Crystal Structure of Nucleoporin Nic96 Reveals a Novel, Intricate Helical Domain Architecture*\(^{2}\)

Received for publication, July 5, 2007, and in revised form, September 21, 2007 Published, JBC Papers in Press, September 25, 2007, DOI 10.1074/jbc.M705479200

Sandra Jeudy and Thomas U. Schwartz

From the Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

The nuclear pore complex (NPC) is an elaborate protein machine that mediates macromolecular transport across the nuclear envelope in all eukaryotes. The NPC is composed of nucleoporins that assemble in multiple copies around an 8-fold symmetry axis. Homology modeling suggests that most architectural nucleoporins are composed of simple β-propeller and α-helical repeat domains. Here we present the crystal structure of Nic96, the Nup93 homolog in *Saccharomyces cerevisiae*, one of the major components of the NPC. This is the first structure of an α-helical nucleoporin domain. The protein folds into an elongated, mostly α-helical structure. Characteristically, non-canonical architectural features define the Nic96 structure. Sequence conservation among Nup93 homologs across all eukaryotes strongly suggests that the distinct topology is evolutionarily well maintained. We propose that the unique Nic96/Nup93 fold has a conserved function in all eukaryotes.

The hallmark of the eukaryotic cell is a nucleus containing the genetic material enclosed by the nuclear envelope (NE).\(^{2}\) Nuclear compartmentalization requires permanent shuttling of macromolecules across the NE. Nucleocytoplasmic transport is exclusively mediated by nuclear pore complexes (NPCs), large proteinaceous channels spanning the NE (1–3).

The NPC is composed of about 30 proteins, termed nucleoporins (Nups), each of which is present in multiple copies (4, 5). Electron microscopy studies of yeast and frog NPCs show that despite differences in overall size, the basic architecture of the NPC is well conserved between species. In all cases, a channel is formed in the center of the 8-fold rotational symmetrical NPC. At about 6 nm resolution, the best current reconstruction of the NPC is its structure from the slime mold *Dictyostelium discoideum* obtained with cryoelectron tomography (6). Three basic elements stand out as follows: the nuclear basket, the central core, and the cytoplasmic fibrils. About one-third of all nucleoporins contain FG repeat domains, long filamentous peptides bearing characteristic phenylalanine-glycine (FG) repeats. These FG repeats adopt no defined structure but extend into the central channel of the NPC and form a barrier, limiting the diffusion of macromolecules larger than 20–40 kDa, which can only traverse the NPC efficiently if facilitated by adaptor proteins called karyopherins.

Karyopherins can interact with both transport substrates and nucleoporins (7). A concentration gradient of the small G protein Ran enforces the directionality of karyopherin-mediated transport. Ran is highly enriched in its GTP-bound form in the nucleus, and GTP hydrolysis by Ran is directly coupled to the import/export cycle (8). The karyopherin-mediated protein transport cycle is mechanistically well understood in large part because of structures of intermediates provided by x-ray crystallographic studies (9, 10). Other facilitated transport processes such as the nuclear export of RNA and ribosomal subunits are less understood, but their regulation appears to be at least partially different from that governing protein traffic across the NE (1, 11).

To understand nucleocytoplasmic transport in detail, structural information of the NPC itself is required. The complexity and size of the NPC make this a significant endeavor. However, recent studies of the assembly and the organization of the NPC components indicate that the apparent structural complexity of the NPC is reduced by multiple layers of modularity (12). First, there is the limited set of nucleoporins that assemble in multiples of eight (4, 5), reflecting the rotational symmetry apparent from electron micrographs. Second, among the ~30 nucleoporins, only a subset is stably attached to the NPC to form the structural scaffold, whereas others associate transiently (13). Third, recent crystal structures and computational analysis have indicated that most architectural nucleoporins consist of simple domain structures, predominantly β-propellers, α-helical solenoid domains, or combinations of the two (10, 14–17).

The stable architectural core of the NPC includes several nucleoporins that are organized in distinct subcomplexes. The nine proteins of the Nup107–160 subcomplex are symmetrically localized on the cytoplasmic and nucleoplasmic face of the NPC and form two outer rings that sandwich a second scaffold module called the inner ring. This module is built from a subcomplex that contains at least four proteins, Nup205, Nup188, Nup53, and Nup93/Nic96 (18–20). Nup93/Nic96 directly interacts with the heterotrimeric Nup62 subcomplex, which resides at the center of the NPC and consists of FG repeat

\(^{2}\) The abbreviations used are: NE, nuclear envelope; NPC, nuclear pore complex; BisTris propane, 1,3-bis[tris(hydroxymethyl)methylamino]propane; MSA, multiple sequence alignment; TPR, tetratricopeptide; r.m.s.d., root mean square deviation; Nup, nucleoporin; SeMet, selenomethionine; SAD, single-wavelength anomalous dispersion; GFP, green fluorescent protein.
containing nucleoporins, held together by coiled-coil domains (21, 22).

