The Three-Dimensional Structure of the Autoproteolytic, Nuclear Pore-Targeting Domain of the Human Nucleoporin Nup98

Alec E. Hodel,1,4 Mary R. Hodel,1 Eric R. Griffis,2,3 Krista A. Hennig,1 Gary A. Ratner,1,3 Songli Xu,2 and Maureen A. Powers2

1Department of Biochemistry
2Department of Cell Biology
3Graduate Program in Biochemistry, Cell, and Development Biology
Emory University School of Medicine
Atlanta, Georgia 30322

Summary

Nup98 is a component of the nuclear pore that plays its primary role in the export of RNAs. Nup98 is expressed in two forms, derived from alternate mRNA splicing. Both forms are processed into two peptides through autoproteolysis mediated by the C-terminal domain of hNup98. The three-dimensional structure of the C-terminal domain reveals a novel protein fold, and thus a new class of autocatalytic proteases. The structure further reveals that the suggested nucleoporin RNA binding motif is unlikely to bind to RNA. The C terminus also contains sequences that target hNup98 to the nuclear pore complex. Noncovalent interactions between the C-terminal domain and the cleaved peptide tail are visible and suggest a model for cleavage-dependent targeting of hNup98 to the nuclear pore.

Introduction

Communication between the nucleus and cytoplasm of a eukaryotic cell is mediated by the nuclear pore complexes (NPCs), which act as selective molecular gateways (Gorlich and Kutay, 1999; Ryan and Wente, 2000). Through these gateways, RNAs and proteins are exported from and imported into the nucleus. The last decade has seen enormous advances; the full complement of yeast nucleoporins have been identified, and the number of vertebrate nucleoporins is growing rapidly (Rout et al., 2000; Vasu and Forbes, 2001). Three-dimensional reconstructions of the NPC revealed a detailed architecture that is essentially conserved between yeast and vertebrates (Allen et al., 2000).

The next level in understanding the structure of the NPC will require identification of biochemical partners within subcomplexes of the pore, followed by structural characterization of the physical interactions both between partners and between subcomplexes. Identification of subcomplexes of the pore has progressed over recent years (Doye and Hurt, 1997; Vasu and Forbes, 2001). However, while structural information on the soluble nuclear transport factors has accumulated (Conti and Izaurralde, 2001), structural characterization of nucleoporins has been lacking. Toward the long-term goal of understanding the physical interactions within the NPC, we undertook what to our knowledge is the first three-dimensional characterization of a pore-interaction domain from a nucleoporin, that of the human nucleoporin, Nup98.

Human Nup98 is encoded by two alternatively spliced mRNA transcripts (Figure 1A) (Fontoura et al., 1999). One transcript encodes a 98 kDa, 920 amino acid protein termed hNup98 which is posttranslationally cleaved after Phe863 into a 90 kDa N-terminal fragment and an 8 kDa C-terminal fragment. A second transcript encodes a 190 kDa polyprotein which is also posttranslationally cleaved after Phe863 to form the 90 kDa N-terminal protein (hNup98) and the C-terminal hNup96. Of the three related nucleoporins in yeast (scNup100, scNup116, and scNup145), only scNup145 undergoes a similar proteolytic event (Fabre et al., 1994; Wente and Blobel, 1994); scNup145 is cleaved into scNup145N and scNup145C which are homologous to hNup98 and hNup96, respectively.

The C-terminal domains of hNup98, scNup100, scNup116, and scNup145N appear to be involved in targeting these proteins to the NPC (Bailer et al., 1998; Ho et al., 2000). Although in vitro the cleaved N-terminal and C-terminal peptides of hNup98 and scNup145 remain noncovalently associated, the proteolytic cleavage appears to be critical in the localization of the resulting proteins (Fontoura et al., 1999). When cleavage-deficient mutants of either hNup98 or hNup98/hNup96 were visualized in cultured cells, both were localized in the nucleoplasm but not at the NPC, suggesting that cleavage is important in targeting these proteins to the NPC. When the 90 kDa N-terminal cleavage product was expressed alone, it localized to the NPC, suggesting that localization to the NPC was somehow blocked by the uncleaved C-terminal fragment.

Both hNup98 and ScNup145 are processed by autoproteolysis (Rosenblum and Blobel, 1999; Teixeira et al., 1999). Deletion experiments demonstrated that the minimal proteolytic domain of hNup98 consisted of residues 715 to 863 with some specific requirements for residues on the C-terminal fragment, in particular Ser864 adjacent to the cleavage site. The proposed self-cleavage mechanism is similar to those of other autoproteolytic enzymes including the self-splicing protein inteins, Hedgehog, and the NTN family of enzymes (Paulus, 2000). All three classes of enzymes utilize acyl shift chemistry with the side chains of cysteine, threonine, or serine acting as the nucleophile. In addition, the inteins and the NTN enzymes all contain a conserved histidine residue positioned two residues N-terminal from the catalytic C/T/S side chain.

In addition to autoproteolysis, the domain corresponding to residues 715–863 in hNup98 has been ascribed other functions. hNup98 and its three yeast homologs have been implicated in the transport of RNA across the pore (reviewed in Vasu and Forbes, 2001). Consistent with this function, the C-terminal domain of scNup145N was suggested to bind to RNA homopolymers in vitro (Fabre et al., 1994). A conserved octapeptide sequence found in the C-terminal domains of all hNup98 homologs
Figure 1. The In Vivo Screen for Independently Folding Domains in hNup98

(A) Domain architecture of hNup98 and its homologs in S. cerevisiae.
(B) Illustrates the method described in the text based on the screening for soluble, folded domains using a C-terminal fusion of green fluorescent protein (GFP).
(C) A quantitative FACS analysis of the relative fluorescence of each isolate as a function of the molecular weight of the C-terminal hNup98 fragment fused at the N terminus of GFP.
(D) A Western blot analysis of the soluble fraction of bacterial lysates from a number of fluorescent isolates is shown. Protein fragments were detected using an antibody specific for the 6× His tag located at the C terminus of the hNup98-GFP fusion protein. GFP alone runs at ~30 kDa. Full-length hNup98-GFP suffered heavy degradation but still maintained a minute amount of protein at ~130 kDa. The lanes corresponding to the most fluorescent isolates are labeled by the number of the most N-terminal residue that appeared in the hNup98 fragment-GFP fusion. These included fragments that started with residues 676, 686, 646, and 673.
(E) The expression constructs for crystallization targets derived from three of these isolates are schematically illustrated. The crystal structure presented here is of the 676-Tail-His fragment whose name is underlined.

(residues 753–760 in hNup98) weakly resembles one half of the ribonucleoprotein (RNP) motif found in a number of RNA binding proteins. The conservation of this octapeptide and the homopolymer binding activity of the domain led to the term nucleoporin RNA binding motif, or NRM, to identify the octapeptide residues. However, binding to RNA in vivo has never been demonstrated, and, indeed, Xenopus Nup98 did not show significant binding to RNA homopolymers in vitro (Ullman et al., 1999). Thus, the role of the NRM in RNA export has remained unclear. Because of its suggested multifunctional nature, the C-terminal domain of hNup98 provides an especially interesting target for detailed structural analysis of a nucleoporin. Here, we present the three-dimensional structure of the C-terminal, pore-targeting/autoproteolytic domain of hNup98. We developed a GFP-based screen for in vivo folding to identify a stable domain of hNup98 that could be expressed in E. coli. The structure was determined to 3.0 Å for residues 710–870 of hNup98 and reveals a novel protein fold. Observations include details of the proteolytic active site and the noncovalent
association of the C-terminal fragment. The structure suggests a model for how this domain facilitates and regulates pore targeting in the assembly of the NPC.

Results

**A Screen for Stably Folded Domains**

We found that full-length hNup98 could not be stably expressed at a high level in bacteria. Of the individual domains of hNup98, only the C-terminal 40 kDa portion could be produced in a soluble form; however, it was not fully stable and frequently degraded into multiple species during purification. As a consequence, we set out to delineate an optimal C-terminal domain of hNup98 that would remain stable when expressed at a high level in *E. coli*. We developed an in vivo folding assay to identify domains of a protein that form independent folding units amenable to expression (see Figures 1B–1D). Our screen depends on the observation that, in a fusion of two proteins, a C-terminal green fluorescent protein (GFP) will assume its native structure and form a functional fluorophore only if the N-terminal protein can be expressed in a correctly folded, soluble form (Sachdev and Chirgwin, 1998; Waldo et al., 1999). Thus, there is a correlation between the solubility of an N-terminal polypeptide fused to GFP and the fluorescence of the fusion protein. We have applied these principles to develop a screen for soluble domains within an insoluble protein.

To carry out the screen, the gene for GFP was fused in frame at the 3’ end of the coding region of hNup98. Nested N-terminal deletions of the hNup98 region were produced by treatment with a 5’ exonuclease to generate a library of variable C-terminal hNup98 fragments fused to GFP (Figure 1B). The library was transformed into *E. coli*, and fluorescent colonies were selected for further screening. First, the relative fluorescence per cell was determined by flow cytometry and compared to the basal level obtained from full-length hNup98-GFP (Figure 1C). In addition, the soluble protein fraction from fluorescent strains was analyzed for protein expression (Figure 1D).

A variety of fragment sizes of hNup98-GFP were recovered, and most exhibited the same basal fluorescence as the full-length protein. Typically, these colonies contained large hNup98 fragments and produced very little intact protein (Figure 1D). In contrast, a minority of colonies exhibited relatively bright fluorescence, contained a variety of 20–25 kDa hNup98 fragments fused to GFP, and did not suffer any significant degradation. The plasmids encoding three unique protein fragments corresponding to the C terminus of hNup98 starting at residues 646, 673, and 676 were isolated from fluorescent strains.

To generate proteins for crystallography, these three fragments of hNup98 were expressed in bacteria with C-terminal hexahistidine tags in the place of GFP. As illustrated in Figure 1E, each fragment was expressed in two different forms. In one set, the Nup98 fragment was truncated at the proteolytic cleavage site after residue 863 (referred to as 646-His, 673-His, and 676-His). A second set of constructs contained the C-terminal coding sequence of hNup98 through residue 920 followed by the hexahistidine tag (referred to as 646-Tail-His, 673-Tail-His, and 676-Tail-His). The yield of soluble purified protein for each of these variants was between 50 and 100 mg per liter of culture; thus, our in vivo screen had successfully selected for domains that could be readily expressed in a soluble form.

Each of the C-terminal fragments of hNup98 demonstrated autoproteolytic activity when the tail sequence following the cleavage site was present (data not shown). However, the initial step of purification isolated both fragments of each protein through a hexahistidine tag present only in the C-terminal product. This copurification of both cleavage products is consistent with previous observations that the C-terminal tail remains bound to hNup98 after cleavage (see Figure 3C, lane 1).

**Structure Determination**

The six protein variants were subjected to crystallization trials, and diffraction quality crystals were readily obtained using the 676-His and the 676-Tail-His variants, each in a unique crystal form. The structure of these two variants was determined using a combination of multiple isomorphous replacement and multiwavelength anomalous diffraction methods (Table 1, Experimental Procedures). A model of the 676-Tail-His fragment was refined to a final R factor of 24.7% and R-free of 27.1% with fairly good geometry.

Subsequent to the determination of this structure, crystals were grown from an autoproteolysis-deficient mutant of hNup98. This fragment, containing residues 676–670 where the catalytic Ser864 was replaced with an alanine (S864A), yielded crystals that diffracted to 2.8 Å. The structure of this uncleavable variant domain of hNup98 was readily solved through molecular replacement. The model for the uncleaved S864A variant was then refined to 2.9 Å with an R factor of 24.2% and an R-free of 28.3%.

**Overall Structure**

The C-terminal autoproteolytic domain of hNup98 adopts a half-open, β sandwich-like fold anchored by a large β sheet that expands across the entire length of the domain. Figure 2 illustrates the architecture of the domain along with the nomenclature used here for the various surfaces of the structure. Three β sheets, comprised of 6, 3, and 2 strands, form a loose, prism-shaped “tent” around the hydrophobic core. Each end of the prism is capped by one or more α helices.

The polypeptide chain of the wild-type sequence is discontinuous: as expected, the peptide bond between Phe863 and Ser864 is broken (Figure 3A). However, the N-terminal residues of the resulting tail are visible in the electron density at the right-posterior end of the domain and participate as the final strand (β11) in the right face of the β sandwich (Figures 2 and 4). The broken peptide bond is situated at the dorsal bulge of the β sandwich juxtaposed with the long β7–β8 loop.

The S864A variant structure is nearly identical to the structure of the wild-type, cleaved form with the exception that the polypeptide chain is uncleaved and continuous through residues 863–864 (Figure 3B). The root-
Table 1. Data Collection Statistics

<table>
<thead>
<tr>
<th>Protein</th>
<th>676-His</th>
<th>676-Tail-His</th>
<th>S864A</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Native</td>
<td>Pt</td>
<td>SeMet</td>
</tr>
<tr>
<td>Space group</td>
<td>P2,2,2</td>
<td>P4,2,2</td>
<td></td>
</tr>
<tr>
<td>Cell dimensions (Å)</td>
<td>a = b = 91.5, c = 136.7, α = β = γ = 90°</td>
<td>a = b = 89.9, c = 203.1, α = β = γ = 90°</td>
<td>a = 90.3, c = 134.3, α = β = γ = 90°</td>
</tr>
<tr>
<td>Molecules per asymmetric unit</td>
<td>4</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>X-ray source</td>
<td>NSLS</td>
<td>NSLS</td>
<td>NSLS X12B</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.1</td>
<td>1.1</td>
<td>0.9497 (Remote)</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>30.0–2.9</td>
<td>30.0–3.3</td>
<td>30.0–3.4</td>
</tr>
<tr>
<td>Total measurements</td>
<td>463,563</td>
<td>337,607</td>
<td>418,959</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>26,325</td>
<td>15,740</td>
<td>16,916</td>
</tr>
<tr>
<td>Redundancy</td>
<td>7.1</td>
<td>6.2</td>
<td>10.2</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.7</td>
<td>97.3</td>
<td>86.7</td>
</tr>
<tr>
<td>Rsym (%)</td>
<td>5.2</td>
<td>7.0</td>
<td>7.2</td>
</tr>
<tr>
<td>Rcullis (%)</td>
<td>90</td>
<td>73</td>
<td>75</td>
</tr>
<tr>
<td>Phasing power</td>
<td>0.5</td>
<td>1.3</td>
<td>1.3</td>
</tr>
<tr>
<td>FOM</td>
<td>0.44</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2. The Architecture of the C-Terminal Domain of hNup98

The structure is roughly similar to a half-open β sandwich with helices capping two ends of the sheets. (A–C) Ribbon diagrams of the secondary structure from three views: (A) right, (B) left, and (C) dorsal. The faces of the domain surface are arbitrarily labeled to aid in discussion as Dorsal (D), Ventral (V), Anterior (A), Posterior (P), Right (R), and Left (L). Images were rendered using the SPDBV program with β sheets in blue, α-helices in yellow, and coiled segments in gray. In (A), the side chains near the proteolytic active site are rendered as stick models where carbons are green, nitrogens are blue, and oxygens are red. The position of the NRM octapeptide is indicated in (B) colored in red. (D) Secondary structure elements are schematically diagramed to illustrate the unique topology of this domain. The view is roughly from the Dorsal side where β sheets are shown as triangles, helices are shown as circles. The cleavage site is indicated by an X.
mean-square-distance between the α carbons of both models is 0.6 Å. Thus, cleavage of the 863–864 peptide bond apparently has very little effect on the overall structure of this domain except in the residues immediately juxtaposed to the cleavage site (see below).

Neither the three-dimensional arrangement of the secondary structures nor the topology of the protein fold is significantly similar to any known protein structure. The TOPS protein topology database could find no similar arrangement of secondary structure in any known protein fold (Gilbert et al., 1999). Using the Dali server to search for similar folds in three-dimensional space, the highest similarity Z score for a known protein was 3.5, where identity would score 29 and 2.0 indicates dissimilarity (Holm and Sander, 1995). Thus, we conclude that this domain of hNup98 represents a novel protein fold.

The Site of Proteolysis and Analysis of Autoproteolytic Function

The ordered part of the structure corresponds closely to the minimal domain of hNup98 found to be necessary and sufficient for autoproteolysis. Based upon the known mechanisms of inteins, Rosenblum and Blobel (1999) proposed that His862 is positioned near the cleavage site to act as a general base for Ser864. We find that Phe863 is buried in a hydrophobic pocket formed by the β sandwich, thus placing the side chain of His862...
in a position to hydrogen bond both with the backbone oxygen of Phe863 and with the side chain of Ser864 (Figure 3C). The position of Ala864 in the S864A variant can be used to model the interactions between the catalytic Ser864 and the rest of the proteolytic active site prior to cleavage. As shown in Figure 3C, when Ser864 is placed in the most favored rotamer, it is positioned to hydrogen bond to His862. The side chain hydroxyl is poised for nucleophilic attack on the backbone carbonyl carbon of Phe863. The cleavage site is juxtaposed to a long loop formed from residues 786–796. Positioned here is the conserved Lys791, whose side chain nitrogen participates in a hydrogen bond with the carbonyl oxygen of Phe863 before the cleavage reaction and remains bound to this oxygen at the new C terminus after cleavage (Figure 3C). Using the S864A variant structure as a model, this lysine is positioned too distant (~4.6 Å) from the modeled serine side chain to make a strong interaction with the serine hydroxyl. Thus, the side chain of Lys791 could provide the positive charge that functions to stabilize the oxanion of the reaction intermediate. In addition to Lys791, the loop also presents Asn799 in a position to hydrogen bond to the backbone nitrogen of Phe863, where it most probably functions to maintain the active site geometry.

