Integrative Structure Determination Using Proteomics Data: Application to the Nuclear Pore Complex

Phillips, Jeremy

2012

Integrative Structure Determination Using Proteomics Data: Application to the Nuclear Pore Complex

by

Jeremy L. Phillips

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biological and Medical Informatics

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

[redacted]

Acknowledgements

Graduate school is both an enlightening experience and a long, tough slog, and many people have helped me to make it through with my sanity intact; I'll try to get to most of them.

I'd first like to thank my advisor, Andrej Sali, who has taught me an enormous amount about how to be a good scientist. I joined his lab with barely the slightest idea of what I was doing, and have learned to identify important scientific problems, propose and carry out a research project on my own from start to finish, work effectively and happily with several collaborators and coworkers at once, and to always keep my eye on good science, even when competitors, research setbacks ("we still don't have enough data"), and the often ridiculous politics of academia get in the way. Andrej has an uncanny ability to always keep the big picture in mind, a skill that he imparts on just about everyone who works with him. He also grants his students an enormous amount of freedom, which, although sometimes it allows for temporary failure, has really helped to make me a better scientist.

Beyond just Andrej, the Sali lab is a group of incredibly smart, hilarious, eccentric, and fascinating people who I've loved working with for the last 5 years. Aside from the fact that I've had the privilege to work with these amazing scientists, the lab's ridiculous catalog of inside jokes, penchant for crazy conversations and various obsessions with things like high-end coffee, high-fidelity audio, and octopi have all made my life far more enjoyable in grad school, one way or another. I specifically have to thank a few members of the group.

Thanks to Libusha Kelly for being my rotation advisor, helping me to get started in the group and in grad school, and for putting up with sitting next to me and listening to occasionally cynical views on life; she'll be a phenomenal PI someday. Thanks to Keren Lasker for all the scientific collaboration and help over the years, and to both her and Avner Schlessinger for the being great friends and supporters, including many, many "hostile coffees" in which we discussed too many crises which turned out to be insignificant but seemed hugely important at the time; Avner and Keren will also both be great PIs. Thanks also to Dave Barkan and Javi Velazquez for being great friends, and for the all the late-night discussions about science, sports, and life in general. Thanks to Mike Kim for convincing me to come to UCSF for grad school in the first place. Thanks to Frank Alber for setting the groundwork for my project, to Daniel Russel for the enormous amount of work that's he's put into developing IMP, and to Ben, Ursula, and Elina for developing and maintaining all of the software that none of us could live without. And, of course, thanks to Robert (TTT), Vadim, Conchita, Karl, and possibly others for truly expanding my imagination. Thanks to many others in the lab who've helped along the way: Min-Yi, Dave Eramian, Eswar, Madhu, Peter (and all of the hard-working farmers of Slovenia), Charles, Pat, SJ, Hao, Riccardo, Max, GQ, Elina, Martin, Jeremy Horst, Natalia and others. And finally, thanks to our administrators, Hilary Mahon and Karin Asensio, for keeping the lab and Andrej running smoothly all these years.

Also, thanks to my undergraduate advisor, David States, and my friends and coworkers from the Michigan Bioinformatics program: Nick, Yuri, Carlos, Heather, Damien, Vasile, and Abhik. They imparted me with just the right amount of scientific knowledge and cynicism to help me get through my Ph.D. intact.

I'd also like to thank many collaborators both inside and outside of UCSF. Thanks especially to Mike Rout and Javi Fernandez-Martinez at Rockefeller University, with whom I've had a long and very productive collaboration. They've done a truly enormous amount of experimental work that has made my computational modeling work possible. They've also been highly engaged in my work, been incredibly helpful anytime they've had a chance, very understanding when I haven't quite produced the models they were hoping to see, and great hosts anytime I visit them in New York. Thanks also to Brian Chait, Bob Stroud, Steve Burley, Partha Sampathkumar, Steve Almo and Mark Field.

I also have to thank several additional UCSF professors and administrators. Tanja Kortemme and Robert Fletterick have been great thesis committee members and provided invaluable advice. Thanks to Chao Tang and Yifan Cheng, who served on my orals committee and helped me to get my projects started. Thanks to Matt Jacobson and Bruce Conklin for letting me learn about protein structure and SNP arrays, respectively, as a relatively clueless rotation student. Thanks to Patsy Babbitt, who provided great advice and was always willing to listen during my first few years of grad school. Finally, thanks to Julia Molla and Rebecca Brown, for keeping the grad program up and running and making sure I do all of the things I'd forget about otherwise.

Outside of the group I have to thank all of my friends and roommates who've helped to distract me and keep me sane. Special thanks to Sheel Dandekar, Sai Duriseti, Rebeca Choy, and Steve Guerrero, also known as "Vat Chunks, inc." or "Los Malandros," with whom I've enjoyed many ski and hiking trips, dinners out commiserating about life in science, and discussions about utterly hopeless startup ideas. Thanks to my roommates over the years including, Sheel (again), David, Jody, Sam, Itai, Robin, Courtney, and Shiloh and Nicala the dogs for putting up with my strange hours and work habits. Thanks also to Dan Black and Josh Katz, old friends who I'm grateful to have around in the Bay Area.

Finally, thanks to my family, especially to my parents, Randy and Beverly, and my sister, Emily, who've provided an incredible amount of love and support even though I didn't always call often enough or visit home as much as they would have liked. Thanks also to my grandfather Perry, and my grandmothers, who didn't quite make it this far, Bea and Berta, for their incredible support, along with my whole extended family. My family's love and encouragement helped me enormously in getting to UCSF in the first place, and in getting through my six years in grad school; I can never thank them enough.

Abstract

Proteomics techniques have been used to generate comprehensive lists of protein interactions in a number of species. However, relatively little is known about how these interactions result in functional multiprotein complexes. This dissertation describes a computational procedure that bridges this gap by combining low-resolution data from affinity purification experiments with data from a heterogenous set of standard structural biology techniques, including electron microscopy (EM) and X-ray crystallography. One such assembly is the Nuclear Pore Complex (NPC), which serves as the sole mediator of nucleocytoplasmic exchange in eukaryotic cells. We used our method to determine the structure of the ~600 kDa heptameric Nup84 complex, an essential component of the NPC. This work demonstrates that integrative approaches based on low-resolution data can generate functionally informative structures at intermediate resolution.

List of Publications Associated with this Dissertation

1. Lasker, K.*, **J. Phillips*,** D. Russel, J. Velazquez-Muriel, D. Schneidman-Duhovny, E. Tjioe, B. Webb, A. Schlessinger, and A. Sali. "Integrative Structure Modeling of Macromolecular Assemblies from Proteomics Data." *Mol Cell Proteomics* 9.8 (2010): 1689-702.

2. Sampathkumar, P., S. A. Ozyurt, J. Do, K. T. Bain, M. Dickey, L. A. Rodgers, T. Gheyi, A. Sali, S. J. Kim, **J. Phillips,** U. Pieper, J. Fernandez-Martinez, J. D. Franke, A. Martel, H. Tsuruta, S. Atwell, D. A. Thompson, J. S. Emtage, S. R. Wasserman, M. P. Rout, J. M. Sauder, and S. K. Burley. "Structures of the Autoproteolytic Domain from the Saccharomyces Cerevisiae Nuclear Pore Complex Component, Nup145." *Proteins* 78.8 (2010): 1992-8.

3. Sampathkumar, P., T. Gheyi, S. A. Miller, K. T. Bain, M. Dickey, J. B. Bonanno, S. J. Kim, **J. Phillips,** U. Pieper, J. Fernandez-Martinez, J. D. Franke, A. Martel, H. Tsuruta, S. Atwell, D. A. Thompson, J. S. Emtage, S. R. Wasserman, M. P. Rout, A. Sali, J. M. Sauder, and S. K. Burley. "Structure of the C- Terminal Domain of Saccharomyces Cerevisiae Nup133, a Component of the Nuclear Pore Complex." *Proteins* 79.5 (2011): 1672-7.

4. Webb, B., K. Lasker, D. Schneidman-Duhovny, E. Tjioe, **J. Phillips,** S. J. Kim, J. Velazquez-Muriel, D. Russel, and A. Sali. "Modeling of Proteins and Their Assemblies with the Integrative Modeling Platform." *Methods Mol Biol* 781 (2011): 377-97.

5. Fernandez-Martinez, J.*, **J. Phillips*,** M. D. Sekedat, R. Diaz-Avalos, J. Velazquez-Muriel, J. D. Franke, R. Williams, D. L. Stokes, B. T. Chait, A. Sali, and M. P. Rout. "Structure-Function Mapping of a Heptameric Module in the Nuclear Pore Complex." *J Cell Biol* 196.4 (2012): 419-34.

6. Sampathkumar, P., S. J. Kim, D. Manglicmot, K. T. Bain, J. Gilmore, T. Gheyi, **J. Phillips,** U. Pieper, J. Fernandez-Martinez, J. D. Franke, T. Matsui, H. Tsuruta, S. Atwell, D. A. Thompson, J. Spencer Emtage, S. R. Wasserman, M. P. Rout, A. Sali, J. Michael Sauder, S. C. Almo, and S. K. Burley. "Atomic Structure of the Nuclear Pore Complex Targeting Domain of a Nup116 Homologue from the Yeast, Candida Glabrata." *Proteins* (2012).

*authors contributed equally

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

The cell contains hundreds of functional macromolecular assemblies responsible for performing critical cellular processes [1, 2]. These include, among others, the ribosome (translation) [3, 4], chaperonins (protein folding) [5, 6], RNA polymerase (RNA synthesis) [7], and the proteasome (protein degradation) [8-11]. A macromolecular machine is often built around a stable core of proteins that defines the basic function of the complex. A structural description of an assembly generally facilitates a mechanistic understanding of the corresponding process [3, 12, 13]. Thus, a critical challenge in structural biology is to determine the structures of macromolecular assemblies at the highest possible resolution.

Integrating data to determine the structures of macromolecular assemblies

The structures of macromolecular assemblies generally cannot be resolved to atomic resolution by a single technique [14]. While X-ray crystallography remains the most powerful approach for visualizing a static snapshot of a complex at atomic resolution, it is limited to samples that can be purified in large quantities and crystallized [15]. Similarly, Nuclear Magnetic Resonance (NMR) spectroscopy results in an ensemble of structures of a system in solution [16-18], but the technique is limited by the size of the complex and sample availability. Electron microscopy (EM) techniques provide an alternative approach for visualizing multiple conformations of complexes in vitro and even within cells [19-

21]. However, in most cases, the resolution of an electron density map is too low to provide a full mechanistic description of a protein complex. Additional techniques, such as high-throughput proteomic methods [22], small-angle X-ray scattering (SAXS) [23, 24], and fluorescence resonance energy transfer (FRET) spectroscopy [25], are generally limited by low resolution [14], and at times by low accuracy [26-28], of the corresponding structural information.

Use of multiple data sources in structure determination

The limitations in the resolution, accuracy, and coverage of individual experimental methods can be bridged by simultaneous consideration of multiple types of information. Examples of techniques that specialize in integrating a few types of experimental data include (i) combining electron density maps of complexes with atomic structures of protein components to build high-resolution structures of protein complexes [29-31], (ii) using atomic models to estimate the phases required for converting diffraction data into electron density maps [32], (iii) inferring the binary interaction map of a complex from affinity purification, mass spectrometry, and comparative modeling data [33] , and (iv) incorporating NMR-derived data into protein structure prediction [34, 35].

A number of macromolecular structures have been resolved by such integrative methods. For instance, the constituent proteins in the Nuclear Pore Complex (NPC) were localized based on the shape and symmetry of the NPC from cryo-EM, positions of the proteins from immuno-EM, relative proximities of proteins from affinity purification, and the shapes of proteins from ultracentrifugation [13, 36]. An atomic model of the AAA-ATPase ring of the 26S proteasome was determined primarily by fitting comparative models of subunits into a single-particle cryo-EM map, subject to protein interactions identified by proteomics [10]. A structural model for a complete clathrin lattice [37] as well as a mechanistic model of the clathrin lattice assembly–disassembly cycle driven by chaperone Hsc70 [38] were suggested by combining data obtained by X-ray crystallography and single particle cryo-EM. The architecture of RNA polymerase II in complex with its initiation factors was determined by chemical cross-linking coupled to mass spectrometry [39]. An NMR solution structure for the interface between two subunits in the HIV-1 capsid was fitted to an electron density map of the whole complex, revealing a relative orientation of subunits different from that in the corresponding crystal structure [40].

Integrative structure determination by optimization of spatial restraints

The shortcomings of individual methods can be minimized by simultaneous consideration of all available information about a given assembly [41]. This information may vary greatly in terms of its accuracy and resolution, and includes data from both experimental and computational methods, such as those mentioned above. A four stage scheme has been described to formulate structure determination as an optimization problem for which the solutions are models of a given complex that are consistent with all available information about its composition and structure. These four steps are i) data gathering, ii) data representation and translation into spatial restraints, iii) sampling of good scoring structures, and iv) ensemble analysis (see Figure 3-5 for an example). First, structural data are generated by experiments, such as cryo-EM and affinity purification of subcomplexes. Second, representations are chosen for the target system's components (such as atomic, sphere-per-residue, sphere-per-protein) the data and theoretical considerations are expressed as spatial restraints that enforce properties such as the observed symmetry or shape of an assembly, or the proximities of the constituent proteins. Third, an ensemble of structural solutions that satisfy the data is obtained by minimizing the violations of the spatial restraints. Fourth, the ensemble is clustered into sets of distinct solutions, and analyzed in different representations, such as protein positions or protein–protein contacts.

This integrative approach to structure determination has several advantages. First, the combination of the input data minimizes the drawbacks of incomplete, inaccurate and/or imprecise data sets. Each individual restraint contains little structural information, but by concurrently satisfying all restraints derived from independent experiments, the ambiguity of structural solutions can be markedly reduced. Second, this approach has the potential to produce all structures that are consistent with the data, not just one structure. Third, the variation between the structures that are consistent with the data allows an assessment of whether there are sufficient data and how precise the representative structure is. Last, this approach can make the process of structure determination more efficient, by indicating which measurements would be the most informative.

Integrative structure determination has been applied to a number of macromolecular assemblies including the 26S proteasome [11] and chromatin structure [42, 43]. Here, we focus on an application to a single molecular assembly, the Nuclear Pore Complex.

The Nuclear Pore Complex (NPC)

Gatekeeper to the nucleus: Structure and function of the NPC

The defining component of a eukaryotic cell is its nucleus, which contains the cell's DNA and is separated from the cytoplasm by a double lipid bilayer called the nuclear envelope. The nuclear envelope surrounds and protects the chromatin, forming a barrier that spatially separates transcription from translation, and shielding the chromatin from biochemical activity in the cytoplasm. Nuclear pore complexes (NPCs) reside in the nuclear envelope and are the sole mediators of exchange of macromolecules between the nucleoplasm and cytoplasm in eukaryotic cells.

Cryo-electron tomography and studies of NPC composition [44-46] have shown that the NPC's architecture and protein makeup are relatively consistent across eukaryotes, though ranging in size from ~40 to 70 MDa depending on the organism. Here, we focus on the NPC in the yeast *S. cerevisiae*, in which the NPC has been most extensively studied.

The yeast NPC is an approximately 50 MDa, ~450 protein complex formed by ~30 unique proteins called nucleoporins (nups) [44, 47], each present in multiple copies and connected in biochemically stable subcomplexes that act as building blocks for the NPC [48, 49]. The complex has a donut-shaped structure, containing of eight spokes arranged radially around a central channel that serves as the conduit for macromolecular transport. Each NPC spans the nuclear envelope, connecting the outer nuclear membrane (ONM), which faces the cytoplasm and is connected to the rough ER, and the inner nuclear membrane (INM), which faces the nucleoplasm (Figure 1-1).

There are several classes of nucleoporins: transmembrane, core, linker, and FG nucleoporins. Transmembrane nucleoporins span the nuclear membrane and are thought to help anchor the NPC to the nuclear envelope. Core scaffold nucleoporins form the Nup84 (inner ring) and Nup170 (outer ring), complexes, which form the cage-like scaffold within the NPC that coats the nuclear membrane. FG nucleoporins are characterized by natively disordered domains that include multiple phenylalanine–glycine (FG) dipeptide repeats. Linker nucleoporins connect the FG nucleoporins to the core scaffold (Figure 1-1).

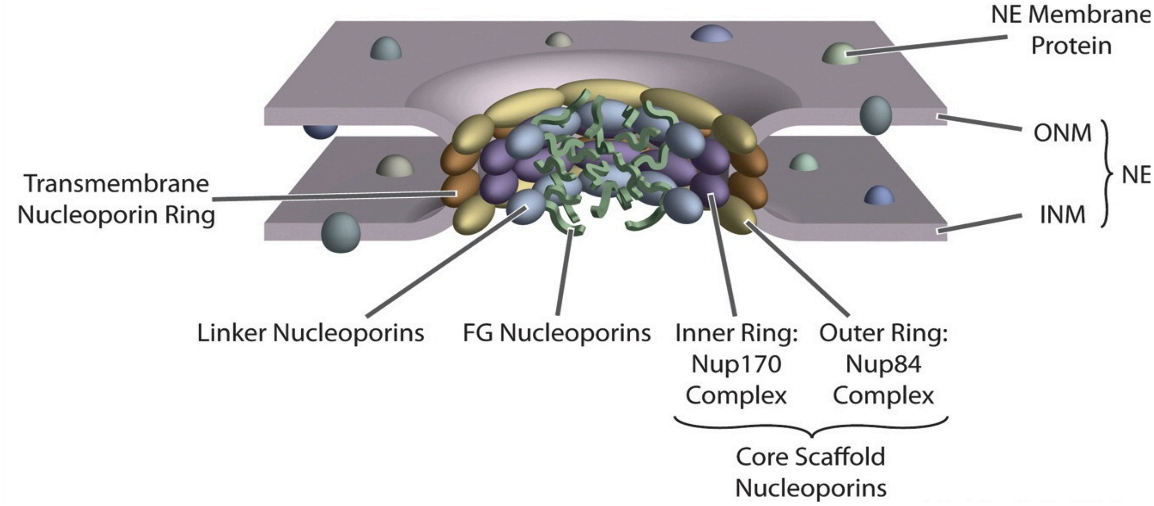


Figure 1-1. The structure of the Nuclear Pore Complex.
Major features of the NPC structure described in the main text are indicated [50].

The transport function of the NPC is primarily mediated by the FG nups. Their disordered domains fill the channel of the NPC, preventing most macromolecules from crossing the central tube. Transport across the NPC is a highly regulated process that involves many transport factors, primarily those belonging to a family of proteins called karyopherins (kaps). Kaps bind to specific import signals (nuclear localization signals, NLS) or export signals (nuclear export signals, NES) in their cargos. The disordered regions of FG nups provide binding sites for transport factor–cargo complexes and facilitate their passage across the NPC. Transport across the NPC is mediated by diffusion and stochastic interactions and is not thought to require major changes in the NPC's scaffold structure [50, 51].

Nuclear transport is powered by the Ran cycle, involving the 25Kda protein Ran. The enzyme Ran GTPase creates a gradient of GTP-bound Ran (RanGTP) that provides the energy for Kap-mediated transport. In the nucleoplasm, Ran is maintained in its GTP-bound form by the chromatin-associated nucleotide exchange factor RanGEF; in the cytoplasm, GTP hydrolysis by Ran is stimulated by RanGAP, resulting in higher RanGTP concentrations in the nucleus. During import, a Kap binds to its NLS-bearing cargo in the cytoplasm and transits the NPC. Once in the nucleus, RanGTP binds to the Kap, causing cargo release. During export, Kaps bind their NES-bearing cargo in the presence of RanGTP. Once the export complex passes through the NPC into the cytoplasm, conversion of Ran to its GDP-bound state triggers cargo release.

Models of NPC transport

The exact mechanism of transport regulation by the FG-nups is the subject of intense debate[52-55]. In one model for this process, known as the “virtual gate model,” the FG-repeat-containing filaments fill the NPC channel and exclude cargo not bound by nuclear transport receptors in an entropy-driven process. That is, any molecule that enters the NPC reduces the available space for the disordered regions of the FG nups, resulting in an entropic, and thus energetic cost. For kap-bound cargos, this energetic cost of entering the pore can be balanced by energetically favorable binding interactions between the kap-cargo complexes and the FG nups, such that kap-bound cargos are allowed to pass across the pore [56]. A second model, the “selective phase” model [57] suggests that the disordered portions of the FG nups interact with each other to form a selective three-dimensional sieve that excludes large molecules. A third model, the “saturated hydrogel” model [58] suggests that the FG-repeat-containing filaments form a “saturated hydrogel” within the NPC, forming a maximum number of FG-FG interactions to generate a highly ordered mesh with even pore sizes. In both this and the “selective phase” model, the kaps are thought to temporarily dissolve the FG sieve and mesh, respectively, facilitating transport across the pore [53].

Although NPC transport is primarily mediated by the disordered regions of the FG nups, determining a high-resolution structure of the nuclear pore core scaffold would help to localize the structured anchoring sites for the FG nups. Precise locations of these anchoring sites would be helpful in visualizing the organization of the FG nups and in constructing simulations of transport across the NPC, which in turn could help to substantiate one of the transport models mentioned above.

Evidence for shared ancestry with coating complexes

A relationship between the components of the NPC and the components of membrane coats, specifically COPI, COPII and clathrin/adaptin has been suggested by crystal structures of components of these complexes and by bioinformatics fold predictions. Fold predictions first predicted that the NPC and coating complexes contain proteins consisting of one or two iterations of a β-propeller fold, an α-solenoid-like fold or both in the order β-α [59]. The relationship is most evident for Sec13, a protein shared by both COPII and the NPC. Specifically, the core scaffold of the NPC, which comprises ~50% of the NPC mass, is entirely composed of β-propeller and α-solenoid-like proteins. Based on these similarities, the protocoatomer hypothesis proposed that NPCs and clathrin, COPI, and COPII vesicle coats share a common evolutionary origin in an early membrane-curving module, the ‘protocoatomer’ [59-62]. Protein structures determined by X-ray crystallography and electron microscopy have strongly supported this hypothesis, which greatly reduces the likelihood that these complexes arose via convergent evolution [37, 63-68, 69 2009, 70-74]. Furthermore, similar architectures are shared with additional complexes associated with coating systems; for example, the α-solenoid is present in NPC-interacting karyopherins [60] and even subunits of the retromer complex, involved in Golgi/late endosomal transport, though their evolutionary origins are currently unclear [75]

The presence of multiple paralogous families in membrane-associated systems and their importance in defining organelle identity provides a potential modular route for evolution of new compartments, as described by the organellar paralogy model [76]. The model proposes that individual components arise by paralogous expansion and may function within pre-existing complexes, but diverge into new complexes by step-wise subunit replacement/sequence divergence, creating diversity and new compartments. The protocoatamer may have diverged into current coating complexes in this manner.

By determining a high-resolution structure of the NPC scaffold, we will be able to further compare the NPC-coating meshwork to the quaternary structures of other coatamers and achieve of deeper understanding of their evolutionary relationships. In addition, any added understanding of the structure and formation of these coating complexes may also have theraputic benefits. For instance, clathrin is directly involved in endocytosis; an better understanding or ability to control the formation of clathrin coated vesicles could have profound impacts on the delivery of drugs to cytoplasmic drug targets.

The NPC in human health and disease

Because transport across the NPC is selective and highly regulated, it is a key focal point for controlling critical cell functions. Not surprisingly, several nucleoporins have been shown to be associated with cancers and human genetic disorders. In addition, almost all viruses of eukaryotic organisms co-opt the machinery of the NPC to inject genetic material into the host cell's nuclei. Finally, recent evidence has emerged suggesting that NPC scaffold degradation may play a role in aging.

The most prominent example of association of nup variants with cancer involves the human FG nup Nup98, which is involved in at least 15 gene fusions producing oncogenic fusion proteins that may function by overactivating transcriptional machinery [77]. Human Nup214 is also involved in oncogenic gene fusions [78]. In addition, Nup88 overexpression has been associated with several cancer types [79].

Specific host-pathogen interactions involving NPC proteins have been identified in the cases of viral infection. For instance, adenoviruses, which are too large to enter the nucleus intact, dock at the NPC; Nup214 has been identified as receptor for some adenovirus capsids [79]. In addition, several NPC proteins, including Nup98, Nup153, and Nup214 have been shown to interact with proteins in the HIV pre-integration complex, which carries RNA from the virus into the nucleus [80, 81]. These protein-protein interactions may be viable targets for disrupting the virus lifecycle and preventing infection; understanding their structures and roles in the NPC may aid in targeting them with therapeutics.

Finally, recent evidence has suggested that there is very little turnover among scaffold nucleoporins in non-dividing cells in worm[82] and rat brains[83]. These extremely long-lived proteins may be subject to buildup of aberrant chemical modifications and oxidative damage over time. In fact, degradation of these proteins appears to create increasingly leaky NPCs over time. Specifically, fluorescence microscopy detected buildup of labeled dextran, a large polysaccharide, and tubulin BIII, a strictly cytoplasmic protein, in the nuclei of cells in older *C. elegans* and rat brains, whereas almost none of these macromolecules could be detected in the nuclei of younger cells [82]. Defective protein-protein interactions that result from age-dependent protein degradation may prove to be viable drug targets in the future; determining a high-resolution structure of the NPC should help to identify critical protein-protein interactions lost to degradation.

