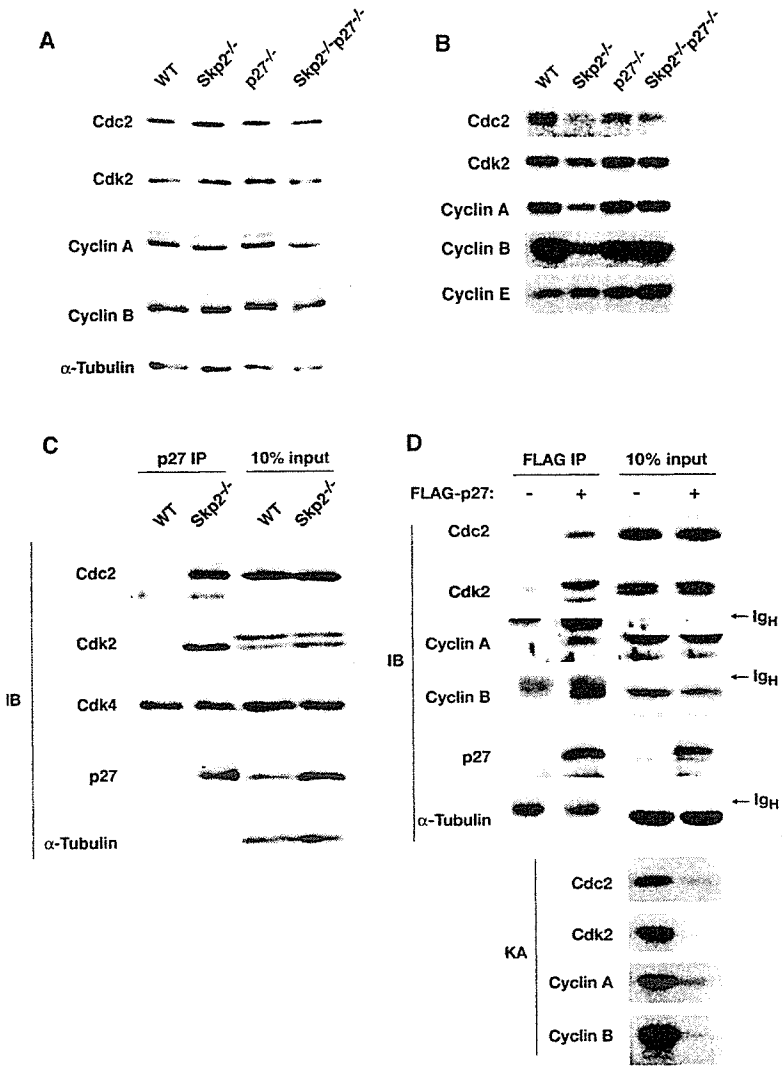

Figure 5. Reduced Cdc2-, Cyclin A-, or Cyclin B-Associated Kinase Activity in *Skp2*$^{-/-}$ Cells and Interaction of p27 with Cdc2, Cdk2, Cyclin A, and Cyclin B

(A and B) Lysates of MEFs from wild-type, *Skp2*$^{-/-}$, *p27*$^{-/-}$, and *Skp2*$^{-/-}$ *p27*$^{-/-}$ mice were subjected either to immunoblot analysis with antibodies to the indicated proteins (A) or to in vitro assays of Cdc2-, Cdk2-, cyclin A-, cyclin B-, or cyclin E-associated kinase activity (B).

(C) Lysates of MEFs from wild-type and *Skp2*$^{-/-}$ mice were subjected to immunoprecipitation (IP) with antibodies to p27, and the resulting precipitates were subjected to immunoblot (IB) analysis with antibodies to the indicated proteins (left panels). A portion (10%) of the input lysates was also subjected directly to immunoblot analysis with the same antibodies (right panels).

(D) COS-7 cells were transfected with an expression vector for FLAG-tagged p27 (or with the empty vector) and with a vector for CD19, and CD19-expressing cells were then isolated with the use of magnetic beads conjugated with antibodies to CD19. Lysates of the CD19-positive cells were subjected to immunoprecipitation with antibodies to FLAG, and the resulting precipitates were subjected to immunoblot analysis with antibodies to the indicated proteins (left panels). Portions (10%) of the input lysates were also subjected directly to immunoblot analysis with the same antibodies (upper right panels) and to in vitro assays of kinase activity (KA) associated with the indicated proteins (lower right panels). Ig_H, immunoglobulin heavy chain.

at least in part, by associated Cdk2, it was possible that the accumulation of cyclin E in the cells of *Skp2*$^{-/-}$ mice was an indirect consequence of p27 accumulation. To investigate this possibility, we examined the amount of cyclin E in MEFs, thymus, testis, and liver of *Skp2*$^{-/-}$ *p27*$^{-/-}$ mice. The abundance of cyclin E in these cells and tissues of the double mutant, however, was markedly greater than that in their wild-type counterparts and similar to that for *Skp2*$^{-/-}$ mice (Figure 7A), suggesting that the accumulation of cyclin E in *Skp2*$^{-/-}$ cells is not dependent on p27 accumulation. The amount of phosphorylated cyclin E in MEFs was similar for the four genotypes (Supplemental Figure S2A at http://www.developmentalcell.com/cgi/content/full/6/5/661/DC1), suggesting that the cyclin E that accumulates in *Skp2*$^{-/-}$ mice is the nonphosphorylated form. The p27-related CKIs p21 and p57, both of which are also targets of Skp2-mediated ubiquitylation (Bornstein et al., 2003; Kamura et al., 2003), were also examined for their binding to cyclin E in wild-type, *Skp2*$^{-/-}$, *p27*$^{-/-}$, and *Skp2*$^{-/-}$ *p27*$^{-/-}$ MEFs. Like p27, more p21 was associated with cyclin E in *Skp2*$^{-/-}$ MEFs than in wild-type MEFs (Supplemental Figure S2B); the increase in the amount of p21 bound to cyclin E was also observed in *Skp2*$^{-/-}$ *p27*$^{-/-}$ MEFs. Association of p57 with cyclin E was not detected in MEFs of any of the four genotypes (data not shown). We therefore are not able to exclude completely the possibility that accumulation of p21 or of other proteins results in inhibition of cyclin E-associated kinase activity and thereby stabilizes cyclin E by blocking its phosphorylation. However, such a possibility seems unlikely because cyclin E-associated kinase activity was not reduced in *Skp2*$^{-/-}$ MEFs (Figure 5B).

We also measured the rate of cyclin E degradation by treatment of MEFs with the protein synthesis inhibitor cycloheximide. In asynchronously cycling cells, cyclin E appeared to be more stable in *Skp2*$^{-/-}$ and *Skp2*$^{-/-}$ *p27*$^{-/-}$ cells than in wild-type or *p27*$^{-/-}$ cells (Figures 7B and 7C). This result was verified by pulse-chase experiments (Supplemental Figure S3). Furthermore, cyclin E was rapidly degraded after the release of wild-type and *p27*$^{-/-}$ MEFs from S phase arrest induced by aphidicolin treatment (Figures 7D and 7E); in contrast, the degradation of cyclin E after release from aphidicolin block was markedly delayed in *Skp2*$^{-/-}$ and *Skp2*$^{-/-}$ *p27*$^{-/-}$ MEFs. In addition, our previous biochemical

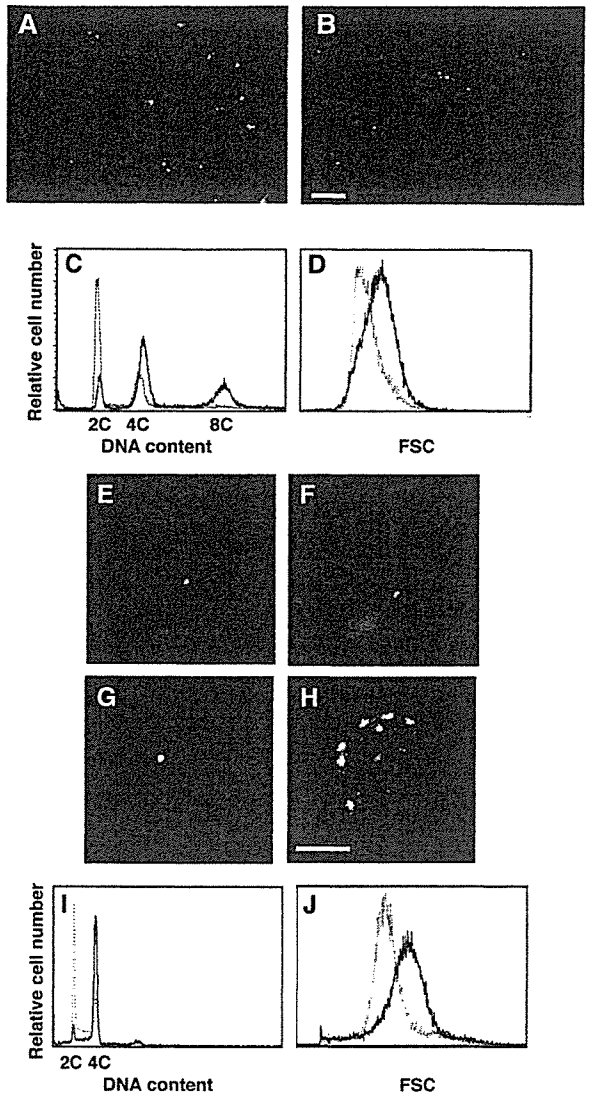

Figure 6. Induction of Nuclear Enlargement by Inhibition of the Kinase Activity of Cdc2

(A–D) Wild-type MEFs were incubated for 72 hr with dimethyl sulfoxide (vehicle control) (A) or with 80 μM butyrolactone I (B), after which they were subjected to immunostaining with antibodies to pericentrin (green) or to β-tubulin (red). Blue staining represents Hoechst 33258 labeling of nuclear DNA. Scale bar, 20 μm. The DNA content (C) and forward light scatter (FSC) (D) of the cells cultured in the absence (dotted lines) or presence (solid lines) of butyrolactone I were also measured by flow cytometry.

(E–J) FT210 cells (G and H), which express a temperature-sensitive mutant of Cdc2, and the parental FM3A cells (E and F) were cultured at 32°C (permissive temperature) (E and G) or 39°C (restrictive temperature) (F and H), after which they were subjected to immunostaining with antibodies to pericentrin (green) and to staining of DNA with propidium iodide (red). The images were taken by laser-scanning confocal microscopy. Scale bar, 10 μm. FT210 cells cultured at the permissive (dotted lines) or restrictive (solid lines) temperature were also subjected to flow cytometric analysis of DNA content (I) or forward light scatter (J).

studies indicated that Skp2 specifically interacts with cyclin E and thereby promotes its ubiquitylation and degradation both in vivo and in vitro (Nakayama et al., 2000). Collectively, these findings suggest that free cyclin E is a candidate substrate of SCF^{Skp2}, and that accumulation of cyclin E alone is not the cause of the cellular abnormalities of *Skp2*$^{-/-}$ mice.

Discussion

p27 Is a Major Physiological Target of SCF^{Skp2}

Protein degradation by the ubiquitin-proteasome pathway plays a fundamental role in determining the abundance of important regulatory molecules. The E3 ubiquitin ligases are thought to determine the substrate specificity of this pathway, and many diverse E3 molecules are therefore thought to exist, although it does not appear to be unusual that several proteins serve as substrates for a given ubiquitin ligase. Skp2, the substrate-recognizing component of an SCF ubiquitin ligase, has thus been shown to recognize p27 (Carrano et al., 1999; Sutterluty et al., 1999; Tsvetkov et al., 1999; Nakayama et al., 2000), the p27-related CKIs p21 (Yu et al., 1998; Bornstein et al., 2003) and p57 (Kamura et al., 2003), cyclin A (Nakayama et al., 2000), cyclin D1 (Yu et al., 1998), free cyclin E (Nakayama et al., 2000), E2F-1 (Marti et al., 1999), p130 (Tedesco et al., 2002; Bhattacharya et al., 2003), ORC1 (Mendez et al., 2002), Cdt1 (Li et al., 2003), Cdk9 (Kiernan et al., 2001), c-Myc (Kim et al., 2003; von der Lehr et al., 2003), and B-Myb (Charrasse et al., 2000). Of these proteins, cyclin A, cyclin D1, E2F-1, and Cdt1 do not accumulate in *Skp2*$^{-/-}$ cells (Nakayama et al., 2000) (our unpublished data), suggesting either that they are not bona fide substrates of Skp2 or that there is redundancy that allows for their ubiquitylation in the absence of Skp2.

Most of the cellular abnormalities apparent in *Skp2*$^{-/-}$ mice are not evident in *Skp2*$^{-/-}$ *p27*$^{-/-}$ double mutant mice. This observation constitutes genetic evidence that the cellular phenotype of nuclear enlargement with polyploidy and overduplication of centrosomes in *Skp2*$^{-/-}$ mice results from the deregulated expression of p27 during the cell cycle, and that p27 is likely the main target of the SCF^{Skp2} ubiquitin ligase. However, the observation that the phenotype of *Skp2*$^{-/-}$ *p27*$^{-/-}$ mice is not completely identical to that of *p27*$^{-/-}$ mice suggests the existence of additional target proteins for Skp2-mediated ubiquitylation.

Role of Skp2 in Prevention of p27 Accumulation during S and G_2 Phases

The concentration of p27 is relatively high in quiescent (G_0) cells, decreases on entry of cells into the cell cycle, and is controlled predominantly by the rate of p27 degradation (Pagano et al., 1995; Shirane et al., 1999) as well as by translational regulation (Hengst and Reed, 1996). This degradation has been thought to require Skp2, which binds to p27 when the latter is phosphorylated on Thr^{187} by the cyclin E–Cdk2 complex (Sheaff et al., 1997; Vlach et al., 1997; Carrano et al., 1999; Montagnoli et al., 1999; Sutterluty et al., 1999; Tsvetkov et al., 1999). The primary function of SCF^{Skp2} might therefore be to render quiescent cells competent to reenter the cell cycle by mediating the degradation of p27. However, there are several inconsistencies with this notion. Skp2 and cyclin E are not expressed until the G_1-S transition of the cell cycle, unequivocally later than the onset of

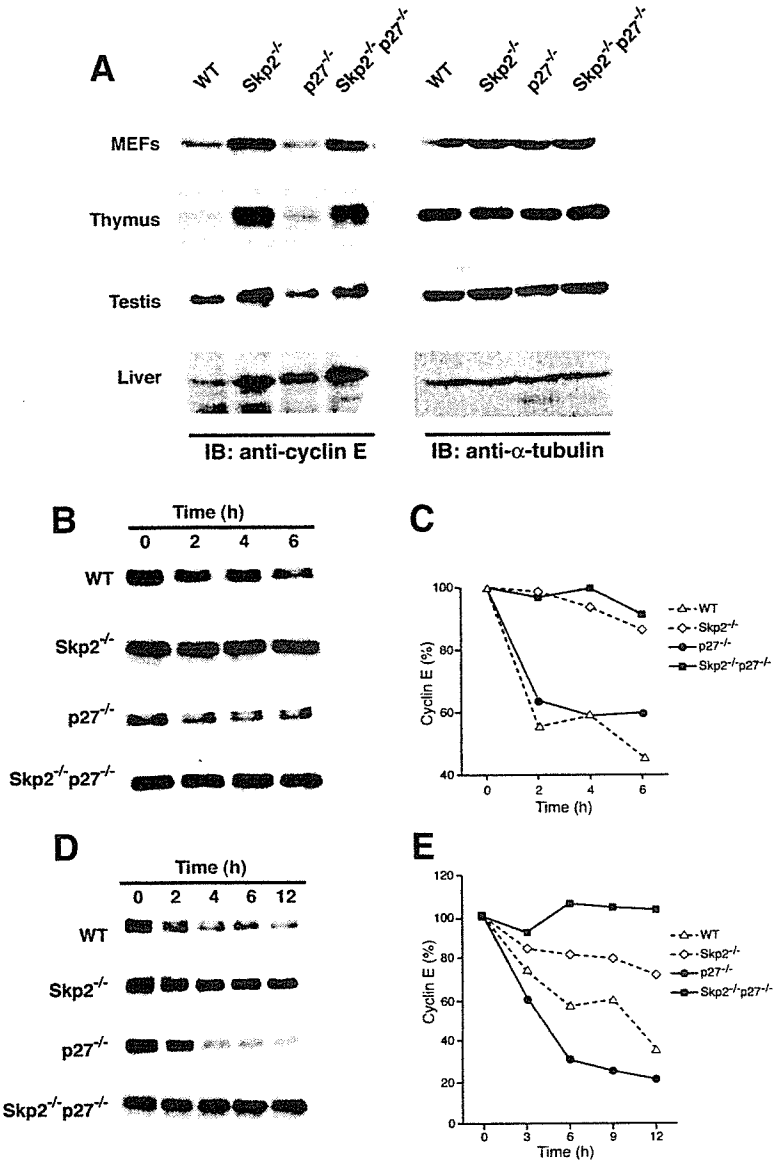

Figure 7. Impaired Degradation of Cyclin E in *Skp2*$^{-/-}$ Mice Is Independent of p27

(A) Immunoblot analysis of cyclin E (left panels) and α-tubulin (control; right panels) in lysates prepared from MEFs, thymus, testis, and liver of wild-type, *Skp2*$^{-/-}$, *p27*$^{-/-}$, and *Skp2*$^{-/-}$ *p27*$^{-/-}$ mice.

