster (7); in the multi-scale representation, a pair of residues are in contact when the distance between the centers of the corresponding highest resolution beads is less than 30 Å (Supplemental Figure S4).

Correlation between the number of cross-links and the accuracy of dimer models

As a benchmark, we modeled the Sec13-Nup145c dimer of known structure (Figure 5A) using up to 10 DSS and 5 EDC inter-molecular cross-links. Each of Sec13 and Nup145c was represented as a rigid-body; we used the same representation, scoring function, and sampling method as described above. First, we modulated the sparseness of the cross-linking data, by varying the number of DSS cross-links, from 1 to 10, without considering the EDC cross-links. For each case, up to 10 different random cross-link sets were chosen. We also modeled the dimer with EDC cross-links alone as well as

with all EDC and DSS cross-links. A total of 93 independent modeling runs were performed, producing 30,000 models each. To assess the accuracy of the solutions, C_{α} dRMSD (Figure 5B-C and Supplemental Figure S2) was calculated between every produced model and the crystallographic dimer (PDB code 3IKO (50)).

RESULTS

Development of a workflow for CX-MS of endogenous protein complexes

Our goal was to develop a robust and sensitive methodology for determining chemical cross-links for structural characterization of endogenous protein complexes. The overall strategy is summarized in Figure 1, as follows. The endogenous protein assembly is isolated from a cryogenically milled whole cell lysate by single-step affinity purification using antibody-coupled magnetic beads (14, 70). This approach produces highly enriched, relatively homogenous complexes with high recovery, in buffers that can be optimized for diverse downstream applications (14, 70). The complex is natively eluted from the affinity matrix and cross-linked in solution using two complementary chemical cross-linkers, DSS and EDC. DSS is primary amine specific (lysines and amino termini of proteins), with a spacer arm of 11.4 Å (α -carbon), while EDC cross-links amines to carboxylic acids (aspartic acids, glutamic acids, and carboxyl termini of proteins) and is generally considered as a “zero” length cross-linker. The relatively high prevalence of carboxylic acid and lysine residues in the protein sequences (~12% and ~6%, respectively, <http://prowl.rockefeller.edu/aainfo/struct.htm>), with these charged residues generally present on protein surfaces) makes EDC a useful complement to the strictly amine specific cross-linkers. The cross-linked complex is digested in solution as well as separated by SDS-PAGE for in-gel proteolysis. In both cases, the digested peptides are then fractionated by peptide size exclusion chromatography (71) and peptides corresponding to the large molecular weight fractions (~2.5 - 8 kD) are analyzed by LC-MS. Putative cross-linked peptides are identified by pLink (40), after which the spectra are filtered using additional stringent criteria and manually verified (Experimental

Procedures). Finally, each of the resulting cross-links is translated into a spatial restraint for integrative modeling (Figure 4).

Chemical cross-linking and MS analysis of the Nup84 complex

The workflow described above was applied to the endogenous heptameric Nup84 complex. Both DSS and EDC cross-linking reactions were optimized for the downstream CX-MS analysis as follows (also see Supplemental Figure S1). We controlled the cross-linking reaction such that the majority of the proteins migrated as diffusely staining regions into the uppermost part (>220 kD) of an SDS-PAGE gel. The protein samples isolated from these uppermost regions of a gel were enriched in cross-linked peptides. We generated a high-quality cross-linking dataset that included 163 unique DSS and 104 EDC cross-linking peptides (Supplemental Tables S1 and S2; Supplemental Figures S9 and S10). The overall connectivity patterns of the DSS and EDC cross-links were similar (Figure 2, A and B), showing that they provide complementary information on similar conformers. Even though 15-20% of the cross-linked peptides were in the high molecular mass range (4,500 - 7,000 Da), use of high-resolution and high mass accuracy MS allowed us to pinpoint the cross-linking sites for virtually all the DSS cross-links and the majority of EDC cross-links, including those for large peptide species (Figure 2, C-F, and Supplemental Tables S1 and S2). We were unable to unambiguously locate the conjugation sites for ~20% of the EDC cross-links (Supplemental Table S2), because many result from peptides containing consecutive and/or adjacent carboxylic acids (Supplemental Table S2 and Supplemental Figure S5C). These ambiguities were accounted for in our modeling calculations.

Importantly, all previously identified interfaces of the complex (using a variety of different approaches including X-ray crystallography, domain deletions, and affinity purifications (14, 33)) were recapitulated by the cross-links (Figure 2, A and C). As expected, we observed a high correlation between the coverage of cross-links and the size of interfaces within the complex; while the majority of the cross-links were mapped to the large, intricate yet previously poorly described region of the pentameric structural hub (Nup145c, Nup85, Seh1, Sec13, and the C-termini of Nup120), relatively fewer were identified in the stalk of the complex (*e.g.*, between Nup84 and Nup133 (46, 47)) where smaller interfaces were expected.

To further evaluate the fidelity of the cross-linking data, we mapped the Euclidean C_{α} - C_{α} distances between the cross-linked residues onto several domains of the Nup84 complex proteins where high-resolution crystal structures have previously been determined (49, 50) (Figure 3A). The distributions of distance differences between either DSS or EDC cross-links and the crystal structure are narrow, and are notably distinct from the distributions of randomly connected residues of the same type (Figure 3A). All the measured DSS cross-links fall within the expected maximum reach threshold of ~ 30 Å (72). Although the great majority of the EDC cross-links fall within the expected reach threshold of 17 Å (= 6 Å for the lysine side chain + 5 Å for the carboxylic acid side chain + 6 Å for flexibility of backbones (20)), interestingly, the distances for $\sim 30\%$ of EDC cross-links are longer than expected (Figure 3A). These discrepancies may arise from differences in flexibility of the subunits at the optimum pHs for EDC

versus DSS, the longer times of incubation for the EDC experiments, or some other as yet unknown cause. Nevertheless, the EDC and DSS cross-links provide mutually confirmatory as well as complementary spatial information (Figure 3B).

Correlation between the number of cross-links and the accuracy of dimer models

We estimated the number of cross-links required to model a subunit dimer with required accuracy. The estimate was possible because we have a large number of experimentally determined cross-links for several heterodimers of known atomic structure. We chose to focus on the Nup145c-Sec13 dimer (49, 50, 73) because of the abundance of inter-molecular DSS and EDC cross-links identified for this dimer. By varying the number of DSS cross-linking restraints used to model the dimer (Figure. 5), we found that at least 4 or 5 cross-links (either DSS or EDC) are needed to determine the dimer structure with an accuracy better than 10 Å C_α dRMSD. Increasing the number of cross-links beyond 5 (up to 15), irrespective of their type, does not further improve the accuracy, although it increases the precision of the resulting ensemble of solutions (Figure 5C and Supplemental Figure S2).

Molecular architecture of the endogenous Nup84 complex revealed by integrative modeling

We computed the configuration of the seven subunits of the budding yeast Nup84 complex (Figure 6), using crystallographic structures of 7 constituent domains (Supplemental Figure S3) as well as 163 DSS and 123 EDC cross-links (Supplemental Tables S1 and S2; Supplemental Figures S9 and S10) and one negative-stain electron

microscopy (EM) class average (14) (Figure 4). To improve the precision of the resulting models, the 3 crystallographically defined interfaces, namely Nup85-Seh1 (31, 53), Nup145c-Sec13 (49, 50), and Nup145c-Nup84 (49, 50), were included as constraints in the modeling calculation; the constrained dimers were simply represented as rigid bodies in the configurational sampling.

The molecular architecture of the Nup84 complex was computed from an ensemble of 6,520 solutions, shown as a localization density map representing probability of any volume element being occupied by a given protein (Figure 6A). The sampling procedure was thorough, as indicated by the similarity of two ensembles of solutions independently calculated using the same cross-link dataset (Supplemental Figure S6).

Next, we validated the solutions against a previously published molecular architecture of the Nup84 complex. The current solutions are in agreement with the Nup84 complex structure determined primarily using the domain mapping data and the EM class average in our earlier study (14) (Supplemental Figure S8) as well as the density map from the single particle EM reconstruction (36) (Supplemental Figure S8).

We also validated the ensemble of solutions against the data used to compute it. First, the ensemble satisfied 86.5% of the DSS cross-links and 83.6% of the EDC cross-links (Experimental Procedures; Supplemental Figure S4 and Supplemental Table S3). Second, 99% of the 6,520 solutions satisfied the excluded volume and sequence connectivity restraints (Experimental Procedures). Finally, the solutions also fit the EM

class average, with the average cross-correlation coefficient of 0.9.

Next, we quantified the precision of the ensemble of solutions (Experimental Procedures). The clustering analysis identified 2 dominant clusters of similar structures in the hub, including 1,257 and 1,010 solutions, respectively (Figure 6C-D and Supplemental Table S4). The precision of the ensemble is sufficient to pinpoint the locations and orientations of the constituent proteins and even domains (Figure 6 and Supplemental Tables S4 and S5), as described in Discussion. The hub region was determined at the highest precision of 12.7 Å (Supplemental Table S4). Furthermore, even the independent modeling calculation without crystallographic interface constraints (Supplemental Figure S7) was able to recapitulate the structures of the Nup145c-Sec13, Nup85-Seh1, and Nup84-Nup145c crystallographic interfaces with an accuracy of 4.0, 12.0, and 7.5 Å, respectively, in the most-populated cluster 1 (Table 1). These accuracies demonstrate the quality of the cross-link data as well as validate the application of the 3 crystallographic interfaces as constraints. Notably, the accuracy of the Nup145c-Sec13 dimer (C_{α} dRMSD of 4.0 Å) in the model of the entire complex is higher than the accuracy of modeling the Nup145c-Sec13 dimer on its own (C_{α} dRMSD of 6.5 Å, Figure 5B-C). This observation underscores the synergy between orthogonal data, thus demonstrating the premise of integrative modeling.

Finally, the proximities of any two residues within each of the two dominant solution clusters were measured by their relative 'contact frequency', which is defined by how often the two residues contact each other in the cluster (7). The contact frequencies are

in remarkable agreement with the cross-link dataset (Supplemental Figure S4).