Previous efforts to improve structural knowledge of the NPC

Investigations of NPC structure by individual experiments

Electron micrographs that detected nuclear pores were first shown in 1959 [84]. Since then many studies have led to increases in knowledge and understanding the NPC structure. Electron microscopy has been the most prominently used single experimental technique used to observe the overall structure of the NPC. Scanning electron microscopy experiments have showed the general NPC architecture and 8-fold symmetry around the pore channel in *Xenopus laevis* [85], *Drosphila melanogaster* [86], and *S. cerevisaie* [87]. Cryo-electron tomography reconstructions have provided further visualizations of the NPC structure in *X. laevis* [88, 89], *Dictyostelium discoideum* [90], and human [91] describing the NPC structure at a maximum resolution of about 6nm. Coarse nup localizations were also provided by electron microscopy of NPCs with immunogold labeled components in yeast [92].

Outside of electron microscopy, an inventory of yeast NPC protein composition was generated more than a decade ago [44] by fractionation and mass spectrometry. Since then, X-ray crystallography has been used to determine structures of many of the individual proteins, along with some dimers and trimers in the NPC scaffold [37, 63-74]. The combination of these high-resolution structures of small components and coarse descriptions of the whole structure provides obvious opportunity to improve structural knowledge with data integration.

Integrative determination of the molecular architecture of the NPC

Some of this experimental data was used as input data to the four-stage integrative structure determination technique described previously, to produce a map of the molecular architecture of the NPC in 2007 [13, 36]. This map describes the relative positions of the nups in the native yeast NPC in a localization map, determining the positions of the proteins at about ~5nm precision.

To determine the molecular architecture of the NPC, a large and diverse set of biophysical and biochemical data was collected and translated into spatial restraints on the nups, including data from ultracentrifugation experiments, stoichiometry data from quantitative immunoblotting experiments, affinity purification data, data from overlay assays, and data from electron microscopy experiments. The optimization relied on conjugate gradients and molecular dynamics optimization with simulated annealing, starting with a random configuration of nups and then iteratively moving these nups so as to minimize violations of the restraints. The final localization map was generated from an ensemble of 1,000 independently calculated structures that satisfied the input restraints [13, 36].

Competing models of the NPC scaffold structure

The 2007 molecular architecture of the NPC is currently one of 3 prevailing models for the overall structure of the NPC core scaffold. According to this computationally determined model, groups of eight Nup84 complexes each self-assembles into a cytoplasmic and a nucleoplasmic ring (the “outer rings” of the NPC, see Figure 1-1). Sandwiched between the two outer rings are two eight-membered rings of the 4-protein Nup170 complex, the “inner rings.” In a second model for NPC structure, termed the “fence pole model,” heterooctameric poles of the nucleoporins Nup145C, Sec13, Nup85, and Seh1 observed in crystal structures organize four rings of 8 Nup84 structures each [67]. These four rings form a cylinder adjacent to the transmembrane nups on the membrane side and, moving inward toward the transport channel, suggest two additional concentric cylinders forming layers of “adaptor nups,” (inner ring nups in the compuatational model), followed by channel nups (linker nups). This model assumes an alternative stoichiometry to the one used based on quantitative immunoblotting data in the computational model. In a third model, the “lattice model,” 8 Nup84 complexes each coat the pore on the cytoplasmic and nucleoplasmic sides of the pore, oriented vertically, with their longest axes in parallel to the pore channel, in contrast to the horizontal rings suggested by the computational model [68, 93]. This model was generated in attempt to mimic the β-propeller - β-propeller interactions in the COPII coat. More information and an improved NPC models are needed to distinguish among the currently competing models.

Strategy to determine a high-resolution NPC structure

Further elucidation of the evolutionary origin, transport mechanism, and assembly pathway of the NPC requires higher resolution information, encompassing the atomic structures of nups and their intermolecular arrangements. To improve upon the resolution and accuracy of the current NPC structural model, it is necessary to incorporate crystallographic and modeled atomic structures of nups into the low-resolution model of the molecular architecture of the NPC, along with lower-resolution data.

We have opted to break down this process by solving the structures of well-established subcomplexes of the NPC individually, integrating high-resolution data from X-ray crystallography applied to individual proteins and protein dimers with EM and proteomics data, and predictions from bioinformatics techniques such as comparative structure modeling[94]. Our goal is to determine the NPC structure at the resolution sufficient to provide relative orientations for the constituent protein domains, and provide a clearer picture of the organization of NPC membrane coating core scaffold. We have worked toward this goal using Integrative Modeling Platform (IMP) software package [95], which provides a framework to carry out the four-stage integrative structure determination scheme described earlier. Once we have the structures of the individual NPC subcomplexes, we can move on to integrate all available information and solve for a higher-resolution structure of the full complex.

Chapter 2 of this dissertation discusses novel methods for introducing proteomic data into the four-stage integrative structure determination scheme, and Chapter 3 discusses the use of integrative structure determination to determine the structure of a key NPC subcomplex, the Nup84 complex.

Chapter 2. Integrative Structure Determination Using Affinity Purification Data

Introduction: affinity purification experiments as sources of structural data

Affinity purification is among the most commonly applied general techniques in proteomics. Various forms of affinity purification have been used to identify interacting sets proteins on a massive scale, notably to catalog protein interactions in yeast organism-wide [96-100]. However, the spatial information provided by affinity purification experiments is rarely considered. Here, we describe a method to exploit this information with spatial restraints, for use in protein assembly structure determination.

In affinity purification experiments, one protein in a complex (the bait) is tagged. The bait and its noncovalently associated partners (the subcomplex) are first purified by affinity chromatography against the tag, using a strong binder such as an antibody or ligand, and then identified by SDS-PAGE and mass spectroscopy [44, 47, 101, 102]. Each affinity purification experiment, in principle, provides some information about spatial relationships among the subunits in the subcomplex.

Each experiment defines a set of directly or indirectly interacting proteins, or a “composite.” At least one copy of each protein in a set directly interacts with at least one copy of another protein in the set; however, affinity purification data does not provide information on the stoichiometry of the proteins in the set, the number of complexes with distinct stoichiometry and configuration, nor exactly which binary interactions occur, thus resulting in a great deal of ambiguity in the structural interpretation of the results.

Ambiguous conditional connectivity restraints for affinity purification data

Ambiguous information must be translated into a restraint that considers all possible structural interpretations of the data. Such a restraint might select the components to be restrained before each structure optimization step. We refer to such restraints as *conditional restraints* [103].

To account for the ambiguity of binary interactions in an affinity-purified composite, we encode each composite set as a specialized conditional connectivity restraint, which allows for all possible combinations of binary interactions that might satisfy the requirements of the composite. These restraints allow for simultaneous optimization of the connectivity graph of the proteins and the distances between directly interacting proteins. The conditional restraint first enforces the requirement that selected interactions must form a spanning tree of the “composite graph,” a graph that defines all possible binary interactions within the composite. The edges of the composite graph are scored using a harmonic upper bound score on the two proteins in each plausible binary interaction; given the current configuration of a system, this restraint score is highest for proteins that are furthest away from an ideal protein-protein interaction distance. The restraint enforces the spanning tree requirement by calculating a minimal spanning tree on the graph, thereby picking the smallest putative set of binary protein-protein interactions that least violate the restraints. Once the minimal spanning tree is chosen, the scores along the edges of the tree can be used to optimize the protein-protein distances for the binary interactions, bringing proteins closer to protein-protein interaction distance. During the optimization of a system, the connectivity of the complex can be chosen repeatedly at each optimization step, such that the connectivity restraints always account for the most recently calculated state of the target protein assembly [13, 36]

To illustrate our integrative modeling approach and the use of affinity purification data within this approach, we have used real experimental data to determine the known architecture of the human RNA polymerase II.

Integrative determination of the known structure of RNA Polymerase II

The structure and composition of RNA polymerase II

The eukaryotic RNA polymerase II (RNAPII) is a central multi-protein machine that synthesizes messenger RNAs and small nuclear RNAs. It is composed of 12 protein subunits, with a total molecular weight of 514 kDa (Figure 2). Ten subunits (Rpb1, Rpb2, Rpb3, Rpb5, Rpb6, Rpb8, Rpb10, Rpb11, and Rpb12) form a structurally conserved core, while the Rpb4-Rpb7 heterodimer is located on the periphery [104, 105]. While the atomic structure of the *Saccharomyces cerevisiae* RNAPII has been solved by X-ray crystallography [106], the human

RNAPII has not been determined at the atomic resolution, mostly due to difficulties in obtaining sufficient quantities of pure sample. However, the molecular architecture of the human RNAPII can be informed by that of its yeast homolog, based on the homology between their constituent proteins [107]. Below, we demonstrate that our integrative structure determination procedure can be used to accurately model the known architecture of human RNAPII (HRNAPII), using only proteomics-derived protein interactions, an EM map at 20 Å resolution, comparative models of the protein subunits based on yeast and human crystallographic structures, and geometric complementarity between the interacting subunits. Below, we describe the input data used for the modeling, the translation of this data into spatial restraints, an optimization procedure for determining the model architectures that satisfy the restraints, and an analysis of the resulting set of solutions. We then use a previously determined crystallographic structure of the full complex in yeast [108] to evaluate our results. If the crystallographic structure of yeast RNAPII had not been previously determined, the model resulting from our integrative structure determination procedure below would be the highest resolution description of HRNAPII, built without any crystallographic structures of any multiprotein complexes.

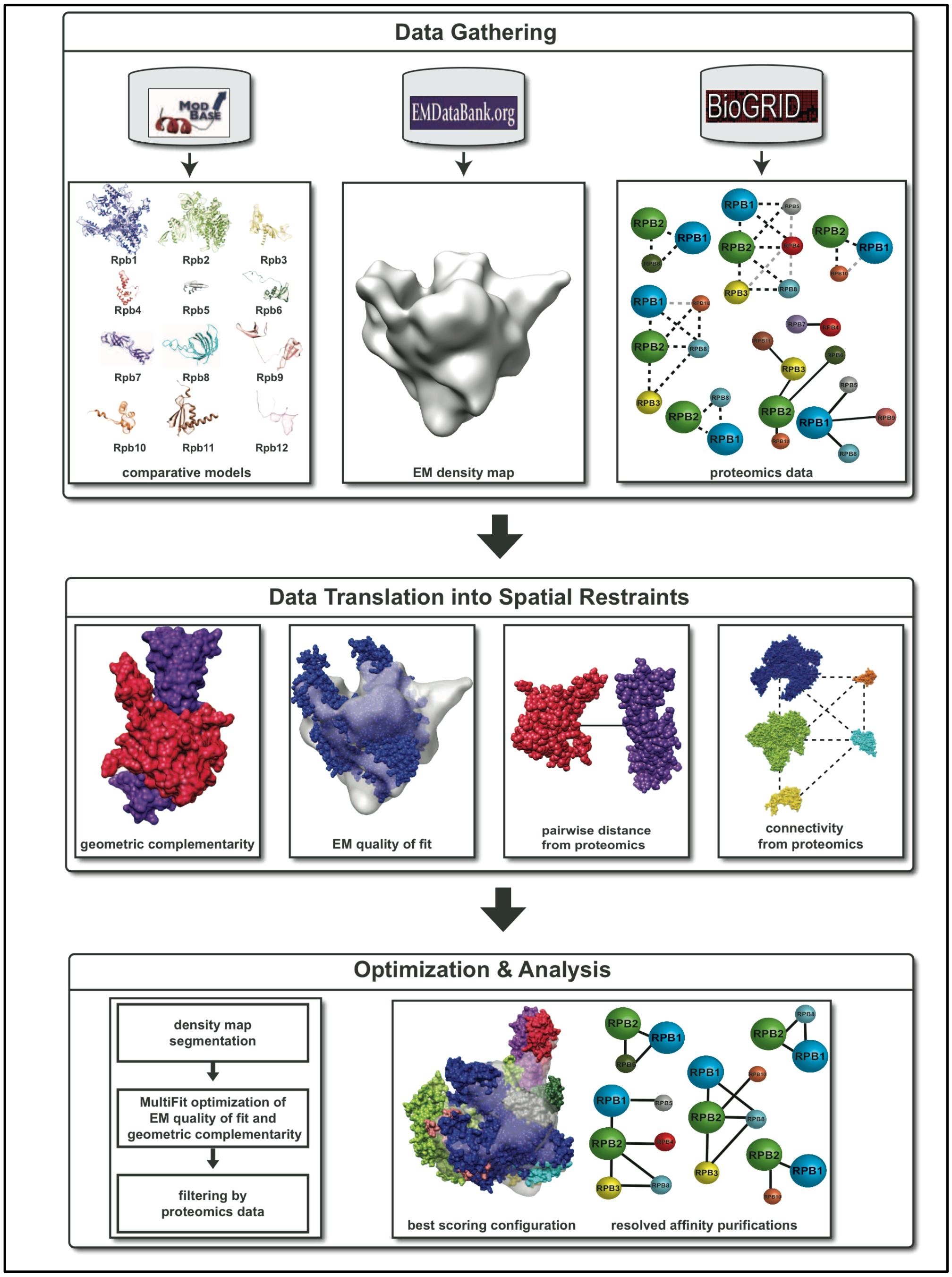


Figure 2-1: Determining the molecular architecture of the human RNAPII. Data gathering: Comparative models of the HRNAPII subunits were obtained from the ModBase database [109]. A density map of HRNAPII at 20 Å resolution was obtained from EMDB [110]. Proteomics data for *S. cerevisiae* RNAPII subunits was obtained from BioGRID (Table 2-2) [111]. Pairwise direct interactions are visualized in a single graph with full edges, and each pulldown experiment is presented as a graph with dash edges to indicate the missing underlying binary interaction network. Gray edges indicate interactions present in BioGRID but not in the yeast RNAPII crystal structure. Scoring: The scoring function is the sum of the distance, connectivity, EM, and geometric complementarity restraints. Optimization: The configuration of the subunits in HRNAPII was optimized using an extension of the MultiFit protocol to incorporate proteomics based restraints. The optimization procedure resulted in a single model that satisfied all of the input restraints.

Data gathering

To determine the molecular architecture of HRNAPII, we have constructed a realistic scenario in which we use structural homologs of individual human protein subunits found in the ModBase database [109] (Table 2-1), proteomics data for yeast RNAPII subunits extracted from the BioGRID database [111] (Table 2-2), and an assembly density map of HRNAPII determined at 20 Å resolution by single-particle cryo-EM [107] deposited in the EM databank [110].

Representing and translating data into spatial restraints

Representation of the RNAPII components

Given the availability of high-accuracy comparative models for the subunits, we represented the structures of the HRNAPII subunits at atomic resolution. We used atomic models found in the ModBase database of comparative models for domains in ~2.3 million protein sequences that are detectably related to known structures[109] (Table 2-1).

Subunit (name, UniProt id)	% Sequence identity	Number of residues, residue range	Template (PDB id and chain id, residue range)
Rpb1, P24928	55	1970, 11-1475	1i6hA, 7-1445
Rpb2, P30876	63	1174, 15-1171	2vumB, 20-1216
Rpb3, P19387	47	275, 7-264	1twfC, 6-263
Rpb4, O15514	100	142,14-142	2c35A, 14-142
Rpb5, P1938	52	210, 146-209	1hmjA, 10-73
Rpb6, P61218	100	127,1-127	1qklA, 801-927
Rpb7, P62487	100	172,1-171	2c35B, 1-171
RPB8, P52434	100	150,1-150	2f3iA, 1-150
Rpb9, P36954	47	125, 15-124	1twfI, 5-111
Rpb10, P62875	73	67,1-64	1twfJ, 1-65
Rpb11, P52435	52	1-105, 117	1twfK, 1-105
Rpb12, P53803	38	1-70, 70	2e2hL, 70

Table 2-1: Representation of RNAPII components

Distance restraints from proteomics

Eight direct physical interactions between pairs of eukaryotic RNAPII subunits were determined by yeast two-hybrid (Y2H) system [112-120], protein complementation assays [121], co-localization [104], and complex reconstitution experiments [122, 123] (Table 2-2); these data were retrieved from the BioGrid database. There are additional "interaction" data in BioGrid. However, because BioGrid does not annotate which interactions are physical as opposed to indirect, we encoded as contact distance restraints only those experimentally measured interactions that have been detected by "pairwise" methods listed above.

Our distance restraint is a harmonic function of the restrained feature, of the form $k \cdot x^2$, where *x* is the distance from the mean and *k* is the force constant of the restrained feature. In general, distance restraints may operate on multiple scales, ranging from the distance between two atoms or residues to the distance between two protein centers in an assembly. For example, if a direct interaction between two proteins has been identified, we may apply a restraint that penalizes deviations from the distance between the two protein centers. This distance restraint scores equally all relative orientations between the two proteins with the same inter-center distance. When the shape of the interacting proteins is known, we can achieve a more accurate score at the cost of additional computational time by restraining the distance between the closest pair of particles across the protein-protein interface, as done here.

Connectivity restraints from proteomics

Five sets of physically interacting RNAPII subunits are revealed by affinity purification and mass spectrometry (Table 2-2) [97, 98]. Each purification defines a set of directly or indirectly interacting proteins. At least one copy of each protein in a set directly interacts with at least one copy of another protein in the set; however, affinity purification data does not provide information on the stoichiometry of the proteins in the set, the number of complexes with distinct stoichiometry and configuration, nor exactly which binary interactions occur, thus resulting in a great deal of ambiguity in the structural interpretation of the results. Because of this ambiguity, each affinity-purified set is encoded as a connectivity restraint, which optimizes the assignment of binary interactions to proteins in the set along with the configuration of proteins [36]. A putative binary interaction network for the proteins that best satisfies all available data for the system is assigned (as a set of distance restraints) during the evaluation of a restraint at each optimization step.

Interacting subunits	Source method
Rpb1, Rpb2, Rpb10	Affinity captured MS [97]
Rpb1, Rpb2, Rpb3, Rpb4, Rpb5, Rpb8	Affinity captured MS [97]
Rpb1, Rpb2, Rpb8	Affinity captured MS [97]
Rpb1, Rpb2, Rpb3, Rpb8, Rpb10	Affinity captured MS [97]
Rpb1, Rpb2, Rpb6	Affinity captured MS [97]
Rpb1, Rpb5	Y2H [112, 113]
Rpb1, Rpb8	Y2H [114]
Rpb1, Rpb9	Y2H [112]
Rpb2, Rpb3	PCA [121]
Rpb2, Rpb6	PCA [121]
Rpb2, Rpb10	Y2H [112]
Rpb3, Rpb11	Y2H [112] Reconstituted-Complex [122] PCA [121]
Rpb4, Rpb7	Y2H [115-120], Reconstituted-Complex [123]

Table 2-2: Proteomics data used for modeling the architecture of RNAPII

Quality of fit restraint from an EM map

The configurations of the HRNAPII subunits are restrained to fit an electron density map of the yeast RNAPII complex [107]. The fit of a model into an assembly density map is assessed by a cross-correlation measure between the assembly density and the model smoothed to the resolution of the map [124].

Excluded volume restraint

Molecules take up space that can't be occupied by other molecules. This space filling property provides a key restraint on the conformations of the assembly. When the structure of a molecule is not known, it can be represented by a sphere; the volume of the sphere can be estimated from its composition (*e.g.*, the number of residues in a protein [125]). If the atomic structure is known, as is the case for HRNAPII, the van der Waals radius for each atom is used to define the excluded volume [126].

Geometric complementarity restraint from first principles

Protein-protein interfaces are typically geometrically complementary, characterized by tight packing with little space between them. The corresponding geometric complementarity is commonly used as a restraint in protein-protein docking [127, 128]. Because atomic models are used for HRNAPII subunit structures, this consideration was enforced with an explicit restraint [127]. The geometric complementarity restraint is more error-prone if used at coarser representations, because it is overly sensitive to shape approximation errors.

Sampling good scoring structures

To find structures of HRNAPII that fit the data, we use the sum of the distance, connectivity, EM quality of fit, and geometric complementarity restraints described above as a scoring function to optimize the structure and generate good scoring solutions.

The configuration of the subunits in HRNAPII was optimized using an extension of the divide-and-conquer MultiFit protocol (Figure 2-1) [31]. We began by segmenting the EM density map into 12 regions, each one of which served to localize one of the 12 constituent HRNAPII proteins. We then flagged neighboring pairs of regions that could potentially contain interacting pairs of proteins. This procedure resulted in 479,001,600 (12!) possible HRNAPII subunit configurations. Next, we eliminated all HRNAPII subunit configurations that did not satisfy a majority of the proteomics restraints (Table 2-2), keeping only 2,576 configurations for further refinement. We then refined each of these 2,576 configurations to optimize the EM quality of fit and geometric complementarity, using the standard MultiFit protocol [31]; 63 of the 2,567 configurations resulted in refined models with “good” scores. These models had equivalent positions for Rpb1, Rpb2 and Rpb3; however, the models varied in the positions of the remaining subunits. Finally, we filtered the 63 models by all proteomics restraints, resulting in a single model that satisfied all proteomics restraints as well as the EM quality of fit and geometric complementarity restraints (Figure 2-2).

Analyzing and assessing the ensemble of structures

Precision

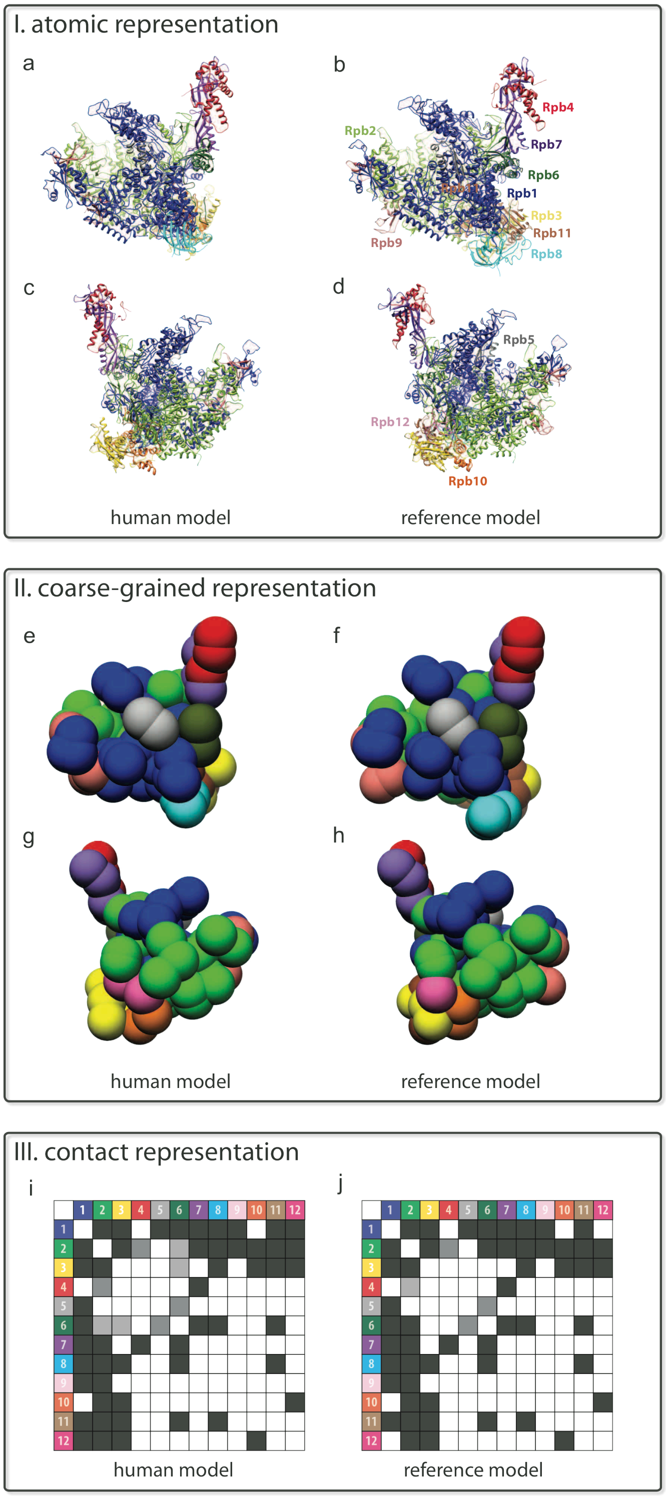
There are three possible outcomes of an optimization procedure. First, if only a single structural model satisfies all restraints, and thus all input information, there is probably sufficient data for prediction of the unique native state. Second, if two or more different models are consistent with the restraints, the data are insufficient to define the single native state, or there are multiple native structures. If the number of distinct models is small, the structural differences between the models may suggest additional experiments to narrow down the possible solutions. Third, if no models satisfy all input information, the data or their interpretation in terms of the restraints are incorrect. For example, it might be that a complex exists in several functional states and that the available data covers more than one of them.

In the case of our HRNAPII model, optimization resulted in a single model that satisfied all the data. Therefore, sufficient information was available to predict the positions and orientations of the HRNAPII subunits. The set of possible models in the absence of proteomics data was much larger (2,576 coarse configurations), and defined the structure far less precisely. In other words, low resolution proteomics data was crucial for providing an unambiguous determination of a precise molecular architecture of HRNAPII.

Accuracy

Assessing the accuracy of a structure is important and difficult [103]. The accuracy of a model is defined as the difference between the model and the native structure. Therefore, it is impossible to know with certainty the accuracy of the proposed structure without knowing the real native structure. Nevertheless, our confidence can be modulated by five considerations: (a) self-consistency of independent experimental data; (b) structural similarity among all configurations in the ensemble that satisfy the input restraints; (c) simulations where a native structure is assumed, corresponding restraints are simulated from it, and the resulting calculated structure is compared with the assumed native structure; (d) confirmatory spatial data that were not used in the calculation of the structure (*e.g.*, criterion similar to the crystallographic free R-factor [129] can be used to assess both the model accuracy and the harmony among the input restraints); and (e) patterns emerging from a mapping of independent and unused data on the structure that are unlikely to occur by chance [13, 36].

