THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 281, NO. 38, pp. 28278–28286, September 22, 2006

RNA Structural Requirements for the Association of the Spliceosomal hPrp31 Protein with the U4 and U4atac Small Nuclear Ribonucleoproteins*[S]

Received for publication, April 7, 2006, and in revised form, July 14, 2006 Published, JBC Papers in Press, July 20, 2006, DOI 10.1074/jbc.M603350200

Annemarie Schultz[1], **Stephanie Nottrott**[1,2], **Klaus Hartmuth, and Reinhard Lührmann**[3]

From the Department of Cellular Biochemistry, Max Planck Institute for Biophysical Chemistry, D-37077 Göttingen, Germany

The kink-turn, a stem I-internal loop-stem II structure of the 5′ stem-loop of U4 and U4atac small nuclear (sn) RNAs bound by 15.5K protein is required for binding of human Prp31 protein (hPrp31) during U4 and U4atac snRNP assembly. In box C/D snoRNPs a similar kink-turn with bound 15.5K protein is required for selective binding of proteins NOP56 and NOP58. Here we analyzed RNA structural requirements for association of hPrp31 with U4 snRNP *in vitro* by hydroxyl radical footprinting. hPrp31 induced protection of the terminal penta-loop, as well as of stems I and II flanking the kink-turn. Similar protection was found with U4/U6 snRNA duplex prebound with 15.5K protein. A detailed mutational analysis of the U4 snRNA elements by electrophoretic mobility shift analysis revealed that stem I could not be shortened, although it tolerated sequence alterations. However, introduction of a third Watson-Crick base pair into stem II significantly reduced hPrp31 binding. While stem I of U4atac snRNA showed relaxed binding requirements, its stem II requirements were likewise restricted to two base pairs. In contrast, as shown previously, stem II of the kink-turn motif in box C/D snoRNAs is comprised of three base pairs, and NOP56 and NOP58 require a G-C pair at the central position. This indicates that hPrp31 binding specificity is achieved by the recognition of the two base pair long stem II of the U4 and U4atac snRNAs and suggests how discrimination is achieved by RNA structural elements during assembly of U4/U6 and U4atac/U6atac snRNPs and box C/D snoRNPs.

The processing of pre-mRNA in the nucleus is catalyzed by a large RNA-protein complex, the spliceosome (1–3). The spliceosome comprises, apart from the pre-mRNA substrate, the U1, U2, U4, U5, and U6 snRNPs,[4] as well as numerous splicing factors, that is, proteins that do not make up an integral part of the snRNPs. Each snRNP is a stable complex, consisting of a single RNA molecule and various proteins, of which some (the so-called Sm proteins) are common to all snRNPs, while others are specific to a particular snRNP (2). The individual snRNPs interact with one another in a dynamic manner during the splicing cycle. For example, at the beginning of this cycle the U4 snRNP and the U6 snRNP are associated firmly with one another by two extended stretches of RNA-RNA base pairs. Two intermolecular helices of the U4/U6 snRNA are separated by an intramolecular 5′-terminal hairpin loop (5′ stem-loop) of the U4 snRNA (4–7). The 5′-terminal stem-loop of the U4 snRNA (Fig. 1*B*) consists of a stem I, an intramolecular loop and then a stem II. The latter ends in a pentanucleotide loop ("penta-loop"). Thus at this stage the U4 and U6 snRNP form a single particle, the U4/U6 di-snRNP. This complex, in turn, associates with the U5 snRNP and thus the U4/U6·U5 tri-snRNP is formed, which is integrated as such into the precatalytic spliceosome. During the rearrangement of the fully assembled spliceosome to its catalytically active state, the intermolecular base pairing between the U4 and U6 snRNAs becomes disrupted and the U4 snRNP particle is released from the spliceosome. In the subsequent steps, a base-paired region is formed between the U6 snRNA and the U2 snRNA, which leads to the formation of the catalytically active center (8–10). Following the processing of the pre-mRNA, the spliceosome dissociates, and the U4, U5, and U6 snRNPs reassemble ready for a new round of splicing (10, 11).