Nup93/Nic96 is essential, as shown in Saccharomyces cerevisiae, Aspergillus nidulans, Caenorhabditis elegans, and Danio rerio (22–25). Quantification of Nup93/Nic96 in rat liver cells and S. cerevisiae suggests it is the most abundant nucleoporin with 32–48 or more copies per NPC (4, 5, 26). It alone contributes more than 10% of the mass of the NPC. In nuclear reconstitution experiments using Nup93-immunodepleted Xenopus laevis egg extracts, nuclear pore assembly is defective, and the size-exclusion barrier for transport substrates is lowered (18, 24). Thus, Nup93/Nic96 acts in the assembly and homeostasis of the NPC. In addition, it has been linked to chromatin organization and gene regulation. Chromatin immunoprecipitation with Nic96 suggests it, among other nucleoporins, recruits certain actively transcribed genes to the NPC (27) in a process referred to as “gene gating” (28).

Nup93/Nic96 is composed of two predicted structural domains, an N-terminal coiled-coil domain, of about 150 residues, and a C-terminal helical domain, of about 600 residues. Direct interaction between the coiled-coil domain and members of the Nup62 subcomplex was shown in yeast (29). The C-terminal region is suggested to interact with the Nup205 complex (20). Thus, Nup93/Nic96 tethers the Nup62 and the Nup205 subcomplexes.

To understand Nup93/Nic96 on a molecular level and to be able to dissect its various functions, we have solved the crystal structure of the yeast homolog of Nup93, Nic96, lacking the N-terminal coiled-coil region. This structure reveals that Nic96 folds into an elongated, mostly helical domain of distinct topology. Structure-guided primary sequence alignment of Nup93 homologs spanning the phylogenetic spectrum of eukaryotes indicates that its architecture and specific surface features are well conserved. Fluorescence microscopy shows that the helical domain is recruited to the NPC independently of the N-terminal coiled coils. The structure will now allow targeted experiments to address the function of Nup93/Nic96.

**EXPERIMENTAL PROCEDURES**

**Protein Expression**—The coding region for the noncoiled-coil portion of Nic96 (residues 186–839) was amplified from genomic DNA of S. cerevisiae and cloned into the pET28a bacterial expression vector (Novagen) to yield an N-terminally His-tagged protein. The resulting vector was transformed into a bacterial expression vector (Novagen) to yield an N-terminally His-tagged portion of Nic96 (residues 186–839) was amplified from S. cerevisiae grown in LB medium containing 0.5% (w/v) glucose, 25 μg/ml chloramphenicol, and 34 μg/ml kanamycin, and grown at 23 °C. Protein purification was carried out as for the native protein.

**Protein Purification**—The clarified lysate was incubated with nickel-charged HisSelect resin (Sigma). After binding, the resin was washed three times with 10 bed volumes of lysis buffer, loaded onto a disposable column (Pierce), and washed with 5 bed volumes of washing buffer (10 mM Tris/HCl, pH 8.0, 400 mM NaCl, 10 mM imidazole, 5 mM β-mercaptoethanol). The protein was eluted in 5 bed volumes of elution buffer (10 mM Tris/HCl, pH 8.0, 250 mM NaCl, 150 mM imidazole, 5 mM β-mercaptoethanol). The eluate was dialyzed overnight at 4 °C against 10 mM Tris/HCl, pH 8.5, 250 mM NaCl, 0.1 mM EDTA, 1 mM dithiothreitol. The N-terminal His tag was removed by proteolytic cleavage. The protein was further purified twice via gel filtration on a Superdex 200 column (GE Healthcare) and concentrated to 10 mg/ml using a centrifugal concentrator.

**Protein Crystalization**—The protein was crystallized at 16 °C with the hanging drop vapor diffusion method. Initial crystals grew in aggregates in drops mixed from 1.5 μl of protein solution with 1.5 μl of reservoir solution containing 0.1 M Bis-Tris propane, pH 8.0–8.2, 17–20% (w/v) PEG 3350, and 0.2 M KSCN. Single crystals were obtained after adding 1 mM cetyltrimethylammonium bromide to the drop. Crystals were quick-soaked in reservoir solutions containing 10–15% glycerol and flash-frozen in liquid nitrogen. Selenomethionine-labeled protein crystallized in the same conditions.

**Data Collection, Structure Determination, and Refinement**—A SAD data set to 2.6 Å of a SeMet derivative and a native data set to 2.5 Å were collected at 100 K at beamline 24-IDC at the Advanced Photon Source (Argonne, IL). Data were processed with the HKL2000 package (31). The statistics of the data collection are listed in Table 1. Twenty-two selenium sites (out of 24 possible for two molecules per asymmetric unit) were found with the program SHELXD and were used for phasing with SHARP (32). The resulting solvent-flattened electron density map was used to build the initial SeMet model with Coot (33). Molecular replacement with the SeMet model was used to find the native structure. The structure was first refined against the native data using CNS (34). Final refinement, including TLS refinement, was performed with the PHENIX suite (35). The refined structure of Nic96 consists of residues 200–835. Residues 361–365, 375–404, 445–454, 516–532, and 748–752 were not defined in the Fo–Fc electron density map and are absent from the final model. The quality of the final model was validated using Molprobity (36). 96.2% of the residues are in the most favored regions of the Ramachandran, and 3.8% are in the additionally allowed regions. Refinement statistics are listed in Table 1.

**Structure Analysis**—Sequence alignments were performed using MUSCLE. Multiple sequence alignments (MSAs) were edited with JalView (37). Homology modeling was performed with Modeler (38). The sequences of homologs to be modeled were aligned with the Nic96 structure omitting poorly ordered loop regions. Electrostatic surface potentials were calculated with APBS (39) and displayed with PyMOL (DeLano Scientific, Palo Alto, CA).