DISCUSSION

Integrative structural characterization of the endogenous Nup84 complex based on CX-MS data

We present here an optimized CX-MS workflow for integrative structural characterization of native protein complexes. Importantly, this pipeline generates structures of complexes with near-atomic resolution and in a fraction of the time that, in our experience, previous approaches have taken to achieve similar accuracy. To achieve this goal, we have developed an improved CX-MS pipeline for robust and comprehensive analysis of endogenous protein complexes available in limited amounts (10-20 μ g), as demonstrated for both the Nup84 complex and the SEA complex (in an accompanying paper (74)). Our approach also allows rapid and efficient integration of structural information from various sources at different levels of resolution, including complementary cross-linking restraints, crystallographic structures, comparative models, electron microscopy class averages and density maps, as well as affinity purification data. This approach also provides information concerning the heterogeneity and structural flexibility of endogenous complexes, which have been poorly investigated before.

We found that at least 4 or 5 cross-links between two crystallographic structures is sufficient to relatively accurately model the corresponding dimer (Figure 5C and Supplemental Figure S2). Additional cross-links further increase the precision of the final solutions, although they only marginally improve their accuracy. We demonstrate the benefits of using two complementary, commercially available cross-linkers (DSS

and EDC), targeting amine and carboxylic groups, respectively. Further efforts are needed to design cross-linkers of varying length and/or type (28).

We empirically observed that in-gel digestion generally yielded more cross-links than in-solution digestion (Fig. S12). We speculate that one reason for this higher yield is the presence of SDS in the gel, which greatly helps to denature the otherwise rigidified cross-linked complex; efficient proteolysis by trypsin thus becomes of great importance for a successful experiment. In addition, the use of a gel makes it easy to clean up contaminants such as detergents and elution peptides.

Conformational heterogeneity of endogenous protein complex inferred from non self-consistency of structural information

Like many native protein assemblies, the Nup84 complex has been shown to exist in multiple conformational states (14, 30, 35, 36). We noticed that not all cross-links could readily fit into our ensemble of structures calculated by integrative modeling (Supplemental Figure S4 and Supplemental Table S3). We suggest that such inconsistency largely reflects the heterogeneity, flexibility, and/or disorder, as seen previously by EM (14, 30, 35, 36). Examples include the N-terminus of Nup85 and Nup145c, which are located in the Nup85-Seh1 arm that has been suggested to be flexible based on negative-stain EM data (14). Other notable examples are the Nup133(253)-Nup84(340) and Nup133(936)-Nup133(392) cross-links, connecting the β -propeller N-terminal region of Nup133 to either the N-terminal region of Nup84 or the C-terminal region of Nup133, respectively. It has been reported that the Nup133 region of

the stalk is flexible in yeast (14, 30, 36) and, in an accompanying paper (75), we described how the *S. cerevisiae* Nup133 molecule shows significant flexibility in the hinge between its N- and C-terminus. Also recently, negative-stain EM analysis of the human counterpart (Nup107-160 complex (35)) suggested extreme flexibility for the Nup133 homolog. Our cross-linking restraints connecting the base of the stalk with its middle region would imply a similar bending of the yeast Nup133 molecule. Other constituents of the NPC core have been shown to present significant flexibility (76). Together, this flexibility may allow for the dilations of the NPC when accommodating large cargoes and the tensions that the NE suffers during cell growth.

Molecular architecture of the endogenous Nup84 complex revealed by integrative modeling

Our CX-MS Nup84 complex structure recapitulates most of the features already described in our previous structure (14) and by others (30, 36, 77), including the Y-shape, kinked stalk, the key role of the α -solenoid interfaces in establishing the overall architecture, and peripheral locations of the β -propeller domains (Figure 6). The stalk of the Y was shown to be formed by a tail-to-tail connection between Nup133 and Nup84 as well as a head to middle region connection between Nup84 and Nup145c. Nup120 and the Nup85-Seh1 dimer form the two arms of the Y, respectively. The pentameric hub of the complex is formed by an intricate connection between Nup145c, the C-termini of Nup120 and Nup85, and the β -propeller Nups of Seh1 and Sec13. This arrangement is consistent with several previous structures and models (14, 35, 77), but does not agree with docking of crystal structures into a density map from single particle

EM reconstruction of the entire complex, where the Nup85-Seh1 arm is oriented the other way around (36), or with the fence model where homodimeric interactions connect different copies of the complex (53, 73). The overall topology of the complex is conserved across the different clusters of solutions (Figure 6C-D). The exact positions of the domains vary slightly across the clusters, reflecting either intrinsic flexibility or lack of sufficient restraints. The main variability is detected in two regions: the Nup85-Seh1 arm and the stalk of the complex. The Nup85-Seh1 arm adopts slightly different arrangements, depending on whether or not the 3 crystallographic interface constraints are considered (Figure 6A and Supplemental Figure S7). This variability of the arm configuration is consistent with the apparent flexibility observed in our negative-stain raw EM particle images (not shown) and the class averages obtained by an independent analysis (36). The stalk of the complex is kinked in our structure (Figure 6), consistent with the shape of the crystallized Sec13-Nup145c-Nup84 trimer (49, 50). However, solutions calculated without the EM 2D restraint and crystallographic interface constraints (data not shown) include structures compatible with an alternative, extended conformation of the stalk of the complex (30, 36). Structural characterization of the Nup84 complex in the context of the whole NPC may be able to address which of the complex conformations is the predominant one *in vivo*.

The main insight from the newly determined structure is the higher level of detail in the hub region of the complex (Figures 6C-D), indicating a more intricate arrangement than previously described (49). This region has proved refractory to classical high-resolution methods due to its heterogeneity and/or flexibility. Our approach, however, was able to

surpass these challenges. The hub arrangement involves the β -propeller proteins Sec13 and Seh1, the C-terminal α -helical domains of Nup85, Nup120, and Nup145c, as well as the N-terminal region of Nup145c. The C-terminal domain of Nup85 wraps around the Seh1 β -propeller and projects into the hub to interact with the C-terminus of Nup120. The resulting interface serves as a platform for the interaction with the C-terminal domain of Nup145c. The N-terminal domain of Nup145c is inserted between Nup85, Sec13, and Seh1, thus bridging the interaction between the dimers Nup145c-Sec13 and Nup85-Seh1. A remarkable feature is the position of the β -propeller proteins Seh1 and Sec13 (Figure 6C-D). The position of Sec13 is consistent between the different clusters of solutions, forming the lower part of the hub. In contrast, Seh1 moves between two configurations: a distal position, away from Sec13 (Figure 6D, bottom row), and another one proximal, close to Sec13 (Figure 6D, top row). Both configurations are supported by the input data, and they may identify alternative configurations of the Nup85-Seh1 arm. The significant flexibility shown by EM data for the Nup85-Seh1 arm (discussed above) is consistent with the different configurations suggested by our clustering analysis (Figure 6C-D). The proximal localization of Seh1 (Figure 6D, top row) likely implies an interaction with Sec13 that was already indicated by our previous structure (14). The fact that Sec13 and Seh1 are not required for the formation of the complex (14) suggests that the interaction between them is weak, allowing the conformational changes observed for the Nup85-Seh1 arm. Such transient β -propeller arrangements are common in the vertices of evolutionarily related vesicle coating complexes (VCC), such as COPII and COPI, and have been suggested to be important for the assembly of the coats (54, 78). None of our experimental evidence points to a

vertex-like arrangement between Nup84 complex copies, but one interesting possibility is that - in the context of the whole NPC - the Seh1-Sec13 pair are arranged in close contact with other β -propellers in the NPC. There are four other Nups not in the Nup84 complex that contain β -propeller domains (Nup82, Nup159, Nup157, and Nup170). We have already suggested that all of them are in close proximity to the Nup84 complex, based on the coarse structure of the entire NPC (8).

The Nup84 complex as a membrane coating module

The protocoatomer hypothesis suggests that the NPC shares a common evolutionary origin with the VCCs clathrin, COPI, and COPII (32, 79, 80) as well as other membrane-associated assemblies such as the SEA (74, 81) and IFT (82) complexes. The Nup84 complex forms part of the core scaffold of the NPC, and in many ways this scaffold resembles that within the outer coats of VCCs (7, 8). However, although shared components, common protein domain arrangements (32), and structural features (31, 54, 73) clearly reveal this ancient relationship, evolution appears to have shaped the overall architecture of the Nup84 coat and VCC outer coats in distinct ways subsequent to their evolutionary divergence (83). Moreover, the present structure of the Nup84 complex reinforces our previous observations (76) that the NPC carries a mixture of the architectural motifs found among the other VCCs. In addition to previously described COPII-like features, like the Nup145c-Nup84 connection (31, 49) and the Nup145c-Sec13 and Nup85-Seh1 dimers (31, 54, 73), our structure also reveals a trimeric interaction between the helical C-terminal domains of Nup145C, Nup120, and Nup85 in the hub (Figure 6C-D). This arrangement may have an ancient evolutionary relationship

with the trimeric C-terminal interaction of the clathrin triskelion (84), and indeed the first description of the hub region of the Nup84 complex noted its triskelion-like shape (85). Also within this hub, we detect for the first time a heterodimeric interaction between the β -propeller proteins Sec13 and Seh1 (Figure 6C-D). Similar β -propeller arrangements are key for both COPI and COPII coat assembly (83). This mixture of motifs, found scattered among the different VCCs, is consistent with our hypothesis that different architectures previously postulated to be discrete (31, 86) actually evolved together within a single ancestral protocoatmer before the divergence of VCCs and NPCs (76). Recent evidence further underscores this idea, by providing a possible “missing link” between COPI and the adaptin complexes that can mediate VCC interaction with membrane-bound receptors (87).

The protocoatmer hypothesis also suggests that NPCs and VCCs arose through a series of gene duplications to generate architectural complexity from a simpler precursor set of only a few proteins. In the case of the eight-fold NPC, we found that each octagonal symmetry unit can be divided into two parallel columns, in which every Nup in one column contains a similarly positioned homolog in the adjacent column (7, 8). This pattern can be seen in the Nup84 complex. Our structure indicates a similarity in the connections of the clathrin/COPI-like paralogs Nup133 and Nup120, through a region proximal to their C-termini, with the C-termini of their respective cognate COPII-like partners, Nup84 and Nup145c. Such similarities agree with our previous suggestion for the origin of the Nup84 complex as the result of an ancient duplication of a hub-like triskelion module, followed by selective loss of proteins and domains from the Nup133-

containing copy of the hub (14).