A key part in the stability of the tri-snRNP and its formation from the U4/U6 di-snRNP and the U5 snRNP is played by the U4/U6-specific protein hPrp31 (also known as 61K). Studies *in vitro*, using HeLa cell nuclear extract depleted of hPrp31, showed that hPrp31 is essential for the splicing of pre-mRNA and for the formation of the human tri-snRNP particle (12). The same is true in living human cells, for which it was shown by RNA interference-mediated gene silencing of hPrp31 that this protein is needed for the stability of the tri-snRNP particle (13).

Protein hPrp31 performs a dual function in the tri-snRNP particle. On the one hand, hPrp31 binds to the U4 snRNA in the U4/U6 di-snRNP (14), although the exact nature of this interaction, the point of contact and the mode of binding, is largely

* This work was supported by grants from the Deutsche Forschungsgemeinschaft (LU294/12-1), the Ernst-Jung-Stiftung, and the Fonds der Chemischen Industrie (to R. L.). 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 http://www.jbc.org) contains supplemental Figs. S1 and S2 and Refs. 1 and 2.

[1] These authors contributed equally to this work.

[2] Present address: University of Massachusetts Medical School, Program in Molecular Medicine, 373 Plantation St., Worcester, MA 01605.

[3] To whom correspondence should be addressed: Dept. of Cellular Biochemistry, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, D-37077 Göttingen, Germany. Tel.: 49-551-201-1405; Fax: 49-551-201-1197; E-mail: reinhard.luehrmann@mpi-bpc.mpg.de.

[4] The abbreviations used are: snRNP, small nuclear ribonucleoprotein particle; snoRNP, small nucleolar RNP; snRNA, small nuclear RNA; snoRNA, small nucleolar RNA; MBP, maltose-binding protein; EMSA, electrophoretic mobility shift analysis.

unknown. On the other hand, hPrp31 forms a bridge between the U4/U6 di-snRNP and the U5 snRNP by binding to the U5-specific protein hPrp6 (also known as U5–102K; Ref. 12). The importance of the bridging function of hPrp31 between the U4/U6 di-snRNP and the U5 snRNP particle is also supported by the observation that, in the course of the catalytic activation of the spliceosome, the association of the U4 snRNP with the U6 snRNP on the one hand and the U5 snRNP on the other hand is weakened and the hPrp31 dissociates along with the U4 snRNP (15).

Assembly studies *in vitro* have shown that hPrp31 does not bind alone to the U4/U6 di-snRNP. First of all, the protein 15.5K must bind to the intramolecular loop ("kink-turn motif"; Ref. 16) of the 5′-terminal hairpin loop of the U4 snRNA (14). Protein 15.5K is also necessary for the binding of the U4/U6-specific heterotrimeric CypH/hPrp4/hPrp3 (also known as 20/60/90K) protein complex. It thus functions as a nucleation factor. While the binding of CypH/hPrp4/hPrp3 depends on the prior binding of protein 15.5K, the protein hPrp31 and the complex CypH/hPrp4/hPrp3 bind independently of one another. The situation just described for the U4/U6 di-snRNP particle is the same for the binding of hPrp31 to the U4atac/U6atac-snRNAs of the minor U12-type spliceosome (14, 17).

By ultraviolet irradiation of the U4atac-snRNP and the native purified U4/U6·U5 tri-snRNP particle, hPrp31 has been cross-linked to single-stranded nucleotide regions at the three-way junction of the base-paired U4 and U6 snRNAs and the terminal penta-loop of the U4atac and the U4 snRNAs (14, 18). This argues for the existence of direct contacts between hPrp31 and the U4 snRNA. The formation of the U4-15.5K complex suffices for the stable binding of hPrp31. However, it remains unclear which regions of the U4 snRNA are required for this and what factors determine its specificity. The corresponding determinants in hPrp31 are likewise unknown.