**GFP-tagged Nic96 Constructs and Fluorescence Microscopy**—Wild-type S. cerevisiae W303 diploid strain was used in this study. General yeast manipulations were conducted by stand-
and methods with transformation by the lithium acetate method. Cells were grown at 30 °C in YPD medium. The GFP/KanMX4 cassette was amplified from the pYM27 plasmid as described (40) and was fused in-frame either with full-length Nic96 to generate Nic96-GFP or with residues 1–185 to construct Nic96ΔCter-GFP. To generate Nic96ΔNter-GFP, residues 200–835 of Nic96 were amplified in fusion with the GFP/KanMX4 cassette with recombination sites flanking the full-length gene. After transformation, cells were selected on YPD-KanMX4 cassette with recombination sites flanking the full-length gene. After transformation, cells were selected on YPD-G418 (G418/geneticin: Invitrogen) medium (200 mg G418/l). Correct chromosomal integration was confirmed by PCR and fluorescence microscopy.

The cells, grown to log phase at 30 °C, were immobilized on coverslips with a solution of 1% poly-L-lysine. Localization of Correct chromosomal integration was confirmed by PCR and G418 (G418/geneticin: Invitrogen) medium (200 mg G418/l). Length gene. After transformation, cells were selected on YPD

TABLE 1
Data collection and refinement statistics

| Data set          | Native | SeMet |
|-------------------|--------|-------|
| **Data collection** |        |       |
| Space group       | P2₁    | P2₁   |
| Cell dimensions   | 87.1 Å, 82.5 Å, 113.0 Å | 87.8 Å, 112.6 Å, 86.7 Å |
| a, b, c           | 90.0', 91.7', 90.0' | 90.0', 116.5', 90.0' |
| a, b, γ           | 55,097 | 87,727 |
| No. of unique reflections | 50 to 2.5 (2.59 to 2.5) | 50 to 2.6 (2.69 to 2.6) |
| Resolution        | 8.3% (38.1%) | 10.0% (47.7%) |
| Redundancy        | 99.5% (99.0%) | 98.9% (94.0%) |
| Completeness      | 4.8 (4.8) | 3.6 (2.9) |
| I/σ(I)            | 19.9 (3.7) | 20.1 (2.6) |

| Refinement        |        |
|-------------------|--------|
| Resolution        | 20 to 2.5 Å |
| No. of unique reflections | 54,695 |
| R<sub>work</sub> / R<sub>free</sub> | 23.3/28.5% |
| No. atoms         | 9036   |
| Protein           | 123    |
| Water             | 62.2 Å² |
| r.m.s.d.          | 0.005  |
| Bond lengths (Å)  | 0.712  |
| B-factors         | 55.5 Å² |

<sup>a</sup> R<sub>work</sub> = Σ|Fo| - |Fc| / Σ|Fo|, where I is the intensity of the i<sup>th</sup> observation and (I) is the mean intensity of the reflection.
<sup>b</sup> R<sub>free</sub> = Σ|Fo| - |Fc| / Σ|Fo|.
<sup>c</sup> R<sub>free</sub> = R value for a randomly selected subset (5%) of the data that were not used for minimization of the crystallographic residual.

RESULTS

Overall Structure—Nic96 (residues 185–839) from *S. cerevisiae* was recombinantly expressed in *E. coli*, purified to homogeneity, and crystallized. The structure was solved by single-wavelength anomalous dispersion (SAD), using selenomethionine-labeled protein. The experimental SAD electron density map was of excellent quality, allowing for building of the majority of the model. The final structure of the native protein was refined to 2.5 Å resolution. Data collection and refinement statistics are summarized in Table 1.

Selenomethionine and native crystals were both monoclinic, but with different cell parameters. Both crystal forms have two molecules in the asymmetric unit with the same relative orientation to one another. The contacts between the two molecules in the asymmetric unit and between crystallographically related molecules are not extensive and do not suggest dimerization in vivo. Accordingly, size-exclusion chromatography indicates that Nic96 is a monomer in solution (data not shown).

The structure is shown in Fig. 1. Overall, it has an elongated cuboid shape with dimensions of ~140 × 45 × 45 Å. The protein folds into one continuous domain composed of 30 α-helices and one short β-sheet of two strands. Starting at the N terminus, located in the middle of the domain, the first 70 residues form three helices that propagate in a zig-zag pattern to one end of the molecule. The next 140 residues fold into an irregular ensemble of 8 helices (α4–α11), shaping one end of the domain. Helix α9 is poorly ordered suggesting that it is only loosely attached to the body of the domain. It could be partially modeled (Fig. 1) from the electron density map obtained from the selenomethionine data set, but it is absent in the final model of the native protein. The following 7 helices (α12–α18, covering about 140 residues) form one of two tetratricopeptide (TPR)-like segments of the domain (see below) and trace back toward the C-terminal end of the molecule, pairing with helices α1–α3. The following helix α19 is interrupted by a short β-hairpin of 12 residues. Within helices α20 to α30, 4 antiparallel helical pairs form the second TPR-like subdomain that makes up the C-terminal end of the domain. Several loops (connecting helices α7–8, α11–12, α15–16, and α26–27) are not modeled completely because they are poorly resolved in the electron density map, presumably because of their inherent flexibility. Both crystal forms we obtained contain two Nic96 molecules in the asymmetric unit, thus we have a total of four protein copies in different packing environments. A superposition of the respective models shows no significant positional differences except for side chains directly involved in crystal contacts (r.m.s.d. comparing equivalent C-α positions below 0.6 Å for all mutual superpositions). The overall fold is identical in all copies suggesting that the domain is rather rigid.