Putative membrane interacting motifs are localized to the periphery of the Nup84 complex

The human Nup133 has been shown to contain an ArfGAP1 Lipid Packing Sensor (ALPS) motif (residues 245-267) in its N-terminal β -propeller (45, 88, 89).. An ALPS motif generally functions as a membrane curvature sensor and may help anchor the protein to the pore membrane (88). Previous analyses were not able to detect such motifs in fungal Nup133 homologs (88), but in an accompanying paper by Kim *et. al* (75) in this same journal issue, we annotated the presence - and solved the atomic structure - of one putative ALPS motif in the β -propeller domain of ScNup133 (252-270), and annotated the presence of two ALPS motifs in the β -propeller domain of its paralog ScNup120 (135-152 and 197-216). We mapped the positions of these motifs in our Nup84 complex ensemble (Figure 7). All the putative ALPS motifs were localized to peripheral positions of the Nup84 complex and showed similar orientation (although the conformational variability of the complex precludes a precise localization). The positions of these putative ALPS motifs are consistent with the Nup133 and Nup120 tips of the Nup84 complex contacting the NE membrane at the interface with the NPC (8, 14), forming clusters of ALPS motifs that may enhance their membrane binding due to the head-to-tail arrangement of the different Nup84 complex copies in the NPC. In addition, the predicted ALPS location is also entirely consistent with the functional role suggested for Nup120 and Nup133 in our previous work as key hotspots for the stabilization of the NPC membrane curvature (14). We suggest that similar ALPS motif arrangements and

mechanisms are conserved between yeast and vertebrates, as supported by recent cryo-EM tomography observations (35) of close proximity between the equivalent Nup120 and Nup133 tips of the human Nup107-160 (*HsNup107-160*) complex and the NPC membrane.

Structural basis for the difference in size and architecture between yeast and human NPCs

Our groups have previously shown that the yeast Nup84 complex is present in 16 copies per NPC (8, 29), organized into two head-to-tail rings at the cytoplasmic and nuclear sides of the NPC (8, 14). In a recent study (35), cryo-EM tomography revealed that the human NPC contains 32 copies of the ScNup84 complex counterpart (*HsNup107-160* complex) arranged head-to-tail into two concentric rings on each side of the NPC. The resulting model for the architecture of the *HsNup107-160* complex suggests a conserved arrangement for the 7 components that are common to the Nup84 complex, and localizes the two additional β -propeller proteins specific to the *HsNup107-160* complex, *HsNup43* and *HsNup37*, at the Nup85-Seh1 and Nup120 arms, respectively. The cryo-EM map indicates that the main contacts between the concentric rings of *HsNup107-160* complexes are established through the hub and arms of the complex, potentially involving both *HsNup43* and *HsNup37*. We speculate that the absence of these two β -propeller proteins in the ScNup84 complex would not allow a similar oligomerization as the one described in the human NPC, and go a long way to explaining the difference in the Nup84 complex copy number and in the overall size and mass between the yeast and human NPCs (8, 29, 90, 91). This scenario is also

compatible with our previous hypothesis suggesting that the Nup84 complex evolved through a series of duplication, divergence, and secondary loss events (14). We suggest that a wider picture of the NPC composition and arrangement coming from distantly related organisms is the best way to trace the evolution of this molecular machine and the origin of the eukaryotic nuclei, as well as reveal its varied functional roles in NPCs between different organisms.

ACKNOWLEDGEMENTS

We thank Paula Upla and Ruben Diaz-Avalos at NYSBC for their help with the analysis of the EM data. We are also grateful to Daniel Russel, Charles Greenberg, and Benjamin Webb at UCSF for their help with and support of the PMI library and the Integrative Modeling Platform (IMP) package. We are grateful to Andrew N. Krutchinsky and Julio C. Padovan at Rockefeller University, and Shengbo Fan at Institute of Computational Technology, CAS, for their suggestions and comments. Funding for this work was provided by NIH grants P41 GM103314 (B.T.C), R01 GM062427 (M.P.R.), R01 GM083960 (A.S.), U54 GM103511 (A.S., M.P.R., and B.T.C.) and U01 GM098256 (A.S. and M.P.R.).

REFERENCES

1. Gavin, A. C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L. J., Bastuck, S., Dumpelfeld, B., Edelmann, A., Heurtier, M. A., Hoffman, V., Hoefert, C., Klein, K., Hudak, M., Michon, A. M., Schelder, M., Schirle, M., Remor, M., Rudi, T., Hooper, S., Bauer, A., Bouwmeester, T., Casari, G., Drewes, G., Neubauer, G., Rick, J. M., Kuster, B., Bork, P., Russell, R. B., and Superti-Furga, G. (2006) Proteome survey reveals modularity of the yeast cell machinery. *Nature* 440, 631-636
2. Krogan, N. J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A. P., Punna, T., Peregrin-Alvarez, J. M., Shales, M., Zhang, X., Davey, M., Robinson, M. D., Paccanaro, A., Bray, J. E., Sheung, A., Beattie, B., Richards, D. P., Canadien, V., Lalev, A., Mena, F., Wong, P., Starostine, A., Canete, M. M., Vlasblom, J., Wu, S., Orsi, C., Collins, S. R., Chandran, S., Haw, R., Rilstone, J. J., Gandi, K., Thompson, N. J., Musso, G., St Onge, P., Ghanny, S., Lam, M. H., Butland, G., Altaf-Ul, A. M., Kanaya, S., Shilatifard, A., O'Shea, E., Weissman, J. S., Ingles, C. J., Hughes, T. R., Parkinson, J., Gerstein, M., Wodak, S. J., Emili, A., and Greenblatt, J. F. (2006) Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* 440, 637-643
3. Malovannaya, A., Lanz, R. B., Jung, S. Y., Bulyanko, Y., Le, N. T., Chan, D. W., Ding, C., Shi, Y., Yucer, N., Krenciute, G., Kim, B. J., Li, C., Chen, R., Li, W., Wang, Y., O'Malley, B. W., and Qin, J. (2011) Analysis of the human endogenous coregulator complexome. *Cell* 145, 787-799
4. Robinson, C. V., Sali, A., and Baumeister, W. (2007) The molecular sociology of

the cell. *Nature* 450, 973-982

5. Alber, F., Forster, F., Korkein, D., Topf, M., and Sali, A. (2008) Integrating diverse data for structure determination of macromolecular assemblies. *Annu Rev Biochem* 77, 443-477
6. Ward, A. B., Sali, A., and Wilson, I. A. (2013) Biochemistry. Integrative structural biology. *Science* 339, 913-915
7. Alber, F., Dokudovskaya, S., Veenhoff, L. M., Zhang, W., Kipper, J., Devos, D., Suprpto, A., Karni-Schmidt, O., Williams, R., Chait, B. T., Rout, M. P., and Sali, A. (2007) Determining the architectures of macromolecular assemblies. *Nature* 450, 683-694
8. Alber, F., Dokudovskaya, S., Veenhoff, L. M., Zhang, W., Kipper, J., Devos, D., Suprpto, A., Karni-Schmidt, O., Williams, R., Chait, B. T., Sali, A., and Rout, M. P. (2007) The molecular architecture of the nuclear pore complex. *Nature* 450, 695-701
9. Duan, Z., Andronescu, M., Schutz, K., McIlwain, S., Kim, Y. J., Lee, C., Shendure, J., Fields, S., Blau, C. A., and Noble, W. S. (2010) A three-dimensional model of the yeast genome. *Nature* 465, 363-367
10. Boura, E., Rozycki, B., Herrick, D. Z., Chung, H. S., Vecer, J., Eaton, W. A., Cafiso, D. S., Hummer, G., and Hurley, J. H. (2011) Solution structure of the ESCRT-I complex by small-angle X-ray scattering, EPR, and FRET spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America* 108, 9437-9442
11. Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F., and Chen, L. (2012) Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nature biotechnology* 30, 90-98

12. Lander, G. C., Estrin, E., Matyskiela, M. E., Bashore, C., Nogales, E., and Martin, A. (2012) Complete subunit architecture of the proteasome regulatory particle. *Nature* 482, 186-191
13. Lasker, K., Forster, F., Bohn, S., Walzthoeni, T., Villa, E., Unverdorben, P., Beck, F., Aebersold, R., Sali, A., and Baumeister, W. (2012) Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. *Proceedings of the National Academy of Sciences of the United States of America* 109, 1380-1387
14. Fernandez-Martinez, J., Phillips, J., Sekedat, M. D., Diaz-Avalos, R., Velazquez-Muriel, J., Franke, J. D., Williams, R., Stokes, D. L., Chait, B. T., Sali, A., and Rout, M. P. (2012) Structure-function mapping of a heptameric module in the nuclear pore complex. *J Cell Biol* 196, 419-434
15. Tosi, A., Haas, C., Herzog, F., Gilmozzi, A., Berninghausen, O., Ungewickell, C., Gerhold, C. B., Lakomek, K., Aebersold, R., Beckmann, R., and Hopfner, K. P. (2013) Structure and subunit topology of the INO80 chromatin remodeler and its nucleosome complex. *Cell* 154, 1207-1219
16. Greber, B. J., Boehringer, D., Leitner, A., Bieri, P., Voigts-Hoffmann, F., Erzberger, J. P., Leibundgut, M., Aebersold, R., and Ban, N. (2014) Architecture of the large subunit of the mammalian mitochondrial ribosome. *Nature* 505, 515-519
17. Cohen, S. L., and Chait, B. T. (2001) Mass spectrometry as a tool for protein crystallography. *Annual review of biophysics and biomolecular structure* 30, 67-85
18. Young, M. M., Tang, N., Hempel, J. C., Oshiro, C. M., Taylor, E. W., Kuntz, I. D., Gibson, B. W., and Dollinger, G. (2000) High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass

spectrometry. *Proceedings of the National Academy of Sciences of the United States of America* 97, 5802-5806

19. Chen, Z. A., Jawhari, A., Fischer, L., Buchen, C., Tahir, S., Kamenski, T., Rasmussen, M., Lariviere, L., Bukowski-Wills, J. C., Nilges, M., Cramer, P., and Rappsilber, J. (2010) Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. *The EMBO journal* 29, 717-726