To determine the relative importance of the residues near the active site to catalytic function, a number of mutants of this domain of hNup98 were generated. The resulting proteins were expressed, rapidly purified, and analyzed by SDS-PAGE. As shown in Figure 3D, lane 1, the 676-Tail-His fragment of hNup98 is cleaved into a 22 kDa N-terminal product and an 8 kDa C-terminal tail. Cleavage is complete on the time scale of purification (~2 hr at 4°C). Replacing the putative catalytic Ser864 with an alanine results in a single, uncleaved polypeptide of 30 kDa (Figure 3D, lane 6). Substitution of the conserved His862 with either alanine or glutamine resulted in a mixture of cleaved and uncleaved products, suggesting that proteolysis still occurs, but at a reduced rate.

The mutation of the conserved Lys791 resulted in a small amount of uncleaved protein, but the effect was weaker than that of the His862 mutations (Figure 3D, compare lane 2 to lanes 4 and 5). The Asn799 mutation had an effect of a similar magnitude. The Lys865 mutation had no effect on cleavage whatsoever, indicating that this residue is not important in the proteolysis of hNup98. A limitation of this assay is that the actual rate of proteolysis in the wild-type protein cannot be determined. It is important to bear in mind that cleavage of hNup98 is an intramolecular reaction. In vivo, cleavage most likely occurs very rapidly after protein synthesis. Thus, the appearance of a small amount of uncleaved protein appears to be a modest effect, but could well represent a dramatic reduction in the rate of cleavage compared to the wild-type reaction.

Interestingly, when His862 and Lys791, the residues positioned most proximal to the cleaved peptide bond, were both substituted with alanine, the combined mutation had a dramatic effect on cleavage. Only uncleaved protein was observed over the 2 hr time course of purification. Since both residues are completely solvent exposed, it is unlikely that the effect is simply due to local misfolding, as in Asn799. Thus, these data indicate an important role for both His862 and Lys791 in the catalysis of intramolecular cleavage.

**The NRM Domain**

hNup98 and its homologs in yeast are functionally linked with RNA transport through the NPC, and the C-terminal...
domain of these proteins was proposed to have a loosely conserved, partial RNP motif. We find that the three-dimensional structure of this region of hNup98 (residues 753–760) does not resemble an RNA binding motif either in topology or in the details of the molecular surface (Figure 2B). The NRM octapeptide appears as one strand of the extensive left-face β sheet, in a position similar to the RNP1 sequences in the β sheet of RNP proteins. However, the similarities between the RNP proteins and hNup98 are not there. In fact, a specific aromatic residue of the RNP motif that provides an important stacking interaction in two characterized RNA-protein complexes is not conserved in the octapeptide NRM motif (corresponding to Ser757 in hNup98). In contrast to the positively charged surfaces found on most RNA binding proteins (Cusack, 1998), the exposed surface of hNup98 containing the NRM sequence is actually strongly acidic in character. While structural characteristics cannot absolutely exclude RNA binding by this domain, they do suggest that RNA binding by the NRM is highly unlikely unless it were to employ an interaction dissimilar to any seen in current RNA-protein complexes.

Interactions between the N- and C-Terminal Products
In the crystal structure of the cleaved form of the protein, strong electron density appears near the C terminus of Phe863 that could easily be modeled as a 7 residue peptide chain (Figure 4). This model suggests that, after cleavage, the tail remains bound in a conformation identical to that adopted before cleavage. The most ordered segment of the tail, residues 866–868 (Tyr-Gly-Leu), participates in extensive interactions with the N-terminal domain. These residues continue the short β sheet through three hydrogen bonds with the β strand of residues 779–781. Glu867 makes two hydrogen bonds with Gln842 positioned on the large posterior α-helix. Both Tyr866 and Leu868 interact with hydrophobic patches of the N-terminal domain contributed by several discontinuous segments. Significantly, tail residues 866–868 are well conserved among hNup98/96 homologs; the position of Tyr866 is always filled by a large hydrophobic residue, and the Gly-Leu dyad of 867–868 is absolutely conserved in all known homologs.

In addition to these strong and ordered interactions, there is some evidence for less specific, electrostatic interactions between the tail sequences and the 90 kDa domain. In all hNup98/96 homologs, the three residues described above are invariably followed by a series of acidic residues. Although these residues are not visible in the electron density (density was not interpretable beyond residue 870, presumably due to disorder), their position is constrained to be centered on a pocket of the surface of hNup98 that is very basic in character (Figures 4A and 4B). Thus, there is the potential for electrostatic interactions between the basic surface and the acidic residues of the tail. Although the disorder in the tail residues suggests that the interaction is nonspecific, these interactions could be significant in the association between the tail and the N-terminal domain.

Biological Role of the Autoproteolytic Cleavage
Previous studies indicated that cleavage of hNup98 was necessary for proper targeting to the nuclear pore (Fontoura et al., 1999). However, these experiments were performed with hNup98 carrying the double mutation F863S/Y866R which the crystal structure suggests could significantly perturb the three-dimensional structure of the domain. To determine whether the mislocalization of hNup98 was truly due to lack of cleavage or was a result of local misfolding, both the full-length wild-type and the uncleavable S864A mutant were tagged with GFP at the N terminus and expressed in HeLa cells. GFP-Nup98 was found both at the NPC and within the nucleoplasm as previously shown (Griffis et al., 2002) (Figure 5A, panel a). In contrast, the uncleavable variant was not found at the nuclear rim, but was localized only to the nucleoplasm (Figure 5A, panel b), demonstrating that the actual cleavage event is necessary for proper pore targeting.

To elucidate more specifically the role of Nup98 cleavage in targeting to the pore, we assayed the effect of cleavage on Nup98 binding to its target protein complex at the nuclear pore. Nup98 associates with the nuclear pore through binding to a subcomplex consisting of Nups160, 133, 107, 96, and sec13 (Vasu et al., 2001). It is presumed that Nup98 binds directly to Nup96, the downstream product of Nup98-catalyzed autoproteolysis of the Nup98/Nup96 polyprotein. Various forms of the C-terminal domain of Nup98 were fused to GST, and pull-down experiments directly tested the requirement for cleavage in binding between Nup98 and Nup96 (Figure 5B). The Nup98/Nup96 precursor was translated in vitro and underwent rapid autoproteolysis to generate the two individual nucleoporins (Figure 5B, lane 1). GST fusions containing either the wt Nup98 C-terminal domain or a truncated variant missing the 8 kDa tail were both able to bind specifically to Nup96. However, no binding to Nup96 was observed with the uncleavable form of Nup98. Thus, cleavage of the Nup98 C terminus is required for direct interaction with Nup96. To confirm this result in the context of the nuclear pore, we tested the ability of the various GST fusions to bind the endogenous nucleoporin subcomplex (Nups 160, 133, 107, 96, and sec13) from Xenopus egg extract. Binding was monitored with antibodies to both the Nup160 and Nup133 components of the subcomplex. Again, while wild-type or truncated Nup98 bound readily, the uncleavable mutant could not bind the complex (Figure 5C).

Based upon the binding data and the localization of the uncleavable form of Nup98, it appeared that the Nup98 tail must be removed by cleavage in order to free its binding site for interaction with Nup96. To further address this hypothesis, we investigated the localization of the 8 kDa tail peptide in vivo by fusing GFP to either the N terminus or the C terminus of full-length Nup98. These constructs alternately reveal the localization of either the 90 kDa N-terminal fragment (Figure 5D, top panels) or the cleaved 8 kDa tail (Figure 5D, bottom panels). In each case, cells were costained with a low concentration of antibody to Nup98. At this antibody concentration, the signal from endogenous protein is weak, and thus it was possible to identify cells that expressed equivalent amounts of transfected GFP fusion protein. It is clear that while GFP-Nup98 was localized as expected, to both the nuclear interior and the nuclear rim, Nup98-GFP (in which the GFP signal corresponds to the tail peptide) never results in a signal at
Figure 5. Proteolytic Cleavage and Nuclear Pore Targeting

(A) Localization of GFP-Nup98 (panel a) and the uncleavable GFP-Nup98(S864A) variant (panel b) in HeLa cells.

(B) In vitro-translated hNup96 is bound by GST fused to the C terminus of hNup98, and GST-hNup98 truncated at the site of cleavage (F863) (lanes 3 and 4) but not by GST alone (lane 2) or GST fused to the uncleavable S864A variant (lane 5). The products of translation of the 190 kDa Nup98/Nup96 precursor are in lane 1.

(C) GST proteins as described in (B) were used to bind proteins from *Xenopus* egg extracts. Binding of the Nup96-containing complex was assessed by immunoblotting with antibodies to two different complex components, Nup160 and Nup133.

(D) Localization and mobility of the Nup98 N-terminal 90 kDa fragment (GFP-Nup98, panels a–c) and the cleaved C-terminal 8 kDa tail (NUP98-GFP, panels d–f). A Nup98-specific antibody was used to visualize the 90 kDa fragment in panels a and d, and GFP localization in cells expressing equivalent amounts of transfected protein is shown in panels b and e. Panels c and f show the merged images, and the insets represent an enlargement of the nuclear rims in each panel. The graph illustrates time-dependent recovery after photobleaching. Each curve is the average of six independent measurements, and error bars depict standard error of the mean (SEM). The curves for GFP and Nup98-GFP are indistinguishable.

the nuclear rim. Thus, not only cleavage but separation of the two polypeptides must occur in order for the 90 kDa fragment of Nup98 to associate with the nuclear pore.

While it was now certain that the 8 kDa peptide must be separated from Nup98 to permit binding to the nuclear pore, the fate of the 8 kDa peptide tail was still unclear. We have previously shown that Nup98 dynamically associates with the nuclear pore and is quite mobile in the nucleoplasm, moving between intranuclear bodies and the nuclear pore (Griffis et al., 2002). To test the association of the 90 kDa and 8 kDa peptides within the nucleus, we performed fluorescence-recovery-after-photobleaching (FRAP) measurements to assess the relative mobilities of the two GFP-tagged polypeptides. Strikingly, we found that FRAP of the 8 kDa tail-GFP in the nucleoplasm resulted in recovery kinetics indistinguishable from those of free GFP, indicating a mobility close to the rate of diffusion (Figure 5D, graph). In contrast, GFP-Nup98 within the nucleoplasm recovered significantly more slowly, as previously observed. Thus, within the nucleoplasm, the mobilities of these two fragments of Nup98 are very distinct and they are unlikely to be physically associated.

The Divergence of Function in Yeast

Although Nup98 is unique in vertebrates, this protein is homologous to three distinct proteins in yeast (Figures 1 and 6). Thus, the multiple functions of Nup98 could each be independently carried out by one of the three yeast homologs. The strongest indication of this comes from the observation that hNup98 maintains two func-
Structure of a Nucleoporin Pore-Targeting Domain

Figure 6. The Sequence Conservation between Pore-Targeting Domains of hNup98 and Its Three Yeast Homologs

(A) A sequence alignment. Residues that are similar (within 2 units on a PAM250 matrix) to the sequence of hNup98 are boxed in gray. Overall, these four domains exhibit 26% amino acid identity.

(B–D) A GRASP-rendered solvent accessible surface of the N-terminal cleavage product (residues 710–863) is colored according to the positions of residues conserved between the human and yeast proteins. The surface is shown from the dorsal (B), right (C), and left (D) sides. The cleaved tail (residues 864–870) is shown as a stick representation. Residues that are identical in all four proteins are shown in dark blue. Residues that are similar [according to (A)] in all four proteins are shown in lighter blue. Residues that are shared only between the two autoproteolytic domains (hNup98 and scNup145) are shown in red.

To analyze the divergence of hNup98 functions in the yeast homologs, each residue in the hNup98 C-terminal structure was classified as to whether it shared homology with a single yeast homolog or was conserved in all homologs (Figure 6). Marked in red are the residues of hNup98 that are shared with scNup145 but not scNup100 or scNup116. Thus, these residues may mediate functions that are specifically shared between these proteins, including proteolytic activity.

Strikingly, the residues that interact with the C-terminal tail appear to be highly conserved in all three yeast proteins. This includes the basic loop juxtaposed to the approximate position of the acidic tail regions. Although scNup100 and scNup116 do not possess proteolytic activity, their polypeptide chains terminate very near to where the proteolytic cleavage of scNup145 and hNup98 occurs. Thus, there is a possibility that scNup116 and scNup100 could interact with the C-terminal cleavage product of scNup145, scNup145C. A second possibility is that this conservation suggests an as yet unknown binding partner of the three yeast proteins and hNup98. The juxtaposition of this conservation to the C-terminal tail binding site would be consistent with a need to remove the C-terminal products of the cleavage in order to bind to such a partner.

Discussion

The assembly of the nuclear pore requires a complicated orchestration of interactions between many protein constituents. A full understanding of this process will require knowledge of the structure of each of these constituents. We present here what to our knowledge is the first three-dimensional structure of a nuclear pore component, the pore-targeting domain of the nucleoporin hNup98 at 3.0 Å resolution. This structure reveals a novel protein fold and topology. The domain is posttranslationally cleaved into two polypeptides through an autoproteolytic event. Cleavage of the peptide chain is necessary for the proper targeting of hNup98 to the nuclear pore and thus provides a unique mechanism for the regulation of nuclear pore association.

The Mechanism of Autoproteolysis

The atomic details of the proteolytic active site presented here in both pre- and postcleavage forms are consistent with the intein-like mechanism proposed by Rosenblum and Blobel (1999). Ser864 is absolutely conserved in all Nup98 homologs that carry out autoproteol-
ysis, and the presence of a nucleophile (S/C/T) at this position is essential for cleavage. His862 appears to play a prominent role in catalysis, via hydrogen bonding to Ser864 to generate the attacking nucleophile.

A significant addition to the active site identified by the crystal structure is the conserved Lys791. Single mutation of either Lys791 or His 862 produced only a small increase in the amount of uncleaved protein present after purification. However, mutation of Lys791 in combination with the mutation of His862 (H862A/K791A) dramatically reduces the rate of intramolecular proteolysis. Thus, we conclude that each of these side chains makes an important contribution to catalysis that is synergistic with the other. Given the relative pKas and positions, it is probable that His862 acts as a general base to abstract a proton from Ser864, and Lys791 acts to reduce the pKa of the serine hydroxyl and thus stabilize the oxanion intermediate.

Although His862 and Lys791 are very important in catalysis, they are not the only contributors to the cleavage reaction. Indeed, cleavage can still be observed, albeit at a drastically reduced rate, when both of these residues are mutated to alanine (~10% cleavage per day compared to no cleavage in the S864A mutant; our unpublished data). Thus, there must be other less obvious factors built into the active site. One such factor would be the specific geometry of the active site in the precleavage conformation of the protein. This active site geometry could impose strain on the scissile peptide bond that makes it susceptible to attack. Such strain has been experimentally observed in other self-cleaving proteins including the NTN family and the gyr A intein (Klabunde et al., 1998; Xu et al., 1999). In these proteins, the strain is imposed by specific tertiary interactions between the polypeptide flanking the cleavage site and the rest of the protein.

The resolution of our data precludes direct observations of strain; however, the existence of strain in the polypeptide precleavage is consistent with the structural and biochemical data. In hNup98, the active site geometry is anchored by both Phe863 on the N-terminal side and the \( \beta \) strand interactions of the tail on the C-terminal side. We find that the Phe863 side chain is tightly held in a hydrophobic interaction with multiple residues within the hydrophobic core of the Nup98 C terminus. Cleavage was shown to be dramatically reduced when the C-terminal tail beyond Ser864 is replaced by polyalanine (Rosenblum and Blobel, 1999). In this case, the loss of specific interactions between the body of the protein and Tyr866/Leu868 would lead to loss of the geometry required for catalysis. Thus, anchoring of the peptide chain geometry is mediated by a number of residues and makes a significant contribution to catalytic function that can be observed as a basal level of proteolysis in the His862A/Lys791A double mutant.

Assembly of the Nuclear Pore Complex
The structure presented here, with the C-terminal domain of hNup98 and the associated peptide tail, is also a model for a nucleoporin complex between hNup98 and hNup96. The N terminus of hNup96, which is generated from the 190 kDa hNup98/hNup96 polyprotein, is identical to the tail sequences characterized in this structure (Fontoura et al., 1999). It is reasonable, therefore, to expect that the interactions within hNup98/Nup96 complex would be equivalent to those in the hNup98/tail complex. Forbes and colleagues recently demonstrated that the C-terminal domain of hNup98 interacts with a NPC subcomplex consisting of Nups 96, 133, 160, and sec13 (Vasu et al., 2001). This hNup96-containing complex has further been shown to be a stably associated, structural subcomplex found on both nuclear and cytoplasmic faces of the NPC (Belgareh et al., 2001). The stability of the hNup96-containing complex in the pore, along with the dynamic association of hNup98 with the pore (Griffis et al., 2002), suggests that hNup96 serves as a docking point for hNup98 in the nuclear pore. Thus, the structure described here should also represent the three-dimensional structure of a pore-docking event.

It is clear that autoproteolytic cleavage influences the association of hNup98 with the Nup96 complex and the nuclear pore. The simplest interpretation is that cleavage must occur to permit dissociation of the peptide tail, thereby freeing its binding site for interaction with Nup96 at the pore. In the uncleaved form, the attached tail acts as a competitive inhibitor of binding to hNup96 and targeting to the pore. In the absence of the peptide tail, the truncated form of Nup98 binds as readily as the wild-type protein. It is less clear why lack of cleavage would influence targeting of the 190 kDa polyprotein; however, the experiments that showed mislocalization of the uncleaved polyprotein may have been performed on a time scale too short to observe incorporation of hNup96 into pore complexes (Fontoura et al., 1999). It was recently shown that NPC incorporation of the Nup96-containing complex occurred only when the cell passed through mitosis and the nucleus was disassembled and reformed (Belgareh et al., 2001). Thus, transfection and localization experiments that did not allow for at least a complete cell cycle might not reveal Nup96 at the NPC.

Our photobleaching experiments in live cells indicated that the peptide tail does not remain bound to Nup98 in the nucleus. In contrast, we observed that, in vitro, the 8 kDa tail and the Nup98 C-terminal domain persisted as a complex through the multi-step purification from bacterial lysates. Possibly, a posttranslational modification lacking in bacteria, such as phosphorylation within hNup98 or the tail peptide, could mediate dissociation of the tail. Indeed, Nup98 has been shown to be a phosphoprotein that is hyperphosphorylated at mitosis (Macaulay et al., 1995). Alternatively, the persistent association of Nup98 and the tail peptide in vitro may be driven by the very high concentrations achieved in the bacterial expression system. In support of this interpretation, we did observe association of the peptide tail with Nup98-containing intranuclear bodies where Nup98 is enriched. Resolution of these two possibilities will require analysis of potential modification sites and measurement of binding affinities.

In summary, we have presented a detailed picture of a critical step in nuclear pore assembly that provides a foundation for further structural and functional studies of hNup98 and the NPC. This analysis was made possible through the development of a screening strategy to facilitate expression of soluble protein domains. Using this
approach, we obtained a readily crystallizable domain of hNup98, a protein that has proven refractory to soluble expression in bacterial systems. By generating a random library of protein fragments, the screen takes an unbiased approach to the selection of optimal domains for expression in bacteria. Since large numbers of colonies can be initially screened on the agar plate and followed up by a FACS analysis, this method lends itself to high throughput applications. Thus, this assay will be a valuable tool in determination of domain structures and an approach that should be widely applicable to many proteins.

Experimental Procedures

In Vivo Screen for Folding Domains in hNup98

To create a library of C-terminal fragments, the hNup98-GFP fusion gene was subcloned into a modified pET-28a vector. Phe863 was fused to GFP through a 10 amino acid linker, and the GFP coding sequence was followed by a hexahistidine tag. In the resulting vector, the EcoRI site at the 5’ end of Nup98 was just downstream of a PstI site which was in turn positioned just downstream of the ATG start site. To create a library of C-terminal domains, this construct was linearized by digesting with PstI and EcoRI and then treated with Exonuclease III. The 3’ overhang of the PstI site protected the ATG site from Exonuclease III digestion, allowing directional cleavage of the N-terminal sequences of hNup98. Following treatment with Mung Bean nuclease to blunt the ends, the start site was re-ligated to the C-terminal fragments of the hNup98 gene.

The library of C-terminal fragments of hNup98 fused to GFP was then transformed into a variant of BL21(DE3) cells that had been selected for colony formation in the presence of IPTG (BL21(DE3) I + A cells, a generous gift of X. Zhang and W. Studier). Transformants were grown overnight at 30°C due to the temperature sensitivity of GFP expression. Colonies were visually screened using a long wavelength UV lamp. All colonies exhibiting green fluorescence were then subjected to further examination. Each isolate was grown in liquid media, and expression was induced at mid-log phase. Three hours after induction, the cells were analyzed by fluorescence-activated flow cytometry to establish a quantitative measure of the relative fluorescence of each library member. In addition, an aliquot of the cells was broken by sonication and the soluble fraction was analyzed by immunoblotting using an antibody against the 6x His tag (H-15; Santa Cruz Biotechnology). Appropriate candidate proteins for crystallography were then prepared by subcloning as described in Results.

Protein Purification and Crystallization

Unlabeled hNup98 fragments were expressed in BL21(DE3) cells by induction with 0.2 mM IPTG at mid-log phase. Selenomethionine-labeled versions of the proteins were expressed in the strain B834(DE3) grown in defined media containing 50 mg/mL L-selenomethionine (Hendrickson et al., 1990). Cells were broken by french press, then clarified by centrifugation. The protein was purified through nickel chelating affinity purification followed by ion exchange chromatography. Fractions containing the protein were then combined and dialyzed against 4 mM Tris, 25 mM NaCl, and 1 mM DTT.

Crystals were grown at room temperature using hanging drops. Both selenomethionine-labeled and unlabeled 676-His formed crystals above a well solution containing 0.1 M Tris (pH 7.75), 0.2 M NaCl, 25% PEG 4000, 8% Glycerol, and 5 mM DTT. Crystals formed over 24–48 hr in the space group P212121 (a = 91.7 Å, b = 136.7 Å). Crystals of 676-Tail-His formed over 48 hr in a well solution containing 0.1 M Tris (pH 7.9), 0.2 M NaCl, and 30 mM DTT. Crystals formed in the space group P4222, (a = b = 89.9 Å, c = 203.1 Å). Data was collected on each crystal by flash-freezing in a liquid nitrogen stream. 676-His crystals could be frozen directly in the well solutions. 676-Tail-His crystals were first introduced to the well solution containing 20% glycerol before freezing.

Data collection for all native and heavy atom-soaked crystals was performed at the NSLS beam lines X29C and X25 at Brookhaven National Laboratory. A three wavelength MAD experiment was performed using the SeMet-labeled 676-His protein crystals at NSLS beamline X12B. Data was reduced using the program HKL (Otwinowski and Minor, 1997).

Structure Determination

The structure of the 676-His fragment was solved first using experimental phases derived from the SeMet MAD experiment and a single Pt derivative using density modification. The structure was solved independently using CNS (Brünger et al., 1998) and SAD and SAD/PHASES (Terwilliger and Berzhenetz, 1999). After density modification, clear electron density maps were observed at 3.5 Å resolution. Four molecules were found in the asymmetric unit, and, with the aid of noncrystallographic averaging, a model was built containing residues 712–863 of hNup98. Ranges of refinement and rebuilding of this model could not lower the R factor below 30%, presumably due to extreme disorder in one of the four asymmetric molecules.

The rough model built for the 676-His fragment was then used to solve the 676-Tail-His structure through molecular replacement. Using the AMORE algorithm (Navaza, 1994), two molecules were found in the asymmetric unit of the P4222 crystals of 676-Tail-His fragment. The positions of the two molecules were refined by rigid body refinement to an R factor of 45%. Unbiased electron density maps were calculated for rotating subsections of the model by omitting 30–40 residues from the asymmetric unit, calculating phases from the omit-model, then combining those phases with the weak but significant phase information from a mercury derivative of the 676-Tail-His crystals. These combined phases were improved by density modification using CNS to provide an electron density map at 3.0 Å resolution. The omitted area of the model was then rebuilt using this map. After rounds of rebuilding and refinement, a final model including 313 residues in the asymmetric unit was assembled with a free R factor of 27.1% and a working R factor of 24.7%.

The final model has 71% of residues lying in the most favored region of the Ramachandran plot; 27% in the additionally allowed regions, and 2% in the generously allowed regions. The model had an rmsd deviation from ideality of 0.011 Å in bond lengths and 1.5° in bond angles.

Mutagenesis and Cleavage Assay

Site-directed mutagenesis was performed using the Quickchange mutagenesis kit (Stratagene) and verified by sequencing, and mutants were expressed in BL21(DE3) cells. After induction, cells were pelleted, then sonicated on ice. Clarified cell lysate was incubated 30 min at 4°C with Ni-resin (Pharmacia), pelleted, and washed four times. Bound proteins were eluted with SDS and resolved by SDS-PAGE.

The S864A Variant Structure

The DNA encoding residues 676–870 of the S864A variant was subcloned into pGEX 4T-3. The protein was expressed and purified on glutathione resin and released by cleavage with thrombin. Crystals were obtained using 17% PEG 8000, 0.1 M Tris (pH 8.25), 0.1 M MgCl2, and 30 mM DTT in a hanging drop. Crystals were isomorphous to the 676-His form described above. A model containing four molecules in the asymmetric unit was refined with good geometry to a final R factor of 24.2% at 2.9 Å resolution.

Immunofluorescence Localizations and Binding Assays

The Nup98-GFP construct was generated by subcloning the human Nup98 gene into pEGFP-N1 (Clontech). GFP-Nup98 was previously described (Griffis et al., 2002), and the GFP-Nup98 S864A mutant was generated using the Quickchange mutagenesis kit (Stratagene). Plasmids were transiently transfected into HeLa cells using Fugene1 (Roche), and immunofluorescence and photobleaching experiments were performed as described previously (Griffis et al., 2002). Nup98 C-terminal antibody was used at 1:3000.

In vitro binding experiments were performed using GST-Nup98 fusion proteins bound to glutathione beads at 2 μg protein/μl of beads, covalently crosslinked using dimethylpimelimidate, and blocked with 10 mg/ml BSA in PBS. For binding to Nup98, the 190 kDa Nup98/Nup96 precursor was translated using a TNT kit.
Molecular Cell

358

(Promega). 10 μl of translation was diluted 1/26 in PBS, 1 mg/ml BSA, and 0.2% Tween-20 and incubated with 20 μl of GST-Nup98 beads for 2 hr at 4°C. Beads were washed five times with PBS, and bound proteins were eluted with gel sample buffer and resolved by SDS-PAGE. For binding to the Xenopus Nup96 complex, 10 μl of Xenopus egg extract (Smythe and Newport, 1991) was diluted 1/50 in PBS + 2 mg/ml BSA, precleared by incubating for 1 hr at 4°C with 40 μl of beads containing 2 μg/ml GST, then incubated with 40 μl of GST-Nup98 beads for 2 hr at 4°C. Beads were washed five times with PBS and then eluted with 0.1 M glycine (pH 2.5). Eluates were neutralized with 1 M Tris (pH 8.0), concentrated (Microfree concentrators; Millipore), and analyzed by PAGE and immunoblotting. For immunoblots, antibody to Nup160 was used at 1/1000 and antibody to Nup133 was used at 1/4000 in PBS + 2% BSA.

Acknowledgments

The authors would like to thank Drs. Xiong Zhang, John Horton, and Xiaodong Cheng for their generous help and advice, Erica Phillips for expert technical assistance, and Dr. Douglass Forbes for the kind gift of antibodies to Nup133 and Nup160. The authors also thank Dr. Dieter Schneider, Michael Becker, and the staff of the X26C, X25, and X12B beamlines at Brookhaven National Laboratories National Synchrotron Light Source for their hospitality and data collection in oscillation mode. Methods Enzymol.

References


Accession Numbers

The coordinates for the 676-Tail-His structure have been deposited in the Protein Data Bank (accession number 1K08).