In the case of HRNAPII, we can estimate the accuracy directly, because we know the crystallographic structure of the yeast RNAPII, which is likely to be highly similar to that of HRNAPII [107] (*c.f.*, the high degree of sequence similarity between yeast and human subunit orthologs (Table 2-1) and the high correlation coefficient of 0.65 between the crystallographic yeast RNAPII structure and the EM map of HRNAPII). The HRNAPII model clearly recapitulates the molecular architecture of yeast RNAPII (Figure 2-2), preserving all of its protein interactions. More quantitatively, the subunits in the HRNAPII model share a Cα RMSD of only 11.4 Å with the human subunits individually superposed on their orthologs in the yeast RNAPII structure.



	1	2	3	4	5	6	7	8	9	10	11	12
1	□	■	■	□	■	■	■	■	■	□	■	■
2	■	□	■	▒	□	▒	■	■	■	■	■	■
3	■	■	□	□	□	▒	□	■	□	■	■	■
4	□	▒	□	□	□	□	■	□	□	□	□	□
5	■	□	□	□	□	▒	□	□	□	□	□	□
6	■	▒	▒	□	▒	□	■	■	□	□	■	□
7	■	■	□	■	□	■	□	□	□	□	□	□
8	■	■	■	□	□	■	□	□	□	□	■	□
9	■	■	□	□	□	□	□	□	□	□	□	□
10	□	■	■	□	□	□	□	□	□	□	□	■
11	■	■	■	□	□	■	□	■	□	□	□	□
12	■	■	■	□	□	□	□	□	□	■	□	□

human model

	1	2	3	4	5	6	7	8	9	10	11	12
1	□	■	■	□	■	■	■	■	■	□	■	□
2	■	□	■	▒	□	■	■	■	■	■	■	■
3	■	■	□	□	□	□	□	■	□	■	■	■
4	□	▒	□	□	□	□	■	□	□	□	□	□
5	■	□	□	□	□	▒	□	□	□	□	□	□
6	■	■	□	□	▒	□	■	■	□	□	■	□
7	■	■	□	■	□	■	□	□	□	□	□	□
8	■	■	■	□	□	■	□	□	□	□	■	□
9	■	■	□	□	□	□	□	□	□	□	□	□
10	□	■	■	□	□	□	□	□	□	□	□	■
11	■	■	■	□	□	■	□	■	□	□	□	□
12	□	■	■	□	□	□	□	□	□	■	□	□

reference model

Figure 2-2. Comparison of the crystallographic structure of yeast RNAPII and the integrative model of human RNAPII.
(a-d) Atomic representations of the integrative model of HRNAPII and the reference structure in two views; the reference structure is composed of human subunits individually superposed on their orthologs in the yeast RNAPII structure. The configuration of the HRNAPII subunits (a,c) is very similar to that in the reference structure (b,d); the Cα RMSD is only 11.4 Å. (e-h) Coarse representation of the HRNAPII model (e,g) and the reference structure (f,h) in the same two views as in (a-d) further illustrates the high similarity between the model and the reference. In the coarse representation, sets of 30 contiguous residues are shown as a single bead. (i-j) Protein contact maps for the HRNAPII model and the reference structure (white, no contact; gray, weak contact; black, contact). The maps are essentially identical, differing only in the interactions of Rpb6 with Rpb2 and Rpb3.

Discussion

As illustrated in the above example, proteomics techniques can now facilitate the characterization of the structure of macromolecular assemblies *via* integrative modeling. We have demonstrated that by using atomic subunit structures, an EM map of an assembly, and proteomics data restraining relative subunit proximities, we can extend the scope of macromolecular structure determination beyond what is possible with single methods. Specifically, using the RNAPII structure as an example, we have shown that proteomics data, although traditionally not considered a source of formal structural information, could help to reduce ambiguity in a set of possible assembly models.

Integrative structure determination with truncated affinity purification data

A typical affinity purification experiment without provides a group of interacting proteins that can be represented as a composite graph. However, higher resolution information – interactions assigned to smaller pieces than proteins – is needed to achieve a higher resolution structure. To resolve this issue, we have developed modified versions of affinity purification experiments that use protein domains and small protein fragments as baits rather than whole proteins as well as a modified conditional connectivity restraint to convert data from these experiments into spatial restraints. In the experiments, putative domain boundaries are identified, the proteins cut at these boundaries, and the resulting fragments tagged and used as bait in an affinity purification experiment (Figure 2-

3). This experiment produces two types of information. First, the proteins that elute with the tagged fragment produce a composite along with the tagged fragment itself (positive data). Second, we can assume that, if any proteins are missing as compared to the original (full-length bait) composite, interactions between the missing proteins and the truncated protein must have occurred through the truncated region. This creates a second composite, which includes the missing portion of the bait protein and any proteins lost in switching from full-length bait to a fragment bait (positive-negative data).

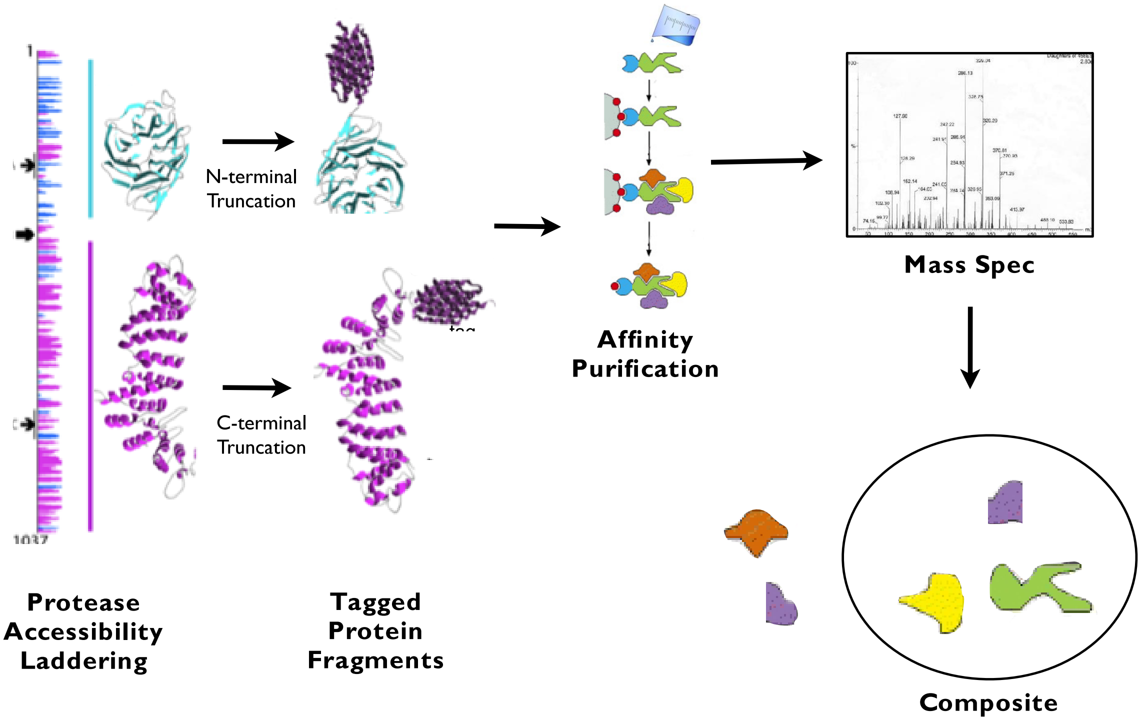


Figure 2-3. Truncated affinity purifications.
Proteins are divided into domains or small fragments based on protease accessibility laddering data[61] or other knowledge localizing flexible linkers in the protein sequence. These domains and fragments are then tagged and affinity purified, producing two types of information. The proteins that elute with the tagged fragment produce a composite along with the tagged fragment itself (positive data, proteins inside the circle). Any proteins that previously eluted with the full-length bait protein, but did not elute with the fragment form a second composite, which includes the missing portion of the protein and the lost proteins (positive-negative data, the orange protein and fragment outside the circle).

In encoding these data as spatial restraints (Figure 2-4), each composite is represented by a graph of nodes connected by edges, as previously. Each node represents a protein or a protein fragment, and each edge represents a possible interaction between connected nodes. Edges are generated for all possible interactions that might be implied by the corresponding pullout experiment. An edge imposes a putative distance restraint, scoring well when the two potentially interacting proteins or fragments are actually interacting in the assessed assembly configuration, and scoring poorly when the proteins (fragments) are too far apart to interact. As previously, a minimal spanning tree (MST) is calculated on the composite graph. Distance restraints for each of the MST edges are then added to the overall scoring function, and thus used to drive the optimization of the complex. Taken together, restraints for many of these truncated affinity purifications can solve for the domain-domain orientation of all subunits in a complex, as will be demonstrated in Chapter 3.

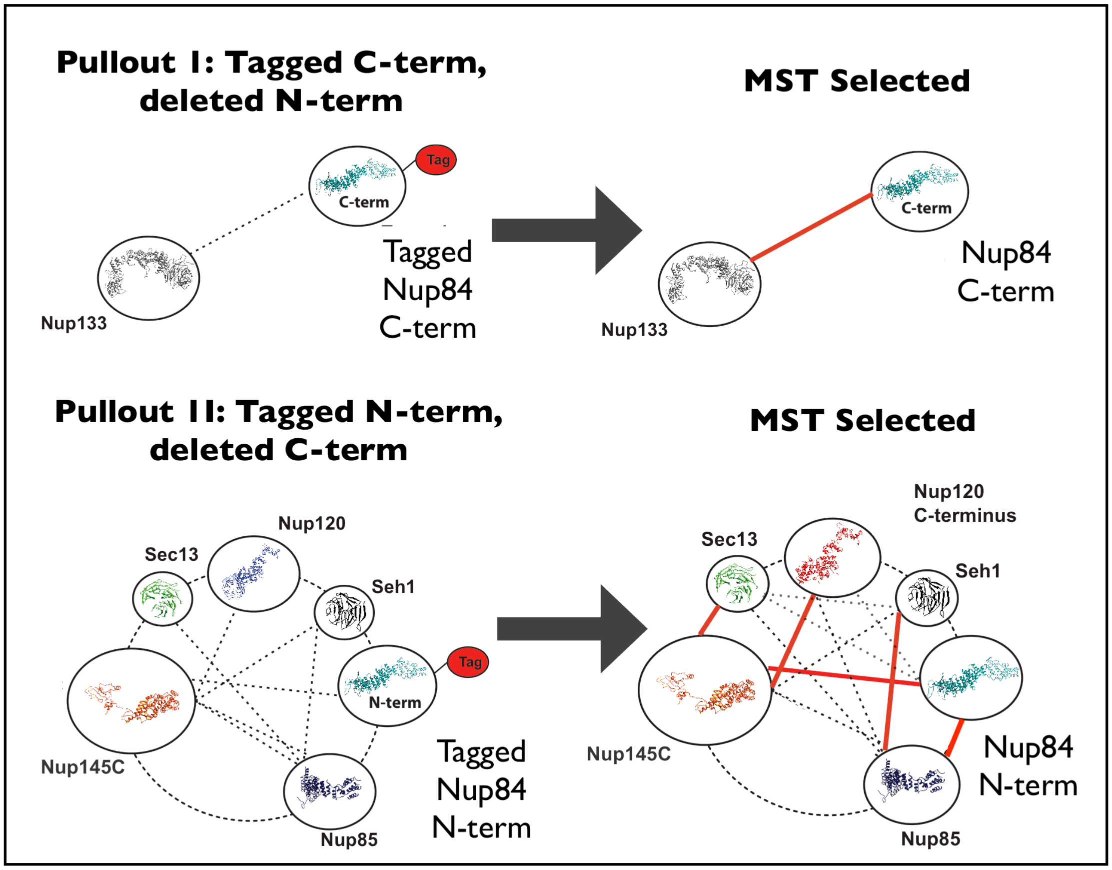


Figure 2-4. Restraints for truncated affinity purification experiments.
These truncated conditional connectivity restraints operate as follows, at each optimization step: 1) give each edge a score based on deviation of the distance between the closest two atoms (x) (or residues, domains, etc.) and the expected distance x0 for a protein interaction $k(x-x_0)^2$. 2) determine a minimal spanning tree (MST) for the composite graph. 3) determine the composite score as the sum of edge scores in the MST. Above are connectivity restraints for two simple and complementary truncated affinity purification experiments on the Nup84 complex (discussed in further detail in chapter 3). In the first (top), only Nup133 elutes with the C-terminus of the protein Nup84, demonstrating that the C-terminus of the protein Nup84 interacts directly with Nup133. In the second experiment, all 6 of the other proteins in the complex elute with the N-terminus of Nup84. Connectivity within this group of 6 proteins plus a segment of the Nup84 protein is enforced by a set of minimally scoring restraints.

Chapter 3. Integrative Structure Determination of the Nup84 subcomplex of the NPC

Introduction

The structural and functional role of the Nup84 complex in the NPC

The NPC structural core is conserved, highly modular and is formed from eight symmetric spokes that connect to form 5 coaxial rings: a membrane ring, two adjacent inner rings, and two outer rings facing respectively the cytoplasmic and nucleoplasmic periphery [13]. Data from both vertebrates and the yeast *S. cerevisiae* [13, 44, 130, 131] indicate that the outer ring of the NPC is comprised of a conserved assembly, which in vertebrates corresponds to a nonameric complex called the Nup107-160 complex [130, 132, 133], and in yeast corresponds to the Nup84 complex which is formed from seven proteins named Nup133, Nup120, Nup145c, Nup85, Nup84, Seh1, and Sec13 [134, 135]. Sec13 is shared with the Sec13/31 COPII vesicle coating complex (VCC) and both Seh1 and Sec13 have recently been found in a coating-related complex termed the SEA complex, underscoring the relationship between coatomers and NPCs [59, 135-137].

The Nup84 complex is the best characterized of the NPC's building blocks, as reflected by the extensive set of genetic, biochemical, and structural data accumulated over the years [93, 138, 139]. Mutations of Nup84 complex nups usually lead to severe phenotypes characterized by fitness defects, mRNA and pre-ribosomal export problems, as well as aberrant NPC biogenesis and distribution (i.e. “clustering” of NPCs into a handful of closely packed groups) within the NE; indeed, the NPC clustering phenotype has been broadly used as a tool to characterize putative NPC-associated proteins [140-144].

Previous structural work on the Nup84 complex

The Nup84 heptamer forms a characteristic Y-shaped assembly, as shown by pioneering EM studies of both isolated complexes and complexes reconstituted *in vitro*; Nup133, Nup84, and Nup145c/Sec13 form the main stalk of the “Y”, with Nup133 at its tip, and Nup85/Seh1 and Nup120 are located in the two short arms of the heptameric assembly [134, 145]. Structural analyses by combined computational and biochemical methods [59, 60] and subsequent crystallographic studies [66-74, 146] have shown that nups within the Nup84 complex are formed almost entirely by a b-propeller fold, an a-solenoid-like (or helix-turn-helix repeat) fold, or a combination of an N-terminal b-propeller and a C-terminal a-solenoid-like folds (termed a b-a fold arrangement), common to vesicle coating complexes (discussed in chapter 1). Despite this wealth of data, we still do not have a full description of the structures or domain interfaces in the Nup84 complex (reviewed in [93]); moreover, differing interpretations of the crystallographic structures have led to models of the arrangement of the Nup84 complex in the NPC [66, 93] (discussed in chapter 1) that do not agree with each other, nor with maps from other groups in the field [130, 131], including our own

[13]. Such ambiguities are frequently observed in the structure determination of protein assemblies [14, 39, 147-149].

An integrative modeling approach was previously used to determine the yeast NPC's molecular architecture based on a diverse set of proteomic and biophysical data (chapter 1). This map defined the approximate positions of the component proteins but not their shape or orientations [13, 36]. We describe an improved experimental and computational approach to determine the subunit configuration of macromolecular assemblies by computational integration of new kinds of biochemical and structural data. We apply this approach to the yeast Nup84 complex, defining the positions and relative orientations of the individual components. The resulting structure allows us to visualize structure-function relationships that provide insights into the assembly and evolution of the Nup84 complex and the NPC as a whole.

Results

A truncated affinity purification protocol (Chapter 2) can in fact restrain protein orientations in addition to relative positions, when the atomic structures or models of the interacting multi-domain proteins are available (as is the case for most of the nups). Here, we conduct an exhaustive set of affinity purification experiments in which the domain closest to either the N-terminus or C-terminus of a protein is truncated. Proteins that elute in a full affinity purification experiment, but do not elute in the same experiment with a truncated region, are assumed to interact *via* the truncated region. A modified implementation of conditional restraints (Figure 2-3) is used below to interpret such data.

Gathering Data

We have used two primary methods to obtain a large quantity of structural information on the Nup84 complex.

Domain Deletion Mapping

Using domain boundary data [59, 61], we expressed domain truncated nups, constructed as in-frame chromosomal deletions under the control of their native promoters and C-terminally tagged with the Protein A (PrA) tag (Table 3-1 and Material and Methods). We determined the set of nups interacting with each of the tagged truncated nups by affinity purification and mass spectrometry (Figure 3-1). Purifications were adjusted so that we initially only probed interactions within the 7-protein complex (Table 3-1). The nups that are affinity purified with a truncated nup bait must necessarily interact (directly or indirectly) *in vivo* with the remaining domain(s) (Figure 3-2A). Loss of an interaction is usually due to loss of a specific interaction domain; however, in rare instances, it can be due to misfolding or mislocalization. For those truncations that caused a gross change in the interaction pattern, the tagged construct was analyzed by Protease Accessibility Laddering (PAL) [61]; in all cases, the PAL maps indicated that the remaining domains had folded correctly (data not shown). All the truncations were also tested for their localization to the NPC *in vivo* (Figure 3-3); mislocalization could have prevented a *bona fide* interaction from forming and so no conclusion as to the loss of an interaction can be drawn in that case (Figure 3-

2B and Figure 3-3). However, if a truncation retained sufficient nup interaction regions to target it to the NPC, any lost interactions with the Nup84 complex components were taken as caused by the loss of the deleted domain(s), allowing us to map at least one interaction point to the deleted domain (Figure 3-2B).

Truncation						**PAL Site**[1]	**Expressed protein**
-	-	Nup133-ProtA	**NPC localization**[2]	**Theoretical MW (kDa)**	**MS detected co-isolating Nups**[3]	**Cell grindate/AP (g)**[5]	**Buffer**
Nup133(1-530)	C	Nup133(1-530)-ProtA	yes	159.2	Nup133; Nup120; Nup145c; Nup85; Nup84; Seh1; Sec13	1.5	20 mM Hepes pH 7.4, 110 mM KOAc, 2 mM $MgCl_2$, 1% Triton X-100, 0.1% Tween 20, 200 mM NaCl.
Nup133(1-772)	D	Nup133(1-772)-ProtA	no	86.1	Nup133(1-530)	2	20 mM Hepes pH 7.4, 110 mM KOAc, 2 mM $MgCl_2$, 1% Triton X-100, 0.1% Tween 20, 200 mM NaCl.
Nup133(1-898)	-	Nup133(1-898)-ProtA	no	114.6	Nup133(1-772)	2.5	20 mM Hepes pH 7.4, 110 mM KOAc, 2 mM $MgCl_2$, 1% Triton X-100, 0.1% Tween 20, 200 mM NaCl.
Nup133(1-1086)	-	Nup133(1-1086)-ProtA	yes	129.5	Nup133(1-898)	3	20 mM Hepes pH 7.4, 2mM MgCl, 0.1%CHAPS, 0.01% Tween-20, 250mM NaCl
Nup133(301-1157)	-	M-loxP-Nup133(301-1157)-ProtA	yes	151	Nup133(1-1086); Nup120; Nup145c; Nup85; Nup84; Seh1; Sec13	3	20 mM Hepes pH 7.4, 2mM MgCl, 0.1%CHAPS, 0.01% Tween-20, 250mM NaCl
Nup133(531-1157)	C	M-loxP-Nup133(531-1157)-ProtA	yes	125.4	Nup133(301-1157); Nup120; Nup145c; Nup85; Nup84; Seh1; Sec13	6	20 mM Hepes pH 7.4, 2mM MgCl, 0.1%CHAPS, 0.01% Tween-20, 250mM NaCl
Nup133(899-1157)	-	M-loxP-Nup133(899-1157)-ProtA	no	99.1	Nup133(531-1157); Nup84	3	20 mM Hepes pH 7.4, 110 mM KOAc, 0.1% CHAPS, 0.02% Tween 20, 20 mM NaCl
-	-	Nup120-ProtA	no	55.6	Nup133(899-1157)	6	20 mM Hepes pH 7.4, 110 mM KOAc, 0.1% CHAPS, 0.02% Tween 20, 20 mM NaCl
Nup120(1-396)	B	Nup120(1-396)-ProtA [4]	yes	146.3	Nup120; Nup133; Nup145c; Nup85; Nup84; Seh1; Sec13	1.5	20 mM Hepes pH 7.4, 110 mM KOAc, 2 mM $MgCl_2$, 1% Triton X-100, 0.1% Tween 20, 150 mM NaCl.
Nup120(1-817)	C	Nup120(1-817)-ProtA					
Nup120(1-972)	D	Nup120(1-972)-ProtA	yes	120.2	Nup120(1-817)	3	20 mM Hepes pH 7.4, 110 mM KOAc, 2 mM $MgCl_2$, 1% Triton X-100, 0.1% Tween 20, 150 mM NaCl.
Nup120(397-1037)	B	M-loxP-Nup120(397-1037)-ProtA	yes	138.8	Nup120(1-972)	3.5	20 mM Hepes pH 7.4, 110 mM KOAc, 2 mM $MgCl_2$, 1% Triton X-100, 0.1% Tween 20, 150 mM NaCl.
Nup120(818-1037)	C	M-loxP-Nup120(818-1037)-ProtA	yes	101.2	Nup120(397-1037); Nup133; Nup145c;	6	20 mM Hepes pH 7.4, 110 mM KOAc, 0.1% CHAPS, 0.01%

					Nup85; Nup84; Seh1; Sec13		Tween-20, 300mM NaCl
Nup120(1-972)-GFP/Nup84-ProtA	D	Nup120(1-972)-GFP / Nup84-ProtA	yes	51	Nup120(818-1037); Nup133; Nup145c; Nup85; Nup84; Seh1; Sec13	6	20 mM Hepes pH 7.4, 110 mM KOAc, 0.1% CHAPS, 0.01% Tween-20, 300mM NaCl
-	-	Nup85-ProtA	yes	138.8	NUP84; Nup133; Nup145c; Sec13	3	20 mM Hepes pH 7.4, 110mM KOAc, 2 mM $MgCl_2$, 0.1% CHAPS, 0.01% Tween-20,100 mM NaCl
Nup85(1-438)	D	Nup85(1-438)-ProtA	yes	110.8	Nup85; Nup133; Nup120; Nup145c; Nup84; Seh1; Sec13	3	20 mM Hepes pH 7.4, 110 mM KOAc, 0.1% CHAPS, 0.02% Tween 20, 300 mM NaCl.
Nup85(233-744)	B	M-loxP-Nup85(233-744)-ProtA	yes	75.8	Nup85(1-438); Seh1	3	20 mM Hepes pH 7.4, 110 mM KOAc, 0.1% CHAPS, 0.02% Tween 20, 300 mM NaCl.
Nup85(1-438)-GFP/Nup84-ProtA	D	Nup85(1-438)-GFP / Nup84-ProtA	yes	84.1	Nup85(233-744); Nup133; Nup120; Nup145c; Nup84; Sec13	3	20 mM Hepes pH 7.4, 110 mM KOAc, 0.1% CHAPS, 0.02% Tween 20, 300 mM NaCl.
-	-	Nup84-ProtA	yes	75.8	Nup84; Nup120; Nup133; Nup145c; Sec13	3	20 mM Hepes pH 7.4, 110mM KOAc, 2 mM $MgCl_2$, 0.1%CHAPS, 100 mM NaCl
Nup84(1-159)	B	Nup84(1-159)-ProtA [4]	yes	109.5	Nup84; Nup133; Nup120; Nup145c; Nup85; Seh1; Sec13	1.5	20 mM Hepes pH 7.4, 110 mM KOAc, 2 mM $MgCl_2$, 1% Triton X-100, 0.1% Tween 20, 175 mM NaCl.
Nup84(1-573)	C	Nup84(1-573)-ProtA					
Nup84(160-726)	B	M-loxP-Nup4(160-726)-ProtA	yes	91.9	Nup84(1-573); Nup120; Nup145c; Nup85; Seh1; Sec13	3	20 mM Hepes pH 7.4, 110 mM KOAc, 2 mM $MgCl_2$, 1% Triton X-100, 0.1% Tween 20, 175 mM NaCl.
-	-	Nup145c-ProtA	no	90.8	Nup84(160-726); Nup133	5	20 mM Hepes pH 7.4, 110 mM KOAc, 2 mM $MgCl_2$, 1% Triton X-100, 0.1% Tween 20, 175 mM NaCl.
Nup145c(1-227)	B	Nup145c(1-227)-ProtA	yes	106.9	Nup145c; Nup133; Nup120; Nup85; Nup84; Seh1; Sec13	1.5	20 mM Hepes pH 7.4, 110 mM KOAc, 2 mM $MgCl_2$, 1% Triton X-100, 0.1% Tween 20, 200 mM NaCl.
Nup145c(1-468)	C	Nup145c(1-468)-ProtA	yes	51.4	Nup145c(1-227)	4	20 mM Hepes pH 7.4, 110 mM KOAc, 2 mM $MgCl_2$, 1% Triton X-100, 0.1% Tween 20, 200 mM NaCl.
Nup145c(1-609)	-	Nup145c(1-609)-ProtA	yes	79.3	Nup145c(1-468); Nup133; Nup84	2.75	20 mM Hepes pH 7.4, 110 mM KOAc, 2 mM $MgCl_2$, 1% Triton X-100, 0.1% Tween 20, 200 mM NaCl.
Nup145c(1-641)	-	Nup145c(1-641)-ProtA	yes	95.2	Nup145c(1-609); Nup133; Nup84; Sec13	4	20 mM Hepes pH 7.4, 110 mM KOAc, 2 mM $MgCl_2$, 1% Triton X-100, 0.1% Tween 20, 200 mM NaCl.
Nup145c(1-670)	-	Nup145c(1-670)-ProtA	yes	98.8	Nup145c(1-641); Nup133; Nup84; Sec13	3	20 mM Hepes pH 7.4, 110 mM KOAc, 2 mM $MgCl_2$, 1% Triton X-100, 0.1% Tween 20, 200 mM NaCl.
Nup145c(1-8)-(228-712)	B	Nup145c(1-8)-loxP-(228-712)-ProtA	yes	102.2	Nup145c(1-670); Nup133; Nup84; Sec13	4	20 mM Hepes pH 7.4, 110 mM KOAc, 2 mM $MgCl_2$, 1% Triton X-

							100, 0.1% Tween 20, 200 mM NaCl.
Nup145c(1-128)-(181-712)	-	Nup145c(1-128)-loxP-(181-712)-ProtA	yes	82.3	Nup145c(D9-227); Nup133; Nup84	4	20 mM Hepes pH 7.4, 110 mM KOAc, 2 mM $MgCl_2$, 1% Triton X-100, 0.1% Tween 20, 200 mM NaCl.
Nup145c(1-316)-(327-712)	-	Nup145c(1-316)-loxP-(327-712)-ProtA	yes	101.2	Nup145c(D129-180); Nup133; Nup120; Nup85; Nup84; Seh1	2	20 mM Hepes pH 7.4, 110 mM KOAc, 2 mM $MgCl_2$, 1% Triton X-100, 0.1% Tween 20, 200 mM NaCl.
Nup145c(1-670)-GFP/Nup85-ProtA	-	Nup145c(1-670)-GFP / Nup85-ProtA	yes	105.9	Nup145c(D317-326); Nup120; Nup85; Seh1; Sec13	2.5	20 mM Hepes pH 7.4, 110 mM KOAc, 2 mM $MgCl_2$, 1% Triton X-100, 0.1% Tween 20, 200 mM NaCl.
Nup145c(1-670)-GFP/Nup85-ProtA	-	Nup145c(1-670)-GFP / Nup85-ProtA	yes				

Table 3-1. Summary of Nup84 truncations and affinity purification conditions. Nup84 complex nucleoporin constructs used in this study, along with their subcellular localization (Figure S1) and affinity purified composites (Figure 1) are summarized.