20. Herzog, F., Kahraman, A., Boehringer, D., Mak, R., Bracher, A., Walzthoeni, T., Leitner, A., Beck, M., Hartl, F. U., Ban, N., Malmstrom, L., and Aebersold, R. (2012) Structural probing of a protein phosphatase 2A network by chemical cross-linking and mass spectrometry. *Science* 337, 1348-1352

21. Leitner, A., Walzthoeni, T., Kahraman, A., Herzog, F., Rinner, O., Beck, M., and Aebersold, R. (2010) Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. *Molecular & cellular proteomics : MCP* 9, 1634-1649

22. Sinz, A. (2006) Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. *Mass Spectrom Rev* 25, 663-682

23. Trnka, M. J., Baker, P. R., Robinson, P. J., Burlingame, A. L., and Chalkley, R. J. (2014) Matching cross-linked peptide spectra: only as good as the worse identification. *Molecular & cellular proteomics : MCP* 13, 420-434

24. Kao, A., Chiu, C. L., Vellucci, D., Yang, Y., Patel, V. R., Guan, S., Randall, A., Baldi, P., Rychnovsky, S. D., and Huang, L. (2011) Development of a novel cross-linking strategy for fast and accurate identification of cross-linked peptides of protein complexes. *Molecular & cellular proteomics : MCP* 10, M110 002212

25. Weisbrod, C. R., Chavez, J. D., Eng, J. K., Yang, L., Zheng, C., and Bruce, J. E. (2013) In vivo protein interaction network identified with a novel real-time cross-linked peptide identification strategy. *Journal of proteome research* 12, 1569-1579
26. Ghaemmaghami, S., Huh, W. K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., O'Shea, E. K., and Weissman, J. S. (2003) Global analysis of protein expression in yeast. *Nature* 425, 737-741
27. Beck, M., Schmidt, A., Malmstroem, J., Claassen, M., Ori, A., Szymborska, A., Herzog, F., Rinner, O., Ellenberg, J., and Aebersold, R. (2011) The quantitative proteome of a human cell line. *Molecular systems biology* 7, 549
28. Leitner, A., Joachimiak, L. A., Unverdorben, P., Walzthoeni, T., Frydman, J., Forster, F., and Aebersold, R. (2014) Chemical cross-linking/mass spectrometry targeting acidic residues in proteins and protein complexes. *Proceedings of the National Academy of Sciences of the United States of America* 111, 9455-9460
29. Rout, Aitchison, Suprpto, A., Hjertaas, K., Zhao, Y., and Chait, B. T. (2000) The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J Cell Biol*, pp. 635-651
30. Lutzmann, M., Kunze, R., Buerer, A., Aebi, U., and Hurt, E. (2002) Modular self-assembly of a Y-shaped multiprotein complex from seven nucleoporins. *The EMBO journal* 21, 387-397
31. Brohawn, S. G., Leksa, N. C., Spear, E. D., Rajashankar, K. R., and Schwartz, T. U. (2008) Structural evidence for common ancestry of the nuclear pore complex and vesicle coats. *Science* 322, 1369-1373
32. Devos, D., Dokudovskaya, S., Alber, F., Williams, R., Chait, B. T., Sali, A., and

- Rout, M. P. (2004) Components of coated vesicles and nuclear pore complexes share a common molecular architecture. *PLoS biology* 2, e380
33. Bilokapic, S., and Schwartz, T. U. (2012) 3D ultrastructure of the nuclear pore complex. *Current opinion in cell biology* 24, 86-91
34. Hoelz, A., Debler, E. W., and Blobel, G. (2011) The structure of the nuclear pore complex. *Annu Rev Biochem* 80, 613-643
35. Bui, K. H., von Appen, A., DiGuilio, A. L., Ori, A., Sparks, L., Mackmull, M. T., Bock, T., Hagen, W., Andres-Pons, A., Glavy, J. S., and Beck, M. (2013) Integrated structural analysis of the human nuclear pore complex scaffold. *Cell* 155, 1233-1243
36. Kampmann, M., and Blobel, G. (2009) Three-dimensional structure and flexibility of a membrane-coating module of the nuclear pore complex. *Nat Struct Mol Biol* 16, 782-788
37. Thierbach, K., von Appen, A., Thoms, M., Beck, M., Flemming, D., and Hurt, E. (2013) Protein interfaces of the conserved Nup84 complex from *Chaetomium thermophilum* shown by crosslinking mass spectrometry and electron microscopy. *Structure* 21, 1672-1682
38. Velazquez-Muriel, J., Lasker, K., Russel, D., Phillips, J., Webb, B. M., Schneidman-Duhovny, D., and Sali, A. (2012) Assembly of macromolecular complexes by satisfaction of spatial restraints from electron microscopy images. *Proceedings of the National Academy of Sciences of the United States of America* 109, 18821-18826
39. Olsen, J. V., Macek, B., Lange, O., Makarov, A., Horning, S., and Mann, M. (2007) Higher-energy C-trap dissociation for peptide modification analysis. *Nature methods* 4, 709-712

40. Yang, B., Wu, Y. J., Zhu, M., Fan, S. B., Lin, J., Zhang, K., Li, S., Chi, H., Li, Y. X., Chen, H. F., Luo, S. K., Ding, Y. H., Wang, L. H., Hao, Z., Xiu, L. Y., Chen, S., Ye, K., He, S. M., and Dong, M. Q. (2012) Identification of cross-linked peptides from complex samples. *Nature methods* 9, 904-906
41. Qin, J., and Chait, B. T. (1995) Preferential Fragmentation of Protonated Gas-Phase Peptide Ions Adjacent to Acidic Amino-Acid-Residues. *Journal of the American Chemical Society* 117, 5411-5412
42. Michalski, A., Neuhauser, N., Cox, J., and Mann, M. (2012) A systematic investigation into the nature of tryptic HCD spectra. *Journal of proteome research* 11, 5479-5491
43. Lasker, K., Phillips, J. L., Russel, D., Velazquez-Muriel, J., Schneidman-Duhovny, D., Tjioe, E., Webb, B., Schlessinger, A., and Sali, A. (2010) Integrative structure modeling of macromolecular assemblies from proteomics data. *Molecular & cellular proteomics : MCP* 9, 1689-1702
44. Russel, D., Lasker, K., Webb, B., Velazquez-Muriel, J., Tjioe, E., Schneidman-Duhovny, D., Peterson, B., and Sali, A. (2012) Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. *PLoS biology* 10, e1001244
45. Berke, I. C., Boehmer, T., Blobel, G., and Schwartz, T. U. (2004) Structural and functional analysis of Nup133 domains reveals modular building blocks of the nuclear pore complex. *J Cell Biol* 167, 591-597
46. Whittle, J. R., and Schwartz, T. U. (2009) Architectural nucleoporins Nup157/170 and Nup133 are structurally related and descend from a second ancestral element. *The*

47. Boehmer, T., Jeudy, S., Berke, I. C., and Schwartz, T. U. (2008) Structural and functional studies of Nup107/Nup133 interaction and its implications for the architecture of the nuclear pore complex. *Molecular cell* 30, 721-731
48. Sampathkumar, P., Gheyi, T., Miller, S. A., Bain, K. T., Dickey, M., Bonanno, J. B., Kim, S. J., Phillips, J., Pieper, U., Fernandez-Martinez, J., Franke, J. D., Martel, A., Tsuruta, H., Atwell, S., Thompson, D. A., Emtage, J. S., Wasserman, S. R., Rout, M. P., Sali, A., Sauder, J. M., and Burley, S. K. (2011) Structure of the C-terminal domain of *Saccharomyces cerevisiae* Nup133, a component of the nuclear pore complex. *Proteins* 79, 1672-1677
49. Brohawn, S. G., and Schwartz, T. U. (2009) Molecular architecture of the Nup84-Nup145C-Sec13 edge element in the nuclear pore complex lattice. *Nat Struct Mol Biol* 16, 1173-1177
50. Nagy, V., Hsia, K. C., Debler, E. W., Kampmann, M., Davenport, A. M., Blobel, G., and Hoelz, A. (2009) Structure of a trimeric nucleoporin complex reveals alternate oligomerization states. *Proceedings of the National Academy of Sciences of the United States of America* 106, 17693-17698
51. Seo, H. S., Ma, Y., Debler, E. W., Wacker, D., Kutik, S., Blobel, G., and Hoelz, A. (2009) Structural and functional analysis of Nup120 suggests ring formation of the Nup84 complex. *Proceedings of the National Academy of Sciences of the United States of America* 106, 14281-14286
52. Leksa, N. C., Brohawn, S. G., and Schwartz, T. U. (2009) The structure of the scaffold nucleoporin Nup120 reveals a new and unexpected domain architecture.