Protein hPrp31 possesses a so-called Nop domain in its central region, which shows clear homology to a domain of the box C/D snoRNP-associated proteins NOP56 and NOP58 (12, 19, 20). Above all, hPrp31 shows functional homology to NOP56 and NOP58. Here too, the binding of protein 15.5K to the box C/D snoRNA, which, like the U4 snRNA, is capable of forming a kink-turn motif, is a prerequisite for the binding of NOP56 und NOP58 to the box C/D snoRNA (21). This raises the question of how the specificity of the binding of hPrp31 to the U4 snRNA, and not to the box C/D snoRNA, is achieved. Furthermore, it has recently been shown that hPrp31 is also of clinical interest. Mutations in its gene (PRPF31) are correlated with the autosomal dominant form of retinitis pigmentosa, a disorder that leads to degeneration of the photoreceptors of the eye (20). Interestingly, mutations associated with this disease are situated in and around the conserved central Nop homology domain.

In this work, we have analyzed the structural requirements for hPrp31 association with the U4 snRNP *in vitro*. We have identified nucleotides within the U4 snRNA that are protected from hydroxyl-radical attack in the presence of protein hPrp31. By employing U4 and U4atac snRNAs deletion and point mutants, we were able to determine the structural requirements within the U4 and U4atac snRNAs for the binding of protein hPrp31 to the U4 and U4atac snRNPs. Our findings provide insight into how binding specificity is achieved in the exclusive interaction of proteins hPrp31 and NOP56/NOP58 with their targets, the U4/U6 and U4/U4atac snRNPs and box C/D snoRNPs, respectively.

EXPERIMENTAL PROCEDURES

Oligoribonucleotides—All oligoribonucleotides were obtained from Dharmacon and are described in the legends to Figs. 2–4, except for the following oligoribonucleotides: U4-9, 5′-AUCGUAGCCAAUGAGGUCGACUUUAUGUCGACCGAGGCGCGAU-3′; U4-10, 5′-GCCGCGCCAAUGAGGUUUAUCCGAGGCGCGGC-3′; U4-11, 5′-AUCGCGCCAAUGAGGUUUAUCCGAGGCGCGAU-3′; U4 12, 5′-AUCGUAGCCAAUGAGGUCGUUUAUCGACCGAGGCGCGAU-3′.

Purification of Recombinant Proteins—Expression and purification of the 15.5K and glutathione *S*-transferase-hPrp31 proteins were performed by the procedures described (22 and 12, respectively). The MBP-hPrp31 fusion protein was created by first subcloning of the hPrp31 cDNA into pETM-41 (EMBL). *Escherichia coli* strain HMS174 was transformed with pETM-41-hPrp31 and grown at 20 °C to an A_{600} of 0.8. isopropyl β-D-thiogalactopyranoside was added to a concentration of 1 mM to induce protein expression. After incubation for 3 h at 20 °C, cells were pelleted. Cells were resuspended in 25 ml of buffer B (50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol) and incubated for 1 h at 4 °C in the presence of 1 mg/ml lysozyme (Sigma) and one tablet of complete protease inhibitor (Roche Applied Science). Cells were then subjected to sonication with a microtip (Branson) and three 20-s bursts at an amplitude of 40%. Insoluble material was pelleted by centrifugation at 10,000 × *g* for 30 min, and the MBP-hPrp31-containing lysate was loaded directly onto 2 ml of amylose-Sepharose (New England Biolabs). Bound MBP-hPrp31 was washed once with 20 ml of buffer C (50 mM Tris-HCl (pH 8.0), 300 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol) and twice with buffer B. The MBP-hPrp31 protein was finally eluted stepwise with buffer B supplemented with 10 mM maltose in 500-μl fractions. Peak fractions were dialyzed against buffer B.

Hydroxyl Radical Footprinting—U4 snRNA (0.6 pmol; prepared by transcription *in vitro* (22)) was incubated for 1 h at 4 °C with 25 pmol of recombinant 15.5K and/or 50 pmol of hPrp31 proteins in a final volume of 40 μl buffer A (20 mM HEPES-KOH (pH 7.9), 150 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.1% Triton X-100). Ribose cleavage of the RNA backbone was initiated by hydroxyl radicals generated with H_2O_2 and Fe(II)-EDTA as described (22), and cleavage products were analyzed by primer extension using a ^{32}P-labeled oligonucleotide complementary to nucleotides 65–82 of human U4 snRNA as before (22).