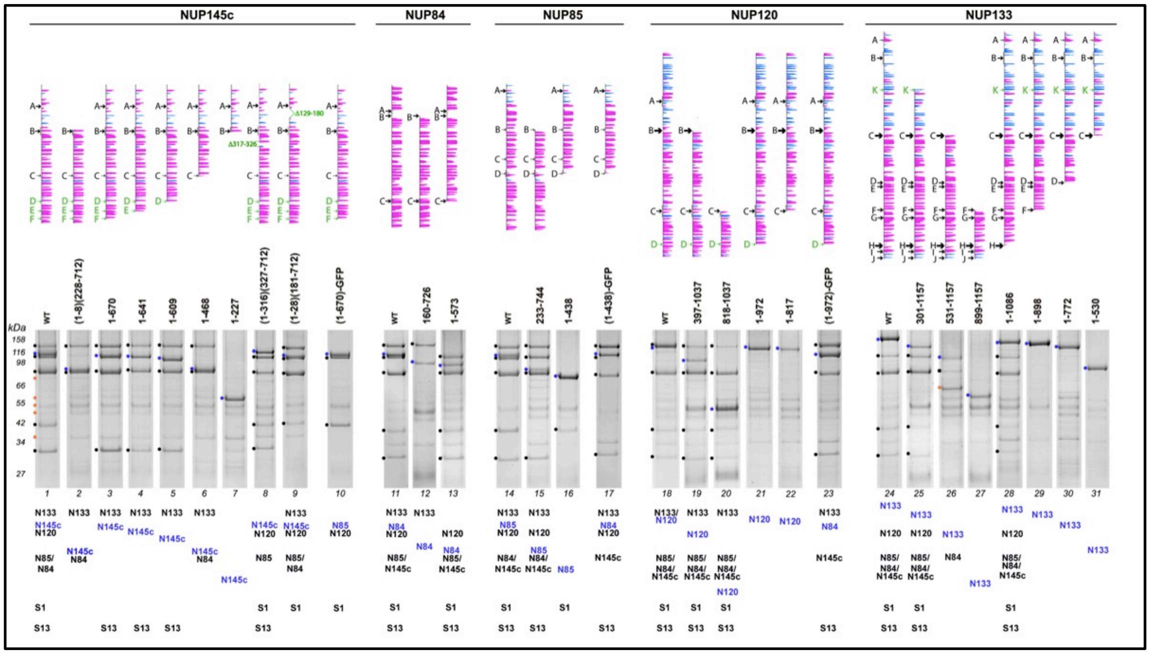
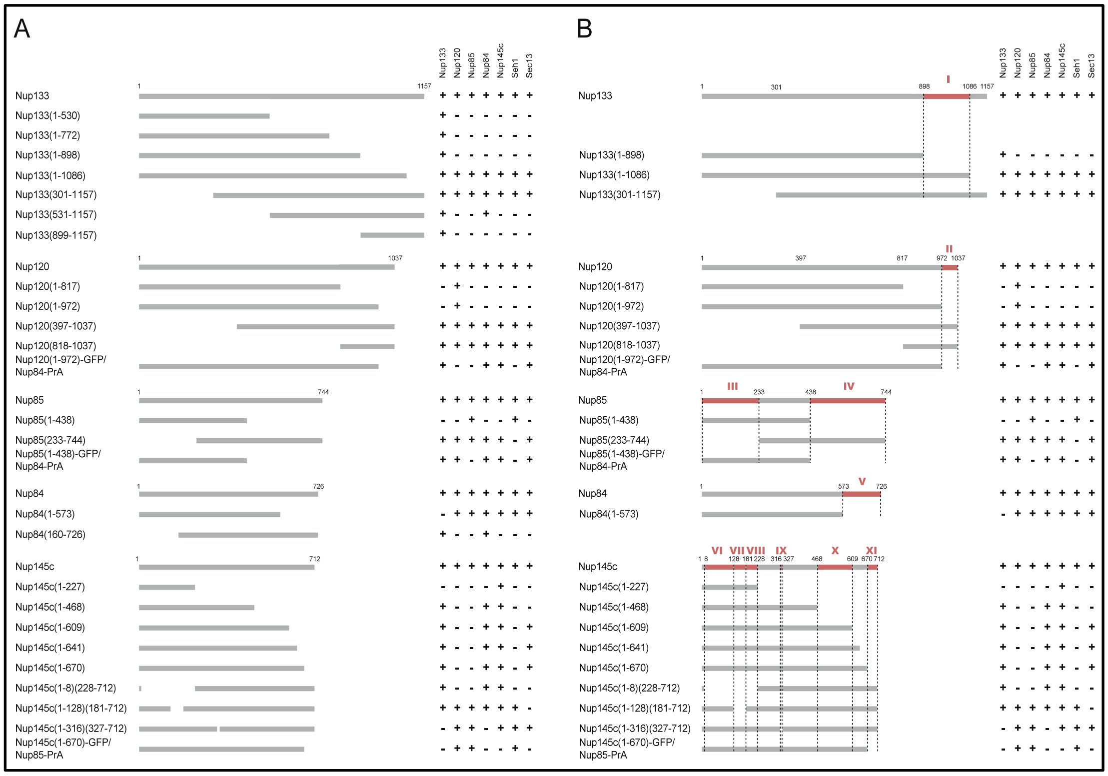


Figure 3-1. Domain mapping of the Nup84 complex.
Top: predicted secondary structure of truncated nup constructs; vertical lines represent the sequence of each nup to scale, and secondary structure predictions are shown as horizontal bars of length proportional to the confidence of the prediction (α-helices, magenta; β-strands, cyan). PAL sites [61] defining the limits of the truncations are indicated with arrowed black letters; truncation points selected by sequence alignment are indicated with arrowed green letters. Bottom, Coomassie stained SDS-PAGE of affinity-purified PrA-tagged truncated Nup84 complex nups. Protein bands marked by filled circles at the left side of the gel lanes were identified by mass spectrometry. The identity of the co-purifying proteins is indicated in order below each lane (PrA-tagged nups, blue; co-purifying nups, black (N=Nup, S=Sec); contaminants, orange).



A

	Nup133	Nup120	Nup85	Nup84	Nup145c	Seh1	Sec13
Nup133	+	+	+	+	+	+	+
Nup133(1-530)	+	-	-	-	-	-	-
Nup133(1-772)	+	-	-	-	-	-	-
Nup133(1-898)	+	-	-	-	-	-	-
Nup133(1-1086)	+	+	+	+	+	+	+
Nup133(301-1157)	+	+	+	+	+	+	+
Nup133(531-1157)	+	-	-	+	-	-	-
Nup133(899-1157)	+	-	-	-	-	-	-
Nup120	+	+	+	+	+	+	+
Nup120(1-817)	-	+	-	-	-	-	-
Nup120(1-972)	-	+	-	-	-	-	-
Nup120(397-1037)	+	+	+	+	+	+	+
Nup120(818-1037)	+	+	+	+	+	+	+
Nup120(1-972)-GFP/ Nup84-PrA	+	-	-	+	+	-	+
Nup85	+	+	+	+	+	+	+
Nup85(1-438)	-	-	+	-	-	+	-
Nup85(233-744)	+	+	+	+	+	-	+
Nup85(1-438)-GFP/ Nup84-PrA	+	+	-	+	+	-	+
Nup84	+	+	+	+	+	+	+
Nup84(1-573)	-	+	+	+	+	+	+
Nup84(160-726)	+	-	-	+	-	-	-
Nup145c	+	+	+	+	+	+	+
Nup145c(1-227)	-	-	-	-	+	-	-
Nup145c(1-468)	+	-	-	+	+	-	-
Nup145c(1-609)	+	-	-	+	+	-	+
Nup145c(1-641)	+	-	-	+	+	-	+
Nup145c(1-670)	+	-	-	+	+	-	+
Nup145c(1-8)(228-712)	+	-	-	+	+	-	-
Nup145c(1-128)(181-712)	+	+	+	+	+	+	-
Nup145c(1-316)(327-712)	-	+	+	-	+	+	+
Nup145c(1-670)-GFP/ Nup85-PrA	-	+	+	-	-	+	-

B

	Nup133	Nup120	Nup85	Nup84	Nup145c	Seh1	Sec13
Nup133	+	+	+	+	+	+	+
Nup133(1-898)	+	-	-	-	-	-	-
Nup133(1-1086)	+	+	+	+	+	+	+
Nup133(301-1157)	+	+	+	+	+	+	+
Nup120	+	+	+	+	+	+	+
Nup120(1-817)	-	+	-	-	-	-	-
Nup120(1-972)	-	+	-	-	-	-	-
Nup120(397-1037)	+	+	+	+	+	+	+
Nup120(818-1037)	+	+	+	+	+	+	+
Nup120(1-972)-GFP/ Nup84-PrA	+	-	-	+	+	-	+
Nup85	+	+	+	+	+	+	+
Nup85(1-438)	-	-	+	-	-	+	-
Nup85(233-744)	+	+	+	+	+	-	+
Nup85(1-438)-GFP/ Nup84-PrA	+	+	-	+	+	-	+
Nup84	+	+	+	+	+	+	+
Nup84(1-573)	-	+	+	+	+	+	+
Nup145c	+	+	+	+	+	+	+
Nup145c(1-227)	-	-	-	-	+	-	-
Nup145c(1-468)	+	-	-	+	+	-	-
Nup145c(1-609)	+	-	-	+	+	-	+
Nup145c(1-641)	+	-	-	+	+	-	+
Nup145c(1-670)	+	-	-	+	+	-	+
Nup145c(1-8)(228-712)	+	-	-	+	+	-	-
Nup145c(1-128)(181-712)	+	+	+	+	+	+	-
Nup145c(1-316)(327-712)	-	+	+	-	+	+	+
Nup145c(1-670)-GFP/ Nup85-PrA	-	+	+	-	-	+	-

Figure 3-2. Identification of interacting regions within the Nup84 complex.
A) Co-purification profile of the different truncations analyzed. Horizontal gray lines represent the amino acid residue length of each protein and truncated version; amino acid residue positions are shown on top of the lines. On the right, MS detected co-purifying Nup84 complex nups are indicated with a "+". Undetected proteins are indicated with a "-".
B) Domain mapping. Truncations showing both nuclear rim localization by immunofluorescence and PAL profile consistent with proper folding were used to interpret a lost interaction as caused by the lost of, at least, one interaction point present in the deleted nup region (represented by red lines on the wild type line). Each region is identified by red Roman numbers. The identified interactions are: Nup133 Region I interacts with Nup120, Nup145c, Nup85, Nup84, Seh1 and Sec13; Nup120 Region II with Nup133, Nup145c, Nup85, Nup84, Seh1 and Sec13; Nup85 Region III with Seh1; Nup85 Region IV, Nup133, Nup120, Nup145c, Nup84 and Sec13; Nup84 Region V with Nup133, Nup120, Nup145c, Nup85, Seh1 and Sec13; Nup145c Region VI with Nup120, Nup85 and Seh1; Nup145c Region VII with Sec13; Nup145c Region VIII with Nup120, Nup85 and Seh1; Nup145c Region IX with Nup133, and Nup84; Nup145c Region X with Sec13; Nup145c Region XI with Nup120, Nup85 and Seh1.

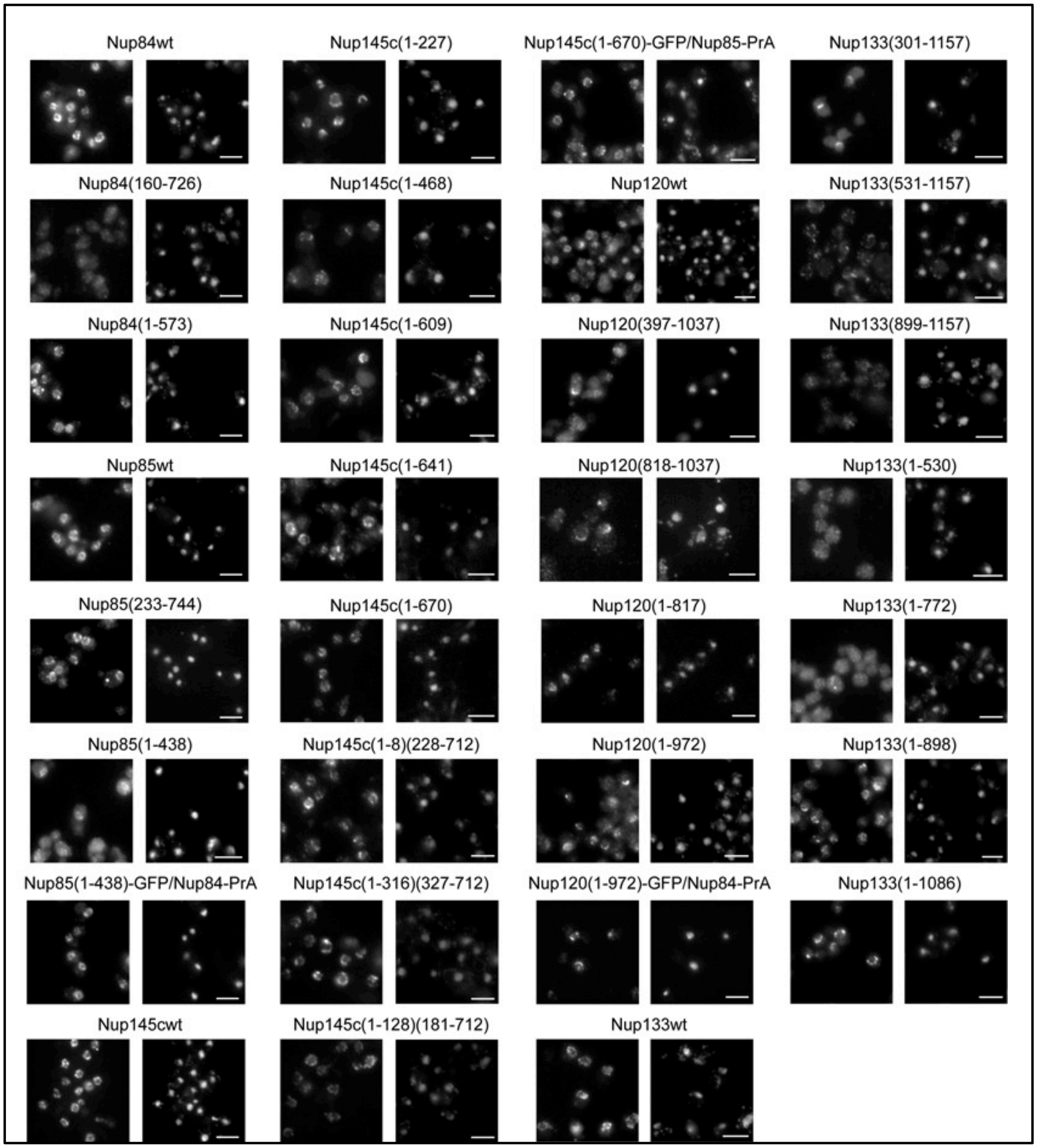


Figure 3-3. Immunofluorescence localization of Nup84 complex truncations.
Cells corresponding to wild type (wt) and truncated versions of Nup84 complex nups were grown to early log-phase, fixed and processed as in [150]. Each column shows the indirect immunofluorescence staining directed against Protein-A tagged proteins (left) and their coincident DAPI staining of the same field of cells (right). Scale bar, 5 μm.

Electron Microscopy

We used negative stain EM to obtain the overall shape and dimensions of the affinity purified Nup84 complex (Figure 3-3). As expected, the observed particles exhibit a characteristic Y shape (Figure 3A) [64, 134]. Other forms of the complex, lacking Nup133 (Figure 3C) [145], and a truncated version, lacking Seh1 and the N-terminus of Nup85 (Figure 3B), were also purified and analyzed to morphologically assign the compositional identity of the Y arms.

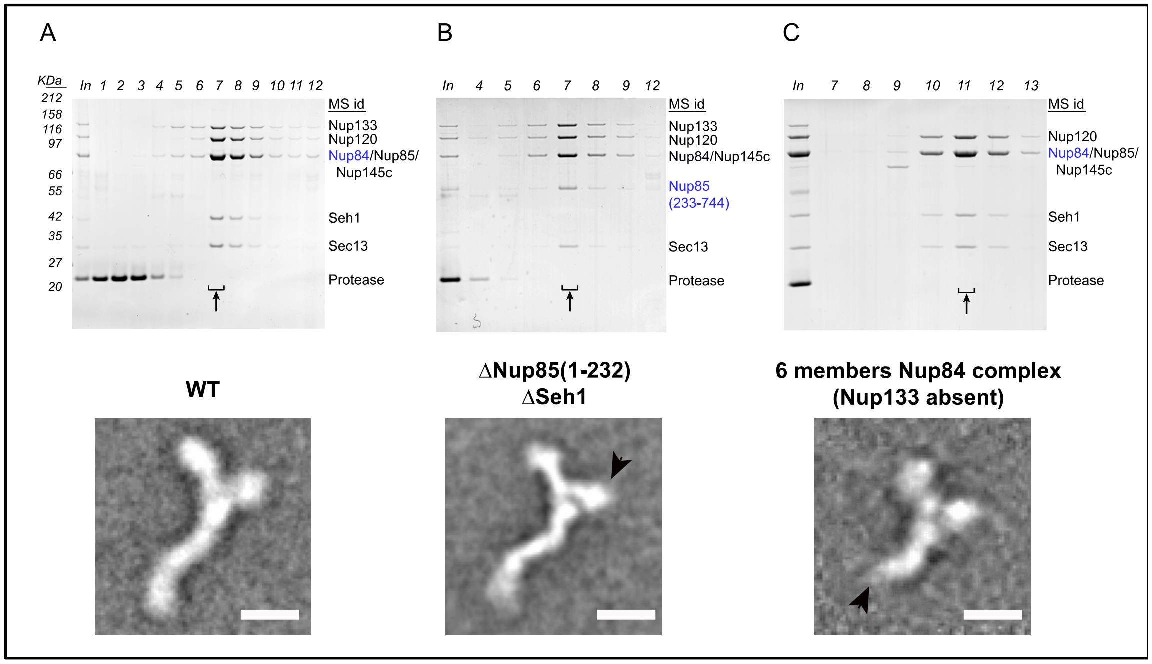


Figure 3-4: EM analysis of the Nup84 complex.
Affinity-purified Nup84 complexes corresponding to the full length complex (A), Nup85(233-744) truncation (B), and the Nup84 complex lacking Nup133 (C) were purified by either sucrose gradient (A and B) or size exclusion chromatography (C). Coomassie-stained SDS-PAGE of the precipitated input (In) and collected fractions are shown. Fractions indicated with an arrow were analyzed by negative stain EM. Class averages for each complex are shown in the lower part of each panel. The missing mass observed in Nup85(233-744) truncation (B) and six-member complex (C) are indicated with an arrowhead. Scale bar, 10 nm.

Representing and Translating Data into Spatial Restraints

Representation of the nups

Because few Nup84 complex components have full atomic resolution coverage, each nup was represented as a hybrid rigid body consisting of its crystallographic data (if available), comparative models (if an accurate alignment could be computed), and/or a string of beads mimicking the coarse shape of a related known fold (Table 3-2); the minimum necessary number of structural modules were used in each case. 56% of residues in the Nup84 complex have been covered by the yeast crystallographic structures of individual domains [67, 68, 70, 74]. We modeled an additional 29% of residues based on alignments to related known structures (Table 3-2). Crystallographic interfaces have not been used in the initial modeling to allow the validation of the approach (Figure 3-6 and below). However, this high resolution information was used to generate an alternative ensemble (Figure S2C). Our previously determined stoichiometry for the complex [44] has since been corroborated [134] and is used here, as was done previously [13, 36].

Nup	Representation[a]	Start residue number	End residue number	Target-template sequence identity[c]	PDB code[b]	Template PDB code[c]
Nup84	Crystal	7	442		3iko	
	Gap	443	461			
	Model	462	723	15% (2zct) 19% (3cqg) 19% (3cqc)		2zcta 3cqga 3cqca
	Gap	724	726			
Nup133	Gap	1	62			
	Model	63	481	16%		1xks
	Gap	482	494			
	Model	495	879	15%		3i4r
	Gap	880	945			
	Crystal	946	1157		3kfo	
Nup85	Crystal	1	564		3ewe	
	Gap	565	604			
	Model	605	744			
Nup120	Crystal	1	729		3hxr	
	Spheres	730	1029			
Nup145c	Model	25	129	19%		3ez1
	Crystal	130	552		3bg0	
	Model	553	712	15% 29%		1w07 3cmu

Table 3-2. Representation of each Nup84 complex protein.
a "Crystal" refers to using an X-ray structure from the PDB; "Gap" refers to an amino acid residue segment represented by spheres; and "Model" refers to a segment represented by an atomic comparative model. b PDB codes are indicated for crystal structures used directly. c PDB codes for templates and sequence identities for target-template alignments are indicated where comparative models were used to represent a fragment.

Domain Mapping Restraints

An affinity purification experiment produces a distinct set of co-isolating proteins, which we term a composite; a composite may represent a single complex of physically interacting proteins or a mixture of such complexes overlapping at least at the tagged protein [36]. We previously described how to encode a composite as "conditional connectivity restraints" based on the proximities between the composite proteins [13, 36, 103, 151] (S_{dm}, see Materials and Methods). We still apply such conditional connectivity restraints here, but now restrain positions of domains and protein fragments, instead of whole proteins, thus orienting as well as localizing protein subunits (discussed in Chapter 2).

Electron Microscopy Restraints

We constructed a class average of 2D EM images of the full Nup84 complex (Figure 3-3A), and used this class average to define a restraint and two filters (Materials and Methods) as follows. We first measured the dimensions of the class average of the full complex to define a bounding triangle restraint on the Nup84 complex (S_{tri}, Materials and Methods). Then, the class average of the full complex was also used as a filter on optimized models (below) (Velazquez-Muriel et. al., unpublished) (S_{em}, Materials and Methods). Next, the class average for the 6-member complex, missing Nup133 (Figure 3-4C), was also used as a filter in the same manner, but compared to the portion of our structures containing the remaining 6 proteins (Nup84, Nup145c, Sec13, Nup85, Seh1, and Nup120). Finally, a filter was used to remove structures in which the longest arm of the complex (as measured by distance from the bottom of the stem) incorrectly contained Nup85 and Seh1 rather than Nup120, as shown by the deletion class average (Figure 3-4B).

Geometric Complementarity and Excluded Volume.