Structure 17, 1082-1091

53. Debler, E. W., Ma, Y., Seo, H. S., Hsia, K. C., Noriega, T. R., Blobel, G., and Hoelz, A. (2008) A fence-like coat for the nuclear pore membrane. *Molecular cell* 32, 815-826
54. Fath, S., Mancias, J. D., Bi, X., and Goldberg, J. (2007) Structure and organization of coat proteins in the COPII cage. *Cell* 129, 1325-1336
55. Jeudy, S., and Schwartz, T. U. (2007) Crystal structure of nucleoporin Nic96 reveals a novel, intricate helical domain architecture. *The Journal of biological chemistry* 282, 34904-34912
56. Lee, J. H., Yi, L., Li, J., Schweitzer, K., Borgmann, M., Naumann, M., and Wu, H. (2013) Crystal structure and versatile functional roles of the COP9 signalosome subunit 1. *Proceedings of the National Academy of Sciences of the United States of America* 110, 11845-11850
57. Soding, J. (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics (Oxford, England)* 21, 951-960
58. Soding, J., Biegert, A., and Lupas, A. N. (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic acids research* 33, W244-248
59. Jones, D. T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *Journal of molecular biology* 292, 195-202
60. Buchan, D. W., Minneci, F., Nugent, T. C., Bryson, K., and Jones, D. T. (2013) Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic acids research* 41, W349-357

61. Ward, J. J., McGuffin, L. J., Bryson, K., Buxton, B. F., and Jones, D. T. (2004) The DISOPRED server for the prediction of protein disorder. *Bioinformatics (Oxford, England)* 20, 2138-2139
62. Sali, A., and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *Journal of molecular biology* 234, 779-815
63. Erzberger, J., Stengel, F., Pellarin, R., Zhang, S., Schaefer, T., Aylett, C., Cimermancic, P., Boehringer, D., Sali, A., Aebersold, R., and Ban, N. (2014) Molecular architecture of the 40S-eIF1-eIF3 translation initiation complex. *Cell*, in press
64. Shen, M. Y., and Sali, A. (2006) Statistical potential for assessment and prediction of protein structures. *Protein Sci* 15, 2507-2524
65. Schneidman-Duhovny, D., Rossi, A., Avila-Sakar, A., Kim, S. J., Velazquez-Muriel, J., Strop, P., Liang, H., Krukenberg, K. A., Liao, M., Kim, H. M., Sobhanifar, S., Dotsch, V., Rajpal, A., Pons, J., Agard, D. A., Cheng, Y., and Sali, A. (2012) A method for integrative structure determination of protein-protein complexes. *Bioinformatics (Oxford, England)* 28, 3282-3289
66. Rieping, W., Habeck, M., and Nilges, M. (2005) Inferential structure determination. *Science* 309, 303-306
67. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics. *Journal of molecular graphics* 14, 33-38, 27-38
68. Heyer, L. J., Kruglyak, S., and Yooseph, S. (1999) Exploring expression data: identification and analysis of coexpressed genes. *Genome research* 9, 1106-1115
69. Levitt, M. (1983) Molecular dynamics of native protein. II. Analysis and nature of motion. *Journal of molecular biology* 168, 621-657

70. 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. *Nature methods* 4, 951-956
71. Leitner, A., Reischl, R., Walzthoeni, T., Herzog, F., Bohn, S., Forster, F., and Aebersold, R. (2012) Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography. *Molecular & cellular proteomics : MCP* 11, M111 014126
72. Merkley, E. D., Rysavy, S., Kahraman, A., Hafen, R. P., Daggett, V., and Adkins, J. N. (2014) Distance restraints from crosslinking mass spectrometry: Mining a molecular dynamics simulation database to evaluate lysine-lysine distances. *Protein Sci*
73. Hsia, K. C., Stavropoulos, P., Blobel, G., and Hoelz, A. (2007) Architecture of a coat for the nuclear pore membrane. *Cell* 131, 1313-1326
74. Algret, R., Fernandez-Martinez, J., Shi, Y., Kim, S. J., Pellarin, R., Cimermancic, P., Cochet, E., Sali, A., Chait, B. T., Rout, M. P., and Dokudovskaya, S. (2014) Molecular architecture and function of the SEA complex, a modulator of the TORC1 pathway. *Molecular & cellular proteomics : MCP*
75. Kim, S. J., Fernandez-Martinez, J., Sampathkumar, P., Martel, A., Matsui, T., Tsuruta, H., Weiss, T., Markina-Inarrairaegui, A., Bonanno, J. B., Sauder, J. M., Burley, S. K., Almo, S. C., Rout, M., and Sali, A. (2014) Integrative Structure-Function Mapping of Nup133, a Scaffold Component of the Nuclear Pore Complex. *Molecular & cellular proteomics : MCP*, Submitted
76. Sampathkumar, P., Kim, S. J., Upla, P., Rice, W. J., Phillips, J., Timney, B. L., Pieper, U., Bonanno, J. B., Fernandez-Martinez, J., Hakhverdyan, Z., Ketaren, N. E.,

- Matsui, T., Weiss, T. M., Stokes, D. L., Sauder, J. M., Burley, S. K., Sali, A., Rout, M. P., and Almo, S. C. (2013) Structure, dynamics, evolution, and function of a major scaffold component in the nuclear pore complex. *Structure* 21, 560-571
77. Brohawn, S. G., Partridge, J. R., Whittle, J. R., and Schwartz, T. U. (2009) The nuclear pore complex has entered the atomic age. *Structure* 17, 1156-1168
78. Lee, C., and Goldberg, J. (2010) Structure of coatamer cage proteins and the relationship among COPI, COPII, and clathrin vesicle coats. *Cell* 142, 123-132
79. Devos, D., Dokudovskaya, S., Williams, R., Alber, F., Eswar, N., Chait, B. T., Rout, M. P., and Sali, A. (2006) Simple fold composition and modular architecture of the nuclear pore complex. *Proceedings of the National Academy of Sciences of the United States of America* 103, 2172-2177
80. Field, M. C., Sali, A., and Rout, M. P. (2011) Evolution: On a bender--BARs, ESCRTs, COPs, and finally getting your coat. *J Cell Biol* 193, 963-972
81. Dokudovskaya, S., Waharte, F., Schlessinger, A., Pieper, U., Devos, D. P., Cristea, I. M., Williams, R., Salamero, J., Chait, B. T., Sali, A., Field, M. C., Rout, M. P., and Dargemont, C. (2011) A conserved coatamer-related complex containing Sec13 and Seh1 dynamically associates with the vacuole in *Saccharomyces cerevisiae*. *Molecular & cellular proteomics : MCP* 10, M110 006478
82. van Dam, T. J., Townsend, M. J., Turk, M., Schlessinger, A., Sali, A., Field, M. C., and Huynen, M. A. (2013) Evolution of modular intraflagellar transport from a coatamer-like progenitor. *Proceedings of the National Academy of Sciences of the United States of America* 110, 6943-6948
83. Faini, M., Beck, R., Wieland, F. T., and Briggs, J. A. (2013) Vesicle coats:

structure, function, and general principles of assembly. *Trends in cell biology* 23, 279-288

84. Fotin, A., Cheng, Y., Sliz, P., Grigorieff, N., Harrison, S. C., Kirchhausen, T., and Walz, T. (2004) Molecular model for a complete clathrin lattice from electron cryomicroscopy. *Nature* 432, 573-579

85. Siniossoglou, S., Lutzmann, M., Santos-Rosa, H., Leonard, K., Mueller, S., Aebi, U., and Hurt, E. (2000) Structure and assembly of the Nup84p complex. *J Cell Biol* 149, 41-54

86. Leksa, N. C., and Schwartz, T. U. (2010) Membrane-coating lattice scaffolds in the nuclear pore and vesicle coats: commonalities, differences, challenges. *Nucleus* 1, 314-318

87. Hirst, J., Schlacht, A., Norcott, J. P., Traynor, D., Bloomfield, G., Antrobus, R., Kay, R. R., Dacks, J. B., and Robinson, M. S. (2014) Characterization of TSET, an ancient and widespread membrane trafficking complex. *eLife* 3, e02866

88. Drin, G., Casella, J. F., Gautier, R., Boehmer, T., Schwartz, T. U., and Antony, B. (2007) A general amphipathic alpha-helical motif for sensing membrane curvature. *Nat Struct Mol Biol* 14, 138-146

89. Drin, G., and Antony, B. (2010) Amphipathic helices and membrane curvature. *FEBS letters* 584, 1840-1847

90. Akey, C. W., and Radermacher, M. (1993) Architecture of the *Xenopus* nuclear pore complex revealed by three-dimensional cryo-electron microscopy. *J Cell Biol* 122, 1-19

91. Yang, Q., Rout, M. P., and Akey, C. W. (1998) Three-dimensional architecture of

the isolated yeast nuclear pore complex: functional and evolutionary implications.

Molecular cell 1, 223-234

FIGURE LEGENDS

Figure 1. CX-MS integrative pipeline for the structural characterization of endogenous complexes

The workflow of the cross-linking and mass spectrometric analysis of endogenous Nup84 complex is summarized. The endogenously tagged protein complex is affinity purified from cell cryolysis, natively eluted, and cross-linked by two complementary cross-linkers DSS and EDC. The cross-linked complex is then subjected to in-solution and in-gel digestion and the resulting peptide mixtures are fractionated using a peptide size exclusion chromatography (SEC). The cross-linked peptides were analyzed by a Velos Orbitrap mass spectrometer using high-resolution HCD and searched by pLink (40). All software-produced identifications were manually inspected (see Experimental Procedures).

Figure 2. CX-MS analysis of the Nup84 complex

(A-B) Cross-linking maps of the Nup84 complex by either DSS cross-linker (A) or EDC cross-linker (B). Straight lines in red connecting residues from different subunits represent DSS inter-subunit cross-links while straight lines in green represent EDC inter-subunit cross-links. Curved, dotted lines represent intra-subunit cross-links. Only intra-subunit cross-links between residues more than 40 positions apart are shown.

(C) A high-resolution HCD MS/MS spectrum ($m/z=1096.169$, $z=6$) of the inter-subunit DSS cross-link connecting Nup85 residue 30 (lysine) and Seh1 residue 1 (N-terminal methionine). The methionines are oxidized. F* indicates an immonium ion of phenylalanine which is frequently observed in HCD spectrum. b and y ion series

including their charge states are labelled. The fragment of y20 is zoomed in and the mass accuracy is labeled in parts per million (ppm).

(D) A high-resolution HCD MS/MS spectrum (m/z 1044.139, $z=6$) of the inter-subunit EDC cross-link connecting Nup85 residue 17 and Seh1 residue 1. The spectrum is labelled as of C).The fragment of y20 is zoomed in and the mass accuracy is labeled in parts per million (ppm).

(E) A high-resolution HCD MS/MS spectrum ($m/z=903.317$, $z=7$) of the intra-subunit DSS cross-link connecting Nup133 residue 506 and Nup133 residue 59. F*, and Y* indicate immonium ions of phenylalanine and tyrosine which are common for HCD fragmentations. The fragment of b14 is zoomed in and the mass accuracy is labeled in parts per million (ppm).

(F) A high-resolution HCD MS/MS spectrum ($m/z= 917.327$, $z=6$) of the intra-subunit EDC cross-link connecting Nup133 residue 506 and residue 562. F*, and Y* indicate immonium ions of phenylalanine and tyrosine which are common to HCD fragmentations. The fragment of y16 is zoomed in and the mass accuracy is labeled in parts per million (ppm). Charge states of the fragments in the high mass region of this spectrum (i.e, 1,100- 1,300 m/z) are not resolved.