Comparison with α-Helical Solenoid Proteins—Solenoid proteins are defined by their buildup from repeating structural units, arranged in such a way that the polypeptide chain forms a continuous superhelix (41). The superhelix can be described by three basic parameters; overall curvature and handedness, as well as twist between the consecutive repeating elements. Furthermore, N and C termini of solenoids are typically localized at opposite ends of the domain. α-Helical solenoids form the most populated subclass, where the repeating unit is made up from two or three helices, prominent examples being HEAT, Armadillo, and TPR repeat proteins. Karyopherins, the mediators of transport through the NPC, are composed of Armadillo or HEAT repeats that form regular elongated or spiral-shaped superhelices (10). Structure predictions based on threading algorithms suggest that many architectural nucleoporins, including Nic96, form α-helical repeat structures similar to karyopherins (17). Our structure does not confirm such predictions. Nic96 does not share the hallmark features of solenoid proteins. N and C termini are not at opposite ends, and an overall repeat pattern is not apparent. Without a repeat pattern, handedness and twist are irrelevant parameters. Furthermore,
the molecule has no overall curvature; it adopts a straight cuboid shape not characteristic of \( \alpha \)-helical solenoids. We performed three-dimensional structure comparison using the Dali and VAST programs. Nic96 has substructures that resemble known proteins, but as a whole, its domain architecture is novel. The substructures expand over two separate regions (Fig. 2). Both substructure modules M1 and M2 have homology to TPR-like domains. The first module M1, comprising helices \( \alpha_{12} - \alpha_{18} \), is most similar to a region within the HAT-C domain (half a TPR) of CstF-77, one of the multiple factors required for polyadenylation and 3' -end cleavage of mammalian pre-mRNAs (r.m.s.d. of 3.1 Å over 100 C-\( \alpha \) atoms; orange in Fig. 2). SMG7, a protein involved in nonsense-mediated mRNA decay is similarly close (data not shown). The second module M2 includes an interrupted portion of the C-terminal region of helices \( \alpha_{20} - \alpha_{30} \) and resembles the prokaryotic TPR protein NlpI (r.m.s.d. 3.7 Å over 112 C-\( \alpha \) atoms; yellow in Fig. 2) and similarly the Bro1 domain. The similarity with the closest matches found in three-dimensional alignments of both substructures appears to be mainly architectural, not functional, as the sequence conservation between Nic96 and its closest structural homologs is very low (below 10%).

M1 contains three TPR-like helical repeats, \( \alpha_{12} - 13, \alpha_{14} - 15 \) and \( \alpha_{16} - 17 \), capped by \( \alpha_{11} \) and \( \alpha_{18} \) on either side. In most TPR repeats, the two helices have approximately equal lengths and are connected via a sharp turn. M1 is more irregular in that the length of the helices varies from 8 to 15 residues. The three repeats are slightly curved to form a shallow concave surface.

Module M2 has four TPR-like repeats, composed of helix pairs \( \alpha_{20} - 21, \alpha_{22} - 23, \alpha_{25} - 26, \) and \( \alpha_{28} - 29 \), respectively. Two short helices, \( \alpha_{24} \) and \( \alpha_{27} \), are interspersed in the repeat pattern. The C-terminal helix of Nic96, \( \alpha_{30} \), packs against the helix hairpin \( \alpha_{28} - 29 \) and forms a hydrophilic cap as seen in other TPR-like structures. Compared with M1, the helices in M2 are longer, and their relative orientation compared with TPR-like domains is more irregular. Most prominently, the typical shallow concave groove formed by TPR-like domains, which is often used for functionally important protein-protein interactions, is closed in Nic96. Helices \( \alpha_{22} \) and \( \alpha_{28} \) roll over helix \( \alpha_{25} \) and directly interact with each other. This way, the concave surface curvature is lost and the module adopts a more compact shape.

**Sequence Comparison Shows Strong Conservation of the Nic96 Architecture**—Apart from the similarities with TPR-like domains of the modules mentioned above, Nic96 has several specific architectural features, two being particularly noteworthy. On one end of the elongated molecule, residues 269 – 444 form an irregular but compact bundle consisting of helices

**FIGURE 1. Stereo view of the overall structure of Nic96 (residues 200 – 835).** The color gradient starts at the N terminus (pale blue) and shifts to the C terminus (dark blue) of the molecule. Helix \( \alpha_{9} \) was poorly defined in the 2\( F_{o} - F_{c} \) map from the native data set and is absent in the deposited model. It was modeled from the selenomethionine data.

**FIGURE 2. Partial similarity to TPR-like helical repeat domains.** Superposition of the closest structural homologs, the HAT-C domain of Cst-77 (Protein Data Bank code 200E, residues 31–154, orange) and the E. coli NlpI protein (Protein Data Bank code 1XNF, residues 58–171, yellow) on Nic96. The similarity spans two separate modules, M1 and M2, within Nic96.
α4–α11 connected by several extended, poorly ordered loops. Helix α6 (residues 333–343) is at the center of the bundle and is encased by helices α4, α5, α7, α10, and α11 (Fig. 3A). Because α6 is completely shielded from the solvent, it is mostly hydrophobic, a feature that was noticed when Nic96 was first described (22). No structural homologs of the helical bundle could be found using either VAST or Dali.