(B) Proliferating wild-type, *Skp2*$^{-/-}$, *p27*$^{-/-}$, or *Skp2*$^{-/-}$ *p27*$^{-/-}$ MEFs were incubated in the presence of cycloheximide (50 μg/ml) for the indicated times, after which cell lysates were subjected to immunoblot analysis with antibodies to cyclin E.

(C) The band intensities in (B) were quantitated by image analysis and expressed as a percentage of the corresponding value for time zero.

(D) Wild-type, *Skp2*$^{-/-}$, *p27*$^{-/-}$, or *Skp2*$^{-/-}$ *p27*$^{-/-}$ MEFs were synchronized at S phase by treatment with aphidicolin (1 μg/ml) for 15 hr and then released into aphidicolin-free medium for the indicated times, after which cell lysates were prepared and subjected to immunoblot analysis with antibodies to cyclin E.

(E) The band intensities in (D) were quantitated and expressed as described in (C).

p27 degradation apparent at mid-G_1 phase (Hara et al., 2001). Moreover, p27 is exported from the nucleus to the cytoplasm at G_0-G_1 (Tomoda et al., 1999; Rodier et al., 2001; Ishida et al., 2002; Connor et al., 2003), whereas Skp2 is restricted to the nucleus (Miura et al., 1999). The discrepancies between the temporal and spatial patterns of p27 expression and those of Skp2 expression suggest the existence of a Skp2-independent pathway for the degradation of p27. Indeed, the downregulation of p27 at the G_0-G_1 transition occurs normally in *Skp2*$^{-/-}$ cells and is sensitive to proteasome inhibitors (Hara et al., 2001). Biochemical analysis of crude extracts of *Skp2*$^{-/-}$ cells revealed the presence in the cytoplasmic fraction of an Skp2-independent E3 activity that mediates the ubiquitylation of p27 (Hara et al., 2001). This ubiquitylation was not dependent on the phosphorylation of p27 on Thr^{187}, which is a prerequisite for Skp2-mediated ubiquitylation. These data indicate that p27 is degraded at the G_0-G_1 transition by a proteasome-dependent, but Skp2-independent, mechanism.

In contrast, the degradation of p27 during S and G_2 phases is impaired in *Skp2*$^{-/-}$ mice, suggesting that the primary function of Skp2 is to prevent both the accumulation of p27 during S and G_2 phases and the consequent inhibition of mitotic cyclin-Cdc2 activity. Consistent with this idea, experimental inhibition of the kinase activity of Cdc2 in budding yeast, fission yeast, and *Drosophila* forces cells that are normally mitotic to become endoreplicative (Edgar and Orr-Weaver, 2001). In the present study, we show that this is also the case, at least in part, in mammals. Cell cycle synchronization of *Skp2*$^{-/-}$ MEFs also revealed a delay at the G_2-M boundary in comparison with wild-type and *Skp2*$^{-/-}$ *p27*$^{-/-}$ cells (Supplemental Figure S4 at http://www.developmentalcell.com/cgi/content/full/6/5/661/DC1). These data suggest that inactivation of Cdc2 by p27 that accumulates as a result of the lack of Skp2 leads to G_2-M block, although the detailed mechanism of this phenomenon remains to be determined. We thus propose that the major target of p27 at G_1 phase is Cdk2, and that at G_2 phase it may

be Cdc2; Skp2 seems to be important for the regulation of the latter.

Endoreplication of Hepatocytes Induced by Deregulation of p27 Degradation

Polyploidy is apparent in certain tissues of *Skp2*$^{-/-}$ mice including the liver. Maintenance of genome ploidy is a fundamental aspect of cell division. Three possible mechanisms might render cells polyploid. (1) Failure of mitosis. Cells enter mitosis normally but anaphase fails to occur, resulting in the subsequent entry of the cells into interphase with a doubled DNA content. (2) Endoreplication. Cells replicate their genomes in S phase, bypass mitosis, and double their DNA content again in the next S phase. (3) DNA rereplication. Cells arrest in S phase and reinitiate DNA replication continuously. We adopted two criteria to determine which of these three mechanisms gives rise to the polyploidy of *Skp2*$^{-/-}$ hepatocytes. First, we examined whether cells entered mitosis. Immunofluorescence analysis with antibodies to phosphorylated histone H3, a marker of early mitosis, revealed that the number of mitotic cells is reduced in the liver of *Skp2*$^{-/-}$ mice, suggesting that failure of mitosis is not likely responsible for the polyploidy in these animals. Second, we measured the DNA content of *Skp2*$^{-/-}$ hepatocytes after mitogenic stimulation in order to determine whether the increase in DNA content occurred in a stepwise manner. The DNA content of *Skp2*$^{-/-}$ cells was shown to increase in multiples of 2C, a characteristic of endoreplication, rather than in a continuous manner, as would be consistent with DNA rereplication (data not shown). Similar results were also obtained after partial hepatectomy in *Skp2*$^{-/-}$ mice (Minamishima et al., 2002). We thus conclude that the polyploidy of *Skp2*$^{-/-}$ hepatocytes is likely attributable to endoreplication, although the possibility of a mitotic defect cannot be formally ruled out.

The regulation of Cdk activity during endoreplication in mammalian cells, with the exception of megakaryocytes and trophoblasts, is poorly understood (Edgar and Orr-Weaver, 2001). Consistent with the association of Skp2 with Cul1 in the SCF ubiquitin ligase complex, the phenotype of trophoblasts of *Cul1*$^{-/-}$ embryos is similar to that of *Skp2*$^{-/-}$ cells (Dealy et al., 1999; Wang et al., 1999). Unlike most cells, trophoblasts undergo multiple rounds of DNA synthesis without an intervening mitosis, resulting in the formation of giant nuclei (Zybina and Zybina, 1996). The lack of Cul1 or Skp2 appears to augment this process, and our present data indicate that the accumulation of p27 may play an important role in its induction. The abundance of cyclin E remains high and that of cyclin B is reduced in a differentiating trophoblast cell line, and p57, which is structurally and functionally similar to p27, is upregulated during S phase in these cells (Hattori et al., 2000). These characteristics resemble those of *Skp2*$^{-/-}$ cells. However, ectopic expression of a form of p57 with a mutation that stabilizes the protein blocks S phase entry in the trophoblast cell line. The gradual increase in the abundance of p27 in *Skp2*$^{-/-}$ cells may give rise to a window in which p27 inhibits mitosis but not entry into S phase, whereas forced expression of p57 at high levels may block S phase entry immediately.

A fission yeast mutant that lacks the F box protein Pop1, which targets the CKI Rum1 for degradation, also exhibits endoreplication and consequent polyploidy (Kominami and Toda, 1997). Rum1 accumulates to high levels in this mutant. Maintenance of ploidy in fission yeast is controlled by a complex of Cdc2 with the mitotic cyclin Cdc13. Inhibition of the kinase activity of Cdc2-Cdc13 as a result of the increased expression of Rum1 in the *pop1* mutant thus likely leads to polyploidy by promoting the bypass of M phase before the next S phase. A mitotic cyclin-Cdk complex in budding yeast prevents endoreplication through multiple overlapping mechanisms, including phosphorylation of the origin recognition complex (ORC), downregulation of Cdc6 activity, and exclusion from the nucleus of the Mcm2-7 complex (Nguyen et al., 2001). Given the similarities in the phenotypes of the fission yeast *pop1* mutant and mouse *Skp2*$^{-/-}$ cells, the accumulation of p27 in *Skp2*$^{-/-}$ cells may functionally correspond to that of Rum1 in the *pop1* mutant. The prevention of endoreplication through degradation of a CKI mediated by an SCF ubiquitin ligase thus appears to be a mechanism that has been well conserved from yeast to mammals.

Free Cyclin E as a Potential Substrate of SCF^{Skp2}

The marked accumulation of cyclin E in the absence of the antagonizing action of p27 in *Skp2*$^{-/-}$ *p27*$^{-/-}$ mice might have been expected either to compromise tissue organization in the developing embryo, and thereby to result in early embryonic death, or to lead to a high incidence of carcinogenesis. The apparent absence of such outcomes suggests that the deregulated expression of cyclin E might not be a direct cause of carcinogenesis, even though altered expression of cyclin E is apparent in many human cancers. The cyclin E that accumulates in *Skp2*$^{-/-}$ mice, however, appears to be the free form of the protein (Nakayama et al., 2000), not that complexed with Cdk2, and therefore does not contribute to the kinase activity of Cdk2. The nature of the interaction between the pool of free cyclin E and the cyclin E-Cdk2 complex is unclear, but our present data suggest that a simple increase in the abundance of free cyclin E does not directly result in the activation of Cdk2. Another possible interpretation of our data is that the increase in the pool of free cyclin E in *Skp2*$^{-/-}$ cells results in an increase in the activity of the cyclin E-Cdk2 complex, but that this effect is antagonized by the accumulated p27.

Our data do not necessarily imply that SCF^{Skp2} is the main mediator of cyclin E turnover. Rather, the mechanisms for the degradation of cyclin E appear complex. The Skp2-Cul1 complex and Cul3 interact with the free, nonphosphorylated form of cyclin E (Dealy et al., 1999; Singer et al., 1999; Wang et al., 1999; Nakayama et al., 2000), thereby mediating its ubiquitylation-dependent proteolysis. In parallel with these pathways, SCF^{Fbw7} is thought to target phosphorylated cyclin E complexed with Cdk2 (Clurman et al., 1996; Won and Reed, 1996; Koepp et al., 2001; Moberg et al., 2001; Strohmaier et al., 2001). However, mice that lack Fbw7, which die in utero during embryogenesis, show neither accumulation of cyclin E nor an increase in Cdk2 activity (Tsunematsu et al., 2004), whereas *Skp2*$^{-/-}$ mice manifest marked

accumulation of cyclin E in the nucleus. Thus, Fbw7 appears to be dispensable for cyclin E degradation, at least until mid-embryogenesis. In contrast, depletion of Fbw7 by RNA interference results in the accumulation of cyclin E in cultured cells (Koepp et al., 2001) (our unpublished observations), suggesting that Fbw7 might be required for cyclin E turnover in adult tissues. These data render unlikely, however, the possibility that the loss of Skp2 might stabilize other proteins that impair the Fbw7-dependent pathway of protein degradation and are thereby responsible for the accumulation of cyclin E.

Given that p27 accumulates to a high level in *Skp2*$^{-/-}$ cells, it is possible that autophosphorylation of cyclin E complexed with Cdk2 is inhibited by this CKI, resulting in stabilization of cyclin E. The accumulation of cyclin E might thus be secondary to the increase in the abundance of p27. Our present data, however, have shown that cyclin E degradation is also impaired in *Skp2*$^{-/-}$ *p27*$^{-/-}$ mice, providing genetic evidence that the altered expression of cyclin E in *Skp2*$^{-/-}$ cells is independent of p27 accumulation. We have previously shown that Skp2 interacts with free cyclin E and promotes its ubiquitylation both in vitro and in vivo, and that cyclin E degradation is impaired, resulting in loss of periodicity of cyclin E expression, in *Skp2*$^{-/-}$ cells (Nakayama et al., 2000). These biochemical observations are therefore consistent with the genetic evidence that Skp2 directly targets cyclin E. It remains possible that the accumulation of CKIs such as p21, p107, and p130 in *Skp2*$^{-/-}$ mice contributes to the stabilization of cyclin E, although our observation that the kinase activity of Cdk2 is unchanged in *Skp2*$^{-/-}$ cells suggests against this possibility.

Experimental Procedures

Mice

Both Skp2-deficient mice and p27-deficient mice were generated in our laboratory (Nakayama et al., 1996, 2000). We have developed polymerase chain reaction-based protocols (details available on request) to identify wild-type and disrupted alleles of *Skp2* and *p27*.

Histological and Immunofluorescence Analyses

Histological analysis and immunofluorescence analysis of centrosomes and microtubules were performed as described (Nakayama et al., 2000).

Preparation of MEFs

Primary MEFs were isolated from embryos on embryonic day 13.5 and cultured as previously described (Nakayama et al., 1996). In asynchronous culture, there was no substantial difference in cell cycle profiles among wild-type, *Skp2*$^{-/-}$, *p27*$^{-/-}$, and *Skp2*$^{-/-}$ *p27*$^{-/-}$ MEFs (Supplemental Figure S4A at http://www.developmentalcell.com/cgi/content/full/6/5/661/DC1).

Flow Cytometry

Flow cytometric analysis of hepatocytes and MEFs was performed as described previously (Nakayama et al., 1996, 2000).

Immunoblot Analysis

Transfection, immunoprecipitation, and immunoblot analysis were performed as previously described (Hatakeyama et al., 1999; Kitagawa et al., 1999). Antibodies used in this study include those to cyclin A (H-432, Santa Cruz Biotechnology), cyclin B (GNS-1, PharMingen), cyclin E (M-20, Santa Cruz Biotechnology), Cdk2 (M2, Santa Cruz Biotechnology), Cdc2 (17, Santa Cruz Biotechnology), p27 (57, Transduction Laboratories), or α-tubulin (TU-01, Zymed), each at a concentration of 0.2 μg/ml.

In some experiments, subconfluent COS-7 cells grown in four 100 mm dishes were transfected with 5 μg of pcDNA3 or pcDNA-FLAG-p27 and with 2 μg of pCD19 (Tedder and Isaacs, 1989) per dish with the use of the FuGENE6 reagent (Roche Molecular Biochemicals). Twenty-four hours after transfection, the CD19-expressing cells were collected with antibodies to CD19 attached to magnetic beads (Dynal Biotech) and were used for immune-complex kinase assays and immunoprecipitation.

Immune-Complex Kinase Assays

Kinase activity associated with Cdk2, Cdc2, or cyclins A, B, or E was measured with an immune-complex kinase assay as described (Nakayama et al., 1996).

Mitogenic Stimulation of Mouse Hepatocytes by Oral Administration of Estriol

Male mice at 8 weeks of age were subjected to oral administration of 0.4 mg of estriol (Wako) per gram of body mass. They were also subjected to daily intraperitoneal injection of BrdU (100 μg/g) (Sigma). Colchicine (1 μg/g) (Wako) was injected intraperitoneally 6 hr before killing. Sections of the liver were stained with hematoxylin-eosin or were subjected to immunostaining with rabbit polyclonal antibodies to phosphorylated histone H3 (Upstate Biotechnology) followed by Cy3-conjugated goat antibodies to rabbit immunoglobulin G (Chemicon). For detection of BrdU, sections were incubated with a rat monoclonal antibody to BrdU (Harlan Sera-Lab) at a dilution of 1:200 followed by biotinylated secondary antibodies; immune complexes were visualized with the use of a streptavidin-biotin-peroxidase detection kit (Vector) and diaminobenzidine (Wako).

Acknowledgments

We thank M. Fujita for helpful discussions, R. Tsunematsu, M. Matsumoto, T. Hara, Y. Yamada, K. Shimoharada, S. Matsushita, N. Nishimura, R. Yasukochi, A. Ito, Y. Ono, and other laboratory members for technical assistance, and M. Kimura and C. Sugita for help in preparation of the manuscript. This work was supported in part by a research grant from the Human Frontier Science Program.

Received: September 25, 2003
Revised: March 19, 2004
Accepted: March 24, 2004
Published: May 10, 2004

References

Bhattacharya, S., Garriga, J., Calbo, J., Yong, T., Haines, D.S., and Grana, X. (2003). SKP2 associates with p130 and accelerates p130 ubiquitylation and degradation in human cells. Oncogene *22*, 2443–2451.

Bornstein, G., Bloom, J., Sitry-Shevah, D., Nakayama, K., Pagano, M., and Hershko, A. (2003). Role of the SCF^{Skp2} ubiquitin ligase in the degradation of p21^{Cip1} in S phase. J. Biol. Chem. *278*, 25752–25757.

Carrano, A.C., Eytan, E., Hershko, A., and Pagano, M. (1999). SKP2 is required for ubiquitin-mediated degradation of the CDK inhibitor p27. Nat. Cell Biol. *1*, 193–199.

Charrasse, S., Carena, I., Brondani, V., Klempnauer, K.H., and Ferrari, S. (2000). Degradation of B-Myb by ubiquitin-mediated proteolysis: involvement of the Cdc34-SCF(p45^{Skp2}) pathway. Oncogene *19*, 2986–2995.