Figure 3. Distance distribution of the cross-links and their mapping on the available crystal structures

(A) We mapped the Euclidean C_{α} - C_{α} distances between the cross-linked residues onto several domains of the Nup84 complex proteins with available crystal structures (49, 50). The C_{α} - C_{α} distance distributions of the cross-linked residues are shown for both

DSS and EDC cross-links. All the measured DSS cross-links fall within the expected maximum threshold of ~ 30 Å (72), and the great majority of EDC cross-links fall within the expected threshold of 17 Å.

(B) Both DSS (in red) and EDC (in green) cross-links are mapped on the crystallographic structure of the Sec13-Nup145c-Nup84 interface (PDB code 3IKO (50)). The EDC and DSS cross-links provide complementary spatial information.

Figure 4. The 4-stage scheme for integrative structure determination of the Nup84 complex

Our integrative approach to determine the Nup84 complex structure proceeds through four stages (7, 13, 14, 43): (1) gathering of data, (2) representation of subunits and translation of the data into spatial restraints, (3) configurational sampling to produce an ensemble of models that satisfies the restraints, and (4) analysis of the ensemble. The modeling protocol (*ie*, stages 2, 3, and 4) was scripted using the *Python Modeling Interface* (<https://github.com/salilab/pmi>), version *be72c15*, a library to model macromolecular complexes based on our open source *Integrative Modeling Platform* (IMP) package (<http://salilab.org/imp/>), version *829c3f0* (44).

Figure 5. Correlation between the number of cross-links and the accuracy of dimer models

(A) Localization density maps of the Nup84 subunits (solid contour surfaces; Fig. 6A) and position of the Nup145c-Sec13 crystallographic dimer (PDB code 3IKO (50)).

(B) Total scores (*i.e.*, the sum of excluded volume and cross-link restraint scores) are

plotted as a function of the C_{α} dRMSD of the Nup145c-Sec13 dimer models with respect to the crystallographic dimer; as the number of cross links increases, the ensemble of models is enriched in accurate structures (i.e., low dRMSD models, left of dotted line).

(C) Accuracy of dimer models as a function of the number and type of cross-links. Each symbol displays the first and third quartile (lower and upper side of the boxes), the median (red line), as well as minimum and maximum (lower and upper limit of the dashed whiskers, respectively) of the C_{α} dRMSD with respect to the crystallographic structure for the 100 best-scoring models. The median and the spread (i.e., difference between the first and third quartile) of the distribution is a measure of the accuracy and precision of the ensemble of models, respectively.

Figure 6. The Nup84 complex molecular architecture revealed by CX-MS integrative pipeline

The localization density maps of the Nup84 subunits (solid contour surfaces) and the entire complex (transparent surfaces) were computed and contoured at the threshold of 2.5 times their volumes estimated from sequence (Supplemental Table S6) (panels A to C).

(A) Front and back views of the localization density maps of the Nup84 subunits and the entire complex.

(B) A representative single Nup84 complex structure (colored ribbon) is shown along with the localization density maps of the individual subunits.

(C) The localization density maps of the two dominant clusters computed on the hub

region (Nup120-CTD, Nup85, Nup145c, Sec13, and Seh1) are shown, along with the representative single structures of the hub region for each of the two clusters, from multiple viewing points.

(D) The positions of Sec13 and Seh1 are presented for each of the two clusters.

(E) The representative model projections in each of the two clusters are shown, along with the EM class average (14).

Figure 7. The putative membrane interacting ALPS motifs

One putative ALPS motif in Nup133 (252-270) and two in Nup120 (135-152 and 197-216) are mapped on the Nup84 complex. All 3 putative ALPS motifs are localized to peripheral positions of the Nup84 complex.

TABLES

Cluster index	Accuracy of crystallographic interface [Å]		
	Nup145c-Sec13	Nup85-Seh1	Nup84-Nup145c
1	4.0 (min 2.4)	12.4 (min 6.5)	7.5 (min 2.0)
2	4.0 (min 2.7)	12.4 (min 4.4)	6.2 (min 2.2)

Table 1. Accuracy of determining the crystallographic interfaces

The accuracy of modeling each crystallographic interface was calculated as the average C_{α} dRMSD between the models in the cluster and the corresponding crystallographic interface. The minimum value of C_{α} dRMSD in the cluster is indicated in parenthesis.

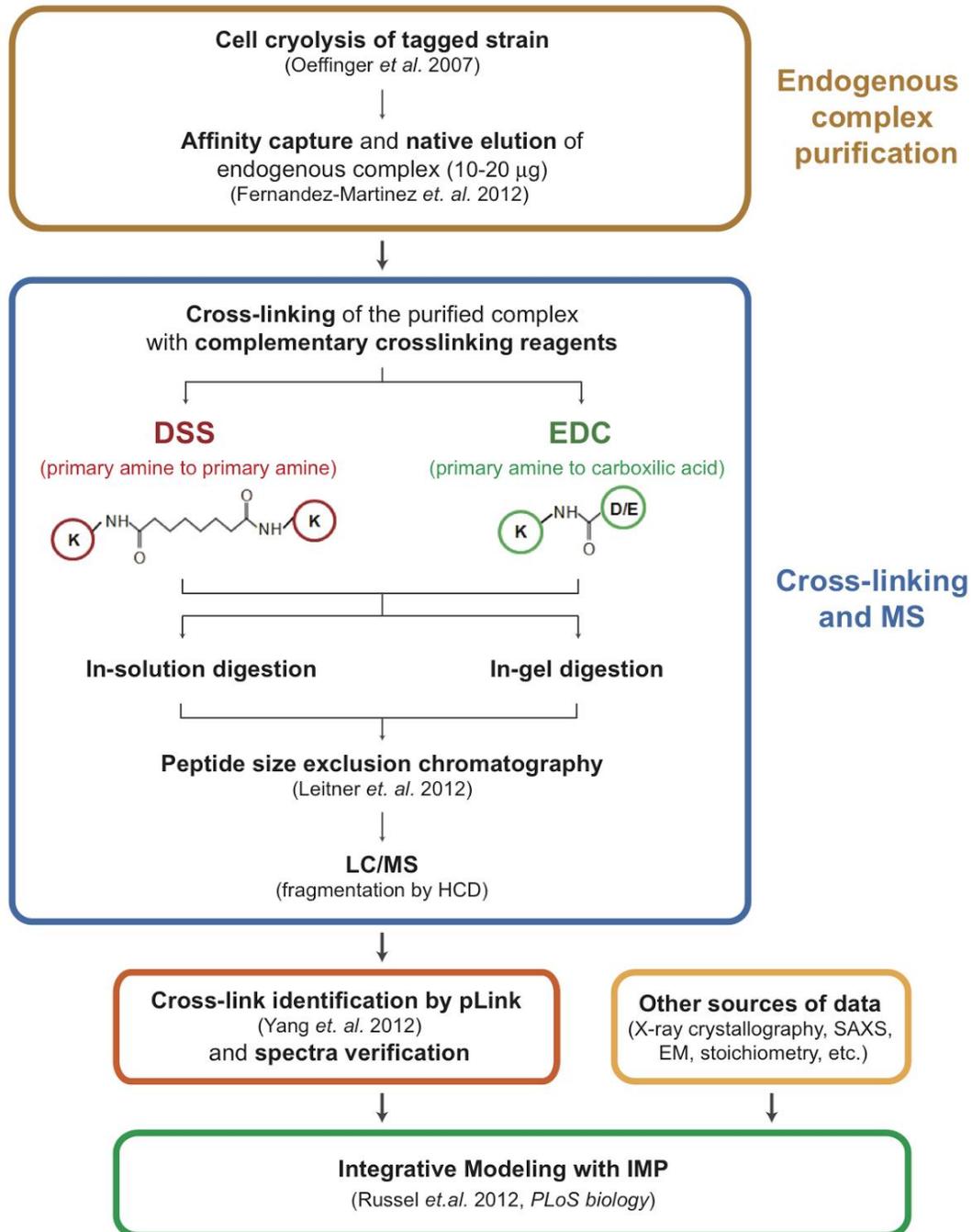


Figure 1. CX-MS integrative pipeline for the structural characterization of endogenous complexes

Suggested location – Results, Development of a workflow for CX-MS of endogenous protein complexes