A second noticeable feature of Nic96 is helix α19, which connects the two TPR-like modules. Helix α19 is interrupted by a 12-residue insertion forming a short β-hairpin. To interrupt the helix, two glycines (Gly-591 and Gly-602) are positioned at the beginning and the end of the insertion (Fig. 3B). These glycines adopt backbone conformations unfavorable for other amino acids and thus specifically enable this unusual feature.

We performed an MSA using 15 sequences from Nup93 homologs spanning the entire eukaryotic phylogenetic spectrum (42). This MSA shows a remarkable conservation of all the residues that are pivotal for the specific topological features of Nic96 (supplemental Fig. 1). Throughout the 635 residues of this structure, 23 positions are strictly conserved. Of these, 20 are structurally important. Helix α6 is strongly conserved in its hydrophobic character. In addition, the strictly conserved Arg-342 at the C-terminal end of α6 forms a structurally important salt bridge to Asp-434 (Fig. 3A). Helices α4 and α11 are clamped together and buttressed against α6 via invariant Trp-437 and Glu-279. Within the first TPR-like module M1, a hydrogen-bonding network is formed between the side chains of His-505, Tyr-541, and Tyr-556 and a water molecule, all being completely buried in the solvent-inaccessible core of the protein (Fig. 3C). These three residues are strictly conserved. The β-hairpin insertion in helix α19 is also conserved, yet shortened from 12 to 9 residues in plants, and is extended in the fish Tetraodon nigroviridis. Gly-591 and Gly-602, which enable the interruption of helix α19 for the reasons mentioned above, are both strictly conserved. Taken together, the pattern of sequence conservation observed across the entire eukaryotic tree of life strongly suggests that the particular architecture of Nup93/Nic96 is well maintained.

We also performed a separate alignment of Nup93/Nic96 sequences from an ensemble of maximally diverged budding yeasts (supplemental Fig. 2). Considering the short life span of its members, this class of organisms is the result of an enormous number of successive generations since the separation from filamentous fungi. As a result, extant budding yeasts have genomes as divergent as those of the entire phylum of chordeutes, even though their life-style is very similar (43). The relative comparison of the two gene trees provides two additional measures. First, it allows for the detection of protein features that distinguish budding yeasts as a class from other organisms. Second, by inspecting the budding yeast gene tree alone, the resistance of a protein against mutation while maintaining a specific function can be understood. For Nup93/Nic96, we find a stronger overall sequence conservation across budding yeasts, as there is no apparent property, like an insertion or deletion in the sequence, that would distinguish budding yeast sequences from those of other eukaryotes.

Surface Conservation Reveals Two Likely Interaction Sites—We mapped the sequence conservation to the surface of Nic96 to detect possibly evolutionary maintained regions (Fig. 4A). Analysis of the conservation pattern revealed that overall, there are two highly conserved surface patches. The same two patches, yet more pronounced, emerge when conservation among six budding yeasts was studied (not shown). These two regions of surface conservation are marked P1 and P2 in Fig. 4A. Patch P1 corresponds to the concave surface of module M1, extended to include neighboring residues. From helix α12 it involves the invariant Gln-463, the adjacent exposed face of Ile-270 of helix α4 and one edge of Tyr-488, positioned in a tight turn between helices α13 and α14. We note that except for Gln-463 the patch is mostly hydrophobic (Fig. 5A).

The second patch P2 corresponds to the extreme C terminus of the protein and includes residues within helices α29, α30, and the turn connecting them. Met-814, Ile-815, Gly-818, Met-819, Tyr-822, Tyr-829, and Ile-833 form a contiguous hydrophobic patch on the surface of the molecule (Fig. 5B). The connecting turn is short, and its residues are well conserved, arguing that the surface architecture is evolutionarily maintained. Strikingly, there is almost complete conservation within this region among budding yeasts (supplemental Fig. 2). Combined with the frequent involvement of TPR-like domains in protein-protein interactions, it is likely that Nic96 connects with one of its potential binding partners via this region.

Electrostatic Surface—The electrostatic surface of Nic96 (Fig. 4B) illustrates that positive and negative charges are not distributed evenly but mostly cluster in distinct regions. The region corresponding to the irregular helix α4–11 bundle is positively
charged, whereas the opposing end and large areas in the central part are negatively charged. Based on our structural analysis, we were confident that threading eukaryotic Nup93 homologs based on the Nic96 structure should result in adequate models. Comparative homology modeling was carried out with Modeler (38), and surface electrostatics were calculated for Nup93 from different species. The distribution seen in *S. cerevisiae* is conserved in many budding yeast species we examined; however, we cannot readily see a similar pattern in other eukaryotes (data not shown). More detailed investigations are currently being pursued. It is possible that the charge distribution has a specific role in budding yeast that might explain features unique to the NPC in these species.