Clurman, B.E., Sheaff, R.J., Thress, K., Groudine, M., and Roberts, J.M. (1996). Turnover of cyclin E by the ubiquitin-proteasome pathway is regulated by cdk2 binding and cyclin phosphorylation. Genes Dev. *10*, 1979–1990.

Connor, M.K., Kotchetkov, R., Cariou, S., Resch, A., Lupetti, R., Beniston, R.G., Melchior, F., Hengst, L., and Slingerland, J.M. (2003). CRM1/Ran-mediated nuclear export of p27(Kip1) involves a nuclear export signal and links p27 export and proteolysis. Mol. Biol. Cell *14*, 201–213.

Correa-Bordes, J., and Nurse, P. (1995). p25rum1 orders S phase

and mitosis by acting as an inhibitor of the p34Cdc2 mitotic kinase. Cell *83*, 1001–1009.

Dealy, M., Nguyen, K.V.T., Lo, J., Gstaiger, M., Krek, W., Elson, D., Arbeit, J., Kipreos, E.T., and Johnson, R.S. (1999). Loss of Cul1 results in early embryonic lethality and dysregulation of cyclin E. Nat. Genet. *23*, 245–248.

Edgar, B.A., and Orr-Weaver, T.L. (2001). Endoreplication cell cycles: more for less. Cell *105*, 297–306.

Fero, M.L., Rivkin, M., Tasch, M., Porter, P., Carow, C.E., Firpo, E., Polyak, K., Tsai, L.H., Broudy, V., Perlmutter, R.M., et al. (1996). A syndrome of multiorgan hyperplasia with features of gigantism, tumorigenesis, and female sterility in $p27^{Kip1}$-deficient mice. Cell *85*, 733–744.

Fujii, H., Hayama, T., and Kotani, M. (1985). Stimulating effect of natural estrogens on proliferation of hepatocytes in adult mice. Acta Anat. *121*, 174–178.

Hara, T., Kamura, T., Nakayama, K., Oshikawa, K., Hatakeyama, S., and Nakayama, K.I. (2001). Degradation of $p27^{Kip1}$ at the G_0-G_1 transition mediated by a Skp2-independent ubiquitination pathway. J. Biol. Chem. *276*, 48937–48943.

Hatakeyama, S., Kitagawa, M., Nakayama, K., Shirane, M., Matsumoto, M., Hattori, K., Higashi, H., Nakano, H., Okumura, K., Onoe, K., et al. (1999). Ubiquitin-dependent degradation of IκBα is mediated by a ubiquitin ligase Skp1/Cul 1/F-box protein FWD1. Proc. Natl. Acad. Sci. USA *96*, 3859–3863.

Hattori, N., Davies, T.C., Anson-Cartwright, L., and Cross, J.C. (2000). Periodic expression of the cyclin-dependent kinase inhibitor $p57^{Kip2}$ in trophoblast giant cells defines a G2-like gap phase of the endocycle. Mol. Biol. Cell *11*, 1037–1045.

Hengst, L., and Reed, S.I. (1996). Translational control of $p27^{Kip1}$ accumulation during the cell cycle. Science *271*, 1861–1864.

Ishida, N., Hara, T., Kamura, T., Yoshida, M., Nakayama, K., and Nakayama, K.I. (2002). Phosphorylation of $p27^{Kip1}$ on serine 10 is required for its binding to CRM1 and nuclear export. J. Biol. Chem. *277*, 14355–14358.

Kamura, T., Hara, T., Kotoshiba, S., Yada, M., Ishida, N., Imaki, H., Hatakeyama, S., Nakayama, K., and Nakayama, K.I. (2003). Degradation of $p57^{Kip2}$ mediated by SCF^{Skp2}-dependent ubiquitylation. Proc. Natl. Acad. Sci. USA *100*, 10231–10236.

Kiernan, R.E., Emiliani, S., Nakayama, K., Castro, A., Labbe, J.C., Lorca, T., Nakayama, K.I., and Benkirane, M. (2001). Interaction between cyclin T1 and SCF^{SKP2} targets CDK9 for ubiquitination and degradation by the proteasome. Mol. Cell. Biol. *21*, 7956–7970.

Kim, S.Y., Herbst, A., Tworkowski, K.A., Salghetti, S.E., and Tansey, W.P. (2003). Skp2 regulates myc protein stability and activity. Mol. Cell *11*, 1177–1188.

Kitagawa, M., Okabe, T., Ogino, H., Matsumoto, H., Suzuki-Takahashi, I., Kokubo, T., Higashi, H., Saitoh, S., Taya, Y., Yasuda, H., et al. (1993). Butyrolactone I, a selective inhibitor of cdk2 and Cdc2 kinase. Oncogene *8*, 2425–2432.

Kitagawa, M., Hatakeyama, S., Shirane, M., Matsumoto, M., Ishida, N., Hattori, K., Nakamichi, I., Kikuchi, A., Nakayama, K.I., and Nakayama, K. (1999). An F-box protein, FWD1, mediates ubiquitin-dependent proteolysis of β-catenin. EMBO J. *18*, 2401–2410.

Kiyokawa, H., Kineman, R.D., Manova-Todorova, K.O., Soares, V.C., Hoffman, E.S., Ono, M., Khanam, D., Hayday, A.C., Frohman, L.A., and Koff, A. (1996). Enhanced growth of mice lacking the cyclin-dependent kinase inhibitor function of $p27^{Kip1}$. Cell *85*, 721–732.

Koepp, D.M., Schaefer, L.K., Ye, X., Keyomarsi, K., Chu, C., Harper, J.W., and Elledge, S.J. (2001). Phosphorylation-dependent ubiquitination of cyclin E by the SCF^{Fbw7} ubiquitin ligase. Science *294*, 173–177.

Kominami, K., and Toda, T. (1997). Fission yeast WD-repeat protein pop1 regulates genome ploidy through ubiquitin-proteasome-mediated degradation of the CDK inhibitor Rum1 and the S-phase initiator Cdc18. Genes Dev. *11*, 1548–1560.

Li, X., Zhao, Q., Liao, R., Sun, P., and Wu, X. (2003). The SCF^{Skp2} ubiquitin ligase complex interacts with the human replication licensing factor Cdt1 and regulates Cdt1 degradation. J. Biol. Chem. *278*, 30854–30858.

Marti, A., Wirbelauer, C., Scheffner, M., and Krek, W. (1999). Interaction between ubiquitin-protein ligase SCF^{SKP2} and E2F–1 underlies the regulation of E2F–1 degradation. Nat. Cell Biol. *1*, 14–19.

Mendez, J., Zou-Yang, X.H., Kim, S.Y., Hidaka, M., Tansey, W.P., and Stillman, B. (2002). Human origin recognition complex large subunit is degraded by ubiquitin-mediated proteolysis after initiation of DNA replication. Mol. Cell *9*, 481–491.

Minamishima, Y.A., Nakayama, K., and Nakayama, K.I. (2002). Recovery of liver mass without proliferation of hepatocytes after partial hepatectomy in Skp2-deficient mice. Cancer Res. *62*, 995–999.

Miura, M., Hatakeyama, S., Hattori, K., and Nakayama, K.I. (1999). Structure and expression of the gene encoding mouse F-Box protein, Fwd2. Genomics *62*, 50–58.

Moberg, K.H., Bell, D.W., Wahrer, D.C., Haber, D.A., and Hariharan, I.K. (2001). Archipelago regulates Cyclin E levels in Drosophila and is mutated in human cancer cell lines. Nature *413*, 311–316.

Montagnoli, A., Fiore, F., Eytan, E., Carrano, A.C., Draetta, G.F., Hershko, A., and Pagano, M. (1999). Ubiquitination of p27 is regulated by Cdk-dependent phosphorylation and trimeric complex formation. Genes Dev. *13*, 1181–1189.

Nagahama, H., Hatakeyama, S., Nakayama, K., Nagata, M., Tomita, K., and Nakayama, K.I. (2001). Spatial and temporal expression patterns of the cyclin-dependent kinase (CDK) inhibitors $p27^{Kip1}$ and $p57^{Kip2}$ during mouse development. Anat. Embryol. *203*, 77–87.

Nakayama, K., Ishida, N., Shirane, M., Inomata, A., Inoue, T., Shishido, N., Horii, I., Loh, D.Y., and Nakayama, K.I. (1996). Mice lacking $p27^{Kip1}$ display increased body size, multiple organ hyperplasia, retinal dysplasia, and pituitary tumors. Cell *85*, 707–720.

Nakayama, K., Nagahama, H., Minamishima, Y.A., Matsumoto, M., Nakamichi, I., Kitagawa, K., Shirane, M., Tsunematsu, R., Tsukiyama, T., Ishida, N., et al. (2000). Targeted disruption of Skp2 results in accumulation of cyclin E and $p27^{Kip1}$, polyploidy and centrosome overduplication. EMBO J. *19*, 2069–2081.

Nakayama, K.I., Hatakeyama, S., and Nakayama, K. (2001). Regulation of the cell cycle at the G_1-S transition by proteolysis of cyclin E and $p27^{Kip1}$. Biochem. Biophys. Res. Commun. *282*, 853–860.

Nguyen, V.Q., Co, C., and Li, J.J. (2001). Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms. Nature *411*, 1068–1073.

Pagano, M., Tam, S.W., Theodoras, A.M., Beer-Romero, P., Del Sal, G., Chau, V., Yew, P.R., Draetta, G.F., and Rolfe, M. (1995). Role of the ubiquitin-proteasome pathway in regulating abundance of the cyclin-dependent kinase inhibitor p27. Science *269*, 682–685.

Rodier, G., Montagnoli, A., Marcotullio, D.L., Coulombe, P., Draetta, D.G., Pagano, M., and Meloche, S. (2001). p27 cytoplasmic localization is regulated by phosphorylation on Ser10 and is not a prerequisite for its proteolysis. EMBO J. *20*, 6672–6682.

Sheaff, R.J., Groudine, M., Gordon, M., Roberts, J.M., and Clurman, B.E. (1997). Cyclin E-CDK2 is a regulator of $p27^{Kip1}$. Genes Dev. *11*, 1464–1478.

Shirane, M., Harumiya, Y., Ishida, N., Hirai, A., Miyamoto, C., Hatakeyama, S., Nakayama, K.I., and Kitagawa, M. (1999). Down-regulation of $p27^{Kip1}$ by two mechanisms, ubiquitin-mediated degradation and proteolytic processing. J. Biol. Chem. *274*, 13886–13893.

Singer, J.D., Gurian-West, M., Clurman, B., and Roberts, J.M. (1999). Cullin-3 targets cyclin E for ubiquitination and controls S phase in mammalian cells. Genes Dev. *13*, 2375–2387.

Stern, B., and Nurse, P. (1996). A quantitative model for the Cdc2 control of S phase and mitosis in fission yeast. Trends Genet. *12*, 345–350.

Strohmaier, H., Spruck, C.H., Kaiser, P., Won, K.A., Sangfelt, O., and Reed, S.I. (2001). Human F-box protein hCdc4 targets cyclin E for proteolysis and is mutated in a breast cancer cell line. Nature *413*, 316–322.

Sutterluty, H., Chatelain, E., Marti, A., Wirbelauer, C., Senften, M.,

Muller, U., and Krek, W. (1999). $p45^{SKP2}$ promotes $p27^{Kip1}$ degradation and induces S phase in quiescent cells. Nat. Cell Biol. *1*, 207–214.

Tedder, T.F., and Isaacs, C.M. (1989). Isolation of cDNAs encoding the CD19 antigen of human and mouse B lymphocytes. J. Immunol. *143*, 712–717.

Tedesco, D., Lukas, J., and Reed, S.I. (2002). The pRb-related protein p130 is regulated by phosphorylation-dependent proteolysis via the protein-ubiquitin ligase SCF^{Skp2}. Genes Dev. *16*, 2946–2957.

Th'ng, J.P., Wright, P.S., Hamaguchi, J., Lee, M.G., Norbury, C.J., Nurse, P., and Bradbury, E.M. (1990). The FT210 cell line is a mouse G2 phase mutant with a temperature-sensitive CDC2 gene product. Cell *63*, 313–324.

Tomoda, K., Kubota, Y., and Kato, J. (1999). Degradation of the cyclin-dependent-kinase inhibitor $p27^{Kip1}$ is instigated by Jab1. Nature *398*, 160–165.

Tsunematsu, R., Nakayama, K., Oike, Y., Nishiyama, M., Ishida, N., Hatakeyama, S., Bessho, Y., Kageyama, R., Suda, T., and Nakayama, K.I. (2004). Mouse Fbw7/Sel-10/Cdc4 is required for notch degradation during vascular development. J. Biol. Chem. *279*, 9417–9423.

Tsvetkov, L.M., Yeh, K.H., Lee, S.J., Sun, H., and Zhang, H. (1999). $p27^{Kip1}$ ubiquitination and degradation is regulated by the SCF^{Skp2} complex through phosphorylated Thr187 in p27. Curr. Biol. *9*, 661–664.

Vlach, J., Hennecke, S., and Amati, B. (1997). Phosphorylation-dependent degradation of the cyclin-dependent kinase inhibitor p27. EMBO J. *16*, 5334–5344.

von der Lehr, N., Johansson, S., Wu, S., Bahram, F., Castell, A., Cetinkaya, C., Hydbring, P., Weidung, I., Nakayama, K., Nakayama, K.I., et al. (2003). The F-box protein Skp2 participates in c-Myc proteasomal degradation and acts as a cofactor for c-Myc-regulated transcription. Mol. Cell *11*, 1189–1200.

Wang, Y., Penfold, S., Tang, X., Hattori, N., Riley, P., Harper, J.W., Cross, J.C., and Tyers, M. (1999). Deletion of the Cul1 gene in mice causes arrest in early embryogenesis and accumulation of cyclin E. Curr. Biol. *9*, 1191–1194.

Won, K.A., and Reed, S.I. (1996). Activation of cyclin E/CDK2 is coupled to site-specific autophosphorylation and ubiquitin-dependent degradation of cyclin E. EMBO J. *15*, 4182–4193.

Yu, Z.K., Gervais, J.L., and Zhang, H. (1998). Human CUL-1 associates with the SKP1/SKP2 complex and regulates p21(CIP1/WAF1) and cyclin D proteins. Proc. Natl. Acad. Sci. USA *95*, 11324–11329.

Zybina, E.V., and Zybina, T.G. (1996). Polytene chromosomes in mammalian cells. Int. Rev. Cytol. *165*, 53–119.

Analysis of Small Human Proteins Reveals the Translation of Upstream Open Reading Frames of mRNAs

Masaaki Oyama,[1] Chiharu Itagaki,[2] Hiroko Hata,[3] Yutaka Suzuki,[1] Tomonori Izumi,[2] Tohru Natsume,[4] Toshiaki Isobe,[2] and Sumio Sugano[1,5]

[1]*Human Genome Center, Division of* [2]*Proteomics Research and* [3]*Cancer Genomics, Institute of Medical Science, University of Tokyo, Minato-ku, Tokyo 108-8639, Japan;* [4]*National Institute of Advanced Industrial Science and Technology, Biological Information Research Center (JBIRC), Koutoh-ku, Tokyo 135-0064, Japan*

To find novel short coding sequences from accumulated full-length cDNA sequences, proteomic analysis of small proteins expressed in human leukemia K562 cells was performed using high-resolution nanoflow liquid chromatography coupled with electrospray ionization tandem mass spectrometry. Our analysis led to the identification of 54 proteins not more than 100 amino acids in length, including four novel ones. These novel short coding sequences were all located upstream of the longest open reading frame (ORF) of the corresponding cDNA. Our findings indicate that the translation of short ORFs occurs in vivo whether or not there exists a longer coding region in the downstream of the mRNA. This investigation provides the first direct evidence of translation of upstream ORFs in human cells, which could greatly change the current outline of the human proteome.

[Supplemental material is available online at www.genome.org.]

In parallel with human genome sequencing projects (Lander et al. 2001; Venter et al. 2001), the accumulation of sequence data of human full-length cDNAs has also been proceeding. The "RefSeq collection" (NCBI) provides us with representative resources of curated human full-length cDNAs, and the protein-coding sequence (CDS) of each cDNA is defined in the RefSeq database (Pruitt and Maglott 2001). Now a total of 19,995 proteins are stored in the RefSeq curated human protein database (as of January 27, 2004), and 19,271 (96.4%) of them are longer than 100 amino acids. This indicates that small proteins with ≦100 amino acids are only a limited fraction of all the proteins annotated in the RefSeq database.