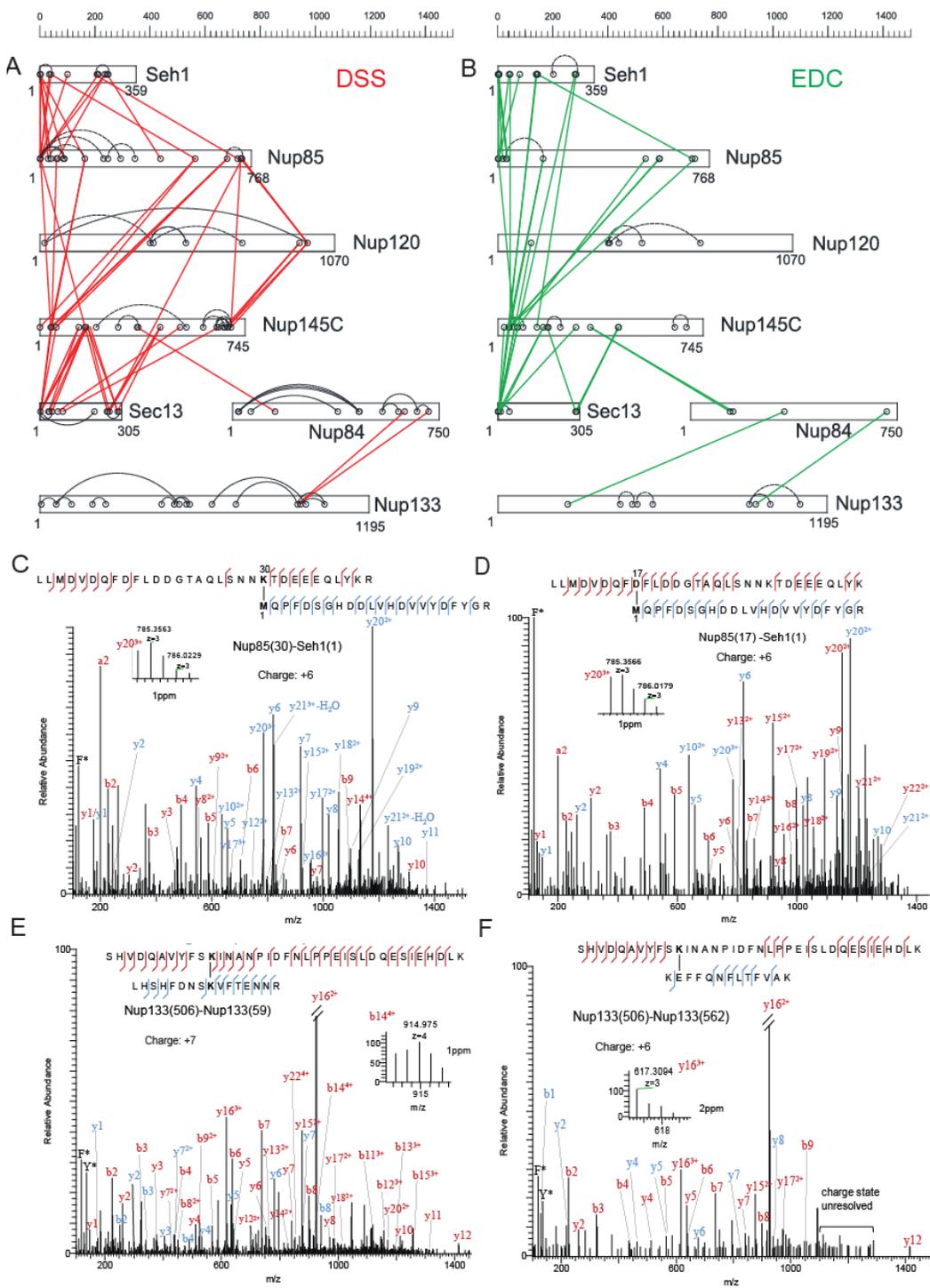


Figure 2. CX-MS analysis of the Nup84 complex

Suggested location – Results, Chemical cross-linking and MS analysis of the Nup84 complex

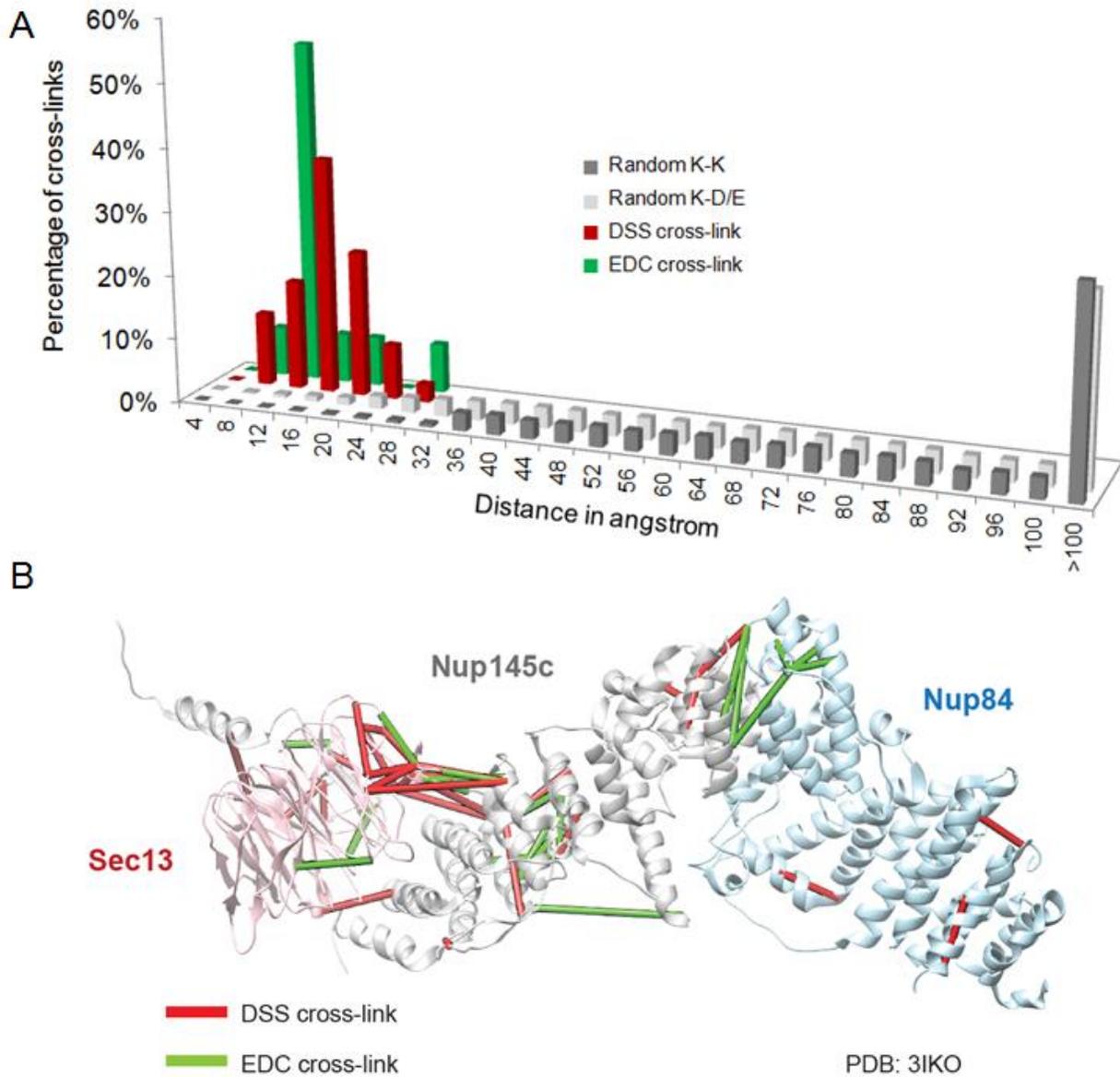


Figure 3. Distance distribution of the cross-links and their mapping into the available crystal structures

Suggested location – Results, Chemical cross-linking and MS analysis of the Nup84 complex

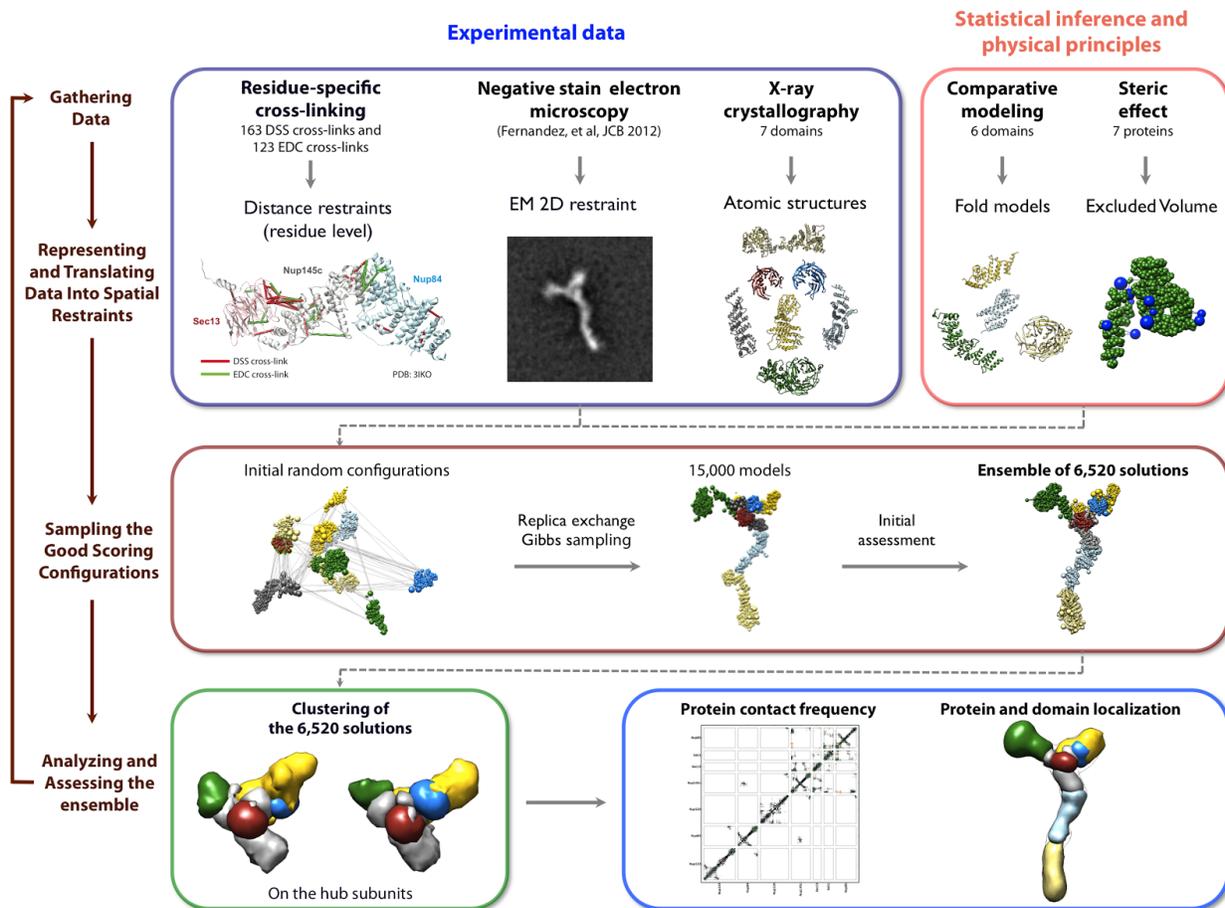


Figure 4. The 4-stage scheme for integrative structure determination of the Nup84 complex

Suggested location – Results, Molecular architecture of the endogenous Nup84 complex revealed by integrative modeling