A low-resolution geometric complementarity term S_{acc} was included in our scoring function to reward shape complementarity and penalize steric clashes [31].

Sampling Good Scoring Structures

Models that satisfied all restraints described above were obtained in four stages (Figure 3-5 and Materials and Methods). In the first, coarse stage, 100,000 coarse models were calculated by a Monte Carlo optimization of the sum of S_{dm} and S_{tri}, each starting from a different random initial configuration. In the second, refinement stage, the models were refined by a Monte Carlo optimization of the sum of S_{acc}, S_{dm}, and S_{tri}, with smaller step sizes. In the third, filtering stage, the 10,000 top scoring models were filtered by requiring sufficiently good alignment to the class averages of the EM images of the full and partial complex (Figure 3A and C), eliminating 521 structures. Finally, the remaining 9479 structures were relaxed by optimizing the sum of S_{acc}, S_{dm}, S_{tri}, and S_{em}, using the simplex algorithm [152].

Analyzing and Assessing the Ensemble of Structures

Characteristics of the final ensemble

The resulting 9479 good scoring models were clustered according to pairwise Ca RMSD, and showed a single dominant cluster of solutions (Figure 3-5 and 3-6). The average RMSD from the center of the cluster is 1.5 nm. Random subsets of

10% of the models also form a single dominant cluster with the average RMSDs from the center of the cluster of 1.4-1.6 nm. Thus, the precision of the ensemble does not significantly change even if only a small fraction of the good scoring solutions are used, demonstrating that our optimization procedure is likely to have exhaustively sampled the set of possible solutions, given the data. The variability in the ensemble of good scoring structures may reflect the heterogeneity (e.g. flexibility) of the sample as well as the lack of information to determine a highly precise structure.

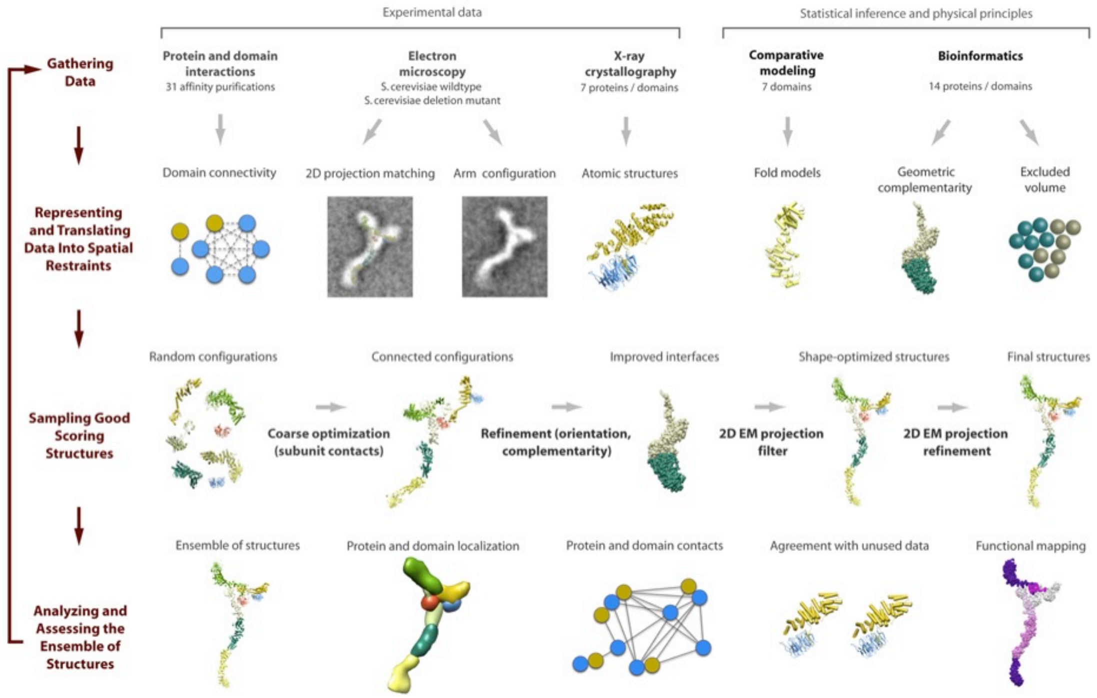


Figure 3-5. Integrative structure determination of the Nup84 complex.
First, structural data and information are generated by various experiments and computational methods. Second, the system is represented by the mostly atomic structures of the subunits and the data are translated into spatial restraints. Third, an ensemble of structures that satisfy the data are obtained by minimizing the violations of the spatial restraints, starting from many different random configurations. Fourth, the ensemble is clustered into distinct sets of structures on the basis of their similarities, and analyzed in terms of geometry and accuracy.

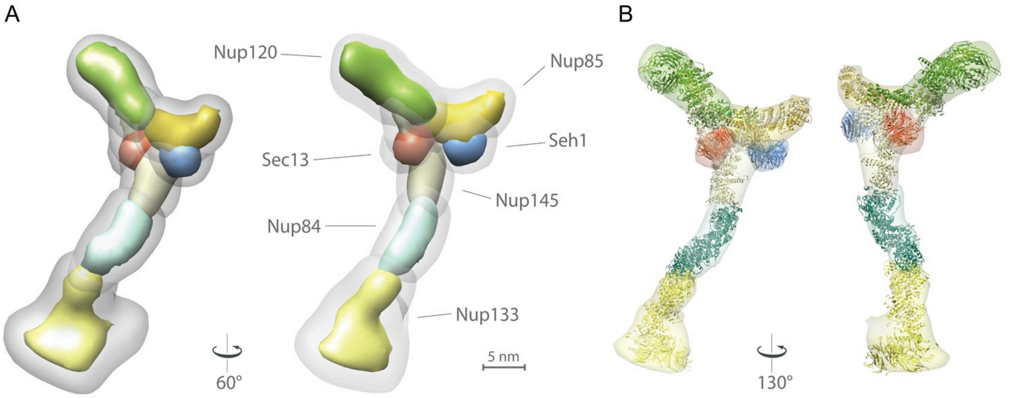


Figure 3-6. Density map for the Nup84 complex ensemble.
A. Two views of the localization density map are shown for the Nup84 complex ensemble. The gray outer envelope represents the density within which 90% of all Nup84 complex structures in the final ensemble were localized, and the colored inner envelope was thresholded by visual inspection to match the volume of an individual structure. B. Two views of the density map, containing two fitted ribbon structures from the ensemble each.

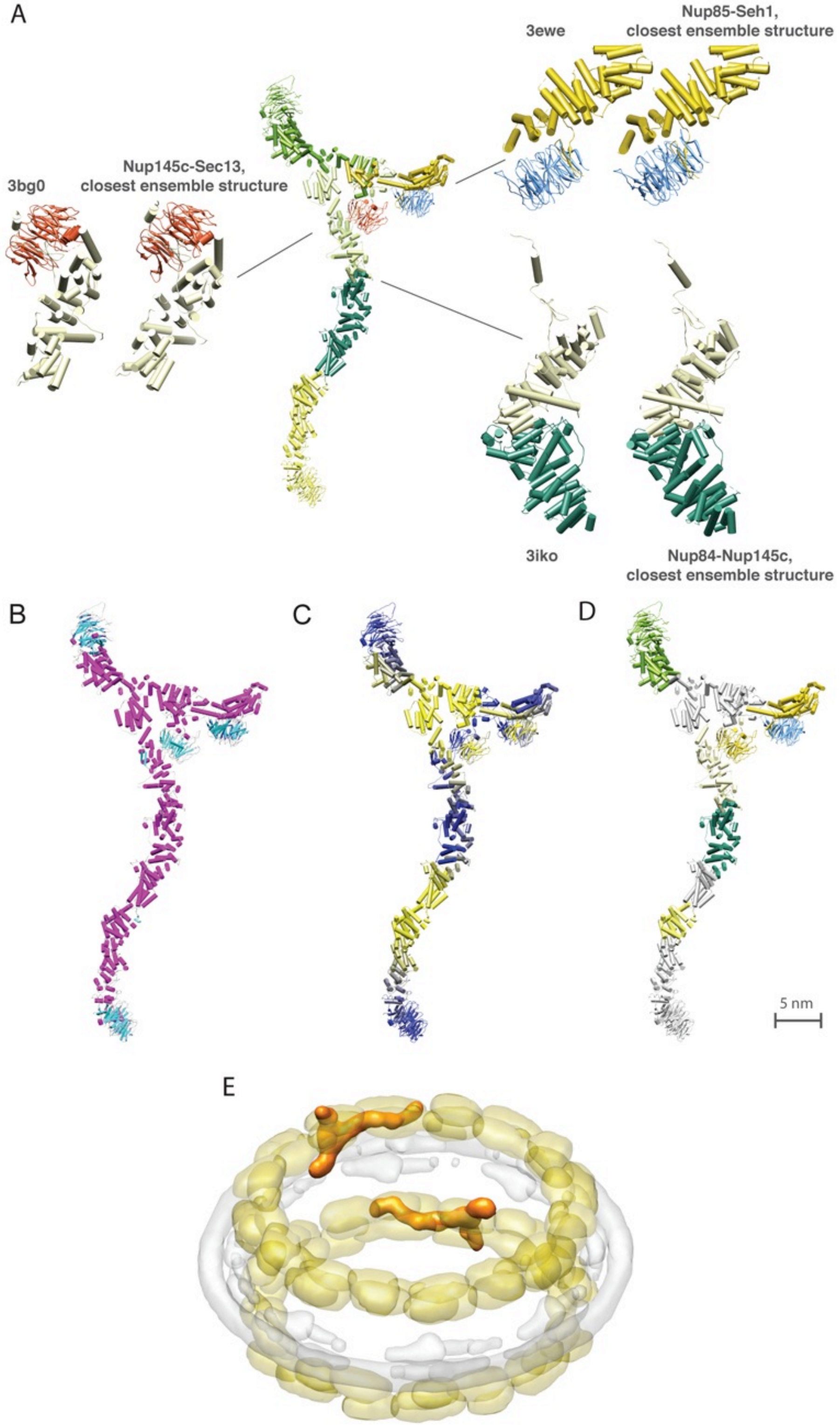

A
3ewe
Nup85-Seh1,
closest ensemble structure
3bg0
Nup145c-Sec13,
closest ensemble structure
3iko
Nup84-Nup145c,
closest ensemble structure
B
C
D
5 nm
E

Figure 3-7. Structural features of the Nup84 complex.
A) Comparison of protein-protein interfaces in our ensemble with crystallographically determined interfaces. Atomic structures for the Nup85-Seh1 (PDB code 3ewe [68]), Nup145c-Sec13 (3bg0, [67]), and Nup84-Nup145c (3iko, [73]) interfaces are shown alongside the closest recapitulation of each interface found among the 9479 ensemble structures. A single full Nup84 complex structure is provided for reference.
B) Secondary structure elements. Helical regions (α-solenoids) are shown in pink. Sheet regions (β-propellers) are shown in cyan.
C) Indication of the N- and C-termini positions. Each of the Nup84 complex proteins is graded from its N-terminus (blue) to the C-terminus (yellow).
D) Crystal structure coverage; where crystal structures were used to represent components of the Nup84 complex structure, the complex is colored; where atomic comparative models were used, fragments are shown in gray.
E) The density of a representative Nup84 complex structure was fitted into our previously published NPC map [13] based on the relative positions of each component within the map. One of 8 instances of the structure in each of the two outer rings of the complex is shown.

Assessing the ensemble's accuracy

We assess our models using the following six criteria [13, 36]. First, our structures satisfy all the input restraints, even though the data were generated by disparate experiments. Second, all the data can be satisfied by a single cluster of structures at a resolution sufficient to determine domain-domain interactions; this demonstrates that the sum of our data has a low level of ambiguity. Third, the ensemble is consistent with the structural data not used to compute it. For example, it recapitulates the three heterodimeric interfaces determined by crystallography [66-69, 73], albeit at a lower resolution (15Å). Amongst the solutions in our ensemble a small number closely mimic the native interfaces, with an Ca RMSD less than 5Å relative to the native state (Figure 3-7A), demonstrating that correct high resolution solutions exist within our ensemble. Where available, we also compared the localization maps for interfaces generated in our 1.5nm precision ensemble with simulated maps of crystal dimers at 1.5nm resolution (Figure 3-8D), showing that our structure is consistent with available data at our defined precision. Fourth, a comparable ensemble is generated when random subsets of 10% of the data are removed (Materials and Methods), demonstrating the redundancy and correctness of the data as well as the structural ensemble. Fifth, the phenotypic data maps onto the structure in a manner that is not expected by chance (below). Sixth, we also computed the structure of the Nup84 complex by treating the 3 crystallographic dimers (above) as rigid bodies while again satisfying spatial restraints implied by our own data (Figure 3-8C). As a result, the precision of the ensemble of the Nup84 complex improved from 1.5 nm to 1.1 nm, without altering the molecular architecture of the complex. The most significant increase in the precision occurred for the b-propeller proteins Sec13 and Seh1 (Figure 3-8C), located closer to each other in the new ensemble. Although our biochemical analysis was not able to detect a direct interaction between these two proteins, we cannot exclude the possibility of interaction between them within the Nup84 complex structure. Further detailed analysis should be able to address this possibility.

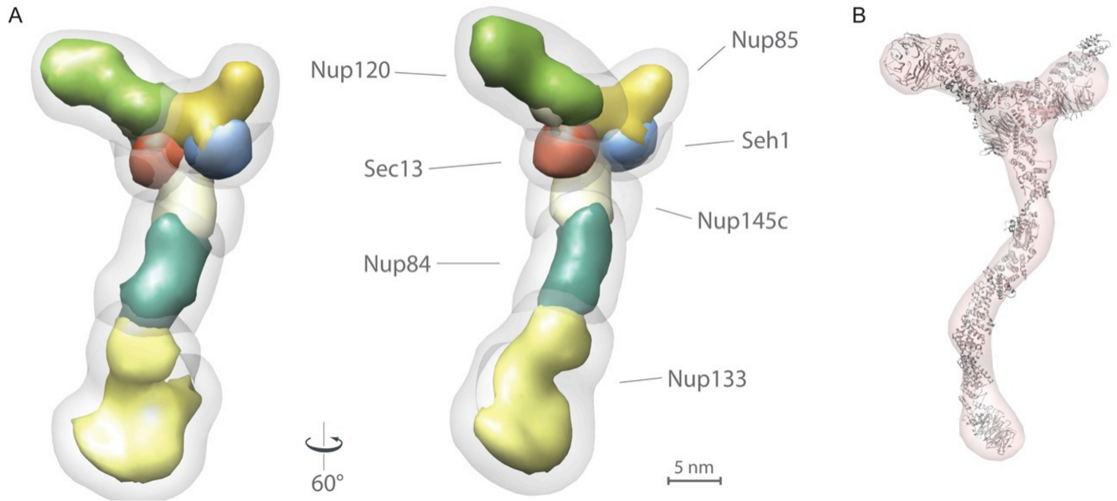


Figure 3-8. Structure of the Nup84 complex built without X-ray structures and fitting of a structure into an EM 3D map.
A) A localization map is shown for an ensemble of the Nup84 complex calculated with comparative models only, instead of crystallographic structures. Comparative models based on non-NPC templates were structurally aligned to the previously calculated ensemble structures using the SALIGN method [153] of MODELLER [94]. The complex structures were then subjected to the refinement stage (stage 3) described in the Materials and Methods section. Structures were aligned as previously described and produced a single cluster of solutions at 1.7 nm precision.
B) A single representative from the ensemble was fit into a previously published 3D negative stain EM map for the Nup84 complex [64]. This map was not used in the calculation of the structural ensemble.

Discussion

Structural Features of the Nup84 Complex

In agreement with previous work [64, 134, 145], the Nup84 complex structure is arranged in the shape of the letter Y (Figures 3-6 and 3-7). The complex is composed almost exclusively of a-solenoid and b-propeller domains, with the majority of inter-subunit interactions occurring through the a-solenoid domains (Figure 3-7B). Although we have used the available crystal structures of the subunits to calculate the Nup84 complex structure, at this resolution their comparative models result in a similar ensemble (Figure 3-8A). The degree of uncertainty in the localization of domains within the ensemble is shown in Figure 3-6. Despite the map being of moderate precision, key features of interest are nevertheless revealed, including domain and fold type arrangements, as well as nup orientations and interactions.

The longer of the two short arms is entirely filled by Nup120, which is oriented with its b-propeller forming the tip of the arm and the a-solenoid forming the main body, the C-terminal region of which connects to the hub (Figures 3-6 and 3-7). The a-solenoid region of Nup120 must form a single extended span along most of the length of the arm to connect the b-propeller of Nup120 to the rest of the complex. We used spheres to represent this region in the structure calculation. Such an extended a-solenoid is seen in the paralog of Nup120, Nup133, and so we have used a portion of Nup133 to represent these residues (Figures 3-4 through 3-8) [72]. Disconnecting Nup120 from the Nup84 complex does not lead to its release from the NPC (Figure 3-3), indicating that Nup120 also directly interacts with nups outside the Nup84 complex. Indeed, it has been shown that Nup157 can connect *in vitro* to the Nup84 complex in a Nup120-dependent fashion [154].

The other short arm is formed by Nup85 and Seh1 (Figures 3-6 and 3-7). Our structure is consistent with crystal structures from a Nup85 fragment showing extensive contacts between the b-sheet N-terminal blade of Nup85 with a cognate groove in Seh1 as well as between Seh1 loops and N-terminal residues of Nup85 (Figure 3-7A and D) [66, 68]. The C-terminal portion of Nup85 forms the hub connection (Figure 3-7), and deletions in this region cause loss of both

Nup85 and Seh1 from the complex (Figure 3-1). Previous fittings of crystal structures to an EM density map of the Nup84 complex [64, 73] have the Nup85/Seh1 dimer oriented in a manner inconsistent with our domain mapping and structure. However, the EM map used for these fittings is entirely consistent with our structure (Figure 3-8B), showing how integration of diverse data sources (here domain mapping and EM) can improve the accuracy of structure determinations. It has been shown that under crystal packing conditions, the N-terminus of Nup85 and Seh1 generate an extensive interaction surface used to connect it with another Nup85-Seh1 dimer [66]. We found no evidence in our data for such connection *in vivo*. However, the orientation of the Nup85-Seh1 dimer in our structure, facing outwards on its arm, allows us to suggest that the crystal packing interaction surface may actually be used by the Nup84 complex to connect to other Nups of the NPC core. This is also consistent with our identification of Nup85 and Seh1 as hot spots for fitness and connectivity in our map (see below).

The two short arms connect to the main stalk of the complex through a hub where an intricate interaction is established. There are no published atomic structures for this interface, but we determined that this hub is formed by the C-terminal extremes of three different a-solenoids (Nup85, Nup120 and Nup145c) and the N-terminus of Nup145c (Figure 3-7B). Though this N-terminal Nup145c region has been predicted to be disordered [67], it actually appears to be a part of the structured hub region, and is essential for maintaining the connection between the short arms and the long stalk. The precision of the ensemble does not allow us to define the exact interaction surfaces comprising the hub.

Although located near the hub (Figure 3-7), Sec13 is not required for the arms-stalk connection to form (see also Figures 3-1 and 3-2A). Similarly, the b-propeller protein Seh1 is also not required for the formation of the hub. The a-solenoid of Nup145c forms the upper part of the stalk, connecting through its middle region to Nup84, as has been previously described [69, 73] (Figure 3-7). However, our data do not support a suggested homodimerization of Nup145c [67], as the putative dimerization region instead forms a connection to Nup84, agreeing with several published crystal structures [69, 73] and point mutants [68]. In our ensemble, the N- and C-terminal regions of Nup145c connect directly with the Sec13 b-propeller (Figure 3-7); domain mapping data shows that both regions of Nup145c are required, but neither is individually sufficient, for a stable Sec13-Nup145c interaction (Figure 3-2B), agreeing with previous data [67].

The Nup84 a-solenoid forms the middle part of the stalk (Figures 3-6 and 3-7), connecting to Nup145c through its N-terminal region (Figure 3-7C). Nup133 and Nup84 connect through the C-terminal end of their a-solenoid domains (Figure 3-7), in a tail-to-tail fashion that has been previously seen in their human homologs [146]. The fact that certain Nup133 mutants that are unable to interact efficiently with Nup84 nevertheless localize to the NPC (Figure 3-3), suggests that proper localization of yeast Nup133 to the NPC is not exclusively driven by its interaction with Nup84 (seemingly unlike its human counterpart [146, 155]).

Position of the Nup84 Complex in the NPC

Our NPC map [13] positioned two sets of 8 copies of the Nup84 complex arranged to form the two outer rings of the NPC. The structure of the Nup84 complex described in the present study is entirely consistent with this arrangement (Figure 3-7E), although none of the data used in the determination of the original NPC map was used in the present Nup84 complex calculation. A “fence-like” model based on crystallographic homodimeric contacts (discussed above) [66] is incompatible with the arrangement of interfaces shown in our structure. Alternatively, a vertical arrangement for the Nup84 complex has been postulated [68]. While our data does not exclude variants of this possibility, alignment of our new data with our previous map points to a horizontal, head-to-tail arrangement of the Nup84 complex to comprise the NPC’s outer rings (Figure 3-7E). A recent fluorescence anisotropy study [156] defines the orientation of the human Nup107-Nup133 dimer as being parallel to the NE plane, in agreement with our arrangement for its yeast counterpart.

Mapping the Connectivity of the Nup84 Complex Within the NPC

It is still not known exactly how the Nup84 complex connects to the rest of the NPC, though it seems to be a stable rather than dynamic interaction [157]. One possibility is that a stable connection could be formed through a few strong, heterodimeric interactions to key neighboring Nups, of the same character that hold the Nup84 complex itself together; alternatively, the complex may be held to the rest of the NPC through multiple, relatively weak, cooperative interactions

[158]. To define the nature of the Nup84 complex connectivity, we have analyzed the stability of the connection between the Nup84 complex and several NPC core components using our set of truncation mutants. We affinity purified the Nup84 complex under conditions that co-purify nups forming the core of the NPC [13] (Figure 3-9A) in all the mutants that showed proper NPC localization (Figure 3-3). We quantified the signal obtained in the gels for five NPC core nups (inner ring nups Nup192, Nup188 and Nup170; linker nup Nic96 and cytoplasmic side nup Nup159; Figure 3-9B). Mutants that showed no connection to other components of the Nup84 complex were considered not informative and not used for quantification (Figure 3-9B, right). These measurements were then converted into a heat map, revealing regions involved in the connection between the Nup84 complex and the analyzed core nups (Figure 3-9C).

Truncations that affect the short arms and the hub ablate co-purification of core nups with the Nup84 complex, whereas truncations in most other regions of the Nup84 complex have a lesser effect on interactions with the analyzed core nups (Figure 3-9). Beta-propeller proteins and domains seem to have a key role on the interaction pattern, consistent with a peripheral location in the complex (below).

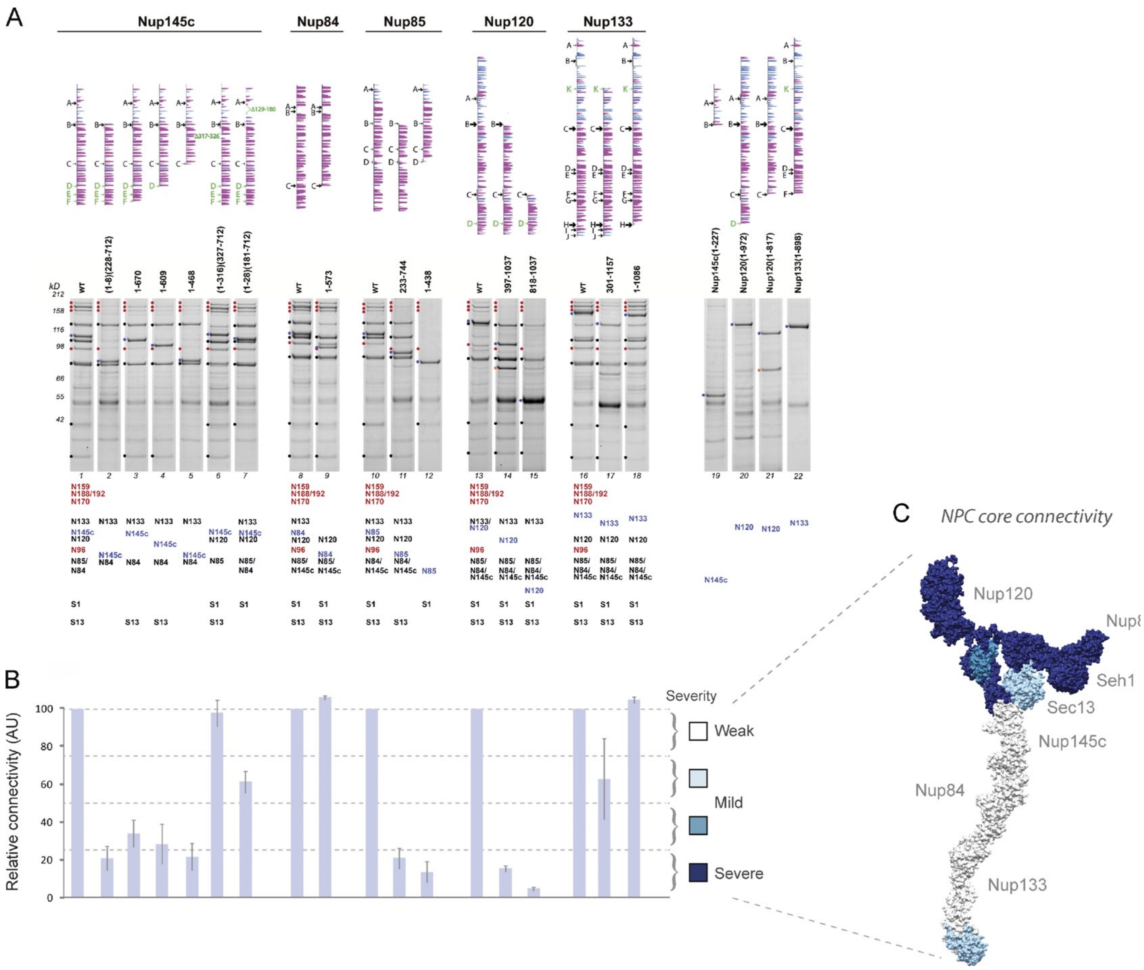


Figure 3-9. Fitness and NPC clustering analysis of the Nup84 complex truncations. A) Growth tests at different temperatures (24, 30 and 37°C) for the full length (wt) and truncated versions of the Nup84 complex nups. Serial ten-fold dilutions of cells were spotted on YPED plates and grown at the indicated temperatures for 1-3 days. Parental strains DF5 and w303 as well as the full length genomically tagged Nup84 complex nups (Nup84, Nup85, Nup133, Nup120, Nup145c) were included as controls. Each growing phenotype was quantified by semi-quantitative methods (Materials and Methods) and the obtained value (in arbitrary units) is shown on the right of each column. Plotted fitness value (mean +/- std error) for each measurement is shown on the right.
B) Left panel on each column shows the localization of a Nup49p-CFP reporter in wild type (wt) and truncation mutants as determined by fluorescence microscopy. A NUP133 gene deletion is also shown as reference for NPC clustering [159]. Right panel shows the DIC image of the same cells. The number shown on the lower left corner of the fluorescence picture represents the measured degree of clustering for each strain (represented by their normalized Coefficient of Variation, multiplied by 100 for representation purposes) in arbitrary units for n=30 cells (Materials and Methods).