According to the typical translation model, a 40S ribosomal subunit is first recruited to the cap structure of mRNA and linearly scans the 5′-UTR for the initiator ATG. When it recognizes the initiator ATG, it pauses until a large 60S subunit joins, and the complete ribosomal complex starts translation (Kozak 1989). Therefore, the most upstream ORF should be translated according to this model, much more with a good context around its ATG codon as previously analyzed (Kozak 1999). Some previous studies have reported that the short ORF in the 5′-untranslated region (UTR) functions as a regulator of the translation of its downstream CDS (Morris and Geballe 2000; Meijer and Thomas 2002). It has been considered that such translational control would be limited to some genes or conditions. However, the previous large-scale analyses focusing on the 5′-UTRs of human full-length cDNA sequences showed that 41%–49% of them had at least one ATG codon upstream of the CDS (Peri and Pandey 2001; Yamashita et al. 2003). This means that there are potential short coding regions in the 5′-UTRs of many genes if this classical model, indeed, represents a general mechanism of translation initiation. To our knowledge, few reports have presented evidence of the translation of upstream ORFs in vivo (Diba et al. 2001). Although there are also some mechanisms by which the ribosomal complex may evade the translation from the first ATG codon, such as leaky scanning (Kozak 1999) and IRES (internal ribosome entry site)-dependent translation (Meijer and Thomas 2002), we expect that the small proteins encoded by upstream ORFs in 5′-UTRs exist in vivo.

With a view to finding novel short upstream CDSs in accumulated cDNA sequences, we performed a proteomic analysis of small proteins expressed in vivo using direct nanoflow liquid chromatography (LC) coupled with the electrospray ionization (ESI)–tandem mass spectrometry (MS/MS) system (Natsume et al. 2002). This LC instrument can separate peptides and introduce them into a mass spectrometer with limited diffusion, leading to more sensitive detection than can be achieved with conventional LC systems. We aimed to identify novel short CDSs by searching not only against the RefSeq curated cDNA database but also against our in-house FLJ-unique cDNA data set, which contained as many as >10,000 full-length cDNA sequences that had no hit against the RefSeq cDNAs (Ota et al. 2004).

Here we report the proteomic analysis of small proteins (≦100 amino acids in length) expressed in human chronic myelogenous leukemia K562 cells. Our analysis led to the identification of 54 proteins in total, including four novel ones. Very intriguingly, these novel small proteins were all derived from the short ORFs in the presumed 5′-UTRs.

RESULTS

To carry out a proteomic analysis of small proteins expressed in K562 cells, we prepared the samples for mass spectrometric analysis by two different methods. Small proteins were isolated by either fractionation through SDS-PAGE or acid extraction. For the proteins resolved by SDS-PAGE, the part of the gel corresponding to the low molecular weight (<~17 kDa) was excised, and the proteins trapped in the gel were digested with proteolytic enzymes (see "MS Sample Preparation 1" in Methods). On the

[5]**Corresponding author.**
E-MAIL ssugano@ims.u-tokyo.ac.jp; FAX 81-3-5449-5416.
Article and publication are at http://www.genome.org/cgi/doi/10.1101/gr.2384604.

14:2048–2052 ©2004 by Cold Spring Harbor Laboratory Press ISSN 1088-9051/04; www.genome.org

Table 1. Novel Short CDSs Identified by Searching Against Human cDNAs

GenBank Accession	Length (bp)	Novel CDS position[a]	Novel CDS length[a] (amino acids)	Novel CDS initiator ATG[a]	Identified peptide[b]	Longest ORF position[a]	Longest ORF length[a] (amino acids)
RefSeq cDNAs							
NM_005770	1408	941...1120	59	6th	QRDSEIMQQK RDDGLSAAAR	1023...1319	98
NM_015532	4107	12...272	86	1st	QPQPAQNVLAAPR GLGAAEFGGAAGNVEAPGETFAQR	127...1233	368
NM_016215	1545	150...401	83	1st	ATPGLQQHQQPPGPGR (N-terminus acetylated)	316...1137	273
FLJ cDNAs							
AK057257	1904	23...280	85	1st	LLPLGASPAGVVGGGLAPPR	654...1013	119

[a]Each data is based on the sequence information of the corresponding cDNA.
[b]Peptides identified from MS samples (see Methods).

other hand, the small proteins enriched by extraction in acid solution were digested directly without PAGE separation (see "MS Sample Preparation 2" in Methods). After concentrating the peptide mixtures prepared by each of the two methods, we applied them to the nanoflow LC-MS/MS system.

We first tried identifying small proteins (≦100 amino acids in length) by searching against the RefSeq curated human "protein" database (NCBI). Accordingly, 36 proteins were identified from the gel-separated samples, and 23 proteins were identified from the acid-extracted samples. In total, 50 proteins (with nine overlaps) were identified out of 724 proteins (≦100 amino acids in length) stored in the RefSeq protein database (as of January 27, 2004; see Supplemental table). The range of amino acid length of the identified proteins was from 44 to 100. The list included various kinds of small proteins, such as ribosomal proteins, transporters, transcriptional regulators, cell cycle regulators, spliceosome components, and proteins involved in energy metabolism. We also found several function-unknown proteins expressed in K562 cells.

Next, to search for novel short CDSs that were not annotated in the RefSeq curated human protein database, all the MS/MS data that had no hit against RefSeq proteins were then searched against all the ORFs (in all three reading frames) of the RefSeq curated cDNA database and of our in-house "FLJ-unique" cDNA data set. As a result, four novel translated ORFs were identified (Table 1). Three of them were derived from RefSeq cDNAs, whereas the other was from an FLJ cDNA. As an example, the MS/MS spectrum matching the NM_015532 novel short CDS is shown in Figure 1A. An intense string of as many as 10 ions from the *y* ion series resulted in an excellent match for the corresponding peptide. The other peptides listed in Table 1 also yielded comparable search results, indicating the translation of these short ORFs. Moreover, the identification of the NM_015532 novel short CDS was also shown by matching of another peptide (Fig. 1B). This evidence gives us additional support for the presence of this novel CDS, which is also the case with the NM_005770 novel CDS (Table 1).

In Figure 2, we show the location of these novel CDSs within each corresponding cDNA. Interestingly, all the novel CDSs are located upstream of the longest ORF. Three of them overlap with each longest ORF, whereas the other is distant. A nucleotide deletion or insertion arising from a sequencing error can cause an erroneous short ORF to be produced from the longest ORF by a frameshift in the reading frame. Also, there might be splicing variants that can result in the fusion of the short ORF to the longest ORF. Therefore, careful confirmation of the corresponding nucleotide sequences was needed.

Here we tried aligning the EST data corresponding to the NM_015532 novel short CDS. As this novel short CDS is located near the 5′-end of the mRNA as shown in Figure 2, we aligned the 5′-end cDNA sequence data provided by the "oligo-capping" method, which was previously established by us for collecting accurate 5′-end nucleotide sequences from the mRNA start site (Suzuki et al. 1997, 2001). The multiple alignment of the sequence data of 11 corresponding cDNAs from 10 different resources showed no alternative splicing pattern over the entire region of this short CDS and a complete match for NM_015532

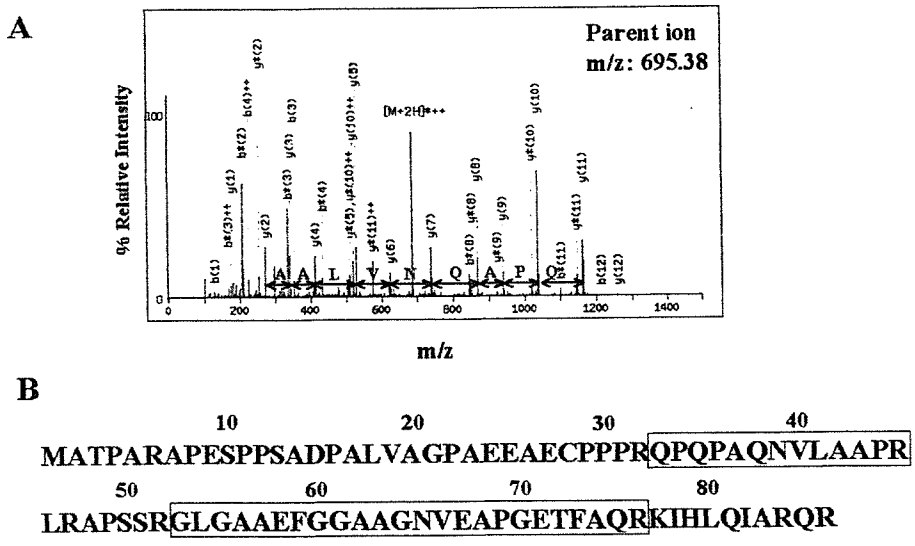

Figure 1 Identification of a novel short coding sequence by mass spectrometry. (*A*) MS/MS spectrum corresponding to the peptide QPQPAQNVLAAPR at *m/z* 695.38 derived from the NM_015532 ORF. The corresponding amino acid differences based on a series of as many as 10 continuous *y* ions are represented. An asterisk (*) indicates an ion that has lost ammonia from its side chain. (*B*) Amino acid sequence of the NM_015532 upstream short ORF. The identified peptides are surrounded by rectangles.

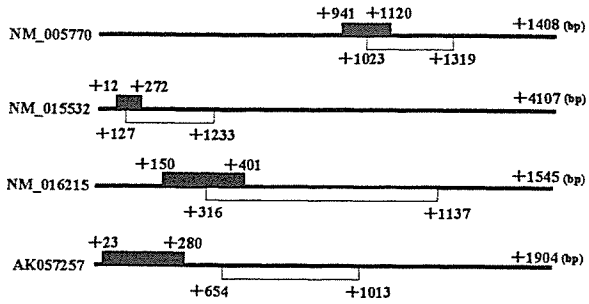

Figure 2 Location of the identified novel CDS (black box) and the longest ORF (white box) of each full-length cDNA. The numbers indicate the positions from each full-length cDNA start site.

around the termination codon as shown in Figure 3A (the detailed sequence data can be seen at DataBase of Transcriptional Start Sites; DBTSS; Release 3.0; http://dbtss.hgc.jp/; Suzuki et al. 2002). As for the other three novel short CDSs, accumulated EST evidence showed that there were also no frameshift or alternative splicing variants that indicated the existence of a fused and longer ORF.

Furthermore, comparative sequence analysis of the NM_015532 novel CDS and its mouse ortholog counterpart showed 86% DNA identity with 16 conservative changes and 15 nonconservative ones over the aligned ORFs (261 nt in length), resulting in 85% identity and 95% similarity across the entire length of their deduced amino acid sequences (86 amino acids in length; Fig. 3B). Evidence of such a high degree of evolutionary sequence conservation indicates functional constraint on this novel short CDS. Table 2 shows that the NM_005770 upstream CDS is also functionally constrained, whereas that of NM_016215 is relatively loosened. As for that of NM_005770, the previous study has indicated that it shares homology with the protein encoded by a candidate modifying gene for spinal muscular atrophy (Scharf et al. 1998).

DISCUSSION

Our proteomic analysis of small proteins expressed in K562 cells has enabled us to reveal the existence of the proteins encoded by upstream ORFs in 5′-UTRs. To our knowledge, this is the first direct evidence of translation of upstream ORFs in human cells. There were only four upstream short CDSs identified in our analysis, while leading to the identification of 50 RefSeq-annotated proteins. One of the reasons might be that some parts of upstream ORFs would not be efficiently translated in K562 cells because of the poor Kozak's context around their ATG codons. The previous studies indicated that 37%–57 % of the upstream ATGs in the 5′-UTRs had an unfavorable Kozak's context around the ATG codon (Kozak 1987; Suzuki et al. 2000). Secondly, our recent analysis also indicated that approximately three-fourths of the upstream ORFs analyzed were shorter than 40 amino acids (Yamashita et al. 2003). Considering that the smallest protein identified from the RefSeq curated database is 44 amino acids in length, it is very possible that such very small proteins were out of the detectable range in our analysis. They would be lost while preparing the samples for MS analysis.

However, these reasons from the statistical point of view cannot by themselves fully explain why there were no more than four proteins identified among thousands of upstream ORFs. One of the other possibilities is that many of the proteins derived from upstream ORFs might be selectively proteolyzed in the cells. Secondly, there might be some mechanisms that allow ribosomes to avoid the translation of upstream ORFs. Although IRES-dependent translation can permit ribosomes to directly enter a downstream site without encountering an upstream ATG, the previous studies have estimated that this mechanism would be applied to a limited fraction of genes (Meijer and Thomas 2002). There might exist another mechanism that enables ribosomes to escape the translation of an upstream ORF. Thirdly, the transcripts expressed in K562 cells might not reflect the corresponding cDNA sequences stored in the database. Our previous large-scale analysis on the 5′-UTRs has shown that the transcription start sites of many genes are more dispersed than was previously believed (Suzuki et al. 2001). Thus, it is likely that there are many genes whose transcripts in K562 cells have a shorter 5′-UTR that lacks an upstream ATG. Further analysis will be required to clarify this point.

Mapping of the novel short CDSs onto the corresponding full-length cDNAs indicates that three of these CDSs (but not the NM_005770 short CDS) use the most upstream ATG as an initiation codon (Table 1). As to the NM_005770 short CDS, the sixth ATG corresponds to its translation start site on the cDNA sequence of NM_005770. However, the accumulated oligo-capped 5′-end cDNA data of this gene obtained from various types of human tissues and cell lines uniformly showed the existence of the short transcript form whose first ATG corresponded to the initiation codon of this novel CDS (see the sequence data at DBTSS [Release 3.0]; http://dbtss.hgc.jp/). This indicates that this

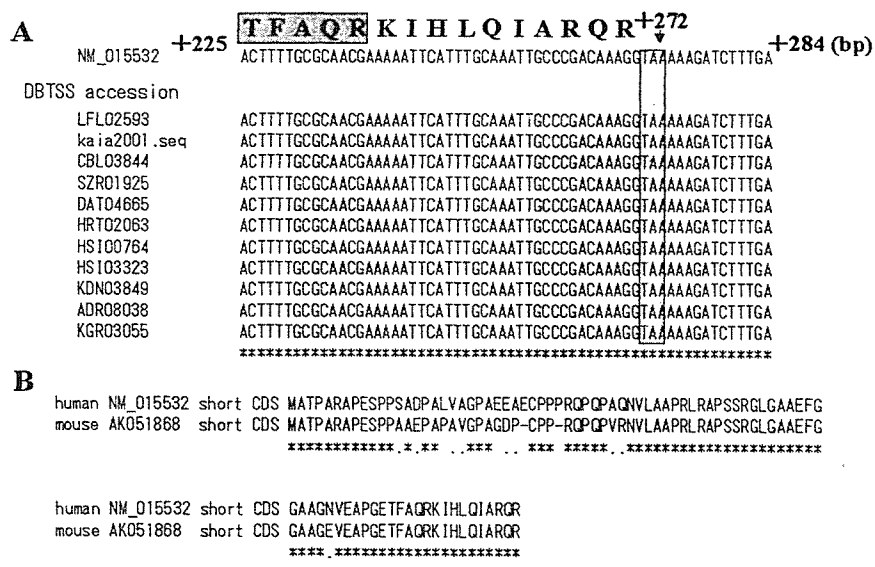

Figure 3 Sequence analysis of the NM_015532 novel short CDS. (*A*) Multiple alignment of the 5′-end EST data around the termination codon. The dark box shows the C terminus of the secondarily identified peptide. An asterisk (*) indicates a complete sequence match between the RefSeq cDNA (NM_015532) and all the EST data. The shaded box indicates the termination codon of this short CDS. (*B*) Alignment of the NM_015532 short CDS with its mouse ortholog at the amino acid level. The amino acid sequence of the mouse ortholog was deduced from the sequence (from +13 to +267) of AK051868. An asterisk (*) and a dot (.) indicate identity and similarity in amino acid sequence, respectively. (DBTSS) DataBase of Transcriptional Start Sites (http://dbtss.hgc.jp/; Suzuki et al. 2002).

Table 2. Sequence Conservation of the Upstream CDS and the Longest ORF of Each RefSeq cDNA

RefSeq ID	Upstream CDS[a] (%)	Longest ORF[a] (%)	Longest ORF RefSeq definition
NM_005770	100	94	Small EDRK-rich factor 2 (SERF2)
NM_015532	95	89	Glutamate receptor, ionotropic, *N*-methyl D-aspartate-like 1A (GRINL1A)
NM_016215	71	92	EGF-like domain, multiple 7 (EGFL7)

[a]Each value represents the rate of similar amino acid residues over the entire region in the alignment of the human protein sequence with that of its mouse ortholog using CLUSTAL W (http://www.ddbj.nig.ac.jp/E-mail/clustalw-j.html).
Each amino acid sequence was deduced from the corresponding nucleotide sequence region of each cDNA. The mouse orthologous regions were extracted from the cDNA data below.
NM_005770 [upstream CDS: NM_011354 (+19–+198); longest ORF: NM_011354 (+101–+397)].
NM_015532 [upstream CDS: AK051868 (+13–+267); longest ORF: AK051868 (+122–+1222)].
NM_016215 [upstream CDS: NM_198724 (+125–+367); longest ORF: NM_198724 (+294–+1130)].

gene has two alternative transcript forms and the majority is the shorter one. Thus, it is very possible that translation of this short CDS initiates from the first ATG of the corresponding short transcript in K562 cells. In the conventional mechanism of translation initiation, the first ATG should be recognized as an initiation codon (the first ATG rule; Kozak 1989). Our finding of these four upstream CDSs is supportive evidence for this rule. In addition, the classification on the probable translation initiation sites of the small RefSeq proteins identified in our analysis shows that 42 (84%) out of the 50 listed proteins use the first ATG as an initiation codon (see Supplemental table). These results indicate that the small proteins relatively abundant in K562 cells were mainly produced according to the first ATG rule. Much more evidence of the translation of upstream ORFs could demonstrate that the translation from the first ATG generally occurs, indeed.