Electrophoretic Mobility Shift Assay (EMSA)—Approximately 0.03 pmol of ^{32}P-labeled U4 snRNA oligoribonucleotides were incubated with 25 pmol of recombinant 15.5K and/or 50 pmol of MBP-hPrp31 proteins and 10 μg of *E. coli* tRNA (Roche Applied Science) for 1 h at 4 °C in a final volume of 20 μl of buffer B. Subsequently, RNA and RNA-protein complexes were resolved on a 6% (80:1) native polyacrylamide gel containing 0.5× Tris borate-EDTA and visualized by autoradiography

as described (22). Additionally, all experiments were quantified with a PhosphorImager.

RESULTS

Hydroxyl Radical Footprinting Studies Reveal Protection of Various Regions of the U4 snRNA by Protein hPrp31—It was previously shown that protein hPrp31 is only in direct contact with the U4 snRNA after protein 15.5K binds to the kink-turn motif of U4 snRNA (14). To narrow down the possible RNA contacts in the ternary 15.5K-hPrp31-U4 snRNA complex, we investigated it by hydroxyl radical footprinting (22). In this method, any ribose of an RNA that is shielded by protein will be protected from cleavage of the RNA backbone induced by hydroxyl radicals. Analysis is by primer extension with reverse transcriptase. Through a comparison with cleavage sites on naked RNA, the absence of a stop before a particular position is considered as a protection at this position in the complex.

We assembled the ternary 15.5K-hPrp31-U4 snRNA complex by incubating full-length U4 snRNA with an excess of recombinant 15.5K and glutathione *S*-transferase-hPrp31 protein (14). Under the conditions used more than 95% of the U4 snRNA was converted into the ternary complex, as assayed by EMSA (data not shown). To distinguish nucleotides protected by protein hPrp31 from those protected by protein 15.5K only, a binary 15.5K-U4 snRNA complex was assembled in a parallel sample; naked RNA was used as a reference. These complexes were treated with H_2O_2 in the presence of Fe(II)-EDTA to generate hydroxyl radicals. The RNA was subsequently analyzed by primer extension (Fig. 1*A*) of the U4 snRNA.

A comparison of the cleavage pattern of naked RNA (*lane 2*) with that of a control omitting the reagent Fe(II)-EDTA (*lane 1*) shows distinct cleavages at all nucleotides (compare *lanes 1* and *2*). *Lane 3* shows the hydroxyl radical cleavage pattern observed in the binary 15.5K-U4 snRNA complex. Strong protection is found in the kink-turn and in the adjacent stems I (A29-G34) and II (G43-C47) (*lanes 2* and *3*; compare Fig. 1*A*, FP 2 and FP 1). This is in agreement with the results of our earlier studies (22, 23) in which these nucleotides were identified as the binding site for protein 15.5K. The cleavage pattern of U4 snRNA in the ternary complex revealed a number of additional protection sites arising from the presence of protein hPrp31. One region extended from G35 to C42 (*lanes 3* and *4*; FP A in Fig. 1*A*), comprising the penta-loop and nucleotides of stem II not protected by protein 15.5K. Furthermore, protection was observed in stem I next to the kink-turn at positions G26, C27, and C28 (FP B) and at nucleotides C16 and G18 (FP C).