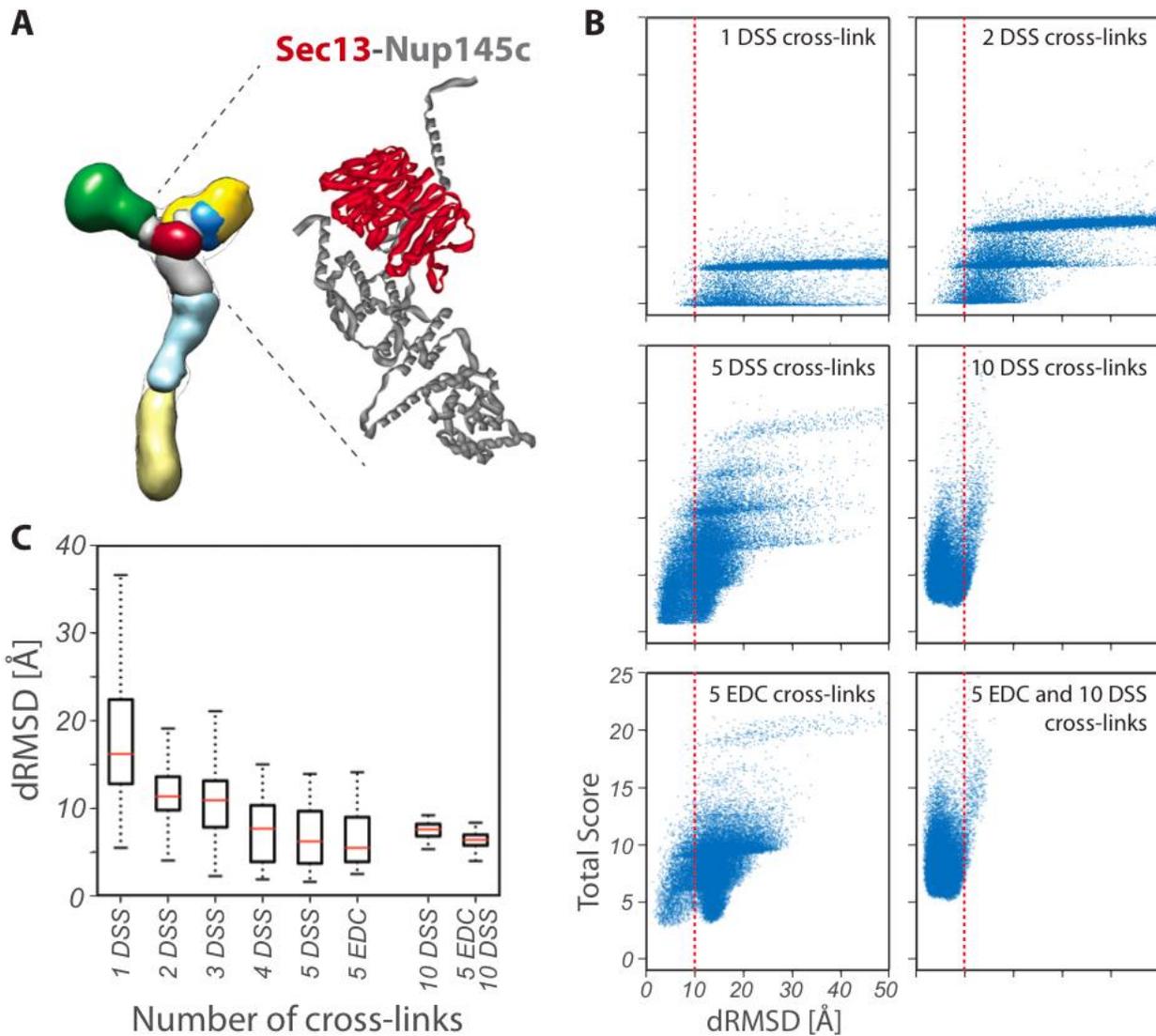


Figure 5. Correlation between the number of cross-links and the accuracy of dimer models

Suggested location – Results, Correlation between the number of cross-links and the accuracy of dimer models

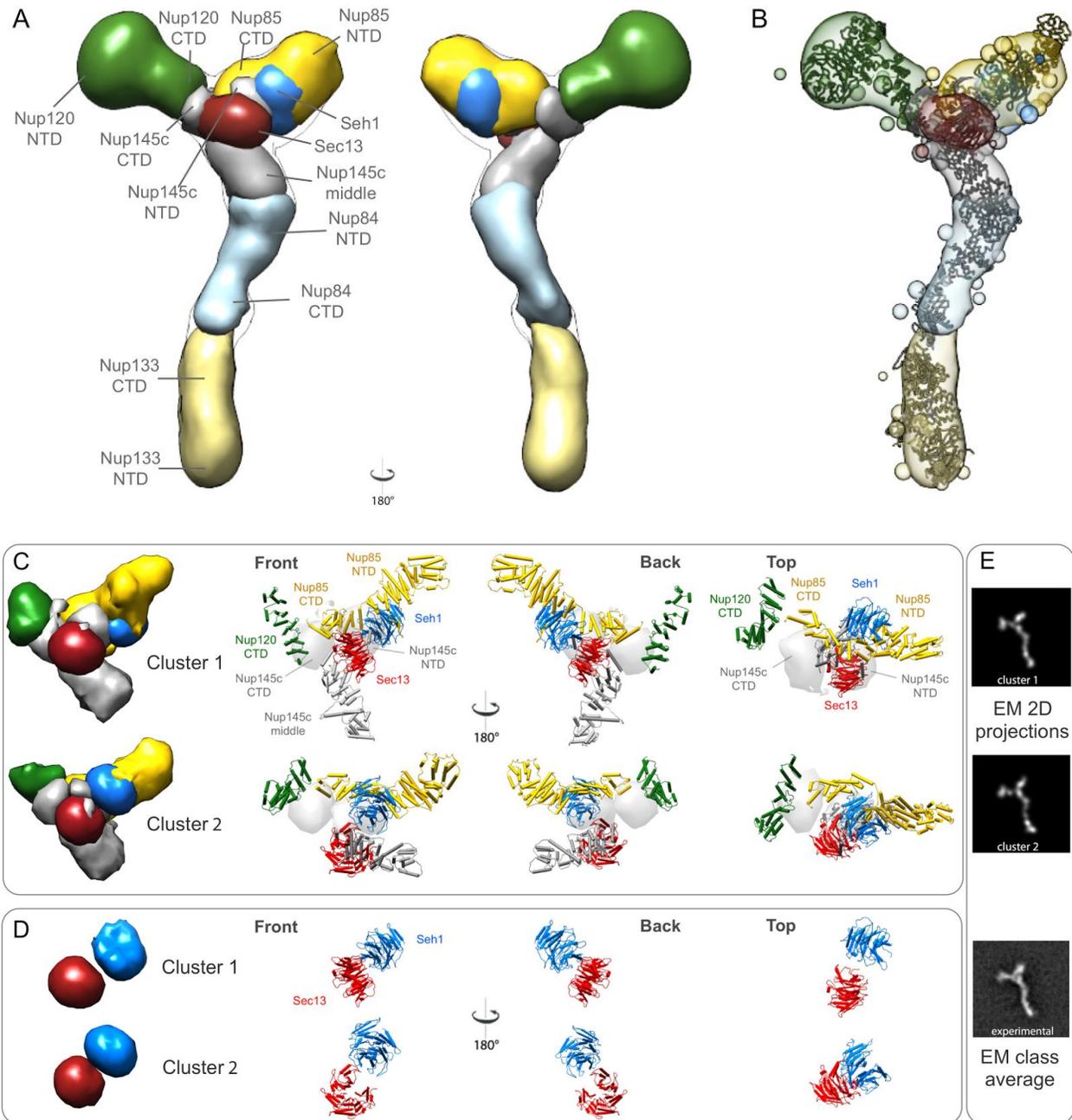


Figure 6. The Nup84 complex molecular architecture revealed by CX-MS integrative pipeline

Suggested location – Results, Molecular architecture of the endogenous Nup84 complex revealed by integrative modeling

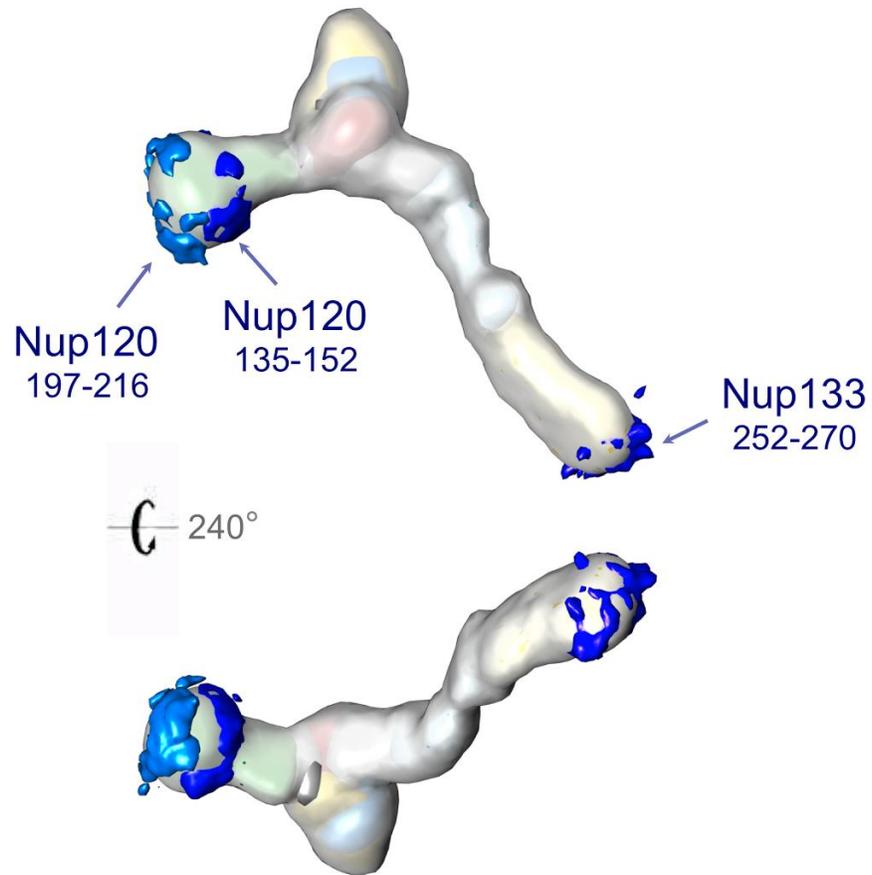


Figure 7. The putative membrane interacting ALPS motifs

Suggested location – Discussion, Putative membrane interacting motifs are localized to the periphery of the Nup84 complex