As for NM_015532, NM_016215, and AK057257, the splicing junctions are left downstream of the translation termination site of each upstream CDS. The nonsense-mediated decay (NMD) pathway triggers the degradation of the transcripts holding exon junction complexes (EJCs), which should be removed by a migrating ribosome during the process of translation (Maquat 2004). Therefore, these transcripts are considered to be susceptible to degradation by this pathway. Translation of the downstream longest ORF can protect the transcripts from degradation through removal of the remaining EJCs. As described in Table 2, the longest ORFs of the three RefSeq genes are highly conserved between human and mouse and the mouse ortholog corresponding to that of NM_016215 has been characterized as an endothelial repressor of smooth muscle cell migration (Soncin et al. 2003). As for the longest ORF of AK057257, it shares strong homology with α-tubulin, which also indicates its functionality (data not shown). The investigation on whether the translation of these downstream longest ORFs occurs in K562 cells will be needed to consider this point.

Further explorations based on mass spectrometric analysis will lead to the identification of more short CDSs through improvement of the method for the fractionation of small proteins or through sophistication of the LC system to acquire more MS/MS spectra. The analysis of small proteins expressed in other cultured cells or tissues will also reveal the existence of those expressed in a tissue-specific manner. Accumulating evidence of the translation of upstream short ORFs will make it possible for us to obtain a clearer outline of translatable regions of numerous mRNA species and help us to determine the real size and contents of the human proteome.

METHODS

Cell Culture

Human chronic myelogenous leukemia K562 cells were grown in RPMI/10% dialyzed FCS to a density of 1×10^6 cells/mL, harvested, and washed three times with PBS.

MS Sample Preparation 1

Harvested K562 cells (5×10^6) were lysed in lysis buffer (50 mM Tris-HCl at pH 7.6, 0.5 % [w/v] Triton X-100, 150 mM NaCl) supplemented with protease inhibitor cocktail Complete mini (Boerhinger Mannheim) and centrifuged for 10 min at 12,000 rpm at 4°C. The obtained supernatant was separated by SDS-PAGE with a 14% lower gel. The part of the lane corresponding to <~17 kDa was cut into small pieces, and the proteins in the gel pieces were digested overnight at 37°C with 25 pmoles of trypsin, sequencing-grade (Roche Diagnostics) in 20 mM Tris-HCl (pH 8.8). These procedures were performed according to the previously described method (Shevchenko et al. 1996). The peptides were extracted from the gel pieces with a total of 300 µL of 50% (v/v) acetonitrile/5% formic acid by sonication and concentrated to ~50 µL using a centrifugal vacuum concentrator. After the sample was desalted using a ZipTip (C_{18}; Millipore), the peptides were eluted with 50 µL of 70% (v/v) acetonitrile/0.1% (v/v) formic acid and again concentrated to a final volume of 5 µL.

MS Sample Preparation 2

Harvested K562 cells (5×10^8) were boiled for 10 min at 95°C to inactivate proteases. The cells were then lysed in 1 M acetic acid using a Dounce homogenizer on ice and centrifuged for 10 min at 12,000 rpm at 4°C. After eliminating salts and other low-molecular-weight contaminants from the supernatant using a PD-10 Column (Amersham Biosciences) filled with Sephadex G-25, one-hundredth of the protein-enriched fraction was digested overnight at 37°C with 25 pmoles of trypsin in 20 mM Tris-HCl (pH 8.8). After the sample was desalted using a ZipTip (C_{18}), it was processed in the same way as described above for MS Sample Preparation 1.

Automated Nanoflow LC-MS/MS Analysis and Protein Identification by Database Search

The peptide mixtures were analyzed using a high-resolution nanoflow reversed-phase capillary LC coupled with an electrospray quadrupole time-of-flight (Q-TOF) tandem mass spectrometer (Q-Tof-2; Micromass Ltd.). The acquired MS/MS spectra were converted to text files of peak lists and processed using the Mascot algorithm (Matrix Science Ltd.) with a maximum tolerance of 500 ppm in MS data and 0.5 Da in MS/MS data against each database. The RefSeq databases were downloaded periodically from the NCBI ftp site (ftp://ftp.ncbi.nih.gov/refseq/). The FLJ-unique cDNA data set was prepared and characterized as previously described (Ota et al. 2004). The results based on the RefSeq data were finally reviewed according to the RefSeq information as of March 5, 2004.

ACKNOWLEDGMENTS

We thank T. Hasui for his technical support in the database construction. We are grateful to E. Nakajima for her critical reading of the manuscript. This work was supported by a Grant-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Science, Sports and Culture of Japan.

REFERENCES

Diba, F., Watson, C.S., and Gametchu, B. 2001. 5'UTR sequences of the glucocorticoid receptor 1A transcript encode a peptide associated with translational regulation of the glucocorticoid receptor. *J. Cell Biochem.* **81:** 149–161.

Kozak, M. 1987. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15:** 8125–8148.

———. 1989. The scanning model for translation: An update. *J. Cell Biol.* **108:** 229–241.

———. 1999. Initiation of translation in prokaryotes and eukaryotes. *Gene* **234:** 187–208.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409:** 860–921.

Maquat, L.E. 2004. Nonsense-mediated mRNA decay: Splicing, translation and mRNP dynamics. *Nat. Rev. Mol. Cell Biol.* **5:** 89–99.

Meijer, H.A. and Thomas, A.A. 2002. Control of eukaryotic protein synthesis by upstream open reading frames in the 5'-untranslated region of an mRNA. *Biochem. J.* **367:** 1–11.

Morris, D.R. and Geballe, A.P. 2000. Upstream open reading frames as regulators of mRNA translation. *Mol. Cell. Biol.* **20:** 8635–8642.

Natsume, T., Yamauchi, Y., Nakayama, H., Shinkawa, T., Yanagida, M., Takahashi, N., and Isobe, T. 2002. A direct nanoflow liquid chromatography–tandem mass spectrometry system for interaction proteomics. *Anal. Chem.* **74:** 4725–4733.

Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., et al. 2004. Complete sequencing and characterization of 21,243 full-length human cDNAs. *Nat. Genet.* **36:** 40–45.

Peri, S. and Pandey, A. 2001. A reassessment of the translation initiation codon in vertebrates. *Trends Genet.* **17:** 685–687.

Pruitt, K.D. and Maglott, D.R. 2001. RefSeq and LocusLink: NCBI gene-centered resources. *Nucleic Acids Res.* **29:** 137–140.

Scharf, J.M., Endrizzi, M.G., Wetter, A., Huang, S., Thompson, T.G., Zerres, K., Dietrich, W.F., Wirth, B., and Kunkel, L.M. 1998. Identification of a candidate modifying gene for spinal muscular atrophy by comparative genomics. *Nat. Genet.* **20:** 83–86.

Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. 1996. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **68:** 850–858.

Soncin, F., Mattot, V., Lionneton, F., Spruyt, N., Lepretre, F., Begue, A., and Stehelin, D. 2003. VE-statin, an endothelial repressor of smooth muscle cell migration. *EMBO J.* **22:** 5700–5711.

Suzuki, Y., Yoshitomo-Nakagawa, K., Maruyama, K., Suyama, A., and Sugano, S. 1997. Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library. *Gene* **200:** 149–156.

Suzuki, Y., Ishihara, D., Sasaki, M., Nakagawa, H., Hata, H., Tsunoda, T., Watanabe, M., Komatsu, T., Ota, T., Isogai, T., et al. 2000. Statistical analysis of the 5' untranslated region of human mRNA using "Oligo-Capped" cDNA libraries. *Genomics* **64:** 286–297.

Suzuki, Y., Taira, H., Tsunoda, T., Mizushima-Sugano, J., Sese, J., Hata, H., Ota, T., Isogai, T., Tanaka, T., Morishita, S., et al. 2001. Diverse transcriptional initiation revealed by fine, large-scale mapping of mRNA start sites. *EMBO Rep.* **2:** 388–393.

Suzuki, Y., Yamashita, R., Nakai, K., and Sugano, S. 2002. DBTSS: DataBase of human transcriptional start sites and full-length cDNAs. *Nucleic Acids Res.* **30:** 328–331.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. 2001. The sequence of the human genome. *Science* **291:** 1304–1351.

Yamashita, R., Suzuki, Y., Nakai, K., and Sugano, S. 2003. Small open reading frames in 5' untranslated regions of mRNAs. *C.R. Biol.* **326:** 987–991.

WEB SITE REFERENCES

ftp://ftp.ncbi.nih.gov/refseq/; NCBI RefSeq ftp site.

http://dbtss.hgc.jp/; DBTSS: DataBase of human Transcriptional Start Sites.

http://www.ddbj.nig.ac.jp/E-mail/clustalw-j.html; CLUSTAL W.

Received January 23, 2004; accepted in revised form April 9, 2004.

THE JOURNAL OF BIOLOGICAL CHEMISTRY

Vol. 279, No. 3, Issue of January 16, pp. 1607–1614, 2004
Printed in U.S.A.

Human Fibrillarin Forms a Sub-complex with Splicing Factor 2-associated p32, Protein Arginine Methyltransferases, and Tubulins α3 and β1 That Is Independent of Its Association with Preribosomal Ribonucleoprotein Complexes*[S]

Received for publication, May 28, 2003, and in revised form, October 16, 2003
Published, JBC Papers in Press, October 28, 2003, DOI 10.1074/jbc.M305604200

Mitsuaki Yanagida‡§¶, Toshiya Hayano‡§, Yoshio Yamauchi§‖, Takashi Shinkawa‖, Tohru Natsume, Toshiaki Isobe§‖, and Nobuhiro Takahashi‡§‡‡**

*From the ‡Department of Applied Biological Science, and Department of Biotechnology, United Graduate School of Agriculture, Tokyo University of Agriculture and Technology, 3-5-8 Saiwai-cho, Fuchu-shi, Tokyo 183-8509, the §Integrated Proteomics System Project, Pioneer Research on Genome the Frontier, Ministry of Education, Culture, Sports, Science & Technology of Japan, the ‖Laboratory of Biochemistry, Graduate School of Science, Tokyo Metropolitan University, 1-1 Minami-ohsawa, Hachiouji, Tokyo 192-0397, and the **National Institutes of Advanced Industrial Science and Technology, Biological Information Research Center, 2-41-6 Ohmi, Kohtoh-ku, Tokyo 135-0064, Japan*

Fibrillarin (FIB, Nop1p in yeast) is an RNA methyltransferase found not only in the fibrillar region of the nucleolus but also in Cajal bodies. FIB is essential for efficient processing of preribosomal RNA during ribosome biogenesis, although its precise function in this process and its role in Cajal bodies remain uncertain. Here, we demonstrate that the human FIB N-terminal glycine- and arginine-rich domain (residues 1–77) and its spacer region 1 (78–132) interact with splicing factor 2-associated p32 (SF2A-p32) and that the FIB methyltransferase-like domain (133–321) interacts with protein-arginine methyltransferase 5 (PRMT5, Janus kinase-binding protein 1). We also show that these proteins associate with several additional proteins, including PRMT1, tubulin α3, and tubulin β1 to form a sub-complex that is principally independent of the association of FIB with preribosomal ribonucleoprotein complexes that co-immunoprecipitate with the sub-complex in human cells expressing FLAG-tagged FIB. Based on the physical association of FIB with SF2A-p32 and PRMTs, as well as the other reported results, we propose that FIB may coordinate both RNA and protein methylation during the processes of ribosome biogenesis in the nucleolus and RNA editing such as small nuclear (nucleolar) ribonucleoprotein biogenesis in Cajal bodies.

Fibrillarin (FIB)[1] is the most abundant protein in the fibrillar regions of the nucleolus where ribosomal RNA transcription and early preribosomal RNA (pre-rRNA) processing take place (1, 2). FIB is also found in Cajal bodies, subnuclear organelles that contain distinct components involved in RNA transcription and editing such as mRNA splicing and small nuclear (nucleolar) ribonucleoprotein (sn(o)RNP) biogenesis (3, 4). FIB is a component of a ribonucleoprotein (RNP) complex that contains U3, U8, and U13 small nucleolar RNAs that exhibit consensus sequence elements denoted box C (5′-UGAUGA-3′) and box D (5′-CUGA-3′) (5). The FIB RNP associates with Nop56p, Nop5p/58p, and a 15.5-kDa protein (a counterpart of yeast Snu13p) to form box C/D snoRNP complexes that function in site-specific 2′-*O*-methylation of pre-rRNA (6–9). FIB is the methyltransferase that catalyzes this 2′-*O*-methylation (10).

FIB, or Nop1p in the yeast *Saccharomyces cerevisiae*, is highly conserved in eukaryotes with respect to sequence, structure, and function (11–17). Deletion of the *Nop1* gene in yeast results in inhibition of 2′-*O*-ribose methylation and pre-rRNA processing at sites A_0 to A_2, indicating that Nop1p is directly involved in both pre-rRNA methylation and processing and ultimately in ribosome assembly (18). Although human FIB is the functional homolog of yeast Nop1p, it only partially complements a yeast nop1-defective mutant (15). Human FIB is a nucleolar autoantigen for the non-hereditary immune disease scleroderma (14). FIB co-localizes with the survival motor neuron (SMN) gene product in both nucleoli and Cajal bodies/gems of primary neurons (19, 20). SMN is linked to one of the most common inheritable causes of childhood mortality, spinal muscular atrophy (19). In fact, a direct interaction between FIB and SMN has been demonstrated, although no functional basis for this interaction has been established, including any involvement of FIB in the pathogenesis of spinal muscular atrophy. Another protein, the nuclear DEAD box protein p68, an RNA-dependent ATPase and RNA helicase, co-localizes with FIB specifically in nascent nucleoli during telophase (21). As with SMN, no physiological role of its interaction with FIB has been established.

Human FIB (~36 kDa) comprises 321 amino acids and three structural domains (14) and is 66% identical to yeast Nop1p. The N-terminal 80 residues comprise a glycine- and arginine-rich (GAR) domain (14) that is also present in Nop1p and *Xenopus* FIB (Fig. 1) but not in *Tetrahymena* FIB (17) or *Methanococcus jannaschii* FIB (22). The GAR domain is methylated at arginine residues, although the arginine methyltransferase responsible for *in vivo* methylation has not been identified (23, 24). The GAR domain is responsible for the interaction

* This work was supported by a Pioneer Research grant on Genome the Frontier from the Ministry of Education, Culture, Sports, Science, & Technology of Japan. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[S] The on-line version of this article (available at www.jbc.org) contains Supplemental Figs. S1 and S2 and Tables SI–SIII.

¶ Present address: Juntendo University, Graduate School of Medicine, BioMedical Research Center, 2-1-1, Hongo, Bunkyo-ku, Tokyo 113-8421, Japan.

‡‡ To whom correspondence should be addressed: Tel./Fax: 81-042-367-5709; E-mail: ntakahas@cc.tuat.ac.jp.

[1] The abbreviations used are: FIB, fibrillarin; snRNP, small nuclear ribonucleoprotein; snoRNP, small nucleolar ribonucleoprotein; LC, liquid chromatography; MALDI-TOF, matrix-assisted laser desorption ionization time-of-flight; MS, mass spectrometry; PBS, phosphate-buffered saline; pre-rRNP, preribosomal ribonucleoprotein; RNase, ribonuclease; RNP, ribonucleoprotein; rRNA, ribosomal RNA; SMN, survival motor neuron; RBD, RNA-binding domain; NLS, nucleolar localization signal; NS, SV40 nuclear localization signal.

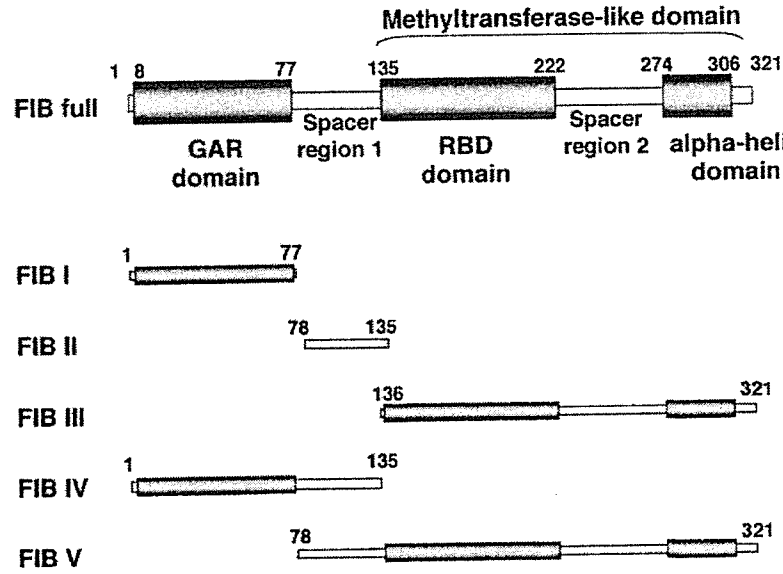

FIG. 1. **Schematic representation of domain structures of FIB and truncation mutants.** Full-length FIB and truncation mutants are shown. The N-terminal glycine- and arginine-rich (*GAR*) domain, RNA-binding domain (*RBD*), and α-helix domain are *shaded in gray*, and the spacer regions 1 and 2 are indicated. The methyltransferase-like domain is composed of the RBD, spacer region 2, and α-helix domain. The residue numbers of FIB and the truncation mutants are indicated above each diagram. A FLAG tag (not shown) was added to the N terminus of each peptide. For FIB truncation mutants, the SV40 antigen nuclear localization signal and the HIV Rex nucleolar localization signal were added between the FLAG tag and the N terminus of each mutant (not shown in the figure).

of FIB with both SMN protein and the DEAD box RNA helicase p68 (19, 21). A centrally located 90-residue sequence resembles an RNA-binding domain (RBD) present in various snRNPs. This RBD together with the C-terminal α-helix domain and the intervening spacer (spacer region 2) constitutes a methyltransferase-like domain that contains an *S*-adenosyl methionine-binding motif (14, 22). Replacement of two residues in the Nop1p *S*-adenosyl methionine-binding motif results in temperature sensitivity and a drastic reduction in nascent rRNA transcript methylation under restrictive conditions (18). Thus, the methyltransferase-like domain is responsible for FIB methyltransferase activity. The C-terminal α-helix domain is composed of ~30 residues, and although this domain probably targets FIB to Cajal bodies, spacer region 2 appears to target FIB to the fibrillar regions. However, the targeting of FIB in both instances occurs only in the presence of the RBD (3). Although it is well established that FIB plays a role in ribosome biogenesis within the nucleolus, its role in Cajal bodies is not understood.

Our recent studies have used proteomic methodology to characterize a series of preribosomal ribonucleoprotein (pre-rRNP) complexes formed in mammalian cells. We have thus far isolated and analyzed the pre-rRNP complexes associated with human nucleolin (25), parvulin (26), and Nop56p (27). These studies demonstrate the applicability of proteomic analysis to the study of human ribosome biogenesis and have identified a number of mammalian counterparts of yeast trans-acting factors involved in this process. Furthermore, several candidate mammalian trans-acting factors were identified that were not previously identified in yeast. Here, we present a proteomic analysis of FIB-associated protein complexes. In addition to the association of FIB with pre-rRNP complexes, we found that this protein interacts with a sub-complex containing a minimal set of proteins including SF2A-p32, PRMT5, tubulin α3, tubulin β1, and PRMT1. The FIB GAR domain and the spacer region 1 interact directly with SF2A-p32, whereas the methyltransferase-like domain interacts with PRMT5. These results provide new clues to the precise functions of FIB not only in ribosome biogenesis but also in sn(o)RNP biogenesis and mRNA processing in Cajal bodies.

EXPERIMENTAL PROCEDURES

Materials—Human kidney cell line 293EBNA, Opti-MEM, and LipofectAMINE were purchased from Invitrogen (Grand Island, NY). Dulbecco's modified Eagle's medium, anti-FLAG M2 affinity gel, FLAG peptide, IGEPAL CA630, RNase A, and α-cyano-4-hydroxycinnamic acid were from Sigma-Aldrich Chemical (Steinheim, Germany). Alkaline phosphatase-conjugated anti-mouse IgG was from Amersham Biosciences (Uppsala, Sweden). Alexa Fluor 488-conjugated rabbit anti-mouse IgG antibody was from Molecular Probes, Inc. (Eugene, OR). Trypsin (sequence grade) was from Promega (Madison, WI) and *Achromobacter lyticus* protease I (Lys-C) was from WAKO Pure Chemicals (Osaka, Japan). ZipTipC18 was from Millipore (Billerica, MA). LA*Taq* DNA polymerase was from Takara (Shiga, Japan). Protease inhibitor mixture Complete Mini was from Roche Diagnostics (Mannheim, Germany). Collagen I-coated Biocoat 8-well culture slides were from BD Biosciences (Franklin Lakes, NJ). All other reagents were from WAKO Pure Chemicals.

Vectors for Epitope-tagged FIB and FIB Truncation Mutants and Expression in 293EBNA Cells—The FIB expression plasmid was constructed using a PCR-amplified DNA fragment comprising the FIB open reading frame with the FLAG tag at its N terminus (Fig. 1). This fragment was introduced between the NheI and the BamHI sites of the mammalian expression vector, pcDNA3.1 (+) (Invitrogen). The PCR primer set was 5′-ATATATCTAGAGCCACCATGGACTACAAGGACGACGACGACAAGAAGCCAGGATTCAGTCCCCGT-3′ and 5′-TATAGGATCCTCAGTTCTTCACCTTGGGGGG-3′, and human placenta cDNA (OriGene Technologies, Inc., Rockville, MD) was used as the template.

The DNA fragment encoding the FLAG tag along with the nucleolar localization signal (NLS) of HIV Rex (TRRRPRRSQRKR) (28) and the SV40 nuclear localization signal (NS) (PKKKRKV) (29) was synthesized by PCR using the oligonucleotide sets 5′-TATAGCTAGCGCCACCATGGACTACAAGGACGACGACGACAAGACCCGTCGGAGGCCCCG-3′ and 5′-TCTTTTTCTTTGGGATCGGCGGGGCCTCCGACGGGT-3′, and 5′-GATCCCAAAGAAAAAGAGCCAGCCCAAAAAAGAAGAGAAA-3′ and 5′-ATATAGGATCCTACCTTTCTCTTCTTTTTTGG-3′. The amplified fragment was subcloned between the NheI and the BamHI sites of pcDNA3.1(+), and the resulting plasmid was designated pcDNA3.1-NLS. All the expression plasmids of the FIB deletion mutants were constructed by introducing PCR-amplified fragments between the BamHI and XhoI sites downstream of the FLAG tag/NLS/NS in pcDNA3.1-NLS. Primer sets used for the amplification of FIB deletion mutants were as follows; 5′-ATATAGGATCCAAGCCAGGATTCAGTCCCCGT-3′ and 5′-TATATCTCGAGTCATTTTCCTCCCCGACCACGACC-3′ for FIB I (residues 2–77), 5′-ATATAGGATCCAGAGGAAACCAGTCGGGGAAG-3′ and 5′-TATATCTCGAGTCATCGGTACTCAAATTTGTCATC-3′ for FIB II (residues 78–135), 5′-ATATAGGATCCGCCTGGAACCCCTTCCGCTCC-3′ and 5′-TATATCTCGAGTCAGTTCTTCACCTTGGGGGG-3′ for FIB III (residues 136–321), 5′-ATATAGGATCCAAGCCAGGATTCAGTCCCCGT-3′ and 5′-TATATCTCGAGTCATCGGTACTCAAATTTGTCATC-3′ for FIB IV (residues 2–135), 5′-ATATAGGATCCAGAGGAAACCAGTCGGGGAAG-3′ and 5′-TATATCTCGAGTCAGTTCTTCACCTTGGGGGG-3′ for FIB V (residues 78–321).

Human 293EBNA cells were cultured in Dulbecco's modified Eagle's medium supplemented with 10% heat-inactivated fetal calf serum, streptomycin (0.1 μg/ml), and penicillin G (100 units/ml) at 37 °C in an

incubator under 5% CO_2. Subconfluent cells in 90-mm dishes were transfected with 10 μg of expression plasmid DNA using LipofectAMINE, and the transfected cells were grown for 48 h at 37 °C.

Isolation of FIB- and Its Truncated Mutant-associated Complexes—At 48 h post-transfection, 293EBNA cells were harvested and washed with PBS and lysed in lysis buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.5% IGEPAL CA630) containing a protease inhibitor mixture on ice for 30 min. The soluble fraction was obtained by centrifugation at 15,000 rpm for 30 min at 4 °C and was incubated with 20 μl of anti-FLAG M2-agarose beads for 4 h at 4 °C for immunoprecipitation of FIB-associated complexes or overnight at 4 °C for deletion mutant-associated complexes. After washing the agarose beads five times with lysis buffer and once with 50 mM Tris-HCl, pH 8.0, 150 mM NaCl, the complexes bound to the agarose beads were eluted with 20 μl of 50 mM Tris-HCl, pH 8.0, 150 mM NaCl containing 500 μg/ml FLAG peptide. The eluted complexes were analyzed by SDS-PAGE.

Ribonuclease Treatment of the FIB- and Truncation Mutant-associated Complexes—The immunoprecipitated FIB and truncation mutant-associated complexes described above were incubated with 50 mM Tris-HCl, pH 8.0, 150 mM NaCl containing 1 μg/ml RNase A for 10 min at 37 °C, washed twice with lysis buffer, and then once with 50 mM Tris-HCl, pH 8.0, 150 mM NaCl, and eluted with buffer containing the FLAG peptide as described above.

Immunocytochemistry—293EBNA cells were grown on Collagen I-coated 8-well culture slides and transfected with expression plasmids using LipofectAMINE. Prior to fixation, cells were washed with PBS followed by incubation with 3.7% formaldehyde in PBS. After several washes with PBS-T (PBS containing 0.05% (w/v) Tween 20), the cells were incubated with PBS containing 0.1% (w/v) Triton X-100 for 5 min at room temperature. The cells were then blocked by 3% (w/v) nonfat dried milk in PBS and were incubated with anti-FLAG for 1 h at room temperature. The cells were rinsed in PBS-T and then incubated with Alexa Fluor 488-conjugated anti-mouse IgG for 1 h at room temperature, followed by three washes with PBS-T. The resulting cells were examined with a confocal laser-scanning microscope TCS (Leica Microsystems AG, Wetzlar, Germany).

Protein Identification by the Peptide Mass Fingerprinting Method—Protein-containing SDS-PAGE gel fragments were subjected to in-gel digestion with trypsin as previously described (25). The resulting peptides were recovered and analyzed for peptide mass fingerprints using a PE Biosystems MALDI-TOF MS (Voyager DE-STR) as described previously (25). Peptide masses were searched with 50 ppm mass accuracy using the data base fitting program MS-Fit (available at prospector.ucsf.edu), and protein identification was performed according to the criteria described previously (25).

Protein Identification by LC-MS/MS Analysis—The FIB- and truncation mutant-associated complexes were digested with Lys-C, and the resulting peptides were analyzed using a nanoscale LC-MS/MS system as described (30). The peptide mixture was applied to a Mightysil-RP-18 (3-μm particle, Kanto Chemical, Osaka, Japan) frit-less column (45 mm × 0.150 mm i.d.) and separated using a 0–40% gradient of acetonitrile containing 0.1% formic acid over 80 min at a flow rate of 50 or 25 nl/min. Eluted peptides were sprayed directly into a quadrupole time-of-flight hybrid mass spectrometer (Q-Tof 2, Micromass Wythenshawe, UK). MS/MS spectra were acquired by data-dependent collision-induced dissociation, and MS/MS data were analyzed using the MASCOT software (Matrix Science, London, UK) for peptide assignment. The criteria were in accordance with the manufacturer's definitions. If necessary, match acceptance of automated batch processes was confirmed by manual inspection of each set of raw MS/MS spectra in which the major product ions were matched with theoretically predicted product ions from the data base-matched peptides. Proteins in the mock eluate from anti-FLAG antibody with FLAG-peptide were analyzed by the same LC-MS/MS method as used for the fibrillarin-associated complexes and then subtracted from the proteins identified in the total fibrillarin-associated complexes. Thus, those proteins identified in the mock eluate were not included in the fibrillarin-associated proteins unless the quantitative increase was confirmed.

Ultracentrifugation of FIB-associated RNP Complexes—The anti-FLAG immunoprecipitate obtained from FLAG-tagged FIB gene-transfected 293EBNA cells after elution with the FLAG peptide was analyzed on a 12–50% sucrose gradient in 50 mM Tris, pH 7.5, 25 mM KCl, 5 mM $MgCl_2$. The gradients were centrifuged in an SW65 rotor at 50,000 rpm (180,000 × *g*) for 3 h at 4 °C. A total of 18 fractions of 300 μl each were collected. The migration of the 40 S/60 S/80 S ribosomal complexes was determined by comparison to the ultraviolet absorption profile of cytosolic ribosomes fractionated by ultracentrifugation under identical experimental conditions.

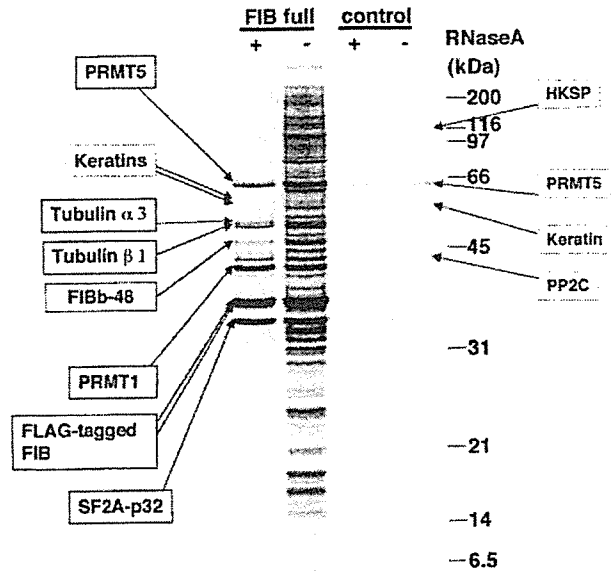

FIG. 2. **Protein components of immunoprecipitated FLAG-FIB-associated complexes and RNA dependence of binding to FIB.** Silver-stained 11% SDS-PAGE gel of FLAG-FIB-associated complexes immunoprecipitated with anti-FLAG after expression of FLAG-tagged full-length FIB. *Lanes 1* and *2*, FIB-associated complexes (*FIB full*) with (+) or without (−) RNase treatment; *lanes 3* and *4*, control immunoprecipitate with or without RNase. Molecular weight markers are indicated to the *right*. The protein bands identified by MALDI-TOF analysis after in-gel digestion of protein bands with protease are indicated on both sides of the gel. *Boxed proteins* denote FIB-associated proteins following RNase treatment. Proteins in *gray* indicate those identified in the control. *PRMT1* and *5*, protein arginine methyltransferases 1 and 5, respectively; *SF2A-p32*, splicing factor-2-associated protein p32; *FIBb-48*, FIB-binding protein 48 kDa; *HKSP*, kinesin-like spindle protein; *PP2C*, protein phosphatase 2C.

RESULTS

Isolation of FLAG-tagged FIB-associated Protein Complexes—Because endogenous FIB is found primarily in the nucleolus of mammalian cells, the subcellular localization of the FLAG-tagged protein in transfected 293EBNA cells was confirmed by immunofluorescence microscopy using an antibody to FLAG (Supplementary Fig. 1). Although weak staining was observed in the cytoplasm and nucleoplasm of the FLAG-FIB-transfected cells, the nucleolus exhibited intense staining thus confirming the correct localization of FLAG-tagged FIB.

Complexes associated with FIB were isolated from FLAG-FIB-transfected cells via immunoprecipitation using the FLAG antibody. A typical silver-stained SDS-PAGE gel of the immunoprecipitated fraction showed that FIB-associated complexes contained many proteins spanning a wide range of molecular weight (Fig. 2). In contrast, only four protein bands were apparent in a mock immunoprecipitate prepared from untransfected control cells (Fig. 2). In addition, when cells were transfected with unrelated FLAG-tagged proteins, an entirely different pattern of protein bands was obtained (data not shown). These results indicated that most of the factors immunoprecipitated from FLAG-FIB-transfected cells represented *de facto* FIB-associated proteins.

RNA Integrity in FIB-associated Complexes—RNA integrity is requisite for protein association in pre-rRNP complexes associated with human nucleolin, parvulin, and Nop56p, all of which may be involved in ribosome biogenesis as reported in our previous studies (25–27). Therefore, the requirement for RNA integrity was also analyzed for FLAG-FIB-associated complexes. RNase treatment prompted the dissociation of the

TABLE I

Putative fibrillarin-associated trans-acting factors involved in ribosome biogenesis

Accession number (GI) and calculated molecular mass are shown. Known function in mammals and yeast were extracted from the NCBI database entry and from the Saccharomyces Genome Database of Stanford University (genome-www.stanford.edu/Saccharomyces/) and David Tollervey's database (homepages.ed.ac.uk/dtoller/processing_components.html), respectively. For proteins that have yeast homologs, the gene names, open reading frame names, and percent identities with the homolog in overlapping amino acid regions are indicated.

Protein	Accession No. (GI)	MW (Da)	Function in mammals	Yeast Homolog	Yeast ORF	Percent identity to yeast homolog	Function in Yeast
Fibrillarin	11425985	33763.4	A component of a snRNP particle thought to participate in the first step of preribosomal RNA processing.	NOP1	YDL014W	66	35S primary transcript processing; rRNA modification.
Nucleolar protein 5A (56 kDa with KKE/D repeat)	12832025	66008.7	Similar to *S. cerevisiae* Sik1p, a nucleolar KKE/D repeat protein involved in pre-rRNA processing.	SIK1(NOP56)	YLR197W	50	35S primary transcript processing; rRNA modification.
Nucleolar protein NOP5/NOP58	7706254	59540.5	Putative snoRNA binding protein	NOP58	YOR310C	47	35S primary transcript processing; rRNA modification.
NHP2 non-histone chromosome protein 2-like 1 (S. cerevisiae)	4826860	14164.6	Binds to the 5' stem-loop of U4 snRNA. Nucleolar protein that associates with the checkpoint protein RAD17; highly similar to *S. cerevisiae* Snu13p.	SNU13	YEL026W	71	Pre-mRNA splicing factor; part of small (ribosomal) subunit (SSU) processosome (contains U3 snoRNA).
Splicing factor, arginine/serine-rich 4	5032089	56759.3	mRNA splicing; Splicing factor 4; member of the SR protein family; has an RNA recognition motif (RRM).	SRP40	YKR092C	32	Nucleocytoplasmic transport; Nopp140 homolog, a nonribosomal protein of the nucleolus and coiled bodies; nucleolar protein.
Hypothetical protein FLJ10774	12697963	116025.8	-	KRE33	YNL132W	55	40S subunit assembly and export to cytoplasm; killer toxin resistance.
Hypothetical protein FLJ10534	8922496	75069.7	Unknown	TSR1	YDL060W	32	-
Hypothetical protein AD034	13899340	64594.2	Unknown	RIO1/RRP10	YOR119C	38	20S pre-rRNA processing
DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 5 (RNA helicase, 68 kDa)	4758138	69104.7	RNA helicase p68	DBP2	YNR112W	56	RNA helicase
DEAD/H box polypeptide 21	2135315	89195.5	RNA helicase Gu	DBP3	YGL078C	32	RNA helicase; 35S primary transcript processing; large ribosomal subunit assembly and maintenance.
Poly(A) binding protein, cytoplasmic 1	3183544	70625.9	Poly(A) binding	PAB1	YER165W	54	Poly(A) binding
Poly(A) binding protein, cytoplasmic 4 (inducible form)	4504715	70738.1	RNA-binding protein, binds to the mRNA poly(A) tail; may play a role in translation and mRNA stability.	PAB1	YER165W	53	Poly(A) binding
Heterogeneous nuclear ribonucleoprotein U (scaffold attachment factor A)	14141163	90423.0	Binds RNA and scaffold attached region DNA; contains an RGG box domain; component of hnRNP complexes; may play a role in hnRNA structure or processing.	NOP3	YDR432W	37	Required for efficient 27S pre-rRNA processing. Nop3p shuttles between the nucleus and the cytoplasm.
Splicing factor, arginine/serine-rich 1 (splicing factor 2, alternate splicing factor)	5902076	27727.8	Splicing factor 10; essential for constitutive pre-mRNA splicing; member of the SR family; alternative splicing factor.	NOP3	YDR432W	26	Required for efficient 27S pre-rRNA processing. Nop3p shuttles between the nucleus and the cytoplasm.
Splicing factor, arginine/serine-rich 6	13653676	39563.4	Splicing factor	NOP3	YDR432W	25	Required for efficient 27S pre-rRNA processing. Nop3p shuttles between the nucleus and the cytoplasm.
Pescadillo homolog 1, containing BRCT domain (zebrafish)	7657455	67960.0	Plays a crucial role in cell proliferation and may be necessary for oncogenic transformation and tumor progression.	NOP7	YGR103W	39	rRNA processing
DKFZP564M182 protein	3668141	58096.6	-	CIC1	YHR052W	26	26S proteasome; ribosome biogenesis.
Nucleolar GTPase	12652715	83629.5	Nucleolar GTPase	NOG2	YNR053C	55	Nuclear/Nucleolar GTP-binding protein 2
Putative nucleotide binding protein, estradiol-induced	7657048	63528.4	Nucleotide binding	NUG1	YER006W	30	Nuclear GTPase
Nucleolin	4885511	76298.2	Ribosome biogenesis	NSR1	YGR159C	32	rRNA processing; small ribosomal subunit assembly.
RNA-binding region (RNP1, RRM) containing 2	4757926	58905.5	Nuclear protein that may be a splicing factor; contains motifs characteristic of splicing factors.	NSR1	YGR159C	23	
Ras-GTPase activating protein SH3 domain-binding protein 2	14916573	54077.8	Probable scaffold protein that may be involved in mRNA transport.	NSR1	YGR159C	25	

TABLE I—*continued*

Protein	Accession No. (GI)	MW (Da)	Function in mammals	Yeast Homolog	Yeast ORF	Percent identity to yeast homolog	Function in Yeast
Hypothetical protein DKFZp762N0610	7959201	60443.4	Unknown	SRM1	YGL097W	27	Nuclear export of rRNA
Nucleophosmin (nucleolar phosphoprotein B23, numatrin)	15214852	32588.8	Nucleophosmin (nucleolar phosphoprotein B23, numatrin, protein B23); RNA-binding nucleolar phosphoprotein.	-	-		-
Other RNA helicases							
DEAD/H box polypeptide 3	13514813	73226.0	RNA helicase	DED1	YOR204W	49	RNA processing
DEAD/H box polypeptide 9 (RNA helicase A, nuclear DNA helicase II; leukophysin)	3915658	140788.0	RNA helicase	-	YLR419W	30	Helicase
DEAD/H box polypeptide 17 (72 kDa)	5453840	72326.0	RNA helicase p72	DBP2	YNR112W	58	RNA helicase
DEAD/H box polypeptide 30	7662362	133854.5	DDX30	-	YLR419W	31	Helicase
DEAD/H box polypeptide 36	14730578	114703.8	DDX36; DEAD/H box polypeptide 36.	-	YLR419W	29	Helicase
Nucleolar protein GU2	13129006	82514.1	A DEAD box enzyme that may be involved in ribosomal RNA synthesis or processing.	DBP3	YGL078C	31	35S primary transcript processing

majority of the protein components of FLAG-FIB-associated complexes. However, at least four major protein-staining bands as well as several minor bands remained associated with FLAG-FIB after RNase treatment (Fig. 2). The four major protein bands appeared to exhibit approximate equivalent stoichiometry with respect to staining intensity and were representative of the more abundant proteins present in the non-RNase-treated complexes (Fig. 2, compare *lanes 1* and *2*). In addition, the proteins released by RNase treatment could not be distinguished from those present in the RNase-untreated complexes except for the four major protein bands and several minor protein bands that remained associated with FLAG-FIB after RNase treatment, as judged by SDS-PAGE analysis (data not shown). These results suggested that at least two distinct groups of proteins are present in FIB-associated complexes, one whose association is dependent on RNA integrity (the FIB-associated RNPs) and another that associates directly with FIB (independent of RNA integrity). However, it was uncertain as to whether these two groups of proteins were associated with each other or present independently.

Identification of Protein Components in FIB-associated RNPs—Given the involvement of FIB in ribosome biogenesis, we expected isolated FIB-associated RNP complexes to contain trans-acting factors involved in this process, as well as ribosomal proteins. We therefore identified the protein components of the FIB-associated RNPs. We previously described a highly sensitive "direct nano-flow LC-MS/MS" system to identify proteins in limited amounts of multiprotein complexes (30). Immunoisolated FIB-associated complexes were digested with Lys-C and analyzed directly using the nano-LC-MS/MS system. In addition to the criteria for match acceptance described under "Experimental Procedures," more stringent criteria were adopted to conclusively identify proteins. Namely, at least two different peptides had to be identified in a single nano-LC-MS/MS analysis, and/or at least one peptide had to be identified at least twice (with highly significant data base matching scores) among four separate analyses. A total of 1426 peptides were identified via sequence data base searches using the collision-induced dissociation spectra obtained from four nano-LC-MS/MS runs of a Lys-C digest of the FIB-associated complexes. These peptide data identified 170 proteins (excluding the bait protein and the proteins present in mock) that met our identification criteria. Although we do not exclude that some of the proteins identified may be nonspecifically associated proteins, we believe most of the protein components identified in the fibrillarin-associated complexes are specifically associated with fibrillarin. Of the 170 proteins, 73 were ribosomal proteins (43 from the large subunit and 30 from the small subunit; Supplementary Table SI) and 97 were non-ribosomal proteins (Supplementary Tables SII and SIII). Of the non-ribosomal proteins, 24 were assigned as probable trans-acting factors involved in ribosome biogenesis based on their homology to yeast proteins known or expected to be involved in ribosome biogenesis (Table I and Supplementary Table SII).

In addition to the putative trans-acting factors that were assigned, the FIB-associated complexes contained 68 more non-ribosomal proteins, including a number of RNA-binding proteins, splicing factors, DNA-topoisomerase, Myb-binding protein, components of DNA-dependent protein kinase, components of the signal recognition particle, nuclear matrix proteins as well as many hypothetical/unknown proteins (Supplementary Table SIII). Of these, at least 44 proteins reportedly localize to the nucleolus or the nucleus (31, 32) (Supplementary Table SIII).

Identification of Proteins Associated with FIB without RNA Integrity—The proteins that remained associated with FIB after the RNase treatment were identified by MALDI-TOF analysis after in-gel digestion of excised SDS-PAGE gel bands with trypsin. The identified proteins were SKB1 homolog (protein arginine-methyltransferase 5, PRMT5) (GI: 5174683; 18 peptides matched; 29.6% peptide coverage; mean error = −3.41 ppm), tubulin $\alpha 3$ (GI: 5174733; 9 peptides matched; 27.2% peptide coverage; mean error = −11.42 ppm), tubulin $\beta 1$ (GI: 135448; 13 peptides matched; 33.3% peptide coverage; mean error = 0.70 ppm), PRMT1 (GI: 2499803; 16 peptides matched; 55.6% peptide coverage; mean error = −4.21 ppm), and splicing factor SF2-associated protein p32 (SF2A-p32) (GI:4502491; 8 peptides matched; 50.7% peptide coverage; mean error = −5.08 ppm) (Fig. 2). These proteins were also identified by nano-LC-MS/MS analysis of the isolated FIB-associated RNPs without

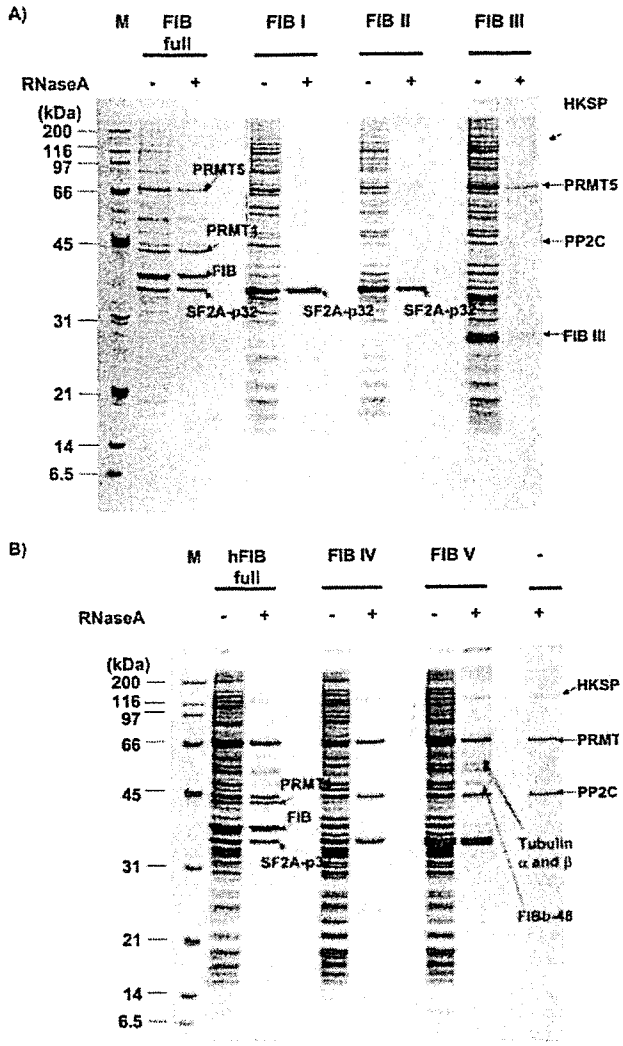

FIG. 3. **Protein components of immunoprecipitated FLAG-FIB- and truncated protein-associated complexes.** Silver-stained 11% SDS-PAGE gel of complexes immunoprecipitated with anti-FLAG after expression of full-length FIB or various truncation mutants. The *arrows* indicate the proteins that were identified. Abbreviations are as in Fig. 2. *A*, *Lane 1*, molecular weight markers; *lanes 2* and *3*, FIB-binding complexes (*FIB full*) with (+) and without (−) RNase treatment; *lanes 4* and *5*, FIB I-binding complexes with or without RNase; *lanes 6* and *7*, FIB II-binding complexes with or without RNase; *lanes 8* and *9*, FIB III-binding complexes with or without RNase. *B*, *Lane 1*, molecular weight markers; *lanes 2* and *3*, FIB-binding complexes (*FIB full*) with or without RNase; *lanes 4* and *5*, FIB IV-binding complexes with or without RNase; *lanes 6* and *7*, FIB V-binding complexes with or without RNase; lane 8, control with (+) RNase treatment.

RNase treatment (Supplementary Table SIII). However, MALDI-TOF analysis failed to identify other minor protein bands, including FIBb-48 (see Fig. 2).

Given that the association of FIB with the above proteins was RNA-independent, they may interact directly with FIB. To clarify the interactions between FIB and these proteins, expression vectors for three FLAG-tagged FIB truncated mutants with nuclear and nucleolar localization signals were initially constructed (Fig. 1) and expressed in 293EBNA cells. Each of the truncated proteins was expressed predominantly in the nucleolus (Supplementary Fig. S1) and exhibited the expected molecular weight, although mutants FIB I and FIB II migrated as multiple bands on SDS-PAGE gels probably due to post-translational modification *in vivo* (Supplementary Fig. S2). The protein complexes associated with each of the truncated proteins expressed in the 293EBNA cells were purified by immunoprecipitation with anti-FLAG-conjugated beads. Silver staining of an SDS-PAGE gel showed that a number of proteins were present in each of the three immunoprecipitates, and many of the bands were common among the three isolated protein complexes as well as in the full-length FIB-associated complex. Still, some protein bands were present uniquely in each of the complexes (Fig. 3*A*). When the immunoprecipitates were treated with RNase A, only SF2A-p32 remained associated with in the FIB I- and FIB II-associated complexes, whereas only PRMT5 remained associated with the FIB III-associated complexes (Fig. 3*A*). This result showed that SF2A-p32 interacts directly with the FIB GAR domain and spacer region 1 and that PRMT5 interacts with the methyltransferase-like domain (Fig. 1). Although PRMT5 was also detected in a mock immunoprecipitate prepared from control cells, its level was clearly elevated upon expression of FIB III. Despite the binding of SF2A-32p and PRMT5 to the corresponding truncated proteins, the binding of PRMT1 and other proteins associated with full-length FIB was not detected.

Two additional FIB truncation mutant expression vectors were constructed. FIB IV contained sequences encoded by FIB I and II, whereas FIB V contained FIB II and III sequences (Fig. 1). Both FIB IV and FIB V were expressed in 293EBNA cells, although FIB IV suffered some degradation that was most likely due to proteolytic cleavage from the C terminus (Supplementary Fig. S2). Immunoisolation using anti-FIB showed that these two truncated mutants were also associated with a number of proteins, most of which became dissociated upon RNase A treatment as was the case for the other truncation mutants (Fig. 3*B*). As expected from the FIB I and FIB II results, FIB IV associated with only SF2A-p32. On the other hand, in addition to SF2A-p32, FIB V also associated with tubulin $\alpha 3$ and $\beta 1$ as well as several other proteins, including one with the molecular mass of 48 kDa (FIBb-48; Fig. 3*B*). PRMT5 and PP2C were detected more strongly in the control (Fig. 3*B*) than in previous experiments (compare with Fig. 3*A*) due to the use of different lots of anti-FLAG M2 affinity gel (purchased from Sigma-Aldrich Chemical, Steinheim, Germany). However, densitometry analysis indicated that PRMT5 was clearly increased in FIB V-associated protein complexes (and full-length FIB-associated complexes) compared with the control; *i.e.* the staining intensities of PRMT5 after RNase treatment were 19,149 (arbitrary values) for full-length FIB, 12,776 for FIB IV, 20,046 for FIB V, and 7,634 for the control (Fig. 3*B*). These results support the assertion that the methyltransferase-like domain corresponding to the FIB III mutant is responsible for the binding of FIB to PRMT5. These results also indicate that the presence of both the spacer region 1 and the methyltransferase-like domain is required for the binding of FIB to tubulin $\alpha 3$ and $\beta 1$ as well as FIBb-48. With respect to the association of FIB with PRMT1, none of the five FIB truncation mutants bound this protein indicating that its binding is dependent on the domain structure of full-length FIB and/or that the binding of the other proteins (*e.g.* SF2A-p32 and PRMT5) is a prerequisite to PRMT1 binding. Together, these results suggest that FIB along with SF2A-p32, PRMT5, tubulins $\alpha 3$ and $\beta 1$, FIBb-48, and PRMT1 constitute a subcomplex that functions as an integrated module.

Fractionation of FIB-associated RNP Complexes by Ultracentrifugation—The above results did not address whether the sub-complex that resulted from RNase A treatment constituted a protein complex that was independent of FIB-associated RNPs. Therefore, the immunoisolated FIB-associated complexes were fractionated by ultracentrifugation through a su-

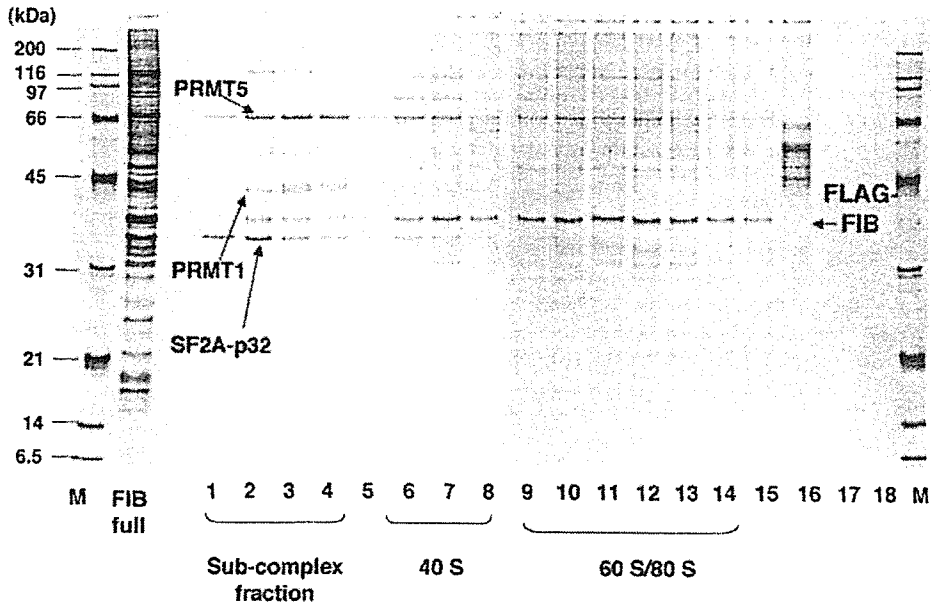

FIG. 4. **Separation of the FIB-associated sub-complex and RNP complexes by ultracentrifugation through a sucrose gradient.** FLAG-tagged FIB gene-transfected 293EBNA cells were subjected to immunoprecipitation using anti-FLAG. The immunoprecipitate was fractionated into 18 fractions by ultracentrifugation on a 12–50% sucrose gradient. The *arrow* below the SDS-PAGE gel indicates the fraction containing the sub-complex. Molecular weights are indicated to the *left*. *M*, molecular weight markers; *FIB full*, components of the immunoprecipitate loaded onto the gradient. Abbreviations are as in Fig. 2.

crose gradient and collected into 18 fractions, each of which was analyzed by SDS-PAGE (Fig. 4). Although FLAG-FIB was observed in nearly all the fractions, its distribution was concentrated in three peaks corresponding to 60 S/80 S ribosomes, 40 S ribosomes and sub-complexes. The majority of FIB was present in largest RNP complexes in the gradient (fractions 9–14). The proteins that were previously shown to remain associated with FIB after RNase treatment were found mostly in the sub-complex fractions (fractions 1–4), indicating that these RNA-independent FIB-associated proteins constitute an independent module. Thus, FIB forms a sub-complex that is primarily independent of the rest of the FIB-associated RNPs. Given that different patterns of protein bands were evident in the different fractions, the protein composition of the sub-complex may be somewhat heterogeneous.

DISCUSSION

FIB-associated protein complexes contain a number of ribosomal proteins and potential trans-acting factors involved in ribosome biogenesis, and the complexes are dependent on RNA integrity. Given the involvement of FIB in pre-rRNA methylation and processing, we conclude that, along with the sub-complex containing SF2A-p32 and PRMTs, the isolated FIB-associated protein complexes contain pre-rRNP complexes involved in ribosome biogenesis. The FIB-associated pre-rRNPs contain mainly 60 S-processing and assembly factors but also contain some 40 S factors suggesting that FIB pre-RNPs most likely form at early to middle stages of 60 S large subunit biogenesis (Table I) (33, 34).

One of the notable findings of this study is the RNA-independent association of FIB with SF2A-p32, PRMT5, tubulin $\alpha 3$, tubulin $\beta 1$, FIBb-48kD protein, and PRMT1. SF2A-p32 binds the FIB GAR domain (corresponding to truncation mutant FIB I) and the spacer region 1 (FIB II), whereas PRMT5 binds the methyltransferase-like domain (FIB III) (Fig. 3). Given that FIB V, but not the FIB III, associated with tubulin $\alpha 3$, tubulin $\beta 1$, and FIBb-48 protein (Fig. 3), both the methyltransferase-like domain and the spacer region 1 are necessary for the association of these proteins with FIB. However, the methyltransferase-like domain and the spacer region 1 were still not sufficient to recruit PRMT1. These results indicate that, in addition to these regions, the FIB GAR domain is required for the association with PRMT1. However, the data do not address whether PRMT1 binding is dependent on the recruitment of SF2A-p32, PRMT5, and the other proteins to FIB. Because these FIB-associated proteins fractionated as a much smaller complex relative to the FIB-associated pre-rRNPs that likely represent a mixture of 40 S and 60S/80S preribosomal particles (Fig. 4), these results indicate that FIB and these associated proteins constitute a sub-complex that is independent of the larger FIB-associated pre-rRNP complexes.

The association of FIB with two protein arginine methyltransferases is very intriguing, because FIB itself is an RNA methyltransferase. Thus, the association provides for the possible coordinated regulation of RNA and protein methylation events such as those in which both the RNA and protein molecules are involved. Two types of arginine methyltransferases were identified as FIB-associated proteins. First, PRMT5 (originally identified as the Janus kinase binding protein 1, or JBP1) is a type II methyltransferase that catalyzes the formation of N^G-monomethylarginine and symmetric N^G,N'^G-dimethylarginine (35) and is the catalytic component of the methylosome that regulates snRNP assembly (36). The second enzyme, PRMT1, is the predominant type I methyltransferase in cells and catalyzes the formation of N^G-monomethylarginine and asymmetric N^G,N^G-dimethylarginine (35). Thus, given that FIB is involved in ribosome biogenesis and is itself an RNA methyltransferase, our results provide biochemical evidence that these three different types of methyltransferases physically associate and therefore may function in concert. Given that, during ribosome biogenesis, not only are pre-rRNAs methylated but several nucleolar and ribosomal proteins as well (37–39), three types of methylation reactions may be required to coordinate certain stages of ribosome biogenesis and/or related processes.

FIB is known to interact with SMN protein (19), and we found that FIB interacts with PRMT5 (JBP1). Because SMN

protein is cooperated with PRMT5 in the methylosome complex during snRNP biogenesis (40), the result coupled with the fact that FIB interacts directly with SMN and co-localizes with SMN in Cajal bodies suggests that FIB may also participate in the regulatory mechanism by which the methylosome and the SMN complex mediate snRNP assembly. Thus, the protein- and RNA-methylation reactions may also occur simultaneously within very close proximity on a single molecular scaffold during snRNP biogenesis.

Another intriguing finding of this study is that FIB also interacts with SF2A-p32, which regulates RNA splicing by inhibiting both the binding of splicing factor ASF/SF2 to pre-mRNA and the phosphorylation of ASF/SF2 (41). Because it may compete with ASF/SF2 for SF2A-p32 binding, mRNA splicing may also be regulated by FIB.

In addition to SF2A-p32 and PRMTs, tubulins $\alpha 3$ and $\beta 1$ were also identified in the FIB-associated sub-complex. These microtubule subunits are heterodimers composed of one polypeptide each of α- and β-tubulin. Therefore, their presence in the sub-complex may provide a link to components of the microtubule cytoskeleton that consists of a highly organized network of microtubule polymers bound to accessory proteins, including microtubule-associated proteins, molecular motors, and microtubule-organizing proteins. Although no other components of microtubule cytoskeleton were found, the tubulins may be involved in transporting the FIB-associated sub-complex among subnuclear domains, including the nucleolus, nucleoplasm, and Cajal bodies.

Despite the preferable subcellular localization of FIB in Cajal bodies, its role in Cajal bodies is completely unknown. However, our results provide some clues on the possible roles of FIB in sn(o)RNP biogenesis and pre-mRNA processing in Cajal bodies. In addition, our finding that FIB associates with protein-arginine methyltransferases is also consistent with the recent report that symmetrical dimethylation of arginine is required to localize SMN to Cajal bodies as well as for its involvement in pre-mRNA splicing (42). FIB may function as a transporter of protein-arginine methyltransferases to Cajal bodies and/or as a scaffold to perform coordinated methylation of both RNA and protein substrates during RNA editing in Cajal bodies as well as in ribosome biogenesis in the nucleolus.

REFERENCES

1. Warner, J. R. (1990) *Curr. Opin. Cell Biol.* **2,** 521–527
2. Eichler, D. C., and Craig, N. (1994) *Prog. Nucleic Acids Res. Mol. Biol.* **49,** 197–239
3. Snaar, S., Wiesmeijer, K., Jochemsen A, G., Tanke, H. J., and Dirks, R. W. (2000) *J. Cell Biol.* **151,** 653–662
4. Spector, D. L. (2001) *J. Cell Sci.* **114,** 2891–2893
5. Smith, C. M., and Steitz, J. A. (1997) *Cell* **89,** 669–672
6. Kiss-Laszlo, Z., Henry Y., Bachellerie J. P., Caizergues-Ferrer, M., and Kiss, T. (1996) *Cell* **85,** 1077–1088
7. Tycowski, K. T., Smith, C. M., Shu, M. D., and Steitz, J. A. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93,** 14480–14485
8. Tyc, K., and Steitz, J. A. (1989) *EMBO J.* **8,** 3113–3119
9. Baserga, S. J., Yang, X. D., and Steitz, J. A. (1991) *EMBO J.* **10,** 2645–2651
10. Omer, A. D., Ziesche, S., Ebhardt, H., and Dennis, P. P. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99,** 5289–5294
11. Schimmang, T., Tollervey, D., Kern, H., Frank, R., and Hurt, E. C. (1989) *EMBO J.* **8,** 4015–4024
12. Henriquez, R., Blobel, G., and Aris, J. P. (1990) *J. Biol. Chem.* **265,** 2209–2215
13. Lapeyre, B., Mariottini, P., Mathieu, C., Ferrer, P., Amaldi, F., Amalric, F., and Caizergues-Ferrer, M. (1990) *Mol. Cell Biol.* **10,** 430–434
14. Aris, J. P., and Blobel G. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88,** 931–935
15. Jansen, R. P., Hurt, E. C., Kern, H., Lehtonen, H., Carmo-Fonseca, M., Lapeyre, B., and Tollervey, D. (1991) *J. Cell Biol.* **113,** 715–729
16. Turley, S. J., Tan, E. M., and Pollard, K. M. (1993) *Biochim. Biophys. Acta* **1216,** 119–122
17. David, E., McNeil, J. B., Basile, V., and Pearlman, R. E. (1997) *Mol. Biol. Cell* **8,** 1051–1061
18. Tollervey, D., Lehtonen, H., Jansen, R., Kern, H., and Hurt, E. C. (1993) *Cell* **72,** 443–457
19. Jones, K. W., Gorzynski, K., Hales, C. M., Fischer, U., Badbanchi, F., Terns, R. M., and Terns, M. P. (2001) *J. Biol. Chem.* **276,** 38645–38651
20. Wehner, K. A., Ayala, L., Kim, Y., Young, P. J., Hosler, B. A., Lorson, C. L., Baserga, S. J., and Francis, J. W. (2002) *Brain Res.* **945,** 160–173
21. Nicol, S. M., Causevic, M., Prescott, A. R., and Fuller-Pace, F. V. (2000) *Exp. Cell Res.* **257,** 272–280
22. Wang, H., Boisvert, D., Kim, K. K., Kim, R., and Kim, S. H. (2000) *EMBO J.* **19,** 317–323
23. Lischwe, M. A., Cook, R. G., Ahn, Y. S., Yeoman, L. C., and Busch, H. (1985) *Biochemistry* **24,** 6025–6028
24. Lischwe, M. A., Ochs, R. L., Reddy, R., Cook, R. G., Yeoman, L. C., Tan, E. M., Reichlin, M., and Busch, H. (1985) *J. Biol. Chem.* **260,** 14304–14310
25. Yanagida, M., Shimamoto, A., Nishikawa, K., Furuichi, Y., Isobe, T., and Takahashi, N. (2001) *Proteomics* **1,** 1390–1404
26. Fujiyama, S., Yanagida, M., Hayano, T., Miura, Y., Uchida, T., Fujimori, F., Isobe, T., and Takahashi, N. (2002) *J. Biol. Chem.* **277,** 23773–23780
27. Hayano, T., Yanagida, M., Yamauchi, Y., Shinkawa, T., Isobe, T., and Takahashi, N. (2003) *J. Biol. Chem.* **278,** 34309–34319
28. Hofer, L., Weichselbraun, I., Quick, S., Farrington, G. K., Bohnlein, E., and Hauber, J. (1991) *J. Virol.* **65,** 3379–3383
29. Goldfarb, D. S., Gariepy, J., Schoolnik, G., and Kornberg, R. D. (1986) *Nature* **322,** 641–644
30. Natsume, T., Yamauchi, Y., Nakayama, H., Shinkawa, T., Yanagida, M., Takahashi, N., and Isobe, T. (2002) *Anal. Chem.* **74,** 4725–4733
31. Andersen, J. S., Lyon, C. E., Fox, A. H., Leung, A. K. L., Lam, Y. W., Steen, H., Mann, M., and Lamond, A. I. (2002) *Curr. Biol.* **12,** 1–11
32. Scheri, A., Coute, Y., Deon, C., Calle, A., Kindbeiter, K., Sanchez, J.-C., Greco, A., Hochstrasser, D., and Diaz, J.-J. (2002) *Mol. Biol. Cell* **13,** 4100–4109
33. Nissan, T. A., Bassler, J., Petfalski, E., Tollervey, D., and Hurt, E. (2002) *EMBO J.* **21,** 5539–5547
34. Takahashi, N., Yanagida, M., Fujiyama, S., Hayano, T., and Isobe, T. (2003) *Mass Spectrom. Rev.* **22,** 287–317
35. Gary, J. D., and Clarke, S. (1998) *Prog. Nucleic Acids Res. Mol. Biol.* **61,** 65–131
36. Friesen, W. J., Paushkin, S., Wyce, A., Massenet, S., Pesiridis, G. S., Van Duyne, G., Rappsilber, J., Mann, M., and Dreyfuss, G. (2001) *Mol. Cell Biol.* **21,** 8289–8300
37. Pintard, L., Kressler, D., and Lapeyre, B. (2000) *Mol. Cell Biol.* **20,** 1370–1381
38. Niewmierzycka, A., and Clarke, S. (1999) *J. Biol. Chem.* **274,** 814–824
39. Whitehead, S. E., Jones, K. W., Zhang, X., Cheng, X., Terns, R. M., and Terns, M. P. (2002) *J. Biol. Chem.* **277,** 48087–48093
40. Pellizzoni, L., Yong, J., and Dreyfuss, G. (2002) *Science* **298,** 1775–1779
41. Petersen-Mahrt, S. K., Estmer, C., Ohrmalm, C., Matthews, D. A., Russell, W. C., and Akusjavi, G. (1999) *EMBO J.* **18,** 1014–1024
42. Boisvert, F. M., Cote, J., Boulanger, M. C., Cleroux, P., Bachand, F., Autexier, C., and Richard, S. (2002) *J. Cell Biol.* **159,** 957–969