MCP Papers in Press. Published on August 19, 2014 as Manuscript M114.040915

Integrative Structure-Function Mapping of the Nucleoporin Nup133 Suggests a Conserved Mechanism for Membrane Anchoring of the Nuclear Pore Complex

Seung Joong Kim[1]*, Javier Fernandez-Martinez[2]*, Parthasarathy Sampathkumar[3]*, Anne Martel[4], Tsutomu Matsui[4], Hiro Tsuruta[4], Thomas Weiss[4], Ane Markina-Inarrairaegui[5], Jeffery B. Bonanno[3], J. Michael Sauder[6], Stephen K. Burley[7], Steven C. Almo[3], Michael Rout[2], and Andrej Sali[1]

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

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

3. Department of Biochemistry, Ullmann Building, Room 409, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461, USA

4. Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, MS 69, Menlo Park, CA 94025, USA

5. Laboratorio de Genetica Molecular de Aspergillus, Departamento de Biología Celular y Molecular, Centro de Investigaciones Biológicas (CSIC). Ramiro de Maeztu 9, 28040, Madrid, Spain.

6. Discovery Chemistry Research and Technologies (DCR&T), Eli Lilly and Company, Lilly Biotechnology Center, 10300 Campus Point Drive, Suite 200, San Diego, CA

92121, USA

7. Center for Integrative Proteomics Research, Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, 174 Frelinghuysen Road, Piscataway, NJ 08854, USA

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

Corresponding authors

Steven C. Almo, Department of Biochemistry, Ullmann Building, Room 409, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461, USA.

tel: +1 718 4302746; fax: +1 718 430 8565; steve.almo@einstein.yu.edu

Michael. P. Rout, Box 213, Laboratory of Cellular and Structural Biology, Rockefeller University, 1230 York Avenue, New York, NY10021, USA.

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

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

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

Running Title: Integrative Structure-Function Mapping of Nup133 in the NPC

Abbreviations

SAXS: small angle X-ray scattering

EM: electron microscopy

NPC: nuclear pore complex

Nup: nucleoporin

*Sc*Nup133 : *Saccharomyces cerevisiae* Nup133

*Vp*Nup133 :*Vanderwaltozyma polyspora* Nup133

*Hs*Nup*133* : *Homo sapiens* Nup133

ALPS : ArfGAP1 lipid packing sensor

GFP: green fluorescent protein

MES : Minimal Ensemble Search

NE: nuclear envelope

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

DSS: disuccinimidyl suberate

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

Summary

The nuclear pore complex (NPC) is the sole passageway for the transport of macromolecules across the nuclear envelope. Nup133, a major component in the Y-shaped Nup84 complex, is an essential scaffold protein of the NPC's outer ring structure. Here, we describe an integrative modeling approach that produces atomic models for multiple states of *Saccharomyces cerevisiae (Sc)* Nup133, based on the crystal structures of the sequence segments and their homologs including the related *Vanderwaltozyma polyspora (Vp)* Nup133 residues 55 to 502 (*Vp*Nup133$^{55\text{-}502}$) determined in this study, small angle X-ray scattering (SAXS) profiles for 18 constructs of *Sc*Nup133 and one construct of *Vp*Nup133, and 23 negative-stain electron microscopy (EM) class averages of *Sc*Nup133$^{2\text{-}1157}$. Using our integrative approach, we then computed a multi-state structural model of the full-length *Sc*Nup133, followed by validation with mutational studies and 45 chemical cross-links determined by mass spectrometry. Finally, the model of *Sc*Nup133 allowed us to annotate a potential ArfGAP1 lipid packing sensor (ALPS) motif in *Sc* and *Vp*Nup133 and discuss its potential significance in the context of the whole NPC; we suggest that ALPS motifs are scattered throughout the NPC's scaffold of all eukaryotes and play a major role in the assembly and membrane anchoring of the NPC in the nuclear envelope. Our results are consistent with a common evolutionary origin of Nup133 with membrane coating complexes (the protocoatomer hypothesis); the presence of the ALPS motifs in coatomer-like nucleoporins suggests an ancestral mechanism for membrane recognition present in early membrane coating complexes.

Introduction

The *Saccharomyces cerevisiae* nuclear pore complex (NPC) is a large macromolecular assembly of ~50 MDa, made of at least 456 protein copies of ~30 distinct proteins called nucleoporins (nups). The NPC is the sole passageway for the exchange of macromolecules across the nuclear envelope (NE) (1). Apart from its main function as the sole mediator of nucleocytoplasmic trafficking, the NPC plays additional roles in numerous essential cellular processes, such as gene expression and chromatin regulation (2), and defects in its components have been implicated in numerous major human diseases (3). The first description of the macromolecular architecture of the NPC (4) was determined by an integrative approach based on a wide variety of experimental data (5). The permeability barrier is formed by FG (phenylalanine–glycine repeat containing) nups, which fill the central channel of the NPC and are anchored to the core scaffold (6). The NPC architectural core is formed by an eight-fold arrangement of symmetric units called spokes that connect to each other, forming coaxial rings: two outer rings (the nuclear and cytoplasmic rings), a membrane ring, and two inner rings (7). In *Saccharomyces cerevisiae*, the membrane ring is mainly formed by the transmembrane nups Pom152, Pom34 and Ndc1; the two adjacent inner rings are formed by large nups Nup192, Nup188, Nup170 and Nup157; and the two outer rings are formed by a radial head-to-tail arrangement of 8 copies of the Nup84 complex (4, 8, 9). The Nup84 complex is a conserved assembly, formed by nine proteins in vertebrates (Nup107-160 complex) and by seven nups in yeast (Nup133, Nup120, Nup145c, Nup85, Nup84, Seh1 and Sec13). The yeast Nup84 complex arranges into a characteristic Y-shaped assembly (10, 11). The stalk of the Y is formed by a tail-to-tail

connection between Nup133 and Nup84 and a head to center connection between Nup84 and the dimer Nup145c-Sec13 (8, 12, 13).

Nup133, a 133 kDa subunit of the Nup84 complex, consists of a N-terminal β-propeller and C-terminal α-solenoid–like folds (14). Nup133 is located at the end of the stalk of the Nup84 complex, through a connection with Nup84 (8). Nup133 has also been suggested to connect through the first 15 residues of its N-terminal domain to the Nup120 copy of an adjacent Nup84 complex heptamer (12). Nup133 is a highly conserved nup that plays key roles in interphase and post-mitotic NPC biogenesis (15, 16) as well as in efficient anchoring of the dynein/dynactin complex to tether centrosomes to the NE in prophase (17). A loop within the N-terminal β-propeller of human Nup133 was suggested to contain an ArfGAP1 lipid packing sensor (ALPS) motif (18), functioning as a membrane curvature sensor. This motif allows human Nup133 to interact with curved membranes both *in vitro* and *in vivo* (15, 18) and has been shown to be required for proper NPC biogenesis during interphase (15). However, previous studies have not been able to detect any membrane interaction motifs in yeast Nup133, leading to the suggestion that the ALPS motif in Nup133 was unique to organisms with open mitosis (18, 19), in turn implying that the ALPS motif is not even a part of the mechanism for membrane association of the NPCs in all eukaryotes. Interestingly, mutations in *Sc*Nup133 cause a characteristic phenotype that leads to clustering of the NPCs into discrete regions of the NE (20). Structure-function mapping of this NPC clustering phenotype suggests that *Sc*Nup133 - and its ancient paralog *Sc*Nup120 - are functionally involved in the stabilization of the NE membrane curvature (8), although the exact mechanism that drives the interaction of these proteins with the

NE is unknown.

Multi-domain, full-length nucleoporins are generally not amenable to X-ray crystallographic structure determination, presumably due to their apparent flexibility. Indeed, the structures of the N- and C- terminal fragments of Nup133 in particular were determined only separately (19, 21-23); the full-length atomic structure has not yet been characterized. Consequently, the relative orientation of the N- and C- terminal domains was depicted only schematically (22). We therefore took an integrative approach to generate the structure and dynamics of full-length *Sc*Nup133, based on the multiple types of data.

Here, we characterize the configuration of the individual domains, defining the shape and populations of the full-length *Sc*Nup133 conformations, based on template structures, X-ray crystallography, SAXS, and EM data, followed by validation with mutational studies and a chemical cross-linking with mass spectrometric readout (CX-MS) dataset. More specifically, we report the crystal structure of the Nup133 N-terminal domain (residues of 55-502) from *Vanderwaltozyma polyspora*, as well as SAXS profiles for 18 constructs of *Sc*Nup133 and one *Vp*Nup133 construct, and 23 negative-stain EM class averages of *Sc*Nup133$^{2\text{-}1157}$. Using our integrative modeling approach described in this study, we then determine atomic models for multiple states of the full-length *Sc*Nup133, based on these new data as well as known structures of *Sc*Nup133$^{944\text{-}1157}$ and a number of Nup133 homologs. The resulting model is subsequently validated by 3 sets of double point mutations at the *Sc*Nup133-*Sc*Nup84 interface and 20 DSS and 25 EDC chemical cross-links determined by mass spectrometry (24).

As a result, the model of the full-length *Sc*Nup133 allows us to annotate a potential ArfGAP1 lipid packing sensor (ALPS) motif in *Sc* and *Vp*Nup133, suggesting that ALPS motifs are scattered throughout the NPC's scaffold of all eukaryotes and play a major role in the assembly and membrane anchoring of the NPC in the NE. Our results are consistent with a common evolutionary origin of Nup133 with membrane coating complexes (the protocoatomer hypothesis); the presence of the ALPS motifs in coatomer-like nucleoporins suggests an ancestral mechanism for membrane recognition present in early membrane coating complexes.

Experimental Procedures

Construct design, cloning, expression and purification of *Sc* and *Vp*Nup133

Nup133 is divided into the N-terminal β-propeller and the C-terminal α-solenoid domains, in an iterative manual process relying on predicted secondary structure, gaps in multiple sequence alignments, and sequence–structure alignment by threading (14)

We cloned, expressed, and purified the resulting 18 constructs of *Sc*Nup133; 7 constructs covering the N-terminal domain, 8 constructs covering the C-terminal domain, and 3 remaining constructs covering the both domains partially or entirely (Supplemental Figure S3 and Supplemental Table S1). Cloning, expression, and purification were performed using a standard protocol as described previously (21) (Supplemental Experimental Procedures). The N-terminal domain of *Vanderwaltozyma polyspora* Nup133 covering residues 55 to 502 (*Vp*Nup133$^{55\text{-}502}$) was also cloned, expressed, and purified following similar procedures (21).

Crystallization and structure determination of *Vp*Nup133$^{55\text{-}502}$

The crystal used for structure determination *via* Se-SAD phasing was obtained by sitting-drop vapour diffusion (*Vp*Nup133$^{55\text{-}502}$ concentration of 10.6 mg/mL) in the presence of 10% PEG3350, 100 mM ammonium sulfate, and 100 mM HEPES (pH 8.2), and flash frozen in liquid nitrogen with 30% (v/v) glycerol. Diffraction data set collected at the LRL-CAT 31-ID (Advanced Photon Source) beamline was processed with XDS (25) and AIMLESS (26), and structure solution was obtained using AutoSol (27) in Phenix (28). An initial model assembled using AutoBuild (29) was further extended with several cycles of density modification using Parrot (30) followed by automated model

improvement with Buccaneer (31, 32), as implemented in CCP4 (33), and manual model building with COOT (34). Final stages of refinement were performed using Refmac5 (35). Illustrations were prepared using PyMol (36) and UCSF Chimera (37).

Small angle X-ray scattering (SAXS)

SAXS profiles of 18 constructs of *Sc*Nup133 and one construct of *Vp*Nup133$^{55-502}$ (Figures 2B and 3B, Supplemental Figures S3 and S4, and Supplemental Table S1) were measured at concentrations of 0.5, 1.0, 2.0 mg/mL and highest possible concentration in the protein storage buffer at 10-15^{o}C, using up to 24 1-10 second exposures at SSRL (Menlo Park, CA) and ALS (Berkeley, CA) beamlines (Supplemental Experimental Procedures). The buffer SAXS profile was obtained in the same manner and subtracted from a protein SAXS profile. The merged experimental SAXS profile of *Vp*Nup133$^{55-502}$ was compared with SAXS profiles calculated using FoXS (38, 39), for the crystal structure of *Vp*Nup133$^{55-502}$ and the "complete" models in which disordered components and 4 Se-Met residues were built using MODELLER 9.13 (40) (Figure 2B).

Negative-stain electron microscopy (EM) for *Sc*Nup133$^{2-1157}$

A specimen of the full length *Sc*Nup133$^{2-1157}$ was prepared for negative-stain EM (41) (Supplemental Experimental Procedures). The 1,976 individual particles were selected interactively from images using Boxer from EMAN (42), and windowed into individual images with a size of 120x120 pixels. The particles were centered and normalized, then subjected to the *Iterative Stable Alignment and Clustering* (ISAC) (43)

technique to produce 23 stable 2D class averages after 10 generations (these class averages comprise 1,530 of the 1,976 particles) (top rows in Figure 3C and column 2 in Supplemental Table S3).