Notably, while truncations in a given Nup84 complex protein cause the absolute loss of a very discrete subset of other Nup84 complex proteins, such deletions did not lead to the loss of discrete subsets of core Nups. Instead, deletions in the hot spots led to a varying degree of loss of all the core Nups simultaneously, while deletions elsewhere had little effect. This is inconsistent with the idea of that there is a single or small number of discrete high affinity interactions, analogous to those comprising the Nup84 complex itself, that hold the complex in the NPC. Rather, as seen recently with the Nic96 complex [158], the distribution and behavior of the hot spots show a pattern consistent with the Nup84 complex connecting through multiple, relatively weak, cooperative interactions to the NPC core. The presence of such cooperative interactions between putative stable building blocks of the NPC (such as the Nup84 complex) might be necessary for the NPC assembly process, as it ensures that the core NPC would only be formed when all the building blocks are located together in a certain region of the NE, and would prevent these building blocks assembling prematurely elsewhere in the cell. In summary, these connectivity experiments support the idea that the short arms of the Nup84 complex are involved in establishing key cooperative contacts with the core components of the NPC.

Structure-Function Mapping of the Nup84 Complex

An advantage of our approach is that the same constructs used for structural domain mapping can also generate phenotypic profiles for functional characterization of those same domains. Integrating both structural and phenotypic information is thus efficient, allowing us to map the regions of the complex that fulfill relevant functional roles. We have analyzed two of the most characteristic phenotypes common to mutations in the Nup84 complex components: fitness (“growth”) and NPC clustering [138, 139]. The different truncations displayed a wide range of phenotypic defects in both fitness (Figure 3-10A) and NPC clustering (Figure 3-10B), from mild to severe. To compare the severity of the different phenotypes in each strain, we performed semi-quantitative measurements (Materials and Methods). These measurements were converted into heat maps of the Nup84 complex depicting the severity of the phenotypes (Figure 3-11 and Figure 3-12). The distribution of hot spots in the two phenotypic maps is non-random, forming distinct patterns in the structure. Severe fitness phenotypes map strongly to the tips of the two short arms of the complex, and to a lesser extent to the tip and middle region of the stalk (Figure 9A). The connectivity heat map shows strong similarities to the fitness phenotype heat map (compare Figures 8C and 9A), but does not strongly resemble the heat map generated for the NPC clustering phenotype (Figure 9B), for which we propose a different molecular basis (below). Taken together, this suggests that the tips of the Nup84 complex are involved in connecting the complex to other components of the NPC scaffold - such as the inner rings and linker nups - and stabilizing the normal functional structure of the NPC’s scaffold. Thus, a major reason for the strong fitness phenotypes observed is likely the loss of connectivity between the Nup84 complex and the NPC core, leading to a functionally compromised NPC.

The clustering phenotype hot spots map mainly to the Nup133 tip of the stalk and the Nup120 short arm (Figure 9B). Strikingly, truncations affecting the short arm component Nup85 do not exhibit such a marked clustered NPC distribution, in agreement with the observed low level of NPC clustering for *nup85* partial gene disruptions and a null allele of its dimeric partner Seh1 [135, 145]. The clustering hotspot-containing proteins Nup133 and Nup120 are known paralogs and resemble vesicle coating complex (VCC) proteins such as clathrin [13, 59]. The NE is formed by two parallel membranes, the outer and inner nuclear membranes; at the outer rings, where we localize the Nup84 complex, these two membranes are deflected to form the pore membrane which anchors the NPC in the NE [13, 46, 89, 160]. We therefore hypothesize that the Nup84 complex, and particularly the clathrin-like Nup120 and Nup133, are key proteins in ensuring the stability of the interface between the NPC and the surrounding outer and inner nuclear membranes. Clustered NPCs are known to result from the migration of pre-existing NPCs into aggregates, rather than from the assembly of new NPCs into a fixed site on the NE [161]. Furthermore, we suggest that clustering of NPCs can result from destabilizing the interface of the NE membranes with the Nup84 complex outer rings. Hence, we might expect an aggravation of the clustering phenotype were we to further destabilize the NE membranes. To test this hypothesis, we used Benzyl alcohol (BA), a reagent that fluidizes and destabilizes membranes [162], which has been extensively used to investigate membrane dynamics and even membrane-NPC interactions in yeast and other fungi [163-165]. Although BA does not affect the stability of the purified Nup84 complex, the degree of clustering in mutants was indeed aggravated by the treatment with BA, while a wild-type strain was not affected, with the BA-induced clustering being fully reversible (Figure 9C). These results support the hypothesis that the b-a nups Nup133 and Nup120 are directly contacting and/or helping in the stabilization of the curvature of the NPC membrane, just as do VCCs.

With the a-solenoids forming the architectural core of the complex (Figure 6B) an obvious question arises: what is the function of the b-propeller domains? As mentioned above, they do not appear to be structural keystones of the Nup84 complex. However, our functional data identifies them as hot spots for structural integrity of the NPC, connectivity to the NPC core and stabilization of the NE membrane (Figures 8 and 9). Altogether these data suggest that the b-propeller domains of the Nup84 complex are involved in NPC-NE membrane interactions and in connecting each Nup84 complex with neighboring complexes in the NPC. Precisely such a role for b-propeller domains in connecting complexes is found in COPI, COPII, and clathrin complexes [63, 65, 166-168].

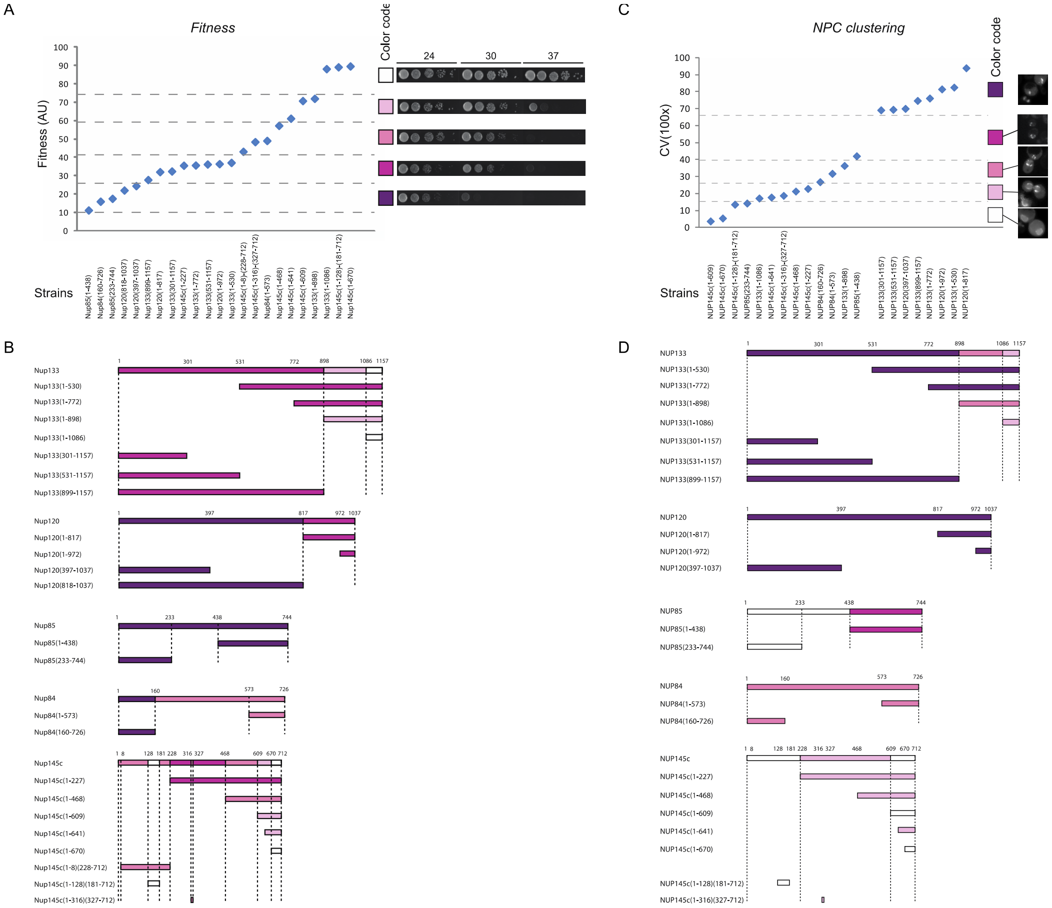


Figure 3-10. Mapping of fitness and NPC clustering phenotypes into the Nup84 complex structure.

A) Quantified fitness phenotype (Materials and Methods and Figure 7A) was plotted (mean value) for each mutant by increasing level of fitness. Proportional divisions (dashed gray lanes) were made and a color code was assigned to each division (right), from dark purple to white (colored squares) to match the increasing level of fitness. Representative examples of strains included in each division are shown on the right.

B) Mapping of the color code described in panel A into the Nup84 complex components. Horizontal lines represent the amino acid residue length of each protein and truncated version; amino acid residue positions are shown on top of the lines. Seh1 and Sec13 color code was defined as the one for minimal truncation causing their loss from the Nup84 complex (Figure 2A).

C) Quantified NPC clustering phenotype (Materials and Methods) was plotted (mean value) for each mutant by increasing level of NPC clustering. Due to the poor linearity of the data, divisions based on the observed leaps of data (dashed gray lanes) were made and a color code was assigned to each division (right), from white to dark purple (colored squares) to match the increasing level of NPC clustering. Representative examples of strains included in each division are shown on the right.

D) Mapping of the color code described in panel A into the Nup84 complex components. Horizontal lines represent the amino acid residue length of each protein and truncated version; amino acid residue positions are shown on top of the lines. Seh1 and Sec13 color code was defined as the one for minimal truncation causing their loss from the Nup84 complex (Figure 2A).

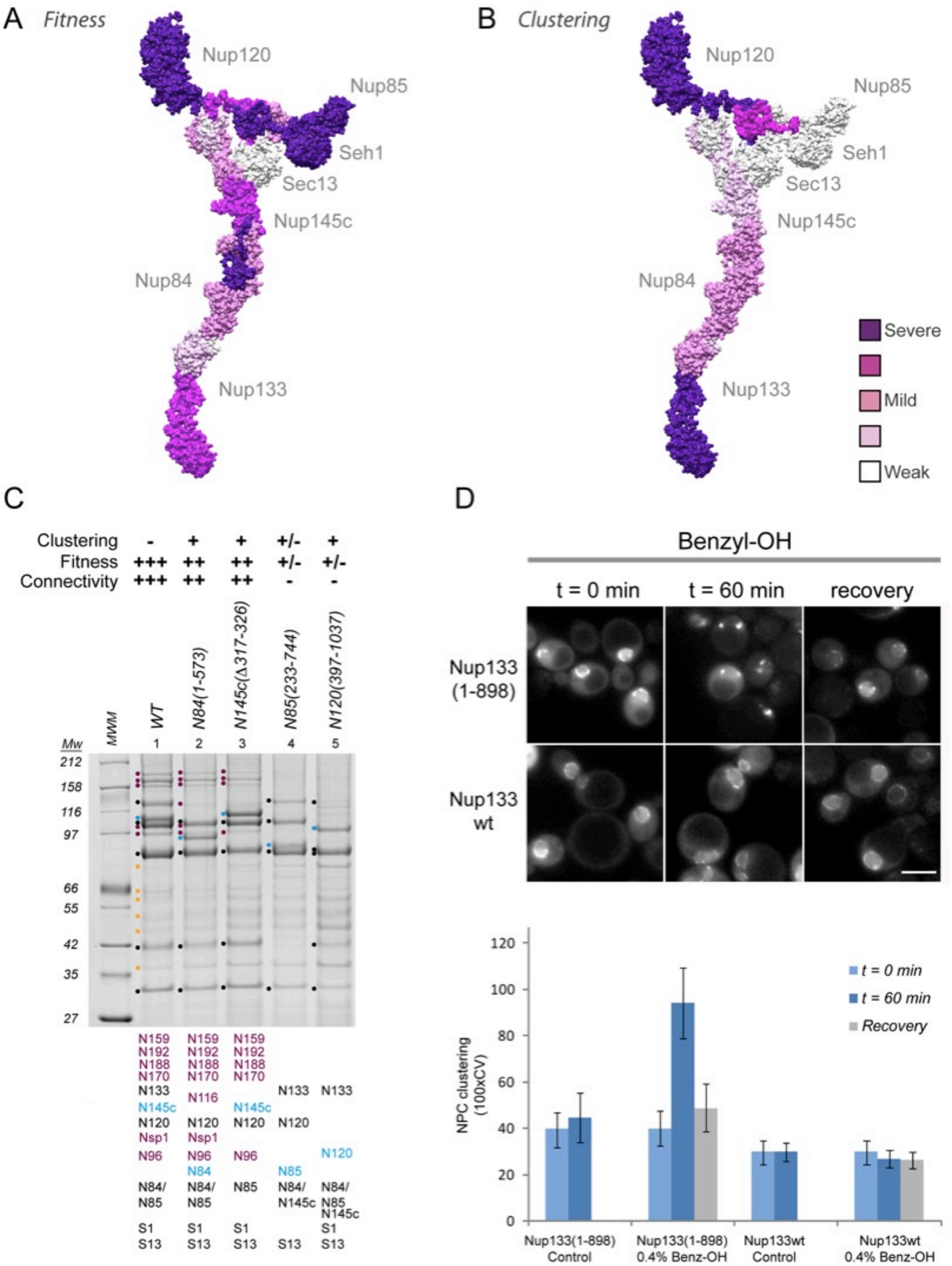

A
Fitness
Nup120
Nup85
Seh1
Sec13
Nup145c
Nup84
Nup133
B
Clustering
Nup120
Nup85
Seh1
Sec13
Nup145c
Nup84
Nup133
Severe
Mild
Weak
C
Clustering - + + +/- +
Fitness +++ ++ ++ +/- +/-
Connectivity +++ ++ ++ - -
Mw
MWM
WT
N84(1-573)
N145c(Δ317-326)
N85(233-744)
N120(397-1037)
1
2
3
4
5
212
158
116
97
66
55
42
35
27
N159 N159 N159
N192 N192 N192
N188 N188 N188
N170 N170 N170
N133 N133 N133
N116
N145c N145c
N120 N120 N120 N120
Nsp1 Nsp1
N96 N96 N96 N120
N84 N85
N84/ N84/ N85 N84/ N84/
N85 N85 N145c N85
N145c
S1 S1 S1 S1
S13 S13 S13 S13 S13
D
Benzyl-OH
t = 0 min
t = 60 min
recovery
Nup133 (1-898)
Nup133 wt
NPC clustering (100xCV)
0
20
40
60
80
100
120
t = 0 min
t = 60 min
Recovery
Nup133(1-898) Control
Nup133(1-898) 0.4% Benz-OH
Nup133wt Control
Nup133wt 0.4% Benz-OH

Figure 3-11. Fitness correlates with loss of interactions between the Nup84 complex and other core NPC nups, while NPC clustering is related to NE membrane stability.
A and B) Effect of truncated regions of the Nup84 complex on organism fitness and NPC clustering. The severity of fitness phenotypes (A) and NPC clustering (B), for specific truncations of the Nup84 complex are shown mapped into a single Nup84 complex structure. Color code is shown on the right and in Figure S3.
C) Coomassie stained SDS-PAGE gel showing the Nup84 complex interacting proteins for wild type (wt) and truncations indicated above each lane. Mass spectrometric identification of bands marked by filled circles are indicated in order below each lane; PrA-tagged nups, blue; Nup84 complex nups, black; core NPC nups, purple; contaminants, orange. Degree of NPC clustering (-, wild type; +/-, weak clustering; +, mild clustering; ++, strong clustering), fitness (-, very sick; +/-, low fitness; ++, medium fitness; +++, wild type level) and core NPC interactions (-, no core nups identified; +, only a subset of core nups identified; ++, all core nups identified but decreased intensity; +++, wild type) are shown above.
D) Membrane destabilizing assay. Benzyl alcohol (0.4%) was added to wild type (wt) or truncated mutant Nup133(1-898) (similar behavior was observed for other mutant strains (not shown)). Time course localization of fluorescent reporter Nup49-CFP after treatment and for non-treated controls are shown in the upper part. Quantification of the level of NPC clustering (Materials and Methods) for each time point is shown in the lower part (mean +/- s.d.). Scale bar, 5 μm.

Evolution of the Nup84 Complex

The connections between Nup84 complex components are established mainly between the a-solenoid domains, in two different fashions (Figures 6, 10) [67-69] first, tail to tail, as in the hub and Nup133-Nup84 connections, reminiscent of the COPI b'-cop and a-cop subunits and the clathrin triskelion arrangement; and second, the connection between the mid a-solenoid domain of Nup145c and the N-terminal head of Nup84 is reminiscent of the dimeric connection between Sec31 units in COPII ([37, 65, 169]. The Nup84 complex is thus an excellent exemplar for the protocoatomer hypothesis [13, 59, 60, 170], having structural and connectivity features found in several of the VCCs and even sharing components with other vesicle coating and tethering complexes [59, 137].

We have previously suggested that the descent of the modern coatomer-like complexes from a single common ancestor occurred through the processes of duplication, divergence, specialization, and secondary loss [13, 36, 59, 60, 170]. In the light of the present structure, we can speculate how evolution shaped the modern Nup84 complex (Figure 3-12). A hypothetical duplication symmetry axis can be drawn between the interface of Nup84 and Nup145c, consistent with our previous data suggesting that each of the eight spokes in the NPC consists of two columns that arose from a genome duplication event, and that the Nup84 complex spans both columns ((Figure 6E) and [13, 36]). On both sides of this axis, paralogous components can be found: the b-a paralogs Nup120/Nup133 and the a-solenoid paralogs Nup145c/Nup84. Crystal structures have shown that the connection between Nup145c and Nup84 is established through a similar U-turn region of their a-solenoids [69, 73]. Both of their C-terminal domains reach back to connect through their C-terminal extremes to a b-a protein (Nup120 and Nup133, respectively; Figure 10). Nup145c makes another connection involving the C-terminus of Nup85, forming a triskelion-like structure that can be reconstituted *in vitro* [134]. We suggest that originally, two triskelion-like subcomplexes were connected through their Nup145c-Nup84 homologs (Figure 10). The current Nup84 complex architecture would have been generated by the simple loss of a hypothetical N-terminal region of the Nup84 ancestor protein and its associated hypothetical Nup85-like protein. The conservation of many of the Nups comprising the Nup84 complex between highly divergent branches of the eukaryotic evolutionary tree [45] suggest that this duplication-loss scenario occurred near or even before the root of the eukaryotic tree. In this proposed duplication scheme, we note that the functional hot spots detected in our analysis are surprisingly well symmetrically located, with the functional hot spots mainly mapping to the boundaries between the duplicate columns and spokes (Figure 9E) [13]. Yet more ancient duplications can be envisioned from the structural similarity between the paralogous components Nup85 and Nup145c and their paralogous partners, Seh1 and Sec13, respectively.

While COPI, COPII, and clathrin building blocks arrange in symmetrical homo-oligomers to form cages that coat membrane vesicles, similar building blocks arrange in asymmetrical hetero-oligomers to form the Nup84 complex. We suggest that this kind of hetero-oligomeric arrangement evolved through duplication and divergence as envisioned above, in order to prevent the Nup84 complex coat from oligomerizing into a continuous cage. Instead, hetero-oligomerization (with one monomer being in this respect monofunctional) ensures that the region of membrane coated by the NPC is sharply limited, both radially and vertically, forming a circumscribed macromolecular structure of the correct size and morphology to form a discrete, precise macromolecular channel.

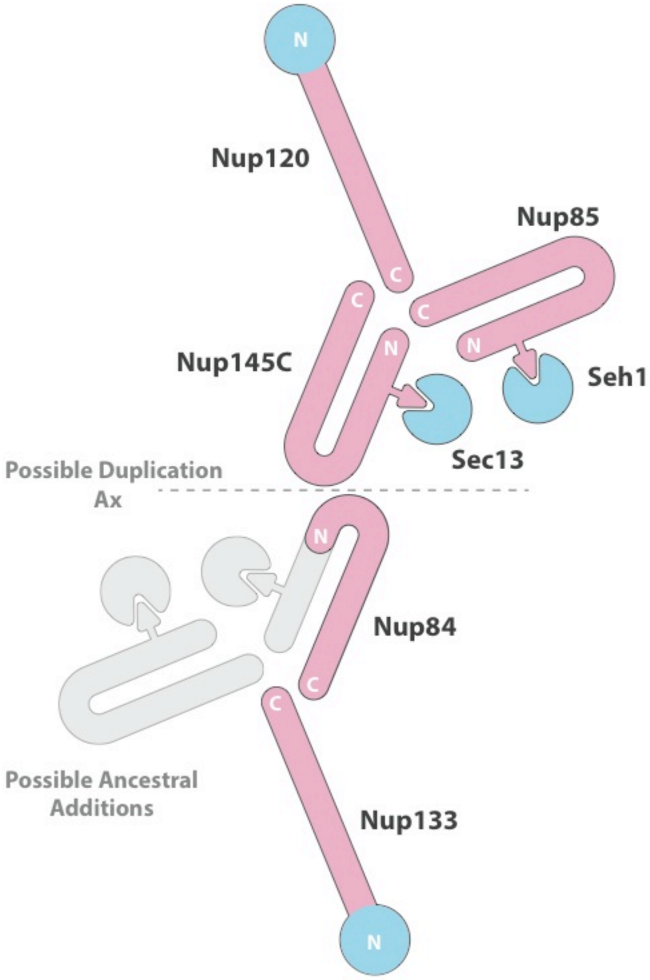


Figure 3-12. Potential origin of the Nup84 complex through ancient duplications and losses.

A diagram of the Nup84 complex nup arrangement is shown. β-propeller domains are colored in cyan, α-solenoid domains are colored in magenta, and invasion domains (Hsia et al., 2007; Brohawn et al., 2008) are indicated with magenta arrows. The N- and C-terminal ends of each protein are highlighted. The hypothetical duplication axis is shown as a dashed gray line dividing the complex between the Nup145c and Nup84 interaction surfaces. Possible ancestral nups lost after the duplication are shown as faded gray domains.

Materials and Methods

Yeast strains and materials.

All yeast strains used in this study are listed in Table S2. Truncated versions of the Nup84 complex nucleoporins were constructed using the DF5 strain (Mata ura3-52 his3-A 200 trp 1-1 leu2-3,112 lys2-801) or W303 (Mata ade2-1 ura3-1 his3-11,15 trp 1-1 leu2- 3,112 can1-100) backgrounds. Strains were grown at 30°C in YPD media (1% yeast extract, 2% bactopeptone, and 2% glucose), except for the thermo-sensitive strains that were grown at 25°C in the same media. The following materials were used in this study: Dynabeads M-270 Epoxy (143.02D; Invitrogen Dynal AS, Oslo, Norway); rabbit IgG (55944; ICN Biochemicals, Costa Mesa, CA); protease inhibitor cocktail (P-8340; Sigma, St. Louis, MO); Asp-N, Lys-C, and trypsin (11420488001, 11420429001, 11418025001; Roche Diagnostics, Indianapolis, IN); HRP-rabbit IgG (011-0303-003; Jackson ImmunoResearch, West Grove, PA); anti-GFP antibody (11814460001; Roche Diagnostics); and anti-mouse IgG-HRP (NA931V; Amersham Biosciences UK Limited, Little Chalfont, Buckinghamshire, UK), Solution P (2 mg Pepstatin A, 90mg PMSF, 5 ml absolute ethanol).

Domain Mapping of the Nup84 complex.

Domains forming the Nup84 complex nucleoporins (nups) were dissected by carboxy and amino-terminal genomic truncation and PrA tagging, so each domain was expressed under the control of its native promoter. Points of truncation for each nup were selected based on the protease sensitive amino acids described by the Protease Accessibility Laddering technique in references [59, 61]. Many of the cleavage sites had been narrowed to a single amino acid residue, and these served as the deletion sites; otherwise, the point of deletion was selected as the midpoint of the corresponding disordered region [59]. Some deletions were also designed to eliminate conserved regions identified by multiple sequence alignment and secondary structure prediction [59]. The carboxy-terminal deletions were constructed by in-frame integration of a PCR-derived DNA fragment encoding flanking gene specific sequences (45 nucleotides) and the IgG binding domains of *Staphylococcus aureus* PrA amplified from the plasmid pProtA/HIS5, which carries the *Schyzosaccharomyces pombe HIS5* gene as its selectable marker [44]. The sequences for all the oligonucleotides used in this study are available upon request. Positive colonies were selected on SC medium lacking histidine. Internal and N-terminal deletions were constructed using the Cre-lox P recombination system, as described previously [171, 172]. Briefly, a PCR-derived DNA cassette, encoding gene specific overhanging sequences (45 nucleotides) and the *Kluyveromyces lactis LEU2* gene marker flanked by LoxP sequences, was amplified from the plasmid pUG73 (Euroscarf). The cassette was used to transform a strain where the target nup was already PrA-tagged [44]. During the PCR, the first LoxP sequence of the cassette was modified to maintain the correct reading frame of the nup by modification of the vector specific sequence in the original 5' oligonucleotide (5'-gccagc**A**gaagcttcgtacgc-3') [171]. Cre-recombinase expression was used to remove the selection marker in positive clones. Tetrads were dissected to obtain haploid strains carrying the mutant copy of the nup, when diploid strains were used for the construction. Proper integration was tested by sequencing and the constructs found to be stably expressed by western-blot analysis were analyzed (Table S1).

The NUP145 gene is transcribed as a precursor polypeptide that is post-translationally self-cleaved to produce two independent peptides, Nup145n and Nup145c [173]. To ensure that the expression of each of our Nup145c deletions was controlled by its native promoter, they were designed to retain the target region of the self-catalyzed cleavage between NUP145n and NUP145c (first 9 amino acids of N145c) and their ability to undergo self-cleavage was tested by inmunoblot analysis. C-terminal tagging and truncations of Nup85 and Nup120 using the GFP tag were performed essentially as described above for the PrA tagged versions, using a PCR-derived DNA fragment encoding the GFP protein and the *E. coli kanr* Geneticin/G418 resistance gene, amplified from the plasmid pFA6-GFP-kanMX6 [174]. Truncated mutants were crossed with strains expressing the wild type copy of Nup84 tagged either with PrA or the ppx-PrA version (see below) to obtain strains carrying a combination of the GFP-tagged truncated nup and a PrA-tagged version of a wild type Nup84 complex nup.

All mutant strains were analyzed by immunofluorescence as previously described [150] to analyze the subcellular localization of the Protein-A tagged truncated nups (Figure S1). Cells were grown to early log phase and then fixed with 4% formaldehyde in phosphate buffer for 5 min. Fixed cells were washed with 1.2M sorbitol-phosphate buffer and spheroplasted using 10% glusulase

(DuPont,Wilmington, DE, USA) and 1mg/ml Zymolyase 20T (Seikagaku). Spheroplasts were immovilized into poly-lysine coated slides and dehydrated by sequential immersion in cold methanol and acetone. Cells were blocked with 2% powdered non-fat milk in TBS buffer for 10 minutes at room temperature. Primary antibody (1:300 of rabbit affinity purified antibody to mouse IgG, ICN Biomedicals) and secondary antibody (Alexa Fluor 568-conjugated donkey anti-rabbit IgG (H+L), Invitrogen) were diluted in blocking solution and incubated overnight at 4°C and 1 hour at room temperature, respectively. Cells were mounted (100 mg/ml p-phenylenediamine in PBS with 0.05 μg/ml of DAPI) and visualized at room temperature with a 63x, 1.4 numerical aperture plan-apochromat objective (Zeiss) using a Carl Zeiss Axioplan 2 microscope equipped with a Hamamatsu Orca ER-cooled CCD camera. The system was controlled with Openlab imaging software (Perkin Elmer). Final images were assembled, and gamma levels adjusted to enhance contrast only, using Adobe Photoshop software.

To analyze the proteins interacting with the wild type nup and each truncated construct we followed the affinity purification procedure described previously by our groups [36, 175]. Briefly, liquid nitrogen frozen yeats cells were cryogenically broken using a planetary ball mill (Retch). Cryogrindate power was resuspended in immunopurification buffer and incubated for 30 minutes with IgG-conjugated Dynabeads (Invitrogen). Magnetic beads were washed 5 times with 1ml of immunopurification buffer, bound proteins were eluted using 0.5M NH4OH, 0.5mM EDTA. Elution was lyophilized in a speed-vac and the resulting pellet resuspended in SDS-PAGE loading buffer. Purifications were optimized to utilize the mildest conditions that isolated only the wild type Nup84 complex (Table S1), and none of the other nups with which it can co-isolate [13, 36], so that we only probe interactions within the 7-protein complex. Several of the mild isolation conditions were explored for each purification to ensure that the maximum number of co-isolating Nup84 components was detected. Isolated proteins were detected by R250 Coomassie staining and identified by MALDI mass spectrometry analysis [13, 36] (data available upon request).

Truncated nups that were not able to interact with other Nup84 complex components were further analyzed by Protease Accessibility Laddering (PAL) to determine whether major misfolding was the main reason for the loss of interaction. PAL was performed as previously described [61]. Briefly, immunopurified proteins (see above) still bound to the magnetic matrix were partially digested using Lys-C or AspN proteases. C-terminal fragments still bound to the matrix after digestion were eluted as described above and analyzed by western-blot against the Protein-A moiety. The band pattern obtained was compared with the pattern generated by a wild type nucleoporin treated in parallel.

Purification of native Nup84 complexes.

To be able to purify the native Nup84 complex, we constructed strains in which the NUP encoding gene was genomically tagged with PrA preceded by the human rhinovirus 3C protease (ppx) target sequence (GLEVLFQGPS). The sequence was introduced by PCR amplification of the transformation cassette from the plasmid pProtA/HIS5. We isolated the complex by affinity purification as described above and released it from the affinity matrix by protease digestion. The recovered sample was then centrifuged at 20.000 g for 10 min. Supernatant (50-100 μl) was loaded on top of a 5-20% sucrose gradient made in digestion buffer plus protease inhibitors. Gradients were ultracentrifuged on a SW55 Ti rotor (Beckman) at 50.000 rpm and 5°C for 7 hours. Gradients were manually unloaded from the top in 12 fractions of 410 μl. Fractions were analyzed by SDS-PAGE, R250 Coomassie staining and mass spectrometry identification of the proteins. The six member Nup84 complex (lacking Nup133 [145]) was isolated as described above and purified by size exclusion chromatography using a TSKgel G4000SW column (Tosoh Biosciences) in 20 mM Hepes/KOH pH 7.4, 110 mM KOAc, 250 mM NaCl, 2 mM MgCl2, 0.5% Triton-X100, 0.1% Tween-20 buffer.

Electron microscopy.

Samples from selected fractions for electron microscopy were diluted to a concentration of ~50μg/ml. A 3 μl drop was applied to a carbon coated grid that had not been previously glow discharged. The drop was blotted after ~30s and stained using 3 drops of 1% uranyl formate. The grid was air-dried before observation in a JEOL-2100F operating at 200kV. The images were collected using low-dose (~6e/Å^2) on a 2kX2k Tietz CCD camera, with a defocus of 3μm.

Images of individual particles were selected and classified using the program EMAN [176]. 1309 individual particles for the whole Nup84 complex, 799 for the

Nup85(233-744) truncation complex and 1973 for the six members complex (lacking Nup133) were selected for two-dimensional image classification.

Representation of the Nup84 complex.

Because few Nup84 complex components have full atomic resolution coverage, each nup was represented as a hybrid rigid body consisting of its crystallographic data (if available), comparative models (if an accurate alignment could be computed), or a string of beads mimicking the coarse shape of a related known fold. Regions not represented by either an atomic X-ray structure or a comparative model included linkers (up to 39 residues in length) and short segments at the termini (up to 62 residues in length); these regions were represented with a sphere for every 50 residues by approximating the average size of a 50 residue globular domain as previously described [36, 177]. Where less than 50 residues needed to be represented, a smaller sphere size corresponding to that number of residues was used.

Conditional connectivity restraints from domain deletion data.

We used affinity purification experiments with domain deletions to identify connected complexes (composites) containing a mix of whole proteins, protein fragments, and domains. In some cases, multiple such deletion composites overlap such that information contained by two or more composites together cannot be derived by considering each composite separately. Here, we consider two such cases. First, if the same protein or set of proteins is lost in two different deletion experiments, where both of the deletions are performed on overlapping regions of a single protein, we deduce that the point of interaction for the lost protein or proteins is in the overlapping region. Second, if the same protein or set of proteins is lost in two different non-overlapping truncations and a third composite contains all the lost proteins and both truncated regions, we deduce that the truncated regions must both interact with the lost proteins in the same location. Additional conditional connectivity restraints are added for these inferred interactions.

Restraint descriptions.

S_{dm}: Each composite is encoded as a graph of nodes connected by edges. Each node represents a protein or a protein fragment, and each edge represents a possible interaction between connected nodes. Edges are generated for all possible interactions that might be implied by the corresponding pullout experiment. An edge imposes a putative distance restraint, scoring well when the two potentially interacting proteins (fragments) are actually interacting in the assessed assembly configuration, and scoring poorly when the proteins (fragments) are too far apart to interact. For each composite graph, the best-scoring set of edges needed to connect the components of the pullout composite is chosen by calculating a minimal spanning tree (MST) on the composite graph. Distance restraints for each of the MST edges are then added to the overall scoring function, and thus used to drive the optimization of the complex. The connectivity of the complex selected by the connectivity restraint can change at each optimization step as the configuration of the complex changes.

S_{tri}: A structure is restrained to a bounding triangle as follows: 1) The two most distant points (points A and B) in the structure are chosen. 2) The most distant point (point C) to a line drawn between A and B (line L) measured along some line perpendicular to L (L_n) is chosen. 3) The three distances between points A, B, and C are measured. The distances are ordered from largest to smallest, and restrained by a harmonic upper bound to match the corresponding distances in the target triangle.

S_{acc}: A Connolly surface is calculated for each component [126]. Interacting surfaces (those with at least 2 points within 5 Å) were scored using a two part term that rewards the total number of surface atom pairs of within a distance cutoff, and penalizes according to a weighted sum of all clashing pairs of atoms[31].

S_{em}: The optimal projection of a model was found by enumerating all possible orientations (3 angles) and positions (2 translations) of the tested model relative to the class average, and selecting the projection with the highest cross-correlation between the projected model density and the class average.

Sampling of good scoring structures.

Models that satisfied all restraints described above were obtained in four stages. Starting positions for the optimization were random configurations of the subunits are were chosen within a cube 3 times the length of the Nup84 complex along its principal component, as determined from the 2D EM class average. In the first, coarse optimization stage, 100,000 coarse models were calculated by a Monte Carlo optimization of the sum of S_{dm} and S_{tri}, each starting from a different random initial configuration. In this stage, an MC move consisted of either a 2 degree rotation or 5 Å translation at a temperature (T) of 273. Moves were accepted or rejected according to the Metropolis criterion: If $S < 0$, the move was always accepted; otherwise, the move was accepted with the probability of $e^{-S/T}$, where S is the total score and ΔS is the difference between the proposed total score and the previous score. Each optimization consisted of 5 MC moves followed by 20 CG refinement steps repeated 100 times, for a total of 500 MC moves and 2000 CG steps. In the second, refinement stage, the models were refined by a Monte Carlo optimization of the sum of S_{acc}, S_{dm}, and S_{tri}, with smaller step sizes; with rotations limited to 1 degree and translations to 3.5Å, with a total of 300 MC steps. In the third, filtering stage, the 10,000 top scoring models were filtered by requiring sufficiently good alignment to the class averages of the EM images of the full and partial complex (Figure 3-3A and C). Finally, the remaining 9479 structures were relaxed by optimizing the sum of S_{acc}, S_{dm}, S_{tri}, and S_{em}, using the simplex algorithm [152].

Ensemble analysis.

Using the MODELLER-9v8 MALIGN3D command, the 9479 good-scoring structures were aligned by minimizing the sum of squared distances between equivalent pairs of $C\alpha$ atoms to an iteratively updated average structure, pruning

$C\alpha$ distances between equivalent pairs longer than 10 Å [94]. The structures were then clustered by pairwise $C\alpha$ RMSD using by single-linkage agglomerative hierarchical clustering showing a single dominant cluster of solutions. UCSF Chimera was used to visualize all molecular structures and density maps, and to generate simulated density maps [178].

Building the Nup84 complex using only comparative models.

Atomic structures of proteins are usually obtained through techniques such as X-ray crystallography and NMR spectroscopy. However, despite sterling efforts, experimentally determined atomic structures are still not available for the vast majority of the eukaryotic proteome. Fortunately, even when an experimentally determined structure is not available, a comparative model based on a related known structure can frequently be computed [179]. In fact, more than two orders of magnitude more sequences can be modeled by comparative modeling than the number of experimentally determined structures [180]. Thus, a key question is how do medium-resolution complex structures computed with crystallographic subunit structures, such as our Nup84 complex structure, differ from those computed with less accurate comparative models. We addressed this question for the Nup84 complex, by computing the ensemble of the Nup84 complex structures using only comparative models of its subunits based on non-nup templates, instead of crystallographic structures; flexible strings of beads were again used where model coverage was not available. All comparative models were generated with the program MODELLER-9v8 [94], using target-template alignments from the fold assignment servers Phyre [181] and pGenThreader [182]. The ensemble of resulting Nup84 complex structures (Figure 3-8A) is clearly similar to that obtained with crystal structures (Figure 6), thus demonstrating that our integrative approach can also be useful when only comparative models are used instead of experimentally determined atomic subunit structures.

Phenotypic analysis.

To analyze the growth at different temperatures strains were grown in liquid YPED media overnight at 25°C, cells were counted and diluted to a final concentration of 20,000 cells/ml. Four 10-fold serial dilutions were made and spotted on YPED plates that were incubated at 25°C, 30°C and 37°C for 1-2 days. Biological replicas of each experiment were performed. Plates were imaged using a Fujifilm LAS-3000 system (linear detection range). Double blinded, semi-quantitative estimation of growth was performed using MultiGauge (Fujifilm) software and ImageJ software (Rasband, W.S., ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, http://rsb.info.nih.gov/ij/, 1997-2009), which rendered similar results. Briefly, the value (in arbitrary units) for each strain growth at each temperature was determined by adding the quantified density of the five 10-fold dilution spots. Values were normalized to each wild type, setting 100 arbitrary units as wild type growth value [183].

To analyze the Nup84 complex connectivity to the NPC core for each truncation mutant, affinity purifications were performed under buffer conditions that preserve those interactions in wild type Nup84 complex PrA-tagged strains. Only mutants that showed NPC localization by immunofluorescence (Figure S1) were used for the analysis. Affinity purifications were performed in parallel using the same buffer composition (20 mM Hepes pH 7.4, 250 mM potassium acetate, 125 mM sodium chloride, 1% Triton-X100, 0.1% Tween-20, 2 mM magnesium chloride, 1mM DTT or 1.2M ammonium acetate pH 7.0, 0.5% Triton, 0.1% Tween-20, 1mM DTT). Coomassie-stained, SDS-PAGE resolved bands were identified by MALDI-MS. Nup159, Nup192, Nup188, Nup170, POM152, Nup157, Nup116, Nup100, Nsp1 and Nic96 were identified as the main bands. Bands corresponding to POM152, Nup116, Nup100, and Nsp1 were not quantified due to extensive overlapping. Band intensities were quantified in two independent experiments for each wild type or mutant construct using the ImageJ (Rasband, W.S., ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, http://rsb.info.nih.gov/ij/, 1997-2009) gel tool. The amount of each protein, calibrated with a bovine serum albumin standard, was transformed into molar amounts and the relative molar ratio relative to the ProteinA-tagged handle for each mutant was normalized to the wild type value and plotted into a bar graph. Divisions of 25% units were defined and assigned a color in the blue palette to generate a heat map for the NPC core connectivity within the Nup84 complex (see Figures 10 and 11).

To analyze the NPC distribution, strains were transformed with plasmid pXYNUP49-CFP [184]. Strains expressing truncations Nup120(397-1037) and Nup133(531-1157), that showed synthetic sickness with plasmid pXYNUP49-

CFP, were transformed with plasmid pBT029 (KAP121-CFP) [185]. Cells were grown in selective minimal media at 25°C to mid-log phase and harvested by centrifugation. Cells were visualized with a 63x 1.4 numerical aperture plan-apochromat objective using a Carl Zeiss Axioplan 2 microscope equipped with a Hamamatsu Orca ER-cooled CCD camera. The system was controlled with Openlab imaging software (Perkin Elmer). Three image sections of several fields of cells in growth media were acquired at 0.5μm increments. For semi-quantitative quantification of the CFP signal distribution, n=30 cells per strain were analyzed, double blinded, by selecting the image section that covered the higher area of nuclear signal for each cell. The NE contour was traced and the cytoplasmic background-subtracted intensity of the CFP signal along the line contour was obtained using ImageJ (Rasband, W.S., ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, http://rsb.info.nih.gov/ij/, 1997-2009). To determine the level of NPC clustering along the contour line we calculated the Coefficient of Variation (CV, standard deviation divided by the mean) of the signal intensity: the CV value decreases when the signal is evenly distributed along the NE, and increases when the signal is clustered. Values were normalized to wild type NPC distribution (CV=0). To analyze the effect of increasing membrane fluidity over NPC clustering strains were grown to early mid-log phase and then transferred to fresh media or 0.4% benzyl alcohol containing fresh media [163]. Cells were grown for 1 hour and then processed for fluorescence microcopy and NPC distribution as described above. Recovery after transfer to fresh media not including BA was tested after 4 hours.

References

1. Alberts, B., D. Bray, J. Lewis, M. Raff, K. Roberts, J.D. Watson, *Molecular Biology of the Cell (3rd Edition)*. 1989, New York and London: Garland Publishing Inc.
2. Abbott, A., *Proteomics: the society of proteins.* Nature, 2002. 417: p. 894-6.
3. Schmeing, T.M. and V. Ramakrishnan, *What recent ribosome structures have revealed about the mechanism of translation.* Nature, 2009. 461(7268): p. 1234-42.
4. Allen, G.S. and J. Frank, *Structural insights on the translation initiation complex: ghosts of a universal initiation complex.* Mol Microbiol, 2007. 63(4): p. 941-50.
5. Horwich, A.L. and W.A. Fenton, *Chaperonin-mediated protein folding: using a central cavity to kinetically assist polypeptide chain folding.* Q Rev Biophys, 2009. 42(2): p. 83-116.
6. Spiess, C., et al., *Mechanism of the eukaryotic chaperonin: protein folding in the chamber of secrets.* Trends Cell Biol, 2004. 14(11): p. 598-604.
7. Cramer, P., et al., *Structure of eukaryotic RNA polymerases.* Annual review of biophysics, 2008. 37: p. 337-52.
8. Cheng, Y., *Toward an atomic model of the 26S proteasome.* Curr Opin Struct Biol, 2009. 19(2): p. 203-8.
9. Murata, S., H. Yashiroda, and K. Tanaka, *Molecular mechanisms of proteasome assembly.* Nat Rev Mol Cell Biol, 2009. 10(2): p. 104-15.
10. Forster, F., et al., *Toward an integrated structural model of the 26S proteasome.* Mol Cell Proteomics, 2010. 9(8): p. 1666-77.
11. Lasker, K., et al., *Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach.* Proc Natl Acad Sci U S A, 2012. 109(5): p. 1380-7.
12. Mitra, K. and J. Frank, *Ribosome dynamics: insights from atomic structure modeling into cryo-electron microscopy maps.* Annu Rev Biophys Biomol Struct, 2006. 35: p. 299-317.
13. Alber, F., et al., *The molecular architecture of the nuclear pore complex.* Nature, 2007. 450(7170): p. 695-701.
14. Robinson, C., A. Sali, and W. Baumeister, *The molecular sociology of the cell.* Nature, 2007. 450(7172): p. 973-982.
15. Blundell, T.L. and L. Johnson, *Protein Crystallography*1976: Academic.
16. Bonvin, A.M.J.J., R. Boelens, and R. Kaptein, *NMR analysis of protein interactions.* Current opinion in chemical biology, 2005. 9(5): p. 501-8.
17. Fiaux, J., et al., *NMR analysis of a 900K GroEL GroES complex.* Nature, 2002. 418(6894): p. 207-11.

18. **Neudecker, P., P. Lundström, and L. Kay, *Relaxation Dispersion NMR Spectroscopy as a Tool for Detailed Studies of Protein ….* Biophysical Journal, 2009.**
19. **Stahlberg, H. and T. Walz, *Molecular electron microscopy: state of the art and current challenges.* Acs Chemical Biology, 2008.**
20. **Chiu, W., et al., *Electron cryomicroscopy of biological machines at subnanometer resolution.* Structure, 2005. 13(3): p. 363-72.**
21. **Lucic, V., A. Leis, and W. Baumeister, *Cryo-electron tomography of cells: connecting structure and function.* Histochem Cell Biol, 2008. 130(2): p. 185-96.**
22. **Berggard, T., S. Linse, and P. James, *Methods for the detection and analysis of protein-protein interactions.* Proteomics, 2007. 7(16): p. 2833-42.**
23. **Svergun, D.I., M.V. Petoukhov, and M.H. Koch, *Determination of domain structure of proteins from X-ray solution scattering.* Biophysical Journal, 2001. 80(6): p. 2946-53.**
24. **Hura, G.L., et al., *Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS).* Nat Methods, 2009. 6(8): p. 606-12.**
25. **Joo, C., et al., *Advances in single-molecule flourescence methods for molecular biology.* Annual Review of Biochemistry, 2008. 77: p. 51-76.**
26. **Hart, G.T., A.K. Ramani, and E.M. Marcotte, *How complete are current yeast and human protein-interaction networks?* Genome Biol, 2006. 7(11): p. 120.**
27. **Collins, S.R., et al., *Toward a comprehensive atlas of the physical interactome of Saccharomyces cerevisiae.* Mol Cell Proteomics, 2007. 6(3): p. 439-50.**
28. **Cusick, M.E., et al., *Literature-curated protein interaction datasets.* Nat Methods, 2009. 6(1): p. 39-46.**
29. **Topf, M., et al., *Refinement of protein structures by iterative comparative modeling and CryoEM density fitting.* J Mol Biol, 2006. 357(5): p. 1655-1668.**
30. **Topf, M., et al., *Protein structure fitting and refinement guided by cryo-EM density.* Structure, 2008. 16(2): p. 295-307.**
31. **Lasker, K., et al., *Inferential optimization for simultaneous fitting of multiple components into a cryoEM map of their assembly.* J Mol Biol, 2009. 388(1): p. 180-194.**
32. **Qian, B., et al., *High-resolution structure prediction and the crystallographic phase problem.* Nature, 2007. 450(7167): p. 259-64.**
33. **Taverner, T., et al., *Subunit architecture of intact protein complexes from mass spectrometry and homology modeling.* Acc Chem Res, 2008. 41(5): p. 617-27.**
34. **Bowers, P.M., C.E. Strauss, and D. Baker, *De novo protein structure determination using sparse NMR data.* J Biomol NMR, 2000. 18(4): p. 311-8.**

35. Raman, S., et al., *NMR structure determination for larger proteins using backbone-only data.* Science, 2010. 327(5968): p. 1014-8.
36. Alber, F., et al., *Determining the architectures of macromolecular assemblies.* Nature, 2007. 450(7170): p. 683-694.
37. Fotin, A., et al., *Molecular model for a complete clathrin lattice from electron cryomicroscopy.* Nature, 2004. 432(7017): p. 573-9.
38. Xing, Y., et al., *Structure of clathrin coat with bound Hsc70 and auxilin: mechanism of Hsc70-facilitated disassembly.* EMBO J. 29(3): p. 655-65.
39. Chen, Z.A., et al., *Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry.* EMBO J, 2010. 29(4): p. 717-26.
40. Byeon, I.-J.L., et al., *Structural convergence between Cryo-EM and NMR reveals intersubunit interactions critical for HIV-1 capsid function.* Cell, 2009. 139(4): p. 780-90.
41. Billinge, S.J. and I. Levin, *The problem with determining atomic structure at the nanoscale.* Science, 2007. 316(5824): p. 561-5.
42. Bau, D., et al., *The three-dimensional folding of the alpha-globin gene domain reveals formation of chromatin globules.* Nat Struct Mol Biol, 2011. 18(1): p. 107-14.
43. Tjong, H., et al., *Physical tethering and volume exclusion determine higher-order genome organization in budding yeast.* Genome Res, 2012. 22(7): p. 1295-305.
44. Rout, et al., *The yeast nuclear pore complex: composition, architecture, and transport mechanism*, in *J Cell Biol*2000. p. 635-51.
45. DeGrasse, J.A., et al., *Evidence for a shared nuclear pore complex architecture that is conserved from the last common eukaryotic ancestor.* Mol Cell Proteomics, 2009. 8(9): p. 2119-30.
46. Yang, Q., M.P. Rout, and C.W. Akey, *Three-dimensional architecture of the isolated yeast nuclear pore complex: functional and evolutionary implications.* Mol Cell, 1998. 1(2): p. 223-34.
47. Cronshaw, J.M., et al., *Proteomic analysis of the mammalian nuclear pore complex.* J Cell Biol, 2002. 158(5): p. 915-27.
48. D'Angelo, M.A. and M.W. Hetzer, *Structure, dynamics and function of nuclear pore complexes.* Trends Cell Biol, 2008. 18(10): p. 456-66.
49. Strambio-De-Castillia, C., M. Niepel, and M.P. Rout, *The nuclear pore complex: bridging nuclear transport and gene regulation.* Nat Rev Mol Cell Biol, 2010. 11(7): p. 490-501.
50. Fernandez-Martinez, J. and M.P. Rout, *Nuclear pore complex biogenesis.* Curr Opin Cell Biol, 2009. 21(4): p. 603-12.
51. Terry, L.J., E.B. Shows, and S.R. Wente, *Crossing the nuclear envelope: hierarchical regulation of nucleocytoplasmic transport.* Science, 2007. 318(5855): p. 1412-6.
52. Kahms, M., et al., *Lighting up the nuclear pore complex.* Eur J Cell Biol, 2011. 90(9): p. 751-8.