**Comparison with Characterized Mutants**—The function of Nup93/Nic96 has been probed in *S. cerevisiae* using a collection of mutants. Two temperature-sensitive mutants, nic96-1 and nic96-2, are well characterized (20, 29, 44). At the permissive temperature these mutants behave fairly normally. However, after shifting to elevated temperatures both, but particularly nic96-1, stall cell growth after 2–3 cell divisions, and the nuclear envelope is depleted of NPCs. Presumably *de novo* NPC synthesis is blocked by the mutants at the nonpermissive temperature (20). In nic96-1 two point mutations are present as follows: Leu-260 in the core of helix α9 is mutated to proline and Pro-332 in a well ordered turn at the N terminus of helix α9 is mutated to leucine. In nic96-2 Trp-334 of the hydrophobic helix α9 is mutated to arginine. Based on the structure, both mutants should severely affect the structural integrity of the protein, a defect likely exacerbated at ele-
Structure of Nic96

FIGURE 6. Nic96-GFP localization. The Nic96-GFP, Nic96ΔNter-GFP, and Nic96ΔCter-GFP proteins were expressed in yeast. The localization of the proteins was visualized by direct fluorescence microscopy. The N- and C-terminal domains of Nic96 integrate into the NPC independently.

vated temperature. The structure does not support the notion that the domain is split as suggested by the C-terminal deletion mutant Nic96Δ532–839 (29). In our studies, the recombinantly expressed mutant protein is poorly behaved and largely insoluble (data not shown). Nevertheless, the functional data reported for Nic96Δ532–839, notably synthetic lethality with Nup188, is consistent with a direct interaction between both proteins, potentially via the conserved surface patch P2 discussed above.

The C-terminal Domain of Nic96 Integrates into the NPC—Nup93/Nic96 is often used as an NPC marker, because its incorporation into the pore is very stable and the protein occurs in a high copy number (13, 21, 45). Additionally, C-terminal fusion of GFP has no measurable effect on NPC function in the commonly used assays. The direct interaction of the N-terminal coiled-coil region of Nic96 with the Nup62 complex of the NPC is well established (22, 29), whereas less is known about the interaction with the additional binding proteins found in other assays (18, 20, 24, 46, 47). We tested full-length Nic96 as well as two truncation mutants for their integration into the NPC in S. cerevisiae. We constructed yeast strains containing Nic96-GFP, Nic96ΔNter-GFP, or Nic96ΔCter-GFP. Nic96ΔNter contains our crystal construct (residues 200–839), and Nic96ΔCter the remainder of the protein. Nic96 is essential for cell growth (22) so we used the W303 diploid strain to make these constructs by substituting one gene copy with the sequence coding for the GFP-labeled fusion protein. Wild-type W303 diploid strain was used as a negative control and full-length Nic96-GFP as a positive control. All GFP-labeled constructs (Nic96-GFP, Nic96ΔNter-GFP, and Nic96ΔCter-GFP) resulted in typical nuclear rim staining, indicating integration into the NPC (Fig. 6). Thus, both the N-terminal and C-terminal domains of Nic96 integrate independently into the NPC.

DISCUSSION

Much has been learned about the function of the NPC as a nucleocytoplasmic transport machine over the past decades. As a result, we now have a general understanding of protein traffic across the nuclear envelope (3, 9). The principles guiding export both of RNA transcripts and of assembled ribosomal subunits are also becoming clear (2, 11, 49). On the other hand, the specific roles that the NPC plays in nuclear organization and the regulation of gene expression are just emerging (50). Progress toward the structure of the NPC, based both on electron microscopy and on tomography of the entire assembly, has been dramatic (6, 51). However, understanding of a complex machine like the NPC will remain fragmentary without knowing its structure in atomic detail. Although a high resolution structure of the assembled NPC is out of reach with today’s technology, the modular architecture of the NPC suggests a possible solution, the overall structure might be solved by x-ray crystallography as a series of overlapping substructures (12).

Toward this end, the crystal structure of Nup93/Nic96, a component of the stable central core of the NPC, has been solved at 2.5 Å resolution. Considered both by mass and by copy number, Nup93/Nic96 is one of the most substantial constituents of the NPC. Without Nup93/Nic96, the cell cannot assemble the NPC. This protein also provides the only known link between two major NPC subcomplexes, the Nup62 and the Nup205 subcomplexes. Knowing its structure therefore aids in understanding how biochemically separable structures within the NPC link to one another, by allowing targeted modification of this individual NPC component.

Based on secondary structure prediction methods paired with limited proteolysis, it was suggested that many of the architectural nucleoporins, including Nup93/Nic96, adopt a simple α-solenoid fold (17). Assuming such simplicity further implies common ancestry and redundancy of nucleoporins. The experimental structure of Nic96 presented here reveals the limitations of the predictive approach. In contrast to regular α-helical repeat domains, Nic96 has a much more complicated architecture that does not share the typical hallmarks of α-helical solenoid proteins (41). Specifically, the formation of a superhelical structure with specific handedness and curvature is not observed. Analogy to α-helical solenoid proteins is limited to the extent that Nic96 has two regions, M1 and M2, that are similar to TPR-like domains in that they contain 3 and 4 TPR-like hairpins, respectively. These regions integrate into the entire domain and serve as architectural modules, but they do not define the overall character of Nic96 in a way that would permit classifying it as an α-solenoid protein. The eponymous 34 residues of canonical TPRs (52) are not found in Nic96. Instead, the helix pairs and the connecting turns vary in length. The commonality between the Nic96 modules M1 and M2 and TPR repeats is reduced to the relative orientation of the hairpin helices to one another and their stacking pattern. The use of substructures built from short TPR repeats integrated into larger domains with an nonrepeat type overall structure is in fact quite common (53).

In contrast, the helical repeats found in karyopherins (Armadillo-type in karyopherin-α and HEAT-type in karyopherin-β) typically occur in extended stacks of 10 or more repeating units and are not combined with other major structural elements. Thus, for karyopherins the overall architecture is defined by the repeat character (10).

One intriguing question is the evolutionary origin of the NPC, because it is one of the few large protein complexes exclusively confined to eukaryotes. Comparative genomics suggests...
that most eukaryote-specific protein structures are in fact α-helical, taking stable, pre-existing helical bundles and crafting new functions onto them by adding further helical segments (54). The structure of Nic96 appears to be built on this principle. The TPR-like modules might have served as prokaryotic precursors onto which the remainder of the domain was added to generate the architecture of extant Nic96. Evaluating the Nic96 structure in comparison to the whole phylogenetic spectrum of Nup93/Nic96 sequences reveals a remarkable correlation between architecturally important residues and evolutionary conservation. 20 of 23 invariant residues have an obvious structural role, strongly suggesting that the Nic96 structure presented here is representative of Nup93/Nic96 of all extant eukaryotes. Because structure defines function, this suggests that the unique Nup93/Nic96 fold has a conserved function common to all eukaryotes.

It is worth mentioning that the sequence conservation of the β-propeller domains of Nup133 and Nup214/yNup159, the other available high resolution structures of core nucleoporins, is radically different from Nup93/Nic96 (14–16). In these β-propeller domains there are specific surface loops that are particularly well conserved, and in both cases the functional significance has been experimentally confirmed (16, 55). The strict absence of sequence-conserved surface loops in Nup93/Nic96 indicates that its function likely consists of the formation of stable protein–protein interactions important for the assembly of a basic, universally conserved NPC core. This function is probably specific to Nup93/Nic96 and cannot be substituted for by other nucleoporins, because not even Nup93/Nic96 homologs can complement each other (18). This is in contrast to the fiber-like FG repeats that line the central channel of the NPC and emanate from a set of different nucleoporins. These structures are apparently redundant, as shown by extensive deletion studies in yeast (48, 56).

This structure provides a basis for focused studies directed at elucidating the role of Nup93/Nic96 in the assembly of the NPC. The interaction with the candidate binding partners can now be examined on a structural basis. Knowing the intricate domain architecture of Nic96 enables us for the first time to generate sensible probes to specifically target individual interaction surfaces on the protein and thus obtain a more detailed understanding of the NPC. The Nic96 structure, the first major α-helical nucleoporin domain known in atomic detail, indicates that the modularity of the NPC is not as simple as anticipated. It remains to be seen whether the individuality of the Nic96 structure is more the rule than the exception regarding α-helical nucleoporins.

Acknowledgments—We thank members of the Schwartz laboratory for critical reading of the manuscript, Matthieu Legendre for technical help on computer scripts, and Chia-Yung Wu and Robert Shivers for technical help on microscopy. We also thank Kanagalaghatta Rajashankar and the staff of NE-CAT beamline 24IDC at Argonne National Laboratory for excellent assistance with data collection.

REFERENCES
1. Tran, E. J., and Wente, S. R. (2006) Cell 125, 1041–1053
2. Mattaj, I. W., and Englemier, L. (1998) Annu. Rev. Biochem. 67, 265–306
3. Weis, K. (2002) Curr. Opin. Cell Biol. 14, 328–335
4. Cronshaw, J. M., Krutchinsky, A. N., Zhang, W., Chait, B. T., and Matunis, M. J. (2002) J. Cell Biol. 158, 915–927
5. Rout, M. P., Aitchison, J. D., Suprapto, A., Hjertaas, K., Zhao, Y., and Chait, B. T. (2000) J. Cell Biol. 148, 653–651
6. Beck, M., Lucic, V., Förster, F., Baumeister, W., and Medalia, O. (2007) Nature 449, 611–615
7. Pemberton, L. F., and Paschal, B. M. (2005) Traffic 6, 187–198
8. Fried, H., and Kutay, U. (2003) Cell. Mol. Life Sci. 60, 1659–1688
9. Cook, A., Bono, F., Jinek, M., and Conti, E. (2007) Annu. Rev. Biochem. 76, 647–671
10. Suel, K. E., Cansizoglu, A. E., and Chook, Y. M. (2006) Methods (Duluth) 39, 342–355
11. Köhler, A., and Hurt, E. (2007) Nat. Rev. Mol. Cell Biol. 8, 761–773
12. Schwartz, T. U. (2005) Curr. Opin. Struct. Biol. 15, 221–226
13. Rabut, G., Lenart, P., and Ellenberg, J. (2004) Curr. Opin. Cell Biol. 16, 314–321
14. Berko, J. C., Boehmer, T., Blobel, G., and Schwartz, T. U. (2004) J. Cell Biol. 167, 591–597
15. Napetschnig, J., Blobel, G., and Hoelz, A. (2007) Proc. Natl. Acad. Sci. U. S. A. 104, 1783–1788
16. Weirich, C. S., Erzerberger, J. P., Berger, J. M., and Weis, K. (2004) Mol. Cell 16, 749–760
17. Devos, D., Dokudovskaya, S., Williams, R., Alber, F., Eswar, N., Chait, B. T., Rout, M. P., and Sali, A. (2006) Proc. Natl. Acad. Sci. U. S. A. 103, 2172–2177
18. Grandi, P., Dang, T., Pane, N., Shevchenko, A., Mann, M., Forbes, D., and Hurt, E. (1997) Mol. Biol. Cell 8, 2017–2038
19. Miller, B. R., Powers, M., Park, M., Fischer, W., and Forbes, D. I. (2000) Mol. Biol. Cell 11, 3381–3396
20. Zabel, U., Doye, V., Tekotte, H., Wepf, R., Grandi, P., and Hurt, E. C. (1996) J. Cell Biol. 133, 1141–1152
21. Baillie, S. M., Balduf, C., and Hurt, E. (2001) Mol. Cell. Biol. 21, 7946–7955
22. Grandi, P., Doye, V., and Hurt, E. C. (1993) EMBO J. 12, 3061–3071
23. Allende, M. L., Amsterdam, A., Becker, T., Kawakami, K., Gaiano, N., and Hopkins, N. (1996) Genes Dev. 10, 3141–3155
24. Galy, V., Mattaj, I. W., and Askaiper, P. (2003) Mol. Biol. Cell 14, 5104–5115
25. Osmani, A. H., Davies, I., Liu, H. L., Nile, A., and Osmani, S. A. (2006) Mol. Biol. Cell 17, 4946–4961
26. Ghaemmaghami, S., Huh, W. K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., O’Shea, E. K., and Weissman, J. S. (2003) Nature 425, 737–741
27. Casolari, J. M., Brown, C. R., Komili, S., West, J., Hieronymus, H., and Silver, P. A. (2004) Cell 117, 427–439
28. Blobel, G. (1985) Proc. Natl. Acad. Sci. U. S. A. 82, 8527–8529
29. Grandi, P., Schaich, N., Tekotte, H., and Hurt, E. C. (1995) EMBO J. 14, 76–87
30. Double, S. (1997) Methods Enzymol. 276, 523–530
31. Otwinowski, Z., and Minor, W. (1997) Methods Enzymol. 276, 307–326
32. La Fortelle, E., and Bricogne, G. (1997) Methods Enzymol. 276, 472–494
33. Emsley, P., and Cowtan, K. (2004) Acta Crystallogr. Sect. D Biol. Crystallogr. 60, 2126–2132
34. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Acta Crystallogr. Sect. D Biol. Crystallogr. 54, 905–921
35. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Joerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacco, A., Sauter, N. K., and Terwilliger, T. C. (2002) Acta Crystallogr. Sect. D Biol. Crystallogr. 58, 1948–1954
36. Lovell, S. C., Davis, I. W., Arendall, W. B., III, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003) Proteins 50, 347–350
37. Clamp, M., Cuff, J., Searle, S. M., and Barton, G. J. (2004) Bioinformatics (Oxf) 20, 426–427
38. Fiser, A., and Sali, A. (2003) Methods Enzymol. 374, 461–491
Structure of Nic96

39. Baker, N. A., Sept, D., Joseph, S., Holst, M. I., and McCammon, J. A. (2001) Proc. Natl. Acad. Sci. U. S. A. 98, 10037–10041
40. Janke, C., Magiera, M. M., Rathfelder, N., Taxis, C., Reber, S., Maekawa, H., Moreno-Borchert, A., Doenges, G., Schwob, E., Schiebel, E., and Knop, M. (2004) Yeast 21, 947–962
41. Kobe, B., and Kajava, A. V. (2000) Trends Biochem. Sci. 25, 509–515
42. Ciccarelli, F. D., Doerks, T., von Mering, C., Creevey, C. J., Snel, B., and Bork, P. (2006) Science 311, 1283–1287
43. Dujon, B. (2006) Trends Genet. 22, 375–387
44. Gomez-Ospina, N., Morgan, G., Giddings, T. H., Jr., Kosova, B., Hurt, E., and Winey, M. (2000) J. Struct. Biol. 132, 1–5
45. Ryan, K. J., and Wente, S. R. (2002) BMC Genet. 3, 17
46. Fahrenkrog, B., Hubner, W., Mandinova, A., Pante, N., Keller, W., and Aebi, U. (2000) Mol. Biol. Cell 11, 3885–3896
47. Hawryluk-Gara, L. A., Shibuya, E. K., and Wozniak, R. W. (2005) Mol. Biol. Cell 16, 2382–2394
48. Zeitler, B., and Weis, K. (2004) J. Cell Biol. 167, 583–590
49. Stewart, M. (2007) Mol. Cell 25, 327–330
50. Akhtar, A., and Gasser, S. M. (2007) Nat. Rev. Genet. 8, 507–517
51. Stoffler, D., Feja, B., Fahrenkrog, B., Walz, J., Typke, D., and Aebi, U. (2003) J. Mol. Biol. 328, 119–130
52. D’Andrea, L. D., and Regan, L. (2003) Trends Biochem. Sci. 28, 655–662
53. Lamb, J. R., Tugendreich, S., and Hieter, P. (1995) Trends Biochem. Sci. 20, 257–259
54. Aravind, L., Iyer, L. M., and Koonin, E. V. (2006) Curr. Opin. Struct. Biol. 16, 409 – 419
55. Drin, G., Casella, J. F., Gautier, R., Boehmer, T., Schwartz, T. U., and Antonny, B. (2007) Nat. Struct. Mol. Biol. 14, 138–146
56. Strawn, L. A., Shen, T., Shulga, N., Goldfarb, D. S., and Wente, S. R. (2004) Nat. Cell Biol. 6, 197–206