Structure and dynamics of *Sc*Nup133 revealed by integrative modeling approach

We developed an integrative modeling approach that produces atomic models for multiple states of a protein, based on EM images of the protein as well as SAXS profiles and crystal structures of the sequence segments and their homologs. We proceeded through three stages (Figure 1): (1) gathering of data, (2) conformational sampling and scoring to produce a minimal ensemble of conformations consistent with SAXS profiles, EM class averages, template structures, and chemical cross-links, and (3) analysis of the ensemble. The integrative modeling protocol was scripted in Python, based on our open source Integrative Modeling Platform package release 2.2 (http://salilab.org/imp) (44). Files for the input data, script, and output models are available at http://salilab.org/nup133.

Stage 1: Gathering of data

19 SAXS profiles and 23 EM class averages were obtained as described above. The atomic structures of *Sc*Nup133$^{944\text{-}1157}$ (PDB code 3KFO) (21), *Vp*Nup133$^{55\text{-}502}$ (4Q9T, Figure 2 and Table 1), and human homologs *Hs*Nup*133*$^{75\text{-}477}$ (1XKS) (19), *Hs*Nup*133*$^{517\text{-}1156}$ (3I4R) (22), and *Hs*Nup*133*$^{935\text{-}1156}$ (3CQC) (23) have been previously determined by X-ray crystallography. In addition, putative homologs of known structure were detected for the first 60 residues of the N-terminal domain by HHpred (45, 46) and

ModWeb (http://salilab.org/modweb) (47). Domain boundaries, secondary structure segments, and disordered regions were predicted by DomPRED (48), PSIPRED (49, 50), and DISOPRED (51), respectively. 18 disuccinimidyl suberate (DSS) and 23 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) intra-molecular chemical cross-links for *Sc*Nup133 (Table 2) as well as 2 DSS and 2 EDC inter-molecular chemical cross-links spanning the *Sc*Nup133 - *Sc*Nup84 interaction interface (Table 3) were obtained from our companion study on the entire *Sc*Nup84 complex (24).

Stage 2: Conformational sampling and scoring to produce a minimal ensemble of conformations consistent with SAXS profiles, EM class averages, template structures, and chemical cross-links

(1) Building an initial atomic model of ScNup133$^{2-1157}$

The shape and size information contained in the SAXS profiles can be used to improve accuracy of atomic comparative models. An initial atomic model of the full-length *Sc*Nup133$^{2-1157}$ was built based on template structures and SAXS profiles for 18 nested *Sc*Nup133 constructs (Supplemental Figures S3 and S4, and Supplemental Table S1) as follows. First, we built 1,000 atomic comparative models for the smallest construct (52-483), based on the crystal structure of *Vp*Nup133$^{55-502}$ (Figure 2) and the closest known structure detected by HHPred (45, 46) and ModWeb (47), using MODELLER 9.13 (40). The theoretical SAXS profile and the χ value of the fit to the experimental SAXS profile of the corresponding construct were calculated for each of the 1,000 comparative models, using FoXS (38, 39). Then, these 1,000 models were ranked by the χ value of the fit to the experimental SAXS profile. Second, the best-

scoring model was used as a template for comparative modeling of the next larger constructs (41-483 and 52-515), supplemented by the additional templates found by HHPred and ModWeb. For each of the two constructs, the resulting models were again ranked based on the corresponding SAXS profile fit. The entire process was repeated until the largest construct of *Sc*Nup133$^{2\text{-}1157}$ was modeled, resulting in the initial model of the full-length *Sc*Nup133$^{2\text{-}1157}$ (Supplemental Figures S3 and S4, and Supplemental Table S1).

(2) Conformational sampling using AllosMod

The initial model of *Sc*Nup133$^{2\text{-}1157}$ was subjected to molecular dynamics-based conformational sampling, using *AllosMod* (http://salilab.org/allosmod) (52), resulting in 7,000 conformations, as follows. The *AllosMod* simulations were short, near-equilibrium trajectories based on an input sequence and the initial model of *Sc*Nup133$^{2\text{-}1157}$. *AllosMod* constructed an energy landscape in which the atomic contacts from the input structure defined the major energetic minima (53), also generally known as a Gō model (54, 55). The energy landscape was then sampled using several constant temperature (at 300-350K) molecular dynamics simulations with short equilibration and a run time of 0.2 nanoseconds using 2 femto-second time steps and velocity rescaling every 200 steps (Supplemental Experimental Procedures).

(3) Scoring and searching for a minimal ensemble of conformations consistent with the SAXS profile, EM class averages, and chemical cross-links

The resulting 7,000 *AllosMod* conformations of *Sc*Nup133$^{2\text{-}1157}$ were pruned to

identify a minimal ensemble of up to five conformations that reproduce both the experimental SAXS profile and EM class averages of *Sc*Nup133$^{2\text{-}1157}$. The pruning was achieved by the *minimal ensemble search* (MES) program (56) that was modified to use a composite score defined as a weighted sum of the "*ensemble* SAXS score" and the "*ensemble em2D* Z-score".

The *ensemble* SAXS score is the χ value for the comparison of the *ensemble* SAXS profile to the experimental profile; the *ensemble* SAXS profile is a weighted average of the theoretical SAXS profiles for the selected subset of conformations, calculated using FoXS (38, 39).

To compute the *ensemble em2D* Z-score, we first calculated individual *em2D* scores for each of the 7,000 conformations matched against each of the 23 EM class averages, using the *EMageFit* application (57) of IMP (44) at 15 Å resolution; the *em2D* score is one minus the cross-correlation coefficient between a class average and the best-matching projection of a conformation (57). Each score was then normalized into a Z-score, by using the average and standard deviation of the scores for the same class average. Finally, the *ensemble em2D* Z-score was obtained by summing the lowest individual *em2D* Z-scores determined for each of the 23 EM class averages in the subset.

Independent fitting of subsets ranging from 1 to 5 conformations showed that a minimal ensemble of 4 conformations were sufficient to explain both the experimental SAXS profile and EM class averages of *Sc*Nup133$^{2\text{-}1157}$ (Figure 3 and Supplemental Table S3). The relative weight of -0.05 for the *ensemble em2D* Z-score in the composite score was determined by trial-and-error to balance the fit of the minimal ensemble to

both the SAXS and EM data.

As the final assessment step, we validated the conformations of *Sc*Nup133$^{2\text{-}1157}$ against each of the 18 DSS and 23 EDC intra-molecular chemical cross-links obtained from our companion study on the entire *Sc*Nup84 complex (24) (Table 2).

Stage 3: Analysis of the minimal ensemble

The most populated conformation in the minimal ensemble of 4 conformations was used as a reference for rigid body least-squares superposition of the remaining 3 conformations. The *ab initio* shape of the full-length *Sc*Nup133$^{2\text{-}1157}$ (a gray envelope in Figure 3A) was generated from the experimental SAXS profile using DAMMIF (58) and DAMAVER (59). UCSF Chimera was used for visualization (37). The TM-scores between the conformations in the minimal ensemble were calculated on web-server (http://zhanglab.ccmb.med.umich.edu/TM-score/) (60).

Each of the 23 EM class averages was attempted to be assigned to one or more of the 4 conformations in the minimal ensemble: An EM class average is assigned to a conformation when its *em2D* *Z*-score is less than -0.95 and the cross-correlation coefficient (c.c.c.) is higher than 0.82 or 0.85 (Supplemental Table S3).

Validation of the *Sc*Nup133 - *Sc*Nup84 interface with mutational analysis and chemical cross-links

The interface between *Sc*Nup133$^{881\text{-}1157}$ and *Sc*Nup84$^{506\text{-}726}$ (in the stalk of the yeast Nup84 complex) was predicted by calculating the difference in solvent accessibility between the unbound and bound models of *Sc*Nup133$^{881\text{-}1157}$ and

*Sc*Nup84$^{506-726}$. The *Sc*Nup133$^{881-1157}$ model was extracted from the initial atomic model of *Sc*Nup133$^{2-1157}$, while the *Sc*Nup84$^{506-726}$ model was built based on the human Nup133-Nup107 complex structure (PDB code: 3CQC) (23) using MODELLER (40). The *Sc*Nup133$^{881-1157}$ - *Sc*Nup84$^{506-726}$ interface model was built by structurally aligning each component into the corresponding chain of the human Nup133-Nup107 crystal structure. The unbound model of *Sc*Nup133$^{881-1157}$ and *Sc*Nup84$^{506-726}$ was obtained by separating the two components by 100 Å. Residue solvent accessibility was calculated using MODELLER. The residues with large solvent accessibility change (residues of L922, L929, and L933) were identified as target interface residues (Figure 4A). Finally, the *Sc*Nup133$^{881-1157}$ - *Sc*Nup84$^{506-726}$ interface model was validated against the 2 DSS and 2 EDC inter-molecular chemical cross-links spanning the *Sc*Nup133 - *Sc*Nup84 interface, which were obtained from our companion study on the entire *Sc*Nup84 complex (24) (Table 3).

To validate the predicted interface experimentally, we designed 3 sets of double point mutants located within and outside of the interface. Maximally disruptive mutations were predicted by program EGAD (61) (Supplemental Table S2), allowing us to propose two sets of double mutants affecting the binding interface (L922Y-L929Y and L929W-L933W) as well as a control mutation at a distal surface position (L1089Y-L1123Y) (Figure 4B).

The three mutants were generated by PCR and cloned into a yeast centromeric expression plasmid under the control of the GAL inducible promoter. Protein A and GFP C-terminal tags were used for IP and subcellular localization, respectively. A *Sc*Nup133 null mutant was transformed with the different constructs and the proteins were

expressed by growth in YP media suplemented with glucose, in which the basal activity of the GAL promoter generated a close to wild type expression of the constructs. Affinity purifications of the *Sc*Nup133 interacting proteins were performed using a buffer formulation that allows exclusive purification of the Nup84 complex components (8). Fitness phenotypic analysis was also performed as described previously (8) (Figure 5). GFP fusions were transformed into a *Sc*Nup133 null, Nup170-mCherry strain and analyzed by fluorescence microscopy. Cells grown in minimal media suplemented with glucose were visualized with a 63x 1.4 NA Plan-Apochromat objective (Carl Zeiss) using a microscope (Axioplan 2; Carl Zeiss) equipped with a cooled charge-coupled device camera (ORCA-ER; Hamamatsu Photonics). The system was controlled with Openlab imaging software (PerkinElmer). Final images were assembled, and gamma-levels were adjusted to enhance contrast only using Photoshop software (Adobe Inc.).

Annotating the potential ArfGAP1 lipid packing sensor (ALPS) motifs

We searched the sequences of *Sc*Nup133, *Vp*Nup133, and *Sc*Nup120 for potential amphipathic helix forming patterns, using the Membrane Protein Explorer (MPEx) package (62). Then, the resulting potential amphipathic helices were analyzed by the HeliQuest web-server (63) to annotate the potential ArfGAP1 lipid packing sensor (ALPS) motif as described previously (18, 64) (Supplemental Experimental Procedures and Supplemental Figure S2).

Results

Crystal structure of *Vp*Nup133$^{55-502}$

While *Vanderwaltozyma polyspora* is considered to be a somewhat distant yeast relative of *Saccharomyces cerevisiae* (65), both species diverged from the same ancestor that underwent a whole-genome duplication event. The sequence identity between *Sc* and *Vp* Nup133, across their entire length, is 37%. As a part of the Protein Structure Initiative project, several fungal Nup133 proteins were screened for crystallization, and the N-terminal domain of *Vanderwaltozyma polyspora* Nup133 yielded diffraction quality crystals. The construct encompassing residues 55 to 502, corresponding to the N-terminal domain of *Vanderwaltozyma polyspora* Nup133 [*Vp*Nup133$^{55-502}$], yielded crystals (resolution ~3.0 Å) in the orthorhombic space group $C222_1$ with two molecules in the asymmetric unit. The structure of *Vp*Nup133$^{55-502}$, determined by Se-SAD was refined to R_{work} and R_{free} values of 21.5 % and 26.8 %, respectively (Table 1). As expected, the overall *Vp*Nup133$^{55-502}$ adopts a disc-shaped, canonical β-propeller fold generated by radial arrangement of seven blades, around a central channel, each containing an anti-parallel β-sheet formed by four strands (Figure 2A; PDB Code 4Q9T, chain B). Notably, the disordered segments 86-98, 139-144, 157-180, and 202-214 occurs within the blades 1 and 2. Within each blades strands A, B, C, and D are arranged from inner-most to outer-most fashion resulting in a top surface decorated with the BC-loops (connecting BC strands within each blade) and DA-loops (loops connecting the D strand of n^{th} and A strand of n+1 blade), and a bottom surface decorated by the AB- and CD-loops. The β-sheets forming the blades 4 and 5, which also has a helical insertion between them, *Vp*Nup133$^{55-502}$ show a greater degree of

curvatures compared to other blades. Interestingly, the 7th blade of the propeller has an additional strand with the A, B, C strands from the C-terminus of the domain interacting with a β-hairpin at the N-terminus of the propeller (Figure 2A). Residues of the E7 and D7 β-strands forming the N-terminal β-hairpin show a notable degree of conservation within fungal Nup133 sequences (Supplemental Figure S1). This type of "velcro" arrangement is important for stability of the circular, β-propeller architectures (66).

Analyses using PISA (https://www.ebi.ac.uk/pdbe/pisa/) (67) suggested that the two monomers of *Vp*Nup133$^{55-502}$ within the asymmetric unit of the crystals do not form a biological complex (total buried surface area = 1690 Å^2) in solution (an important distinction when reconstructing the assembly state of the NPC *in vivo*). In addition, the *DiMoVo* score (http://albios.saclay.inria.fr/dimovo) (68) of -0.140 also indicated that the two monomers are merely the crystallographic dimers. Accordingly, the merged experimental SAXS profile of *Vp*Nup133$^{55-502}$ is well matched (χ = 1.14) to the SAXS profile calculated from the "complete" monomer model, generated by modeling the disordered residues of the crystal structure (blue in Figure 2B and Supplemental Table S1). In contrast, the SAXS profile for the "complete" dimer model, representing the crystallographic asymmetric unit, had an unacceptably high χ value of 12.4 and the R_g value of 35.7 Å (red in Figure 2B). Further, the measured radius of gyration (R_g) of 24.4 ± 0.3 Å, determined from the experimental SAXS profile with AUTORG (69), was consistent with the R_g value of 24.8 Å calculated from the "complete" monomer model of *Vp*Nup133$^{55-502}$. The composition of *Vp*Nup133$^{55-502}$, estimated with OLIGOMER (70), based on the experimental SAXS profile is 100% monomer. Thus, our SAXS analyses of the solution behavior of *Vp*Nup133$^{55-502}$ are fully consistent with the monomer of the

X-ray crystal structure.

A superposition of human Nup133 (*Hs*Nup133$^{75-477}$) (PDB code 1XKS) (19) and *Vp*Nup133$^{55-502}$ crystal structures reveal that the two β-propeller domains show overall similarity in their architecture in-spite of low sequence identity. These two structures agree well resulting in root mean square deviation of 2.2 Å over 293 superposed residues with sequence identity of only 15 % (Figure 2C). Notable differences in the architecture of the two structures include i) the 7th blade of *Hs*Nup133 β-propeller has only four strands with the absence of β-hairpin at its N-terminal region; ii) the *Hs*Nup133 β-propeller has additional helical insertions between blades 7 and 1, and blades 4 and 5 compared to the *Vp*Nup133; and iii) importantly, the residues of "DA$_{34}$-loop", connecting the D-strand of blade 3 to the A-strand of blade 4, which are disordered in the *Hs*Nup133, adopt mostly a random-coil conformation with a 3_{10}-helix near the A4-strand (Figure 2, A and C). Interestingly, the disordered "DA$_{34}$-loop" of *Hs*Nup133 has been shown to contain an ArfGAP1 lipid packing sensor (ALPS) motif, which is likely to facilitate membrane-curvature sensing and formation in the NPC. We have identified a potential ALPS motif within the "DA$_{34}$-loop" *Vp*Nup133 β-propeller domain comprising the sequence 267-LIKPQNSFFFRNLDSSKEIISL-288 (Figure 6B and Supplemental Figure S2B). Notably, an electrostatic potential map of *Vp*Nup133$^{55-502}$ reveals a large positively charged surface adjacent to the "DA$_{34}$-loop" (highlighted in yellow, Figure 2D). This surface is decorated mainly by residues from 3rd and 4th blades, which show a notable degree of conservation (Supplemental Figure S1), of the propeller. It plausible that this positively charged surface being close to the "DA$_{34}$-loop" might play a role facilitating interaction of the potential ALPS motif of *Vp*Nup133 with the membrane.

Structure and dynamics of *Sc*Nup133 revealed by integrative modeling approach

An initial atomic model of ScNup133$^{2-1157}$

An initial atomic model of the full-length *Sc*Nup133$^{2-1157}$ was built based on the nested template structures and 18 SAXS profiles (Supplemental Figures S3 and S4, and Supplemental Table S1). The model contains a single "linker" region (residues 480-495) between the N- and C-terminal domains aligned with the template structures (Supplemental Figures S3), indicating potential variability in the relative orientation of the two domains. Elastic network model analyses using *HingeProt* (71) suggested a long-range motion of the N- and C- terminal domains about the hinge residues of Leu51-His52 (near the beginning of the β-propeller), Glu484-Thr485 (at the flexible linker residues), and Ser895-Tyr896 (near the beginning of the 3KFO template). This motion was also identified as one of the top scoring normal modes by the web server *elNémo* (72). Such conformational dynamics might play a role in regulating the dynamic structure of the NPC. Therefore, we probed the structure and dynamics of *Sc*Nup133 in solution using an integrative modeling approach that benefits from both SAXS profiles and EM micrographs (Figure 1).

Conformational sampling and minimal ensemble search (MES)

The experimentally measured SAXS profile of *Sc*Nup133$^{2-1157}$ in solution (black in Figure 3B) did not match the theoretical SAXS profiles computed from the comparative model (χ = 6.27, red in Figure 3B) (Supplemental Figure S4C and Supplemental Table S1), although each of the N- and C- terminal domains satisfies its

corresponding SAXS profile (χ = 1.36 and 1.71, respectively) (Supplemental Figure S4A-B and Supplemental Table S1). Further, the maximum particle size (D_{max}) of 169.2 Å determined experimentally was 4.3% larger than the maximum dimension of 162.1 Å from the comparative model. EM analysis of *Sc*Nup133$^{2\text{-}1157}$, using the *Iterative Stable Alignment and Clustering* (ISAC) method (43), revealed 23 stable class averages after 10 generations (top rows in Figure 3C). Similarly to the SAXS results, a number of EM classes could not be fit by the comparative model of *Sc*Nup133$^{2\text{-}1157}$.

To study the structure and dynamics of *Sc*Nup133$^{2\text{-}1157}$ in solution, conformational sampling by molecular dynamics was carried out by *AllosMod* (52), starting from the initial atomic model of *Sc*Nup133$^{2\text{-}1157}$. We tested whether or not the resulting 7,000 conformations generated by *AllosMod* are consistent with the experimental SAXS profile and 23 EM class averages, using *minimal ensemble search* (MES) (56). As expected, both the SAXS profile and the 23 EM class averages were not explained simultaneously by any single sampled conformation of *Sc*Nup133$^{2\text{-}1157}$, indicating that *Sc*Nup133 is heterogeneous in solution.

The multi-state structural model of ScNup133$^{2\text{-}1157}$

The analysis of the resulting 7,000 conformations using the *minimal ensemble search* (MES) indicated that a minimal ensemble of 4 conformations (the multi-state model) was sufficient to explain the experimental SAXS profile (χ = 1.54, blue in Figure 3B), most of the EM class averages (Figure 3C), and most of the chemical cross-links determined by mass spectrometry (Table 2). The multi-state model consists of a single major “extended” conformation with a population weight of 0.506 (blue) and three minor

“compact” conformations with population weights of 0.242 (red), 0.202 (cyan), and 0.050 (yellow) (Figure 3A). The maximum particle size (D_{max}) of the “extended” conformation was measured as ~180 Å, while D_{max} of the three “compact” conformations ranged from ~135 Å to ~145 Å. In addition, the root-mean-square deviation (RMSD) and the TM-scores (60) of each pair from the three “compact” conformations ranged 10.6-23.7 Å and 0.44-0.64, respectively, indicating similar folds within the “compact” ones. On the other hand, the RMSD and the TM-scores of the “extended” conformation relative to the rest of the conformations ranged 25.6-29.0 Å and 0.38-0.49, respectively, indicating lower similarity in folds between the “extended” and the “compact” conformations. Importantly, 22 EM class averages (out of the total 23 class averages) could be assigned to at least one of the 4 conformations in the multi-state model with high confidence, covering 95.6% of the 23 EM class averages (Figure 3C and Supplemental Table S3). Moreover, 94.4% of the 18 DSS and 91.3% of the 23 EDC intra-molecular chemical cross-links were satisfied by the multi-state model, within 35 Å and 25 Å thresholds, respectively, independently validating our modeling (Table 2). Furthermore, the variability in the multi-state model supports the long-range motion indicated by *HingeProt* (71) and *elNémo* (72).

The conformational dynamics of *Sc*Nup133 appears to be dominated by the relative motions of the N- and C-terminal domains, connected by a relatively lengthy linker of 16 residues (480-495) (Supplemental Figures S3). The maximal displacement of the N-terminal β-propeller domains was measured as ~110 Å in the multi-state model. In conclusion, it appears that *Sc*Nup133 is a highly dynamic molecule in solution, with the maximal dimension (D_{max}) of the molecule changing from 135 Å to 180

Å in solution.

The *Sc*Nup133 - *Sc*Nup84 interface validated by mutational analysis and chemical cross-links

*Sc*Nup133 connects to the rest of the Nup84 complex through a tail-to-tail interaction with *Sc*Nup84 (8, 22, 23). The model of *Sc*Nup133 allowed us to predict the location and molecular details of this interaction surface (Figure 4A). Accordingly, the inter-molecular chemical cross-links spanning the *Sc*Nup133 - *Sc*Nup84 interface (24) were fully consistent with the *Sc*Nup133$^{881-1157}$ - *Sc*Nup84$^{506-726}$ interface model (Table 3).

To validate the predicted interface experimentally, we designed 3 sets of double point mutants located within and outside of the interface (Figure 4). Combinations of two *Sc*Nup133 leucine residues within the interface were mutated to bulky hydrophobic (tryptophan; L929W-L933W) or bulky polar (tyrosine; L922Y-L929Y) residues in an effort to create steric incompatibility between *Sc*Nup133 and *Sc*Nup84. As a control, two *Sc*Nup133 leucine residues located at the distal surface positions were also mutated to tyrosine (L1089Y-L1123Y). The mutant proteins were expressed in a *Sc*Nup133 null background and tested for their ability to interact with *Sc*Nup84 by affinity purification (8). None of the mutant proteins affecting the *Sc*Nup133 - *Sc*Nup84 interaction surface was able to copurify with any of the *Sc*Nup84 complex components (columns 2 and 3 in Figure 4B), strongly suggesting that the binding between *Sc*Nup133 and *Sc*Nup84 has been disrupted.

Fitness analysis of the same mutant strains (Figure 5A) showed that the mutants

at the *Sc*Nup133 - *Sc*Nup84 interface are not able to fully rescue the fitness phenotype caused by the deletion of *Sc*Nup133. The mutant proteins were also tagged with GFP to analyze their subcellular localization (Figure 5B). All mutant proteins show nuclear rim localization and co-localization with the NPC marker Nup170-mCherry, discarding the possibility of mislocalization as the main cause of the loss of binding to Nup84. While the wild type and the control mutant can effectively rescue the NPC clustering phenotype characteristic of the *Sc*Nup133 null mutation (Figure 5B, upper right panel), the mutants disrupting the *Sc*Nup133 - *Sc*Nup84 interface still show a NPC clustering defect similar to that of the *Sc*Nup133 deletion. In summary, our experimental tests validate the predicted *Sc*Nup84 binding sites on *Sc*Nup133. Moreover, *Sc*Nup133 interactions with other parts of the NPC scaffold are indicated by the relative independence of NPC localization on the *Sc*Nup133 - *Sc*Nup84 interaction.

The ArfGAP1 lipid packing sensor (ALPS) motifs

The seven-bladed β-propellers of Nup133s from *Vanderwaltozyma polyspora* (*Vp*Nup133$^{55-502}$, PDB code: 4Q9T) and *Homo sapiens* (*Hs*Nup133$^{75-477}$, PDB code: 1XKS) (19) share a high degree of similarity in their structural arrangement (Figure 2C). A loop (245-LPQGQGMLSGIGRKVSSLFGILS-267) unresolved in the X-ray structure of *Hs*Nup133$^{75-477}$ has been shown to act as a membrane curvature sensing ALPS motif at the vicinity of membranes (18, 64) (Figure 6A and Supplemental Figure S2A). Notably, *Vp*Nup133$^{55-502}$ contains a clearly resolved loop region (267-LIKPQNSFFFRNLDSSKEIISL-288) with a partial helical segment, including the "DA_{34}-loop" between blades 3D and 4A at the equivalent position to the ALPS motif (245-267)

in *Hs*Nup133$^{75-477}$ (Figures 6B, 1A, and 1C, and Supplemental Figure S2B); According to the helical-wheel representation, the *Vp*Nup133$^{267-288}$ loop region contains a polar face rich in serine residues and displays high level of mean hydrophobicity and hydrophobic moments, similar to those of canonical ALPS motifs (18) (Figure 6B and Supplemental Figure S2B). This observation suggests that *Vp*Nup133 may also contain an ALPS motif in its β-propeller domain.

To verify that the potential ALPS motif in *Vp*Nup133 is a conserved feature, we also analyzed the sequence of *Sc*Nup133 to look for an equivalent motif using the MPEx package (62) and the HeliQuest web-server (63). Similarly to *Vp*Nup133, we found that a loop region of *Sc*Nup133$^{252-270}$ (252-FKLGIWSKIFNTNSSVVSL-270) at the equivalent position to the ALPS motif (245-267) in *Hs*Nup133, also contains a polar face very rich in serine and threonine residues and displays high level of mean hydrophobicity and hydrophobic moments (Figure 6C and Supplemental Figure S2C). Its sequence in addition satisfies most of the biophysical criteria to be identified as an ALPS motif (18).

Also, we annotated two potential ALPS motifs in the β-propeller N-terminal domain of *Sc*Nup120, an ancient paralog protein of *Sc*Nup133 in the yeast Nup84 complex: i) the 2CD loop region *Sc*Nup120$^{135-152}$ (135-LSFLFSSANTLNGEWFHL-152); ii) and the 3D4A loop region *Sc*Nup120$^{197-216}$ (197-LLFNDNSYLKSLTRFFSRSS-216) were identified as potential ALPS motifs (Figure 6D-E and Supplemental Figure S2D-E), similarly to *Vp*Nup133$^{267-288}$ and *Sc*Nup133$^{252-270}$.

Discussion

Structure and dynamics of ScNup133 revealed by integrative modeling approach

Most proteins exist in multiple structural states. Conclusions based only on one state may therefore be incomplete or even misleading. In general, the structure of a protein is best determined based on all available data. Therefore, we developed an integrative structure determination approach for multiple states, that relies on data from X-ray crystallography, SAXS, EM, point mutations, and chemical cross-linking for either component fragments or an entire target protein (or its homologs) (Figure 1).

Components of the NPC have proven to be exemplars of proteins flexing between several states (73, 74), so we applied this approach to determining the structure and dynamics of one such component, full-length *Sc*Nup133$^{2\text{-}1157}$. We computed the minimal ensemble of 4 conformations (the multi-state model) of *Sc*Nup133$^{2\text{-}1157}$ that are consistent with EM images of the protein as well as SAXS profiles and crystal structures of the sequence segments and their homologs (including *Vp*Nup133$^{55\text{-}502}$ (Figure 2) determined in this study) (Figure 3). We validated the resulting model with 3 sets of double point mutations (Figures 4 and 5) and 45 chemical cross-links determined by mass spectrometry (24) (Tables 2 and 3).

The multi-state model consists of a single major “extended” conformation and three minor “compact” conformations (Figure 3A). The conformational dynamics of *Sc*Nup133 appears to be dominated by the relative motions of the N- and C-terminal domains. Notably, the conformational dynamics of *Sc*Nup133 are consistent with 38 of the 41 intra-molecular chemical cross-links (Table 2), corresponding to the typical false positive rate observed for chemical cross-linking (24). Finally, the *Sc*Nup133 dynamics

is also consistent with the extreme flexibility shown for the *Hs*Nup133 region of the Nup107-160 complex by negative stain EM (13).

Similar inter-domain dynamics were shown for the N-terminal domain of Nup192 (73). These dynamics may contribute to the passage of bulky cargoes through the restricted central channel of the NPC. In addition, the conformational dynamics of *Sc*Nup133 may insulate the structure of the NPC from morphological changes of the nuclear envelope during cell division and growth (73). Such flexibility in the NPC has been suggested by prior high-resolution EM studies (74, 75). Also, this flexibility may be required during the biogenesis of NPCs to interlock various nucleoporins in the Nup84 complex (76, 77), through the Nup133-Nup84 and Nup133-Nup120 interfaces (8, 12). A related possibility is that, once assembled into the mature NPC, Nup133 could be preferentially stabilized in one of the 4 conformations described in this study.

*Sc*Nup133 connects to the rest of the Nup84 complex through a tail-to-tail interaction with *Sc*Nup84 (8, 22, 23). The model of *Sc*Nup133 allowed us to predict the location and molecular details of this interaction surface, validated by 3 sets of double point mutations at the *Sc*Nup133-*Sc*Nup84 interface (Figures 4 and 5). We also identified a number of inter-molecular cross-links between *Sc*Nup133 and *Sc*Nup84 (24) that are consistent with our predicted interface (Table 3). In addition, the predicted *Sc*Nup133-*Sc*Nup84 interface is consistent with that of *Hs*Nup133-*Hs*Nup107 identified by X-ray crystallography (22, 23).

The presence of the ALPS motifs indicates a conserved mechanism for assembly and membrane anchoring of the NPC

In our previous structure-function analysis of the Nup84 complex (8), we identified the β-propeller region of the paralog proteins Nup133 and Nup120 as hotspots for the NPC clustering phenotype, an abnormal distribution of NPCs into closely packed groups that occurs in response to mutations in certain nucleoporins. Based on this and additional functional data, we hypothesized that NPC clustering was caused by the inability of the Nup84 complex to interact and stabilize the curvature of the NE membrane at the interface with the NPC. However, the mechanism used by the Nup84 complex to interact with the membrane was not clear, and previous studies have not been able to detect any membrane interaction motifs in yeast Nup133 (18), leading to the suggestion that the ALPS motif in Nup133 was restricted to organisms with open mitosis (18, 19).

In this study, we have now identified a potential ALPS motif located in the β-propeller domain of two related yeast Nup133 proteins (Figure 6B-C), which is inconsistent with the previous suggestion (18, 19) because yeast have a fully closed mitosis. In both yeast species, each of the ALPS motifs closely matches the consensus previously established for this kind of membrane curvature sensing motifs (Supplemental Figure S2B-C) (18). We identified the ALPS motif in a resolved loop in the crystal structure of the *Vp*Nup133 β-propeller (Figure 6B); the equivalent region in the human Nup133 counterpart contains a functional ALPS motif (Figure 6A) (15, 18). We have also detected the ALPS motif in our model of the *Sc*Nup133 β-propeller (Figure 6C). The presence of the ALPS motif in the domain that was previously identified as a hotspot for NPC clustering indicates a conserved mechanism used by yeast Nup133 to interact with the NE membrane and stabilize its curvature.

In addition, the β-propeller of the paralog protein *Sc*Nup120 was also identified as an NPC clustering hotspot, suggesting that a similar ALPS motif-dependent mechanism could apply to *Sc*Nup120. Correspondingly, we detected two potential ALPS motifs in the *Sc*Nup120 β-propeller domain (Figure 6D-E and Supplemental Figure S2D-E). Thus, the identification of the conserved ALPS motifs in the components of the yeast Nup84 complex strongly suggests that these motifs are a common and ancient eukaryotic feature, not restricted to the open mitosis organisms.

Here, we propose that the coordinated action of multiple copies of the ALPS motifs in the NPC (current predictions and nucleoporin stoichiometries (4, 5) indicate 16 copies in *Sc*Nup133s and 32 copies in *Sc*Nup120s, per yeast NPC) results in a network of protein-membrane contacts around the NPC periphery that may help stabilize the NE membrane, both during early stages of the NPC biogenesis in interphase (15) and throughout the NPC life cycle. This hypothesis may explain why in certain organisms, such as the fungus *Aspergillus nidulans*, the presence of transmembrane nucleoporins is not required to form a functional NPC, so long as an intact Nup84 complex is present (78). Other Nups are predicted to carry similar membrane-interacting motifs; both the yeast (79) and the vertebrate Nup53 (80) contain membrane-interacting motifs necessary for correct NPC assembly. The presence of such a conserved network of membrane-interacting motifs spread across the inner face of the NPC's scaffold and facing the pore membrane may be a key mechanism driving the assembly and stable association of the NPC with the NE.

The protocoatomer hypothesis suggests that many of the eukaryotic membrane coating complexes originated from an ancestral coating complex through a series of

duplication, specialization, and secondary loss processes (14, 81). The Nup84 complex has a common evolutionary origin with the outer coats of vesicle coating complexes, including clathrin, COPI, and COPII (14, 81-83). Thus, our results suggesting that the ALPS membrane-interacting motifs are a common feature of the coat-like proteins Nup133 and Nup120 in the Nup84 complex, allow us to refine the protocoatomer hypothesis. No ALPS-like motifs have been identified in the proteins that form the outer coat in clathrin, COPI, and COPII. However, the small GTPase components that mediate the first stages of vesicle coat formation commonly associate with membranes through a membrane anchor, which can be an amphipathic peptide (84, 85). Perhaps the Nup84 complex retains ancestral features, including the ALPS membrane-interacting motifs, that were lost from other outer coats during the development and specialization of vesicle coating complexes.

Acknowledgements

Authors are grateful to the scientists at Eli Lilly & Company, San Diego, CA, who contributed towards cloning and expression of the *Sc* and *Vp*Nup133 constructs during PSI2. We thank Agustin Avila-Sakar and Yifan Cheng for helping with EM data collection at UCSF, and Paula Upla and William Rice at NYSBC for the advice of analyzing the EM data. We thank Dina Schneidman-Duhovny at UCSF for a discussion on SAXS analysis and Roxana Mironska for her assistance with the functional experiments. We are also grateful to Drs. John Tainer and Michal Hammel for help in using the SIBYLS beamline 12.3.1 at ALS. Funding for the NYSGXRC and NYSGRC were provided by NIH Grants U54 GM074945 (S.K.B.) and U54 GM094662 (S.C.A.), respectively. Additional funding for this work was provided by NIH grants R01 GM062427 (M.P.R.), R01 GM083960 (A.S.), U54 GM103511 and U01 GM098256 (A.S. and M.P.R.). Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Basic Energy Sciences. Use of the LRL-CAT beam line facilities at Sector 31 of the APS was provided by Eli Lilly, which operates the facility. Portions of this research were carried out at the SSRL, SLAC National Accelerator Laboratory operated for DOE by Stanford University. The SSRL SMBP is supported by the DOE Office of Biological and Environmental Research, by the NIH, NCRR, Biomedical Technology Program (P41RR001209), and by the NIGMS (P41GM103393). Use of the SIBYLS beamline 12.3.1 at ALS was supported in part by the Office of Science, Office of Biological and Environmental Research, U.S. Department of Energy, under Contract DE-AC02-05CH11231.

References

1. Aitchison, J. D., and Rout, M. P. (2012) The yeast nuclear pore complex and transport through it. *Genetics* 190, 855-883

2. Wozniak, R., Burke, B., and Doye, V. (2010) Nuclear transport and the mitotic apparatus: an evolving relationship. *Cell Mol Life Sci* 67, 2215-2230

3. Grossman, E., Medalia, O., and Zwerger, M. (2012) Functional architecture of the nuclear pore complex. *Annu Rev Biophys* 41, 557-584

4. Alber, F., Dokudovskaya, S., Veenhoff, L. M., Zhang, W., Kipper, J., Devos, D., Suprapto, A., Karni-Schmidt, O., Williams, R., Chait, B. T., Sali, A., and Rout, M. P. (2007) The molecular architecture of the nuclear pore complex. *Nature* 450, 695-701

5. Alber, F., Dokudovskaya, S., Veenhoff, L. M., Zhang, W., Kipper, J., Devos, D., Suprapto, A., Karni-Schmidt, O., Williams, R., Chait, B. T., Rout, M. P., and Sali, A. (2007) Determining the architectures of macromolecular assemblies. *Nature* 450, 683-694

6. Peters, R. (2009) Translocation through the nuclear pore: Kaps pave the way. *Bioessays* 31, 466-477

7. Rout, M. P., Aitchison, J. D., Suprapto, A., Hjertaas, K., Zhao, Y., and Chait, B. T. (2000) The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J Cell Biol* 148, 635-651

8. Fernandez-Martinez, J., Phillips, J., Sekedat, M. D., Diaz-Avalos, R., Velazquez-Muriel, J., Franke, J. D., Williams, R., Stokes, D. L., Chait, B. T., Sali, A., and Rout, M. P. (2012) Structure-function mapping of a heptameric module in the nuclear pore complex. *J Cell Biol* 196, 419-434

9. Szymborska, A., de Marco, A., Daigle, N., Cordes, V. C., Briggs, J. A., and Ellenberg, J. (2013) Nuclear pore scaffold structure analyzed by super-resolution microscopy and particle averaging. *Science* 341, 655-658

10. Lutzmann, M., Kunze, R., Buerer, A., Aebi, U., and Hurt, E. (2002) Modular self-assembly of a Y-shaped multiprotein complex from seven nucleoporins. *EMBO J* 21, 387-397

11. Kampmann, M., Atkinson, C. E., Mattheyses, A. L., and Simon, S. M. (2011) Mapping the orientation of nuclear pore proteins in living cells with polarized fluorescence microscopy. *Nat Struct Mol Biol* 18, 643-649

12. Seo, H. S., Ma, Y., Debler, E. W., Wacker, D., Kutik, S., Blobel, G., and Hoelz, A. (2009) Structural and functional analysis of Nup120 suggests ring formation of the Nup84 complex. *Proc Natl Acad Sci U S A* 106, 14281-14286

13. Bui, K. H., von Appen, A., DiGuilio, A. L., Ori, A., Sparks, L., Mackmull, M. T., Bock, T., Hagen, W., Andres-Pons, A., Glavy, J. S., and Beck, M. (2013) Integrated structural analysis of the human nuclear pore complex scaffold. *Cell* 155, 1233-1243

14. Devos, D., Dokudovskaya, S., Williams, R., Alber, F., Eswar, N., Chait, B. T., Rout, M. P., and Sali, A. (2006) Simple fold composition and modular architecture of the nuclear pore complex. *Proc Natl Acad Sci U S A* 103, 2172-2177

15. Doucet, C. M., Talamas, J. A., and Hetzer, M. W. (2010) Cell cycle-dependent differences in nuclear pore complex assembly in metazoa. *Cell* 141, 1030-1041

16. Walther, T. C., Alves, A., Pickersgill, H., Loiodice, I., Hetzer, M., Galy, V., Hulsmann, B. B., Kocher, T., Wilm, M., Allen, T., Mattaj, I. W., and Doye, V. (2003) The conserved Nup107-160 complex is critical for nuclear pore complex assembly. *Cell* 113,

195-206

17. Bolhy, S., Bouhlel, I., Dultz, E., Nayak, T., Zuccolo, M., Gatti, X., Vallee, R., Ellenberg, J., and Doye, V. (2011) A Nup133-dependent NPC-anchored network tethers centrosomes to the nuclear envelope in prophase. *J Cell Biol* 192, 855-871

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

19. Berke, I. C., Boehmer, T., Blobel, G., and Schwartz, T. U. (2004) Structural and functional analysis of Nup133 domains reveals modular building blocks of the nuclear pore complex. *J Cell Biol* 167, 591-597

20. Doye, V., Wepf, R., and Hurt, E. C. (1994) A novel nuclear pore protein Nup133p with distinct roles in poly(A)+ RNA transport and nuclear pore distribution. *EMBO J* 13, 6062-6075

21. Sampathkumar, P., Gheyi, T., Miller, S. A., Bain, K. T., Dickey, M., Bonanno, J. B., Kim, S. J., Phillips, J., Pieper, U., Fernandez-Martinez, J., Franke, J. D., Martel, A., Tsuruta, H., Atwell, S., Thompson, D. A., Emtage, J. S., Wasserman, S. R., Rout, M. P., Sali, A., Sauder, J. M., and Burley, S. K. (2011) Structure of the C-terminal domain of Saccharomyces cerevisiae Nup133, a component of the nuclear pore complex. *Proteins* 79, 1672-1677

22. Whittle, J. R., and Schwartz, T. U. (2009) Architectural nucleoporins Nup157/170 and Nup133 are structurally related and descend from a second ancestral element. *J Biol Chem* 284, 28442-28452

23. Boehmer, T., Jeudy, S., Berke, I. C., and Schwartz, T. U. (2008) Structural and

functional studies of Nup107/Nup133 interaction and its implications for the architecture of the nuclear pore complex. *Mol Cell* 30, 721-731

24. Shi*, Y., Fernandez-Martinez*, J., Tjioe*, E., Pellarin*, R., Kim*, S. J., Williams, R., Schneidman-Duhovny, D., Sali, A., Rout, M. P., and Chait, B. T. (2014) Rapid structural characterization by cross-linking reveals the detailed architecture of a coatomer-related heptameric module from the nuclear pore complex. *Mol Cell Proteomics*, MCP/2014/041673

25. Kabsch, W. (2010) XDS. *Acta Crystallographica Section D* 66, 125-132

26. Evans, P. R., and Murshudov, G. N. (2013) How good are my data and what is the resolution? *Acta Crystallographica Section D* 69, 1204-1214

27. Terwilliger, T. C., Adams, P. D., Read, R. J., McCoy, A. J., Moriarty, N. W., Grosse-Kunstleve, R. W., Afonine, P. V., Zwart, P. H., and Hung, L. W. (2009) Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr D Biol Crystallogr* 65, 582-601

28. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66, 213-221

29. Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L. W., Read, R. J., and Adams, P. D. (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild

wizard. *Acta Crystallogr D Biol Crystallogr* 64, 61-69

30. Cowtan, K. (2010) Recent developments in classical density modification. *Acta Crystallogr D Biol Crystallogr* 66, 470-478

31. Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* 62, 1002-1011

32. Cowtan, K. (2008) Fitting molecular fragments into electron density. *Acta Crystallographica Section D-Biological Crystallography* 64, 83-89

33. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67, 235-242

34. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66, 486-501

35. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53, 240-255

36. DeLano, W. L. (2002) The PyMOL User's Manual, DeLano Scientific, San Carlos, CA

37. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25, 1605-1612

38. Schneidman-Duhovny, D., Hammel, M., and Sali, A. (2010) FoXS: a web server

for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res* 38, W540-544

39. Schneidman-Duhovny, D., Hammel, M., Tainer, J. A., and Sali, A. (2013) Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophys J* 105, 962-974

40. Sali, A., and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234, 779-815

41. Ohi, M., Li, Y., Cheng, Y., and Walz, T. (2004) Negative Staining and Image Classification - Powerful Tools in Modern Electron Microscopy. *Biological procedures online* 6, 23-34

42. Ludtke, S. J., Baldwin, P. R., and Chiu, W. (1999) EMAN: semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol* 128, 82-97

43. Yang, Z., Fang, J., Chittuluru, J., Asturias, F. J., and Penczek, P. A. (2012) Iterative stable alignment and clustering of 2D transmission electron microscope images. *Structure* 20, 237-247

44. Russel, D., Lasker, K., Webb, B., Velazquez-Muriel, J., Tjioe, E., Schneidman-Duhovny, D., Peterson, B., and Sali, A. (2012) Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. *PLoS Biol* 10, e1001244

45. Soding, J. (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics* 21, 951-960

46. Soding, J., Biegert, A., and Lupas, A. N. (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 33, W244-248

47. Pieper, U., Webb, B. M., Dong, G. Q., Schneidman-Duhovny, D., Fan, H., Kim, S.

J., Khuri, N., Spill, Y. G., Weinkam, P., Hammel, M., Tainer, J. A., Nilges, M., and Sali, A. (2014) ModBase, a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res* 42, D336-346

48. Marsden, R. L., McGuffin, L. J., and Jones, D. T. (2002) Rapid protein domain assignment from amino acid sequence using predicted secondary structure. *Protein Sci* 11, 2814-2824

49. Jones, D. T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 292, 195-202

50. Buchan, D. W., Minneci, F., Nugent, T. C., Bryson, K., and Jones, D. T. (2013) Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res* 41, W349-357

51. Ward, J. J., McGuffin, L. J., Bryson, K., Buxton, B. F., and Jones, D. T. (2004) The DISOPRED server for the prediction of protein disorder. *Bioinformatics* 20, 2138-2139

52. Weinkam, P., Pons, J., and Sali, A. (2012) Structure-based model of allostery predicts coupling between distant sites. *Proc Natl Acad Sci U S A* 109, 4875-4880

53. Wolynes, P. G. (2005) Recent successes of the energy landscape theory of protein folding and function. *Q Rev Biophys* 38, 405-410

54. Schug, A., Whitford, P. C., Levy, Y., and Onuchic, J. N. (2007) Mutations as trapdoors to two competing native conformations of the Rop-dimer. *Proc Natl Acad Sci U S A* 104, 17674-17679

55. Li, W., Wolynes, P. G., and Takada, S. (2011) Frustration, specific sequence dependence, and nonlinearity in large-amplitude fluctuations of allosteric proteins. *Proc*

Natl Acad Sci U S A 108, 3504-3509

56. Pelikan, M., Hura, G. L., and Hammel, M. (2009) Structure and flexibility within proteins as identified through small angle X-ray scattering. *General physiology and biophysics* 28, 174-189

57. Velazquez-Muriel, J., Lasker, K., Russel, D., Phillips, J., Webb, B. M., Schneidman-Duhovny, D., and Sali, A. (2012) Assembly of macromolecular complexes by satisfaction of spatial restraints from electron microscopy images. *Proc Natl Acad Sci U S A* 109, 18821-18826

58. Franke, D., and Svergun, D. I. (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Cryst.* 42, 342-346

59. Volkov, V. V., and Svergun, D. I. (2003) Uniqueness of ab initio shape determination in small-angle scattering. *J Appl Crystallogr* 36, 860-864

60. Zhang, Y., and Skolnick, J. (2004) Scoring function for automated assessment of protein structure template quality. *Proteins* 57, 702-710

61. Pokala, N., and Handel, T. M. (2005) Energy functions for protein design: adjustment with protein-protein complex affinities, models for the unfolded state, and negative design of solubility and specificity. *J Mol Biol* 347, 203-227

62. Snider, C., Jayasinghe, S., Hristova, K., and White, S. H. (2009) MPEx: a tool for exploring membrane proteins. *Protein Sci* 18, 2624-2628

63. Gautier, R., Douguet, D., Antonny, B., and Drin, G. (2008) HELIQUEST: a web server to screen sequences with specific alpha-helical properties. *Bioinformatics* 24, 2101-2102

64. Drin, G., and Antonny, B. (2010) Amphipathic helices and membrane curvature.

FEBS Lett 584, 1840-1847

65. Kurtzman, C. P. (2003) Phylogenetic circumscription of Saccharomyces, Kluyveromyces and other members of the Saccharomycetaceae, and the proposal of the new genera Lachancea, Nakaseomyces, Naumovia, Vanderwaltozyma and Zygotorulaspora. *FEMS yeast research* 4, 233-245

66. Neer, E. J., and Smith, T. F. (1996) G protein heterodimers: new structures propel new questions. *Cell* 84, 175-178

67. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372, 774-797

68. Bernauer, J., Bahadur, R. P., Rodier, F., Janin, J., and Poupon, A. (2008) DiMoVo: a Voronoi tessellation-based method for discriminating crystallographic and biological protein-protein interactions. *Bioinformatics* 24, 652-658

69. Petoukhov, M. V., Konarev, P. V., Kikhney, A. G., and Svergun, D. I. (2007) ATSAS 2.1 - towards automated and web-supported small-angle scattering data analysis. *J Appl Crystallogr* 40, S223-S228

70. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., and Svergun, D. I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J Appl Crystallogr* 36, 1277-1282

71. Emekli, U., Schneidman-Duhovny, D., Wolfson, H. J., Nussinov, R., and Haliloglu, T. (2008) HingeProt: automated prediction of hinges in protein structures. *Proteins* 70, 1219-1227

72. Suhre, K., and Sanejouand, Y. H. (2004) ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement.

Nucleic Acids Res 32, W610-614

73. Sampathkumar, P., Kim, S. J., Upla, P., Rice, W. J., Phillips, J., Timney, B. L., Pieper, U., Bonanno, J. B., Fernandez-Martinez, J., Hakhverdyan, Z., Ketaren, N. E., Matsui, T., Weiss, T. M., Stokes, D. L., Sauder, J. M., Burley, S. K., Sali, A., Rout, M. P., and Almo, S. C. (2013) Structure, dynamics, evolution, and function of a major scaffold component in the nuclear pore complex. *Structure* 21, 560-571

74. Akey, C. W. (1995) Structural plasticity of the nuclear pore complex. *J Mol Biol* 248, 273-293

75. Yang, Q., Rout, M. P., and Akey, C. W. (1998) Three-dimensional architecture of the isolated yeast nuclear pore complex: functional and evolutionary implications. *Mol Cell* 1, 223-234

76. Melcak, I., Hoelz, A., and Blobel, G. (2007) Structure of Nup58/45 suggests flexible nuclear pore diameter by intermolecular sliding. *Science* 315, 1729-1732

77. Solmaz, S. R., Chauhan, R., Blobel, G., and Melcak, I. (2011) Molecular architecture of the transport channel of the nuclear pore complex. *Cell* 147, 590-602

78. Liu, H. L., De Souza, C. P., Osmani, A. H., and Osmani, S. A. (2009) 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* 20, 616-630

79. Marelli, M., Lusk, C. P., Chan, H., Aitchison, J. D., and Wozniak, R. W. (2001) A link between the synthesis of nucleoporins and the biogenesis of the nuclear envelope. *J Cell Biol* 153, 709-724

80. Vollmer, B., Schooley, A., Sachdev, R., Eisenhardt, N., Schneider, A. M.,

Sieverding, C., Madlung, J., Gerken, U., Macek, B., and Antonin, W. (2012) Dimerization and direct membrane interaction of Nup53 contribute to nuclear pore complex assembly. *EMBO J* 31, 4072-4084

81. Devos, D., Dokudovskaya, S., Alber, F., Williams, R., Chait, B. T., Sali, A., and Rout, M. P. (2004) Components of coated vesicles and nuclear pore complexes share a common molecular architecture. *PLoS Biol* 2, e380

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

83. Field, M. C., Sali, A., and Rout, M. P. (2011) Evolution: On a bender--BARs, ESCRTs, COPs, and finally getting your coat. *J Cell Biol* 193, 963-972

84. Faini, M., Beck, R., Wieland, F. T., and Briggs, J. A. (2013) Vesicle coats: structure, function, and general principles of assembly. *Trends Cell Biol* 23, 279-288

85. Antonny, B. (2011) Mechanisms of membrane curvature sensing. *Annu Rev Biochem* 80, 101-123

86. Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, 2577-2637

87. Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* 60, 2256-2268

88. Dolinsky, T. J., Czodrowski, P., Li, H., Nielsen, J. E., Jensen, J. H., Klebe, G., and Baker, N. A. (2007) PDB2PQR: expanding and upgrading automated preparation of

biomolecular structures for molecular simulations. *Nucleic Acids Res* 35, W522-525

89. Baker, N. A., Sept, D., Holst, M. J., and McCammon, J. A. (2001) The adaptive multilevel finite element solution of the Poisson-Boltzmann equation on massively parallel computers. *IBM Journal of Research and Development* 45, 427-438

90. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947-2948

91. Karplus, P. A., and Diederichs, K. (2012) Linking crystallographic model and data quality. *Science* 336, 1030-1033

92. Chen, V. B., Arendall, W. B., III, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D* 66, 12-21

Figure Legends

Figure 1. Integrative modeling approach for *Sc*Nup133

Our integrative modeling approach proceeds through three stages: (1) gathering of data, (2) conformational sampling and scoring to produce a minimal ensemble of conformations consistent with SAXS profiles, EM class averages, template structures, and chemical cross-links, and (3) analysis of the ensemble. The integrative modeling protocol was scripted in Python, based on our open source Integrative Modeling Platform package release 2.2 (http://salilab.org/imp) (44).

Figure 2. Crystal structure of *Vp*Nup133$^{55-502}$

(A) Overall fold of *Vp*Nup133$^{55-502}$ β-propeller domain is shown in cartoon representation as rainbow of blue to red from N- to C- terminus. Terminal residues of the disordered segments are marked in gray spheres. The location of "DA34-loop" containing a potential ALPS motif is indicated by an arrow. Secondary structure elements are shown as defined by the DSSP (86) program. Strands within each propeller blades marked in pink.

(B) Comparison of the merged experimental SAXS profile (black) of *Vp*Nup133$^{55-502}$ with the calculated SAXS profiles from the "complete" dimer model (χ = 12.4, red) and the "complete" monomer model (χ = 1.14, blue) are shown. The lower plot presents the residuals (calculated intensity / experimental intensity) of each calculated SAXS profile. The upper inset shows the SAXS profiles in the Guinier plot with an R_g fit of 24.4 ± 0.3 Å. The maximum particle size (D_{max}) is 77.8 Å determined experimentally.

(C) Superposition of *Vp*Nup133$^{55-502}$ (shown as pink cartoon) and *Hs*Nup133$^{75-477}$ (PDB:

1XKS; shown as light-blue cartoon) (19) structures performed using the SSM (87) routine as implemented in COOT (34).

(D) Electrostatic potential of *Vp*Nup133$^{55-502}$ plotted onto its solvent accessible surface. Missing side chains and charges were assigned for *Vp*Nup133$^{55-502}$ structure on the PDB2PQR web-server (http://nbcr-222.ucsd.edu/pdb2pqr_1.9.0/) (88), and electrostatic surface was calculated using APBS (89) within PyMOL. Negative (-4 kT/e) and positive (+4 kT/e) potentials are shown in red and blue, respectively. The positively charged surface located adjacent to the "DA$_{34}$-loop" (annotated as the potential ALPS motif) is marked by an ellipse in yellow. Phenylalanine at positions 275 and 276 are shown in sticks as reference for location of the "DA$_{34}$-loop".

Figure 3. Structure and dynamics of *Sc*Nup133 revealed by integrative modeling approach

(A) The minimal ensemble of 4 conformations (the multi-state model), comprised of a single major "extended" conformation with a population weight of 0.506 (blue) and three minor "compact" conformations with weights of 0.242 (red), 0.202 (cyan), and 0.050 (yellow) are shown. The most populated conformation (blue) was used as a reference for rigid body least-squares superposition of the remaining 3 conformations. The *ab initio* shape (represented as a gray envelope) computed from the experimental SAXS profile was also superposed for comparison.

(B) Comparison of the merged experimental SAXS profile (black) of *Sc*Nup133$^{2-1157}$ with the calculated SAXS profiles from the *Sc*Nup133$^{2-1157}$ comparative model (χ = 6.27, red) and the ensemble of 4 conformations (χ = 1.54, blue) are shown. The lower plot

presents the residuals (calculated intensity / experimental intensity) of each calculated SAXS profile. The upper inset shows the SAXS profiles in the Guinier plot with an R_g fit of 48.3 ± 0.6 Å. The maximum particle size (D_{max}) is 169.2 Å determined experimentally. (C) The 23 negative-stain EM class averages are shown along with the projections of each of the 4 conformations. 22 EM class averages were assigned to at least one of the 4 conformations with high confidence, as highlighted in colored boxes.

Figure 4. Mutational analysis of the *Sc*Nup133 - *Sc*Nup84 interface

The double mutations at the predicted *Sc*Nup133 - *Sc*Nup84 interface disrupted the association of *Sc*Nup133 with the rest of the *Sc*Nup84 complex.

(A) Alignment of the Nup133 amino acid sequences of different organisms (indicated on the left) corresponding to the predicted Nup84 interaction surface in *Saccharomyces cerevisiae*. The alignment was performed with ClustalW2 (90). The amino acids affected by the designed point mutations are indicated by an asterisk.

(B) Affinity purification of nucleoporins associated to the Nup133-PrA mutants indicated above each gel line. The identity of each protein is indicated in blue letters on the right. Marker protein molecular weights are indicated in kDa on the left side. Immunoglobulin contaminants are identified in gray letters.

Figure 5. Phenotypic analysis of the *Sc*Nup133 point mutants

(A) Fitness analysis of the Nup133 mutants is shown. 10-fold serial dilutions of strains expressing the Nup133-PrA construct indicated on the left side or carrying an empty plasmid (controls DNup133 and parental w303 strains) were spotted in minimal media

plates without triptophan and grown at the indicated temperatures for 2-3 days. A semi-quantitative score, in arbitrary units (AU), normalized to the wild type fitness, is shown on the right.

(B) Nup133 mutants associate with the NPC. NPC association of the Nup133-GFP constructs indicated on the left side was analyzed by live cell direct fluorescence microscopy in a DNup133, Nup170-mCherry mutant background. Cells were grown to early log phase at 25° C. Differential interference contrast (DIC, left) and single-channel fluorescence (central and right columns) are shown. Scale bar = 5 mm.

Figure 6. The potential ArfGAP1 lipid packing sensor (ALPS) motifs

(A-E) Each of the potential ALPS motifs and sequences for *Hs*Nup133 (A), *Vp*Nup133 (B), *Sc*Nup133 (C), and *Sc*Nup120 (D-E) are presented. The potential ALPS motifs (red) are visualized in their corresponding structures (gray) using UCSF Chimera (37), and highlighted by the blue circles. The helical-wheel representations of the potential ALPS motifs are also shown, with an arrow in the center of the helical-wheel representing the direction and strength of the mean hydrophobic moment of the corresponding ALPS motif.

Tables

Table 1. Crystallographic Statistics

Suggested location – Results, Crystal structure of *Vp*Nup133$^{55-502}$

Data Collection	***Vp*Nup133$^{55-502}$**
PDB code:	4Q9T
Space group:	$C222_1$
Unit-cell dimensions (Å):	a=109.2, b=133.8, c=136.8
Matthew's coefficient (Å^3/Da):	2.44
Solvent content (%):	49.6
Wavelength (Å) :	0.9793
Resolution (Å):	38.00-3.00 (3.18-3.00)[a]
Number of unique reflections:	20381 (3224)
Completeness (%):	99.7 (98.8)
R_{merge} (%):	18.9 (254.2)
Mean (I) half-set correlation CC(½) (91):	0.990 (0.668)
Multiplicity:	14.6 (13.9)
< I/s(I) >:	14.2 (1.2)
Refinement	
Number of reflections :	19,268
Number of reflections in test set:	1042
R_{work} (%):	21.5
R_{free} (%):	26.8
Root mean square deviations from ideal values	
Bond length (Å):	0.0080
Bond angles (°):	1.3299
Ramachandran Plot	
MolProbity (92) residues in	
Favored region (%):	90.8
Allowed region (%):	99.6

[a] Values in parenthesis correspond to the highest-resolution shell

Table 2. Validation of the multi-state model of *Sc*Nup133$^{2\text{-}1157}$ with 41 chemical cross-links determined by mass spectrometry

We validated the conformations of *Sc*Nup133$^{2\text{-}1157}$ against each of the 18 DSS and 23 EDC chemical cross-links obtained from our companion study on the entire *Sc*Nup84 complex (24). As a result, 94.4% of DSS and 91.3% of EDC cross-links were satisfied by the multi-state model, within 35 Å and 25 Å thresholds, respectively. Therefore, the conformational dynamics of *Sc*Nup133 are consistent with 38 of the 41 intra-molecular chemical cross-links, corresponding to the typical false positive rate observed for chemical cross-linking (24)

18 DSS crosslinks		**Cα-Cα distance (Å)**				**Result**
residue 1	residue 2	Conformation #1	Conformation #2	Conformation #3	Conformation #4	
59	5	40.9	65.4	27.8	34.1	Satisfied
115	59	15.2	34.9	13.3	26.7	Satisfied
168	191	24.6	17.6	22.1	20.3	Satisfied
171	191	20.6	15.2	17.8	17.8	Satisfied
191	239	17.0	18.1	18.9	19.2	Satisfied
278	239	22.9	17.0	15.6	15.4	Satisfied
488	442	17.7	27.9	16.9	20.1	Satisfied
506	59	42.4	45.0	47.3	38.1	Violated
532	488	23.4	17.3	22.6	27.4	Satisfied
544	488	33.5	11.0	31.4	29.8	Satisfied
913	946	26.1	25.1	22.1	21.7	Satisfied
946	912	24.9	24.6	20.7	19.6	Satisfied
966	963	6.2	5.9	5.2	4.5	Satisfied
973	976	5.5	6.0	5.2	6.7	Satisfied
1034	946	24.4	20.4	26.5	18.7	Satisfied
1072	936	31.3	20.9	35.0	24.4	Satisfied
1072	1034	16.9	18.3	19.4	18.8	Satisfied
1072	1035	17.3	20.1	20.1	18.5	Satisfied
Number of violation		2	2	1	1	**1**
Cross-links satisfied (%)		88.9%	88.9%	94.4%	94.4%	**94.4%**

23 EDC crosslinks		Cα-Cα distance (Å)				Result
residue 1	residue 2	Conformation #1	Conformation #2	Conformation #3	Conformation #4	
103	130	12.4	9.5	12.9	17.8	Satisfied
103	131	14.2	10.1	13.9	15.6	Satisfied
168	131	15.4	11.2	15.8	15.5	Satisfied
171	127	17.5	21.1	17.2	12.8	Satisfied
171	195	10.1	7.1	6.2	8.0	Satisfied
171	197	7.9	4.5	7.3	5.7	Satisfied
187	127	30.7	19.4	19.0	23.7	Satisfied
394	454	14.2	12.2	11.2	14.0	Satisfied
480	63	11.9	13.1	13.1	14.4	Satisfied
484	442	12.8	20.3	12.7	14.1	Satisfied
486	442	11.5	22.3	10.5	14.3	Satisfied
506	562	13.9	14.2	13.0	16.4	Satisfied
912	1097	16.3	25.8	13.6	13.0	Satisfied
913	989	16.8	18.4	16.6	16.3	Satisfied
913	1097	19.7	22.6	16.2	14.7	Satisfied
936	392	131.9	85.4	111.8	115.3	Violated
936	952	22.6	17.1	23.1	23.1	Satisfied
946	935	15.8	17.0	17.8	17.4	Satisfied
946	1097	32.1	28.7	30.6	28.9	Violated
1072	1032	20.9	23.2	24.0	23.5	Satisfied
1142	1068	16.2	15.8	16.6	15.2	Satisfied
1142	1069	16.5	15.9	17.0	15.1	Satisfied
1144	1068	13.1	11.9	12.4	13.6	Satisfied
Number of violation		3	3	2	2	**2**
Cross-links satisfied (%)		87.0%	87.0%	91.3%	91.3%	**91.3%**

Suggested location – Results, The multi-state structural model of ScNup133$^{2\text{-}1157}$

Table 3. Validation of the *Sc*Nup133 - *Sc*Nup84 interface with 4 chemical cross-links determined by mass spectrometry

We validated the *Sc*Nup133$^{881\text{-}1157}$ - *Sc*Nup84$^{506\text{-}726}$ interface model against each of the 2 DSS and 2 EDC chemical cross-links spanning the *Sc*Nup133 - *Sc*Nup84 interface, which were obtained from our companion study on the entire *Sc*Nup84 complex (24). As a result, all the inter-molecular chemical cross-links are fully consistent with the *Sc*Nup133$^{881\text{-}1157}$ - *Sc*Nup84$^{506\text{-}726}$ interface model.

4 crosslinks			**Cα-Cα distance (Å)**	**Result**
Chemical Cross-linker	*Sc*Nup133 residue	*Sc*Nup84 residue	*Sc*Nup133 (881-1157) - *Sc*Nup84 (506-726)	
DSS	936	712	14.6	Satisfied
DSS	966	625	19.0	Satisfied
EDC	935	712	11.6	Satisfied
EDC	936	713	18.3	Satisfied
Number of violation			**0**	
Cross-links satisfied (%)			**100.0%**	

Suggested location – Results, The *Sc*Nup133 - *Sc*Nup84 interface validated by mutational analysis and chemical cross-links

Figures

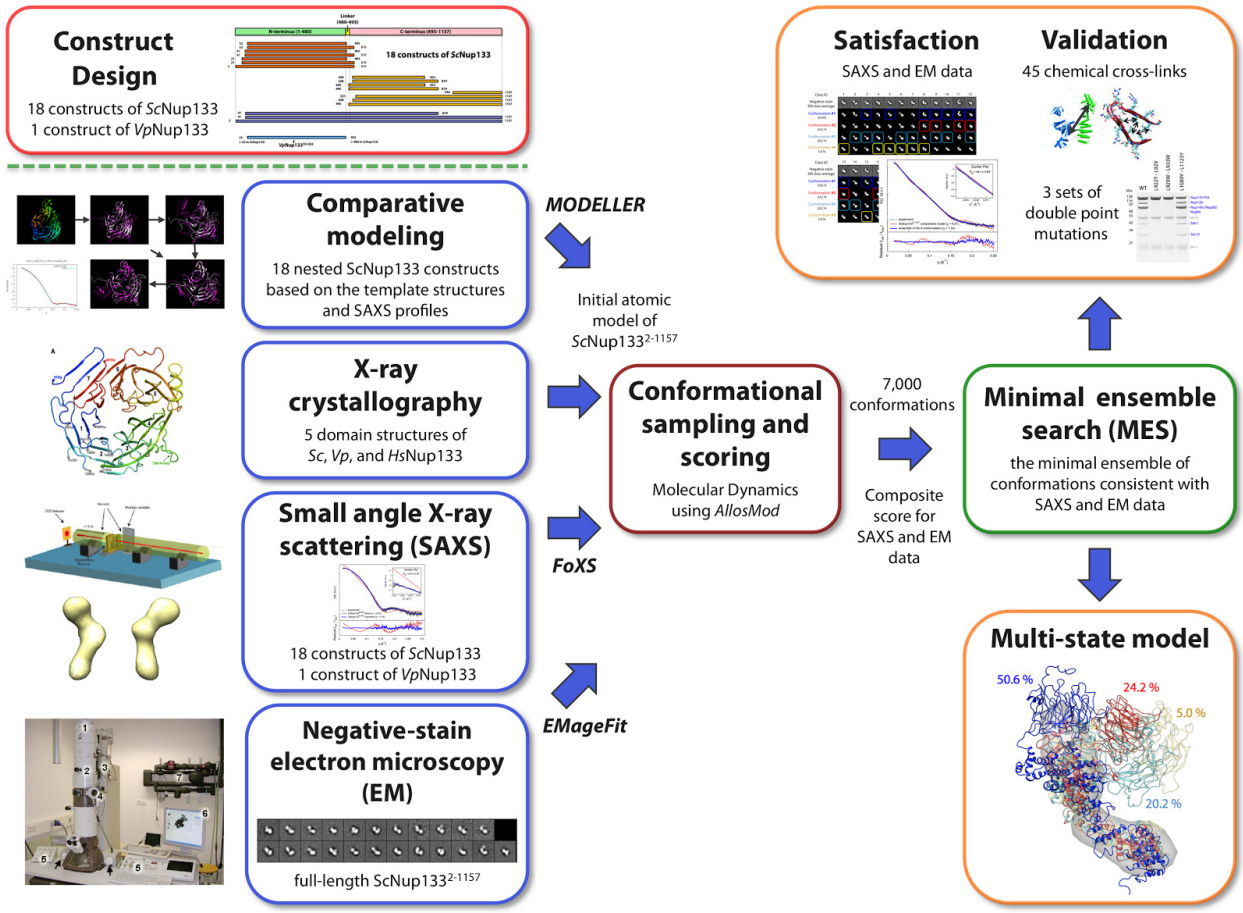

Figure 1. Integrative modeling approach for *Sc*Nup133

Suggested location – Experimental Procedures, Structure and dynamics of *Sc*Nup133 revealed by integrative modeling approach

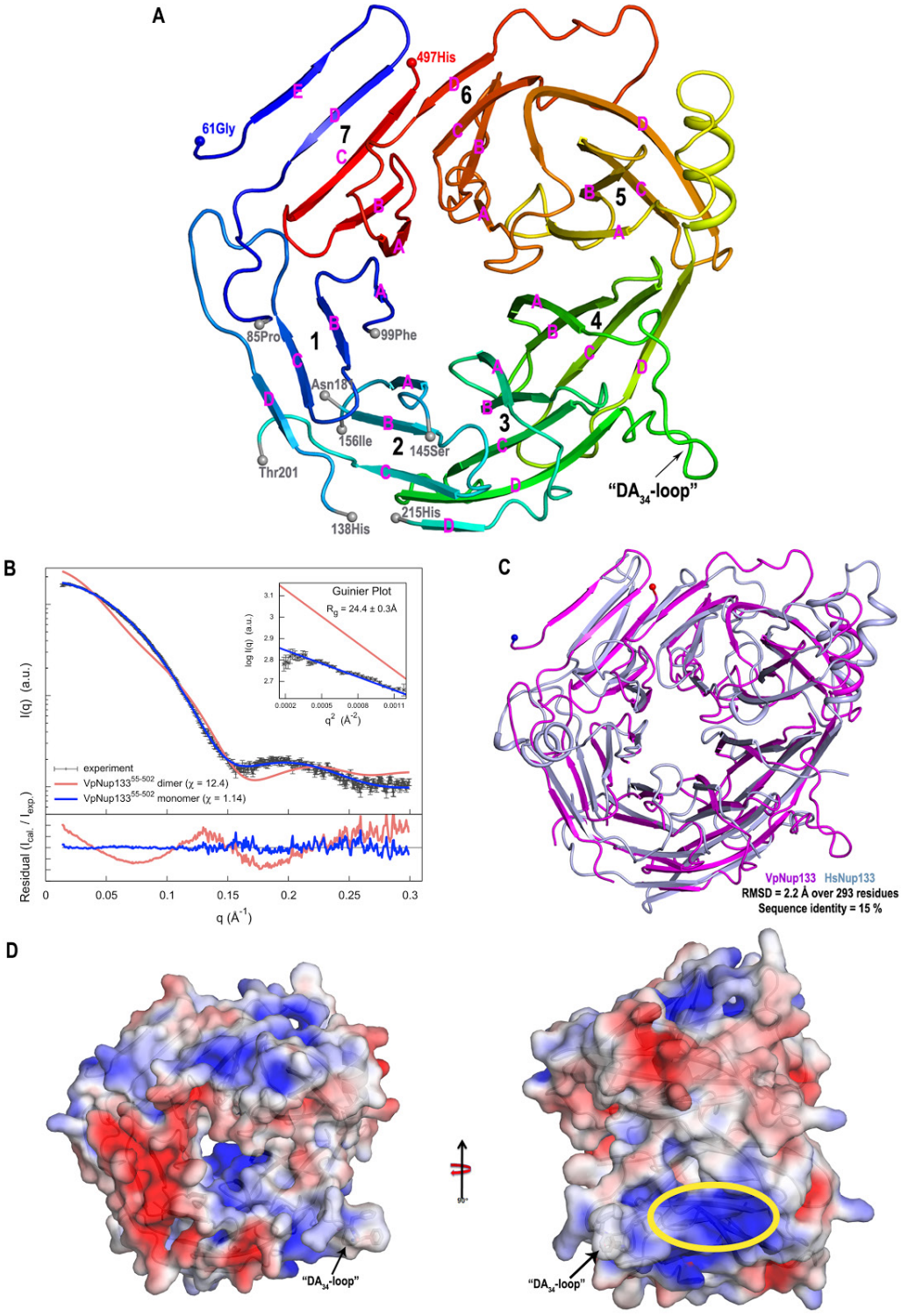

Figure 2. Crystal structure of *Vp*Nup133$^{55-502}$

Suggested location – Results, Crystal structure of *Vp*Nup133$^{55-502}$

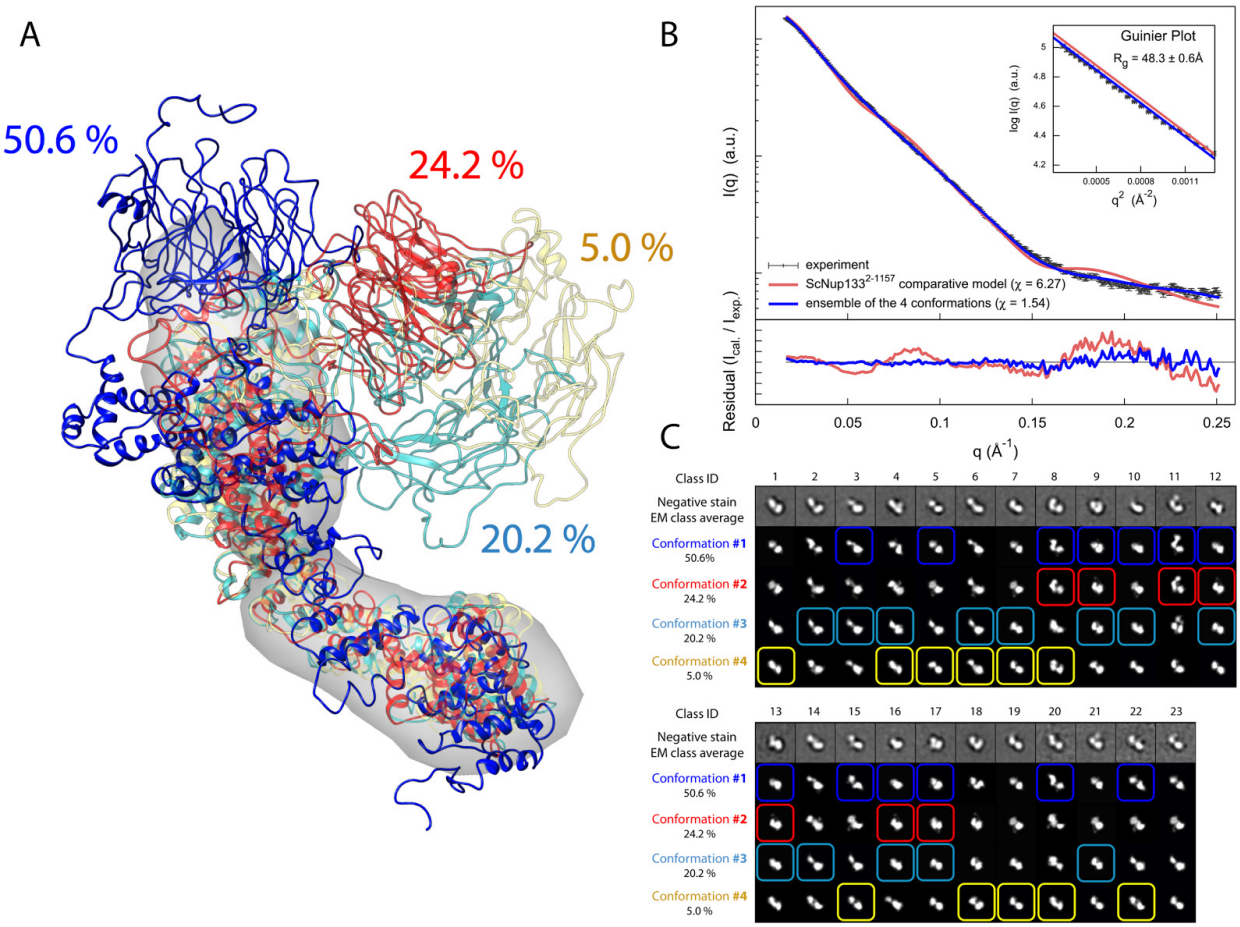

Figure 3. Structure and dynamics of *Sc*Nup133 revealed by integrative modeling approach

Suggested location – Results, The multi-state structural model of ScNup133$^{2-1157}$

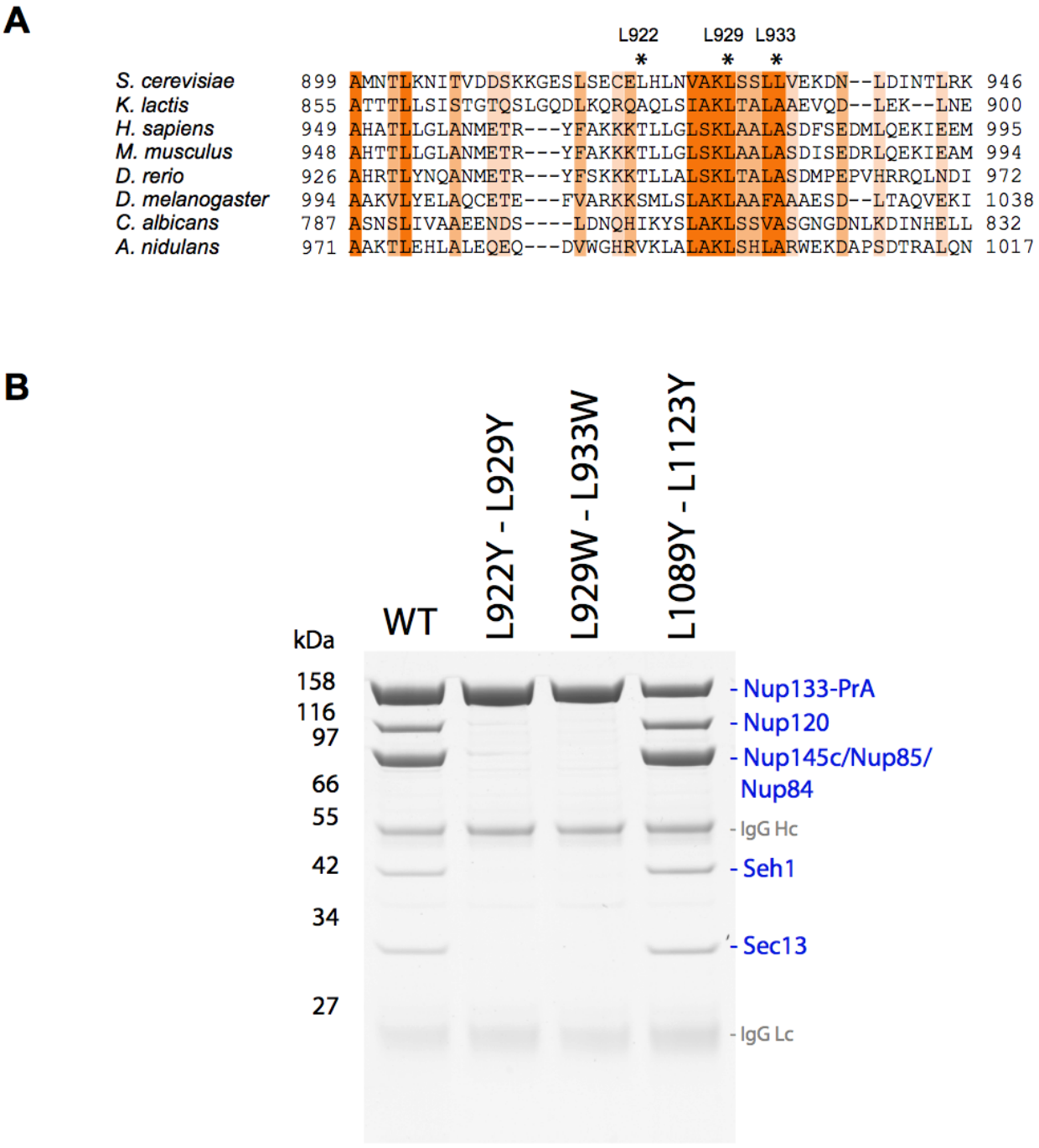

Figure 4. Mutational analysis of the *Sc*Nup133 - *Sc*Nup84 interface

Suggested location – Results, The *Sc*Nup133 - *Sc*Nup84 interface validated by mutational analysis and chemical cross-links

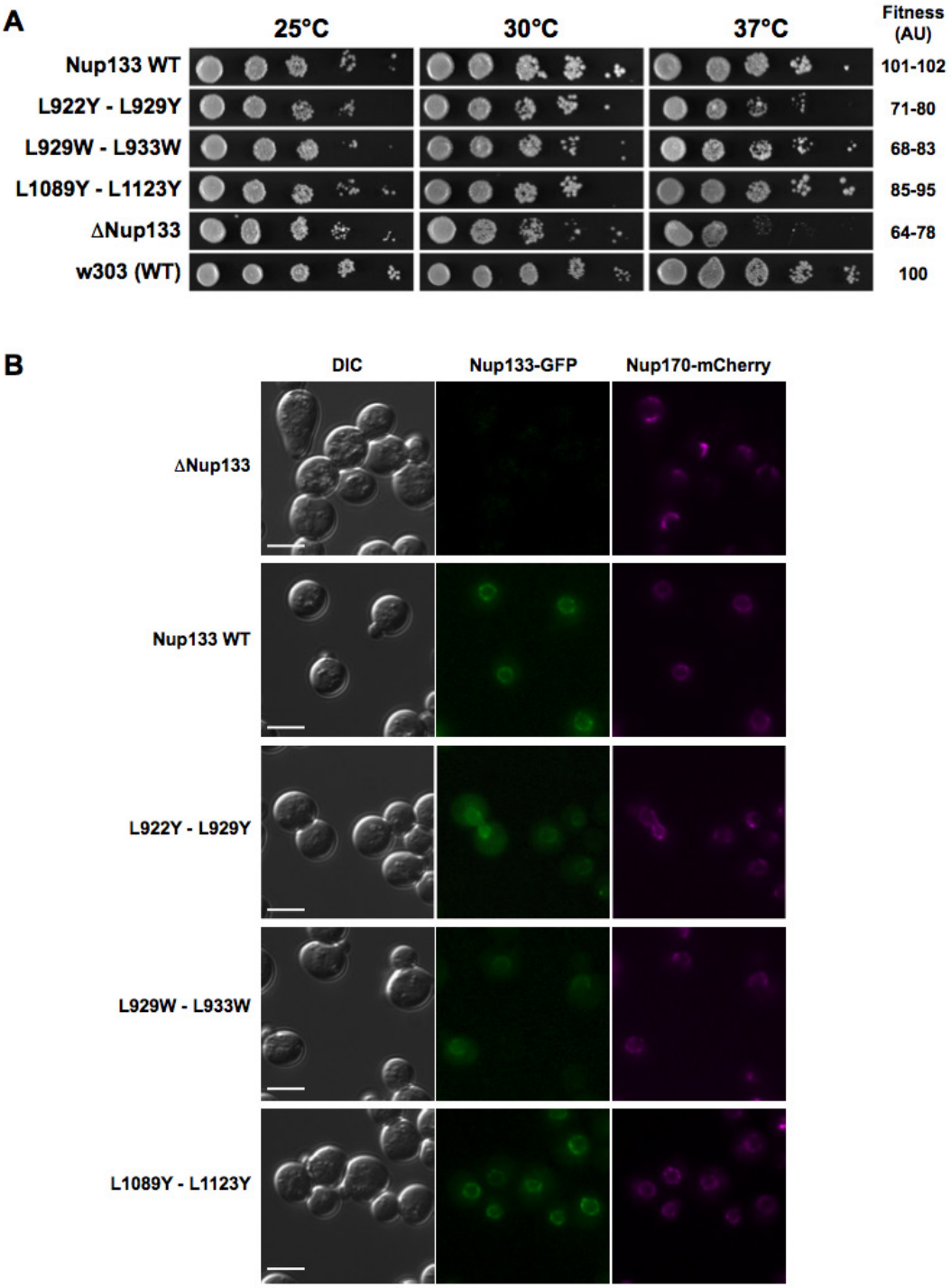

Figure 5. Phenotypic analysis of the *Sc*Nup133 point mutants

Suggested location – Results, The *Sc*Nup133 - *Sc*Nup84 interface validated by mutational analysis and chemical cross-links

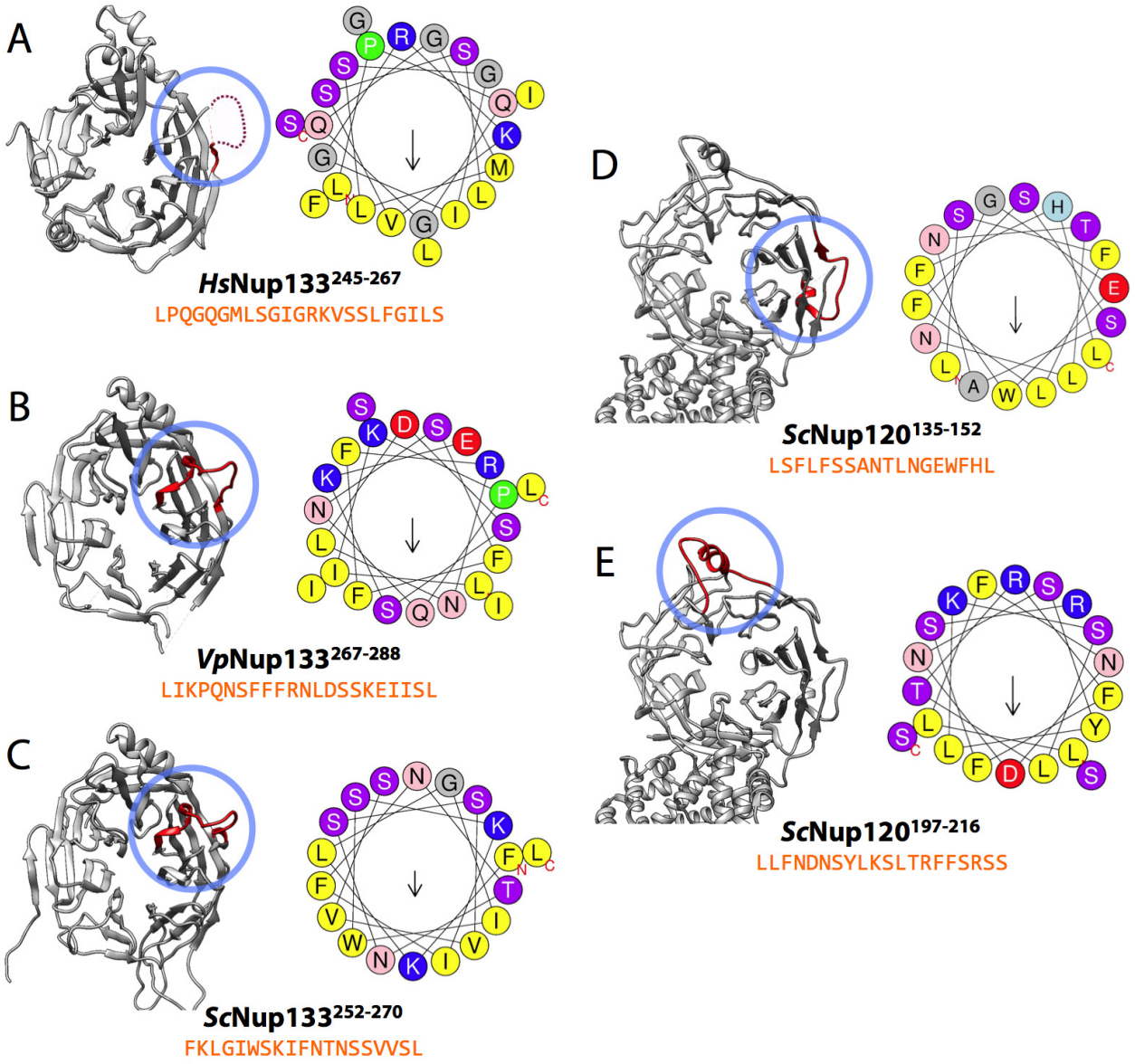

Figure 6. The potential ArfGAP1 lipid packing sensor (ALPS) motifs

Suggested location – Results, The ArfGAP1 lipid packing sensor (ALPS) motifs

Keywords

Nuclear Pore Complex; Nup133; Nup84 complex; Nucleoporin; small angle X-ray scattering (SAXS); electron microscopy (EM); Crosslinking; Mass Spectrometry; ArfGAP1 lipid packing sensor (ALPS) motif; Computational Biology; Integrative Modeling; comparative modeling; Molecular Dynamics; Structural Biology; Protein Structure.