53. **Weis, K., *The nuclear pore complex: oily spaghetti or gummy bear?* Cell, 2007. 130(3): p. 405-7.**
54. **Wente, S.R. and M.P. Rout, *The nuclear pore complex and nuclear transport.* Cold Spring Harb Perspect Biol, 2010. 2(10): p. a000562.**
55. **Fiserova, J. and M.W. Goldberg, *Nucleocytoplasmic transport in yeast: a few roles for many actors.* Biochem Soc Trans, 2010. 38(Pt 1): p. 273-7.**
56. **Rout, M.P., et al., *Virtual gating and nuclear transport: the hole picture.* Trends Cell Biol, 2003. 13(12): p. 622-8.**
57. **Ribbeck, K. and D. Gorlich, *Kinetic analysis of translocation through nuclear pore complexes.* EMBO J, 2001. 20(6): p. 1320-30.**
58. **Frey, S. and D. Gorlich, *A saturated FG-repeat hydrogel can reproduce the permeability properties of nuclear pore complexes.* Cell, 2007. 130(3): p. 512-23.**
59. **Devos, D., et al., *Components of coated vesicles and nuclear pore complexes share a common molecular architecture.* PLoS Biol, 2004. 2(12): p. e380.**
60. **Devos, D., et al., *Simple fold composition and modular architecture of the nuclear pore complex.* Proc Natl Acad Sci U S A, 2006. 103(7): p. 2172-7.**
61. **Dokudovskaya, S., et al., *Protease accessibility laddering: a proteomic tool for probing protein structure.* Structure, 2006. 14(4): p. 653-60.**
62. **Field, M.C. and J.B. Dacks, *First and last ancestors: reconstructing evolution of the endomembrane system with ESCRTs, vesicle coat proteins, and nuclear pore complexes.* Curr Opin Cell Biol, 2009. 21(1): p. 4-13.**
63. **Stagg, S.M., et al., *Structural basis for cargo regulation of COPII coat assembly.* Cell, 2008. 134(3): p. 474-84.**
64. **Kampmann, M. and G. Blobel, *Three-dimensional structure and flexibility of a membrane-coating module of the nuclear pore complex.* Nat Struct Mol Biol, 2009. 16(7): p. 782-8.**
65. **Lee, C. and J. Goldberg, *Structure of coatomer cage proteins and the relationship among COPI, COPII, and clathrin vesicle coats.* Cell, 2010. 142(1): p. 123-32.**
66. **Debler, E.W., et al., *A fence-like coat for the nuclear pore membrane*, in *Mol Cell*2008. p. 815-26.**
67. **Hsia, K.C., et al., *Architecture of a coat for the nuclear pore membrane.* Cell, 2007. 131(7): p. 1313-26.**
68. **Brohawn, S.G., et al., *Structural evidence for common ancestry of the nuclear pore complex and vesicle coats.* Science, 2008. 322(5906): p. 1369-73.**
69. **Brohawn, S.G. and T.U. Schwartz, *Molecular architecture of the Nup84-Nup145C-Sec13 edge element in the nuclear pore complex lattice.* Nat Struct Mol Biol, 2009.**

70. Leksa, N.C., S.G. Brohawn, and T.U. Schwartz, *The structure of the scaffold nucleoporin Nup120 reveals a new and unexpected domain architecture.* Structure, 2009. 17(8): p. 1082-91.
71. Seo, H.S., et al., *Structural and functional analysis of Nup120 suggests ring formation of the Nup84 complex.* Proc Natl Acad Sci U S A, 2009. 106(34): p. 14281-6.
72. Whittle, J.R. and T.U. Schwartz, *Architectural nucleoporins Nup157/170 and Nup133 are structurally related and descend from a second ancestral element.* J Biol Chem, 2009.
73. Nagy, V., et al., *Structure of a trimeric nucleoporin complex reveals alternate oligomerization states.* Proc Natl Acad Sci U S A, 2009. 106(42): p. 17693-8.
74. Sampathkumar, P., et al., *Structure of the C-terminal domain of Saccharomyces cerevisiae Nup133, a component of the nuclear pore complex.* Proteins, 2011. 79(5): p. 1672-7.
75. Cook, A., et al., *Structural biology of nucleocytoplasmic transport.* Annu Rev Biochem, 2007. 76: p. 647-71.
76. Dacks, J.B. and M.C. Field, *Evolution of the eukaryotic membrane-trafficking system: origin, tempo and mode.* J Cell Sci, 2007. 120(Pt 17): p. 2977-85.
77. Lam, D.H. and P.D. Aplan, *NUP98 gene fusions in hematologic malignancies.* Leukemia, 2001. 15(11): p. 1689-95.
78. von Lindern, M., et al., *The translocation (6;9), associated with a specific subtype of acute myeloid leukemia, results in the fusion of two genes, dek and can, and the expression of a chimeric, leukemia-specific dek-can mRNA.* Mol Cell Biol, 1992. 12(4): p. 1687-97.
79. Martinez, N., et al., *The nuclear pore complex protein Nup88 is overexpressed in tumor cells.* Cancer Res, 1999. 59(21): p. 5408-11.
80. Konig, R., et al., *Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication.* Cell, 2008. 135(1): p. 49-60.
81. Jager, S., et al., *Purification and characterization of HIV-human protein complexes.* Methods, 2011. 53(1): p. 13-9.
82. D'Angelo, M.A., et al., *Age-dependent deterioration of nuclear pore complexes causes a loss of nuclear integrity in postmitotic cells.* Cell, 2009. 136(2): p. 284-95.
83. Savas, J.N., et al., *Extremely long-lived nuclear pore proteins in the rat brain.* Science, 2012. 335(6071): p. 942.
84. Watson, M.L., *Further observations on the nuclear envelope of the animal cell.* J Biophys Biochem Cytol, 1959. 6: p. 147-56.
85. Goldberg, M.W. and T.D. Allen, *The nuclear pore complex: three-dimensional surface structure revealed by field emission, in-lens scanning electron microscopy, with underlying structure uncovered by proteolysis.* J Cell Sci, 1993. 106 (Pt 1): p. 261-74.
86. Kiseleva, E., et al., *Steps of nuclear pore complex disassembly and reassembly during mitosis in early Drosophila embryos.* J Cell Sci, 2001. 114(Pt 20): p. 3607-18.

87. Kiseleva, E., et al., *Yeast nuclear pore complexes have a cytoplasmic ring and internal filaments.* J Struct Biol, 2004. 145(3): p. 272-88.
88. Akey, C.W. and M. Radermacher, *Architecture of the Xenopus nuclear pore complex revealed by three-dimensional cryo-electron microscopy.* J Cell Biol, 1993. 122(1): p. 1-19.
89. Stoffler, D., et al., *Cryo-electron tomography provides novel insights into nuclear pore architecture: implications for nucleocytoplasmic transport.* J Mol Biol, 2003. 328(1): p. 119-30.
90. Beck, M., et al., *Snapshots of nuclear pore complexes in action captured by cryo-electron tomography.* Nature, 2007. 449(7162): p. 611-5.
91. Maimon, T., et al., *The human nuclear pore complex as revealed by cryo-electron tomography.* Structure, 2012. 20(6): p. 998-1006.
92. Fahrenkrog, B., et al., *Comparative spatial localization of protein-A-tagged and authentic yeast nuclear pore complex proteins by immunogold electron microscopy.* J Struct Biol, 2000. 129(2-3): p. 295-305.
93. Brohawn, S.G., et al., *The nuclear pore complex has entered the atomic age.* Structure, 2009. 17(9): p. 1156-68.
94. Sali, A. and T.L. Blundell, *Comparative protein modelling by satisfaction of spatial restraints.* J Mol Biol, 1993. 234(3): p. 779-815.
95. Russel, D., et al., *Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies.* PLoS Biol, 2012. 10(1): p. e1001244.
96. Gavin, A.C., et al., *Functional organization of the yeast proteome by systematic analysis of protein complexes.* Nature, 2002. 415(6868): p. 141-7.
97. Krogan, N.J., et al., *Global landscape of protein complexes in the yeast Saccharomyces cerevisiae.* Nature, 2006. 440(7084): p. 637-43.
98. Gavin, A.-C., et al., *Proteome survey reveals modularity of the yeast cell machinery.* Nature, 2006. 440(7084): p. 631-6.
99. Li, X., et al., *Extensive in vivo metabolite-protein interactions revealed by large-scale systematic analyses.* Cell, 2010. 143(4): p. 639-50.
100. Ghaemmaghami, S., et al., *Global analysis of protein expression in yeast.* Nature, 2003. 425: p. 671-672.
101. Aebersold, R. and M. Mann, *Mass spectrometry-based proteomics.* Nature, 2003. 422(6928): p. 198-207.
102. Puig, O., et al., *The tandem affinity purification (TAP) method: a general procedure of protein complex purification.* Methods, 2001. 24(3): p. 218-29.
103. Alber, F., et al., *Integrating diverse data for structure determination of macromolecular assemblies.* Annu Rev Biochem, 2008. 77: p. 443-477.

104. Jasiak, A.J., et al., *Genome-associated RNA polymerase II includes the dissociable Rpb4/7 subcomplex.* J Biol Chem, 2008. 283(39): p. 26423-7.

105. Hahn, S., *Structure and mechanism of the RNA polymerase II transcription machinery.* Nat Struct Mol Biol, 2004. 11(5): p. 394-403.

106. Cramer, P., et al., *Architecture of RNA polymerase II and implications for the transcription mechanism.* Science, 2000. 288(5466): p. 640-9.

107. Kostek, S.A., et al., *Molecular architecture and conformational flexibility of human RNA polymerase II.* Structure, 2006. 14(11): p. 1691-700.

108. Cramer, P., D.A. Bushnell, and R.D. Kornberg, *Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution.* Science, 2001. 292(5523): p. 1863-76.

109. Pieper, U., et al., *MODBASE, a database of annotated comparative protein structure models and associated resources.* Nucleic Acids Res, 2009. 37(Database issue): p. D347-354.

110. Henrick, K., et al., *EMDep: a web-based system for the deposition and validation of high-resolution electron microscopy macromolecular structural information.* J Struct Biol, 2003. 144(1-2): p. 228-37.

111. Stark, C., et al., *BioGRID: a general repository for interaction datasets.* Nucleic Acids Res, 2006. 34(Database issue): p. D535-9.

112. Flores, A., et al., *A protein-protein interaction map of yeast RNA polymerase III.* Proc Natl Acad Sci U S A, 1999. 96(14): p. 7815-20.

113. Zaros, C., et al., *Functional organization of the Rpb5 subunit shared by the three yeast RNA polymerases.* Nucleic Acids Res, 2007. 35(2): p. 634-47.

114. Briand, J.F., et al., *Partners of Rpb8p, a small subunit shared by yeast RNA polymerases I, II and III.* Mol Cell Biol, 2001. 21(17): p. 6056-65.

115. Tan, Q., M.H. Prysak, and N.A. Woychik, *Loss of the Rpb4/Rpb7 subcomplex in a mutant form of the Rpb6 subunit shared by RNA polymerases I, II, and III.* Mol Cell Biol, 2003. 23(9): p. 3329-38.

116. Qi, H. and V.A. Zakian, *The Saccharomyces telomere-binding protein Cdc13p interacts with both the catalytic subunit of DNA polymerase alpha and the telomerase-associated est1 protein.* Genes Dev, 2000. 14(14): p. 1777-88.

117. Sampath, V., et al., *The conserved and non-conserved regions of Rpb4 are involved in multiple phenotypes in Saccharomyces cerevisiae.* J Biol Chem, 2003. 278(51): p. 51566-76.

118. Khazak, V., et al., *Human RNA polymerase II subunit hsRPB7 functions in yeast and influences stress survival and cell morphology.* Mol Biol Cell, 1995. 6(7): p. 759-75.

119. Sareen, A., et al., *Mapping the interaction site of Rpb4 and Rpb7 subunits of RNA polymerase II in Saccharomyces cerevisiae.* Biochem Biophys Res Commun, 2005. 332(3): p. 763-70.

120. Selitrennik, M., et al., *Nucleocytoplasmic shuttling of the Rpb4p and Rpb7p subunits of Saccharomyces cerevisiae RNA polymerase II by two pathways.* Eukaryot Cell, 2006. 5(12): p. 2092-103.
121. Tarassov, K., et al., *An in vivo map of the yeast protein interactome.* Science, 2008. 320(5882): p. 1465-70.
122. Benga, W.J., et al., *Distinct regions of RPB11 are required for heterodimerization with RPB3 in human and yeast RNA polymerase II.* Nucleic Acids Res, 2005. 33(11): p. 3582-90.
123. Orlicky, S.M., et al., *Dissociable Rpb4-Rpb7 subassembly of rna polymerase II binds to single-strand nucleic acid and mediates a post-recruitment step in transcription initiation.* J Biol Chem, 2001. 276(13): p. 10097-102.
124. Frank, J., *Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State* 2006: oxford university press.
125. Shen, M., F. Davis, and A. Sali, *The optimal size of a globular protein domain: A simple sphere-packing model.* Chemical Physics Letters, 2005. 405(1-3): p. 224-228.
126. Connolly, M.L., *Solvent-accessible surfaces of proteins and nucleic acids.* Science, 1983. 221(4612): p. 709-13.
127. Katchalski-Katzir, E., et al., *Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques.* Proc Natl Acad Sci U S A, 1992. 89(6): p. 2195-9.
128. Schneidman-Duhovny, D., et al., *Taking geometry to its edge: fast unbound rigid (and hinge-bent) docking.* Proteins, 2003. 52(1): p. 107-12.
129. Brunger, A.T., *Assessment of phase accuracy by cross validation: the free R value. Methods and applications.* Acta Crystallogr D Biol Crystallogr, 1993. 49(Pt 1): p. 24-36.
130. Belgareh, N., et al., *An evolutionarily conserved NPC subcomplex, which redistributes in part to kinetochores in mammalian cells.* J Cell Biol, 2001. 154(6): p. 1147-60.
131. Krull, S., et al., *Nucleoporins as components of the nuclear pore complex core structure and Tpr as the architectural element of the nuclear basket.* Mol Biol Cell, 2004. 15(9): p. 4261-77.
132. Loiodice, I., et al., *The entire Nup107-160 complex, including three new members, is targeted as one entity to kinetochores in mitosis.* Mol Biol Cell, 2004. 15(7): p. 3333-44.
133. Vasu, S., et al., *Novel vertebrate nucleoporins Nup133 and Nup160 play a role in mRNA export.* J Cell Biol, 2001. 155(3): p. 339-54.
134. Lutzmann, M., et al., *Modular self-assembly of a Y-shaped multiprotein complex from seven nucleoporins.* EMBO J, 2002. 21(3): p. 387-97.

135. Siniossoglou, S., et al., *A novel complex of nucleoporins, which includes Sec13p and a Sec13p homolog, is essential for normal nuclear pores.* Cell, 1996. 84(2): p. 265-75.
136. Salama, N.R., J.S. Chuang, and R.W. Schekman, *Sec31 encodes an essential component of the COPII coat required for transport vesicle budding from the endoplasmic reticulum.* Mol Biol Cell, 1997. 8(2): p. 205-17.
137. Dokudovskaya, S., et al., *A conserved coatomer-related complex containing Sec13 and Seh1 dynamically associates with the vacuole in Saccharomyces cerevisiae.* Mol Cell Proteomics, 2011. 10(6): p. M110 006478.
138. Doye, V. and E.C. Hurt, *Genetic approaches to nuclear pore structure and function.* Trends Genet, 1995. 11(6): p. 235-41.
139. Fabre, E. and E. Hurt, *Yeast genetics to dissect the nuclear pore complex and nucleocytoplasmic trafficking.* Annu Rev Genet, 1997. 31: p. 277-313.
140. Aitchison, J.D., G. Blobel, and M.P. Rout, *Nup120p: a yeast nucleoporin required for NPC distribution and mRNA transport.* J Cell Biol, 1995. 131(6 Pt 2): p. 1659-75.
141. Doye, V., R. Wepf, and E.C. Hurt, *A novel nuclear pore protein Nup133p with distinct roles in poly(A)+ RNA transport and nuclear pore distribution.* EMBO J, 1994. 13(24): p. 6062-75.
142. Heath, C.V., et al., *Nuclear pore complex clustering and nuclear accumulation of poly(A)+ RNA associated with mutation of the Saccharomyces cerevisiae RAT2/NUP120 gene.* J Cell Biol, 1995. 131(6 Pt 2): p. 1677-97.
143. Li, O., et al., *Mutation or deletion of the Saccharomyces cerevisiae RAT3/NUP133 gene causes temperature-dependent nuclear accumulation of poly(A)+ RNA and constitutive clustering of nuclear pore complexes.* Mol Biol Cell, 1995. 6(4): p. 401-417.
144. Pemberton, L.F., M.P. Rout, and G. Blobel, *Disruption of the nucleoporin gene NUP133 results in clustering of nuclear pore complexes.* Proc Natl Acad Sci U S A, 1995. 92(4): p. 1187-91.
145. Siniossoglou, S., et al., *Structure and assembly of the Nup84p complex.* J Cell Biol, 2000. 149(1): p. 41-54.
146. Boehmer, T., et al., *Structural and functional studies of Nup107/Nup133 interaction and its implications for the architecture of the nuclear pore complex.* Mol Cell, 2008. 30(6): p. 721-31.
147. Das, R. and D. Baker, *Macromolecular modeling with rosetta.* Annual review of biochemistry, 2008. 77: p. 363-82.
148. Lasker, K., et al., *Integrative structure modeling of macromolecular assemblies from proteomics data.* Mol Cell Proteomics, 2010. 9(8): p. 1689-702.
149. Zhou, M., et al., *Mass spectrometry reveals modularity and a complete subunit interaction map of the eukaryotic translation factor*

eIF3. Proceedings of the National Academy of Sciences of the United States of America, 2008. 105(47): p. 18139-18144.

150. Wente, S.R., M.P. Rout, and G. Blobel, *A new family of yeast nuclear pore complex proteins.* J Cell Biol, 1992. 119(4): p. 705-23.
151. Alber, F., M. Kim, and A. Sali, *Structural characterization of assemblies from overall shape and subcomplex compositions.* Structure, 2005. 13(3): p. 435-445.
152. Nelder, J.A. and R. Mead, *A Simplex Method for Function Minimization.* The Computer Journal, 1965. 7(4): p. 308-313.
153. Madhusudhan, M.S., et al., *Alignment of multiple protein structures based on sequence and structure features.* Protein Eng Des Sel, 2009. 22: p. 569-574.
154. Lutzmann, M., et al., *Reconstitution of Nup157 and Nup145N into the Nup84 complex.* J Biol Chem, 2005. 280: p. 18442-18451.
155. Boehmer, T., et al., *Depletion of a single nucleoporin, Nup107, prevents the assembly of a subset of nucleoporins into the nuclear pore complex.* Proc Natl Acad Sci U S A, 2003. 100(3): p. 981-5.
156. Kampmann, M., et al., *Mapping the orientation of nuclear pore proteins in living cells with polarized fluorescence microscopy.* Nat Struct Mol Biol, 2011. 18(6): p. 643-9.
157. Dultz, E., et al., *Systematic kinetic analysis of mitotic dis- and reassembly of the nuclear pore in living cells.* J Cell Biol, 2008. 180(5): p. 857-65.
158. Amlacher, S., et al., *Insight into structure and assembly of the nuclear pore complex by utilizing the genome of a eukaryotic thermophile.* Cell, 2011. 146(2): p. 277-89.
159. Belgareh, N. and V. Doye, *Dynamics of nuclear pore distribution in nucleoporin mutant yeast cells.* J Cell Biol, 1997. 136(4): p. 747-59.
160. Beck, M., et al., *Nuclear pore complex structure and dynamics revealed by cryoelectron tomography.* Science, 2004. 306: p. 1387-1390.
161. Bucci, M. and S.R. Wente, *In vivo dynamics of nuclear pore complexes in yeast.* J Cell Biol, 1997. 136(6): p. 1185-99.
162. Gordon, L.M., et al., *The increase in bilayer fluidity of rat liver plasma membranes achieved by the local anesthetic benzyl alcohol affects the activity of intrinsic membrane enzymes.* J Biol Chem, 1980. 255(10): p. 4519-27.
163. Scarcelli, J.J., C.A. Hodge, and C.N. Cole, *The yeast integral membrane protein Apq12 potentially links membrane dynamics to assembly of nuclear pore complexes.* J Cell Biol, 2007. 178(5): p. 799-812.
164. Liu, H.L., et al., *The three fungal transmembrane nuclear pore complex proteins of Aspergillus nidulans are dispensable in the presence of an intact An-Nup84-120 complex.* Mol Biol Cell, 2009. 20(2): p. 616-30.

165. Titus, L.C., et al., *Members of the RSC chromatin-remodeling complex are required for maintaining proper nuclear envelope structure and pore complex localization.* Mol Biol Cell, 2010. 21(6): p. 1072-87.
166. McMahon, H.T. and I.G. Mills, *COP and clathrin-coated vesicle budding: different pathways, common approaches.* Curr Opin Cell Biol, 2004. 16(4): p. 379-91.
167. Miele, A.E., et al., *Two distinct interaction motifs in amphiphysin bind two independent sites on the clathrin terminal domain beta-propeller.* Nat Struct Mol Biol, 2004. 11(3): p. 242-8.
168. ter Haar, E., S.C. Harrison, and T. Kirchhausen, *Peptide-in-groove interactions link target proteins to the beta-propeller of clathrin.* Proc Natl Acad Sci U S A, 2000. 97(3): p. 1096-100.
169. Fath, S., et al., *Structure and organization of coat proteins in the COPII cage.* Cell, 2007. 129(7): p. 1325-36.
170. Field, M.C., A. Sali, and M.P. Rout, *Evolution: On a bender--BARs, ESCRTs, COPs, and finally getting your coat.* J Cell Biol, 2011. 193(6): p. 963-72.
171. Gueldener, U., et al., *A second set of loxP marker cassettes for Cre-mediated multiple gene knockouts in budding yeast.* Nucleic Acids Res, 2002. 30(6): p. e23.
172. Strawn, L.A., et al., *Minimal nuclear pore complexes define FG repeat domains essential for transport.* Nat Cell Biol, 2004. 6(3): p. 197-206.
173. Teixeira, M.T., E. Fabre, and B. Dujon, *Self-catalyzed cleavage of the yeast nucleoporin Nup145p precursor.* J Biol Chem, 1999. 274(45): p. 32439-44.
174. Longtine, M.S., et al., *Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae.* Yeast, 1998. 14(10): p. 953-61.
175. Oeffinger, M., et al., *Comprehensive analysis of diverse ribonucleoprotein complexes.* Nat Methods, 2007. 4(11): p. 951-6.
176. Ludtke, S.J., P.R. Baldwin, and W. Chiu, *EMAN: semiautomated software for high-resolution single-particle reconstructions.* J Struct Biol, 1999. 128(1): p. 82-97.
177. Shen, M.Y. and A. Sali, *Statistical potential for assessment and prediction of protein structures.* Protein Sci, 2006. 15(11): p. 2507-2524.
178. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis.* J Comput Chem, 2004. 25(13): p. 1605-12.
179. Baker, D. and A. Sali, *Protein structure prediction and structural genomics.* Science, 2001. 294(5540): p. 93-6.
180. Pieper, U., et al., *ModBase, a database of annotated comparative protein structure models, and associated resources.* Nucleic Acids Res, 2011. 39(Database issue): p. D465-74.

181. Kelley, L.A. and M.J. Sternberg, *Protein structure prediction on the Web: a case study using the Phyre server.* Nat Protoc, 2009. 4(3): p. 363-71.
182. Ward, J.J., et al., *The DISOPRED server for the prediction of protein disorder.* Bioinformatics, 2004. 20(13): p. 2138-9.
183. Tackett, A.J., et al., *Proteomic and genomic characterization of chromatin complexes at a boundary.* J Cell Biol, 2005. 169(1): p. 35-47.
184. Niepel, M., et al., *The nuclear pore complex-associated protein, Mlp2p, binds to the yeast spindle pole body and promotes its efficient assembly.* J Cell Biol, 2005. 170(2): p. 225-35.
185. Timney, B.L., et al., *Simple kinetic relationships and nonspecific competition govern nuclear import rates in vivo.* J Cell Biol, 2006. 175(4): p. 579-93.

Publishing Agreement

Author Signature

9/10/12
Date