In the U4/U6 snRNA duplex, FP C would be located in the U4/U6 helix II, adjacent to the three-way junction of the U4/U6 duplex. We were therefore interested whether protein hPrp31 would similarly protect these nucleotides when binding to the 15.5K-U4 snRNA complex in the context of the duplex. To investigate this, we first created the duplex by annealing U6 snRNA to U4 snRNA before addition of proteins 15.5K and hPrp31. Formation of the U4/U6 duplex was quantitative, and the duplex was as efficient as the single-stranded U4 snRNA in ternary complex formation, as determined by EMSA (data not shown). Hydroxyl radical footprinting was then performed as above, on (i) the protein-free U4/U6 duplex (Fig. 1*A*, *lane 6*), (ii) the complex formed between the duplex and protein 15.5K (Fig. 1*A*, *lane 7*), and (iii) the complex formed between the preformed 15.5K-U4/U6 snRNA complex and protein hPrp31 (Fig. 1*A*, *lane 8*). Duplex formation on its own did not influence the accessibilities of the riboses in U4 snRNA (Fig. 1*A*, *lanes 6* and *2*), except for a reduced reactivity at A39-G43 in the U4/U6 snRNA duplex. Similarly, the footprint of protein 15.5K on the U4 in the U4/U6 duplex was essentially identical to the one described above (Fig. 1*A*, *lanes 7* and *3*, *FP 1* and *FP 2*). The addition of protein hPrp31 resulted in protection of nucleotides essentially identical to those seen in the absence of U6 snRNA (Fig. 1*A*, *lanes 8* and *4*). In particular, the protection of the penta-loop and the adjacent stem II was unchanged (FP A). Protection at the bottom of stem I (FP B), adjacent to the binding site for protein 15.5K, was more extensive in that A25 also showed protection. Somehow more enlarged was FP C, which was found to extend from G14 to G18. Fig. 1*B* depicts the footprints on the secondary structures of U4 snRNA (24) and the U4/U6 snRNA duplex (4), respectively. The overall pattern of protection of U4 snRNA by protein hPrp31 was thus comparable in the hPrp31-15.5K-U4 snRNA and the hPrp31-15.5K-U4/U6 snRNA complexes, with the only differences observed on the borders of the protected regions in FP B and C. In summary, these data show that *in vitro* protein hPrp31 protects regions of stems I and II, the complete terminal penta-loop, and nucleotides in the helix adjacent to the three-way junction.

Protein hPrp31 Requires the Complete Stem I of the 5′ Stem-loop of U4 snRNA for Binding—We next investigated which of the structural elements of U4 snRNA that were found to be protected from hydroxyl radical attack by protein hPrp31 in the footprinting experiment were required for stable binding of protein hPrp31 to the 15.5K-U4 snRNA binary complex. For this more detailed analysis, we prepared a series of mutant RNA oligonucleotides and assayed them for ternary complex formation by EMSA. To ease solubilization of hPrp31 we used a fusion protein comprising the maltose-binding protein (MBP) fused to hPrp31. This behaves in a manner identical to non-tagged hPrp31 (data not shown). In all experiments, the binary 15.5K-U4 snRNA complex (*B* in Fig. 2) was first assembled with recombinant 15.5K protein and radiolabeled RNA oligonucleotide. Formation of the ternary complex (*T* in Fig. 2) was then initiated by adding MBP-hPrp31 protein. Complexes were then analyzed directly by native PAGE at 4 °C.

Since the single-stranded 5′-region of the U4 snRNA was protected by the protein, we first deleted this segment (U4-2, Fig. 2). No reduction in the amount of ternary complex formed (Fig. 2, *lanes 3* and *7*) was observed compared with the longer deletion mutant U4 snRNA (U4-1), which binds in a manner identical to that of the wild-type U4 snRNA (14). Hence, the contact between the single-stranded 5′ end of the U4 snRNA and protein hPrp31 is not required for stable binding. However, the deletion of five base pairs from stem I significantly depressed formation of stable ternary complex (U4-3; Fig. 2, *lanes 3* and *11*). Residual ternary complexes (in yields of ~20% compared with the U4-2 or U4-1 RNA) were unstable, resulting in an ill-defined smear migrating above the binary 15.5K-U4 snRNA complex. Complex formation could not be restored by extending the truncated three base pair long stem I of mutant

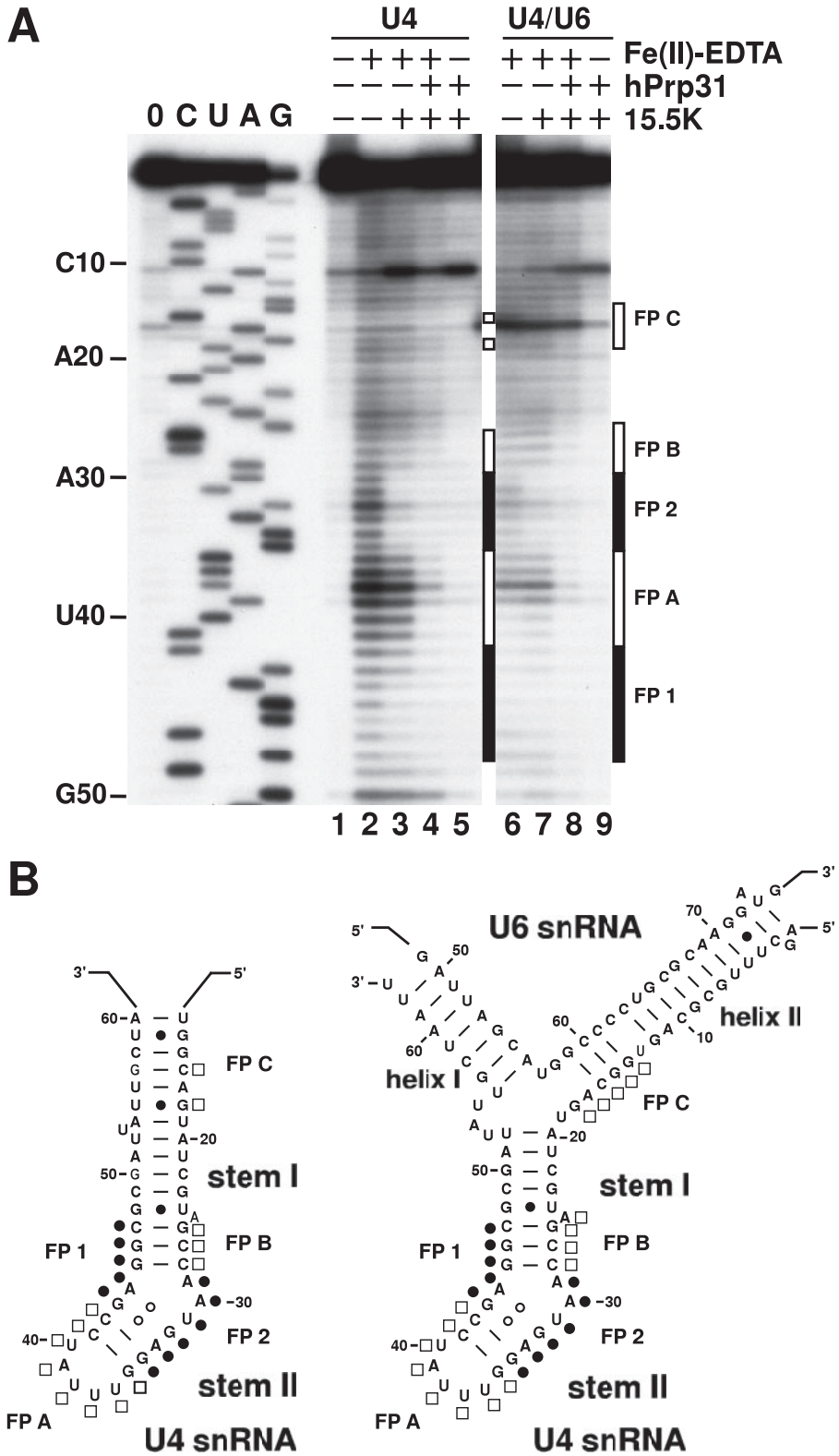

FIGURE 1. **Protein hPrp31 binds directly to U4 snRNA.** *A*, hydroxyl radical footprints of the U4 snRNA-15.5K-hPrp31 and the U4/U6-15.5K-hPrp31 protein complexes analyzed by primer extension, with a primer was complementary to nucleotides 65–82 of the U4 snRNA. Riboses protected from cleavage in the presence of protein 15.5K are indicated by *filled bars* on the *right* (FP 1 and FP 2), and those protected additionally by hPrp31 by *open bars* (FP A, FP B, and FP C). The reverse transcriptase stop is one nucleotide before the actual cleavage site. *Lane 1* contains unmodified U4 snRNA as a control for spontaneous stops by the reverse transcriptase. C, U, A, and G refer to dideoxy sequencing reactions (0, no ddNTPs added). The footprints were additionally analyzed by densitometry of the autoradiographs, and the detailed quantification is shown in supplemental Fig. S1. *B*, superposition of the cleavage data onto the secondary structure models of the human U4 snRNA 5′ stem-loop (Ref. 24; *left*) and the U4/U6 snRNA duplex (Ref. 4; *right*). The *open circles* indicate non-Watson-Crick base pair interactions as found in the crystal structure of the U4 5′ stem-loop bound by protein 15.5K (23). Other non-Watson-Crick base pairs are depicted by *solid dots*. The positions of the footprints of protein 15.5K are indicated by *filled circles* (FP 1 and FP 2) and those of protein hPrp31 by *open squares* (FP A, FP B, and FP C).

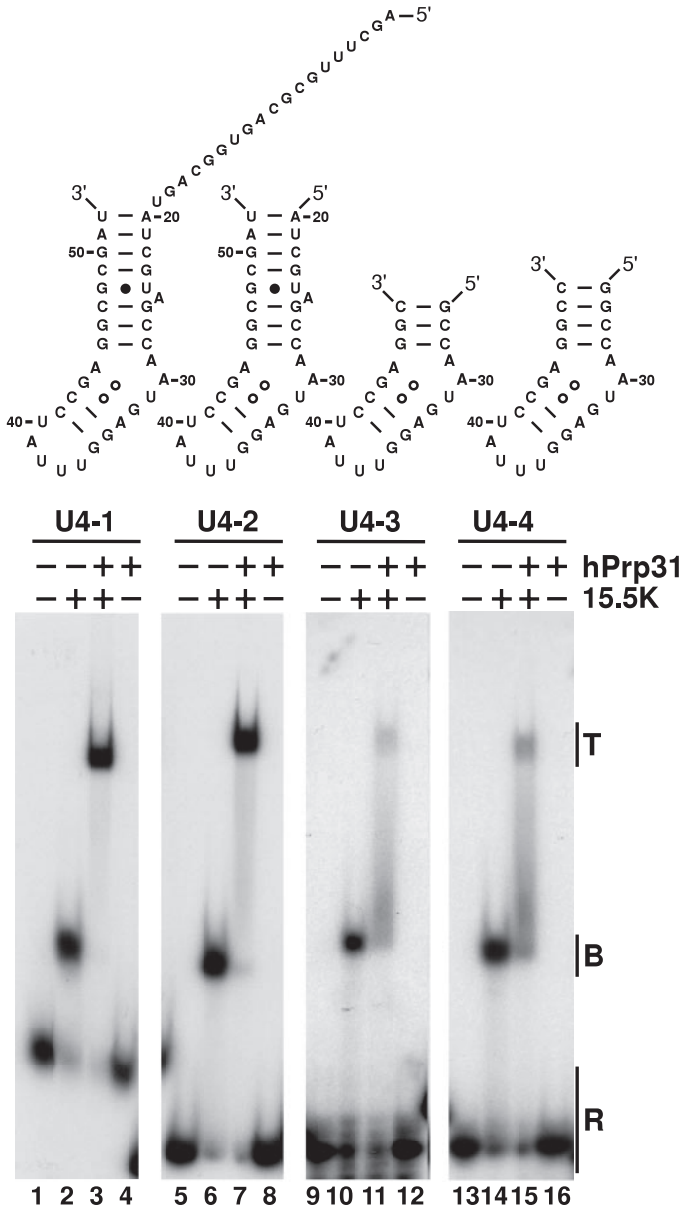

FIGURE 2. **Binding of protein hPrp31 to the U4 snRNP requires the complete stems I and II of the 5′ stem-loop of the U4 snRNA.** The binary complex was first assembled with protein 15.5K and the different mutant U4 snRNAs (*top panels*), and ternary complex formation was initiated by adding protein hPrp31. Complexes were analyzed by EMSA. The positions of the protein-RNA complexes are indicated on the *right*: *R*, free RNA; *B*, binary complex consisting of the RNA and protein 15.5K; *T*, ternary complex, consisting of the RNA, protein 15.5K, and hPrp31. The sequence and structure of the respective U4 snRNA deletion mutant used is indicated above each panel. Base pairing interactions are the same as described for Fig. 1.

U4-3 by a C-G base pair (U4-4, SL1 in Ref. 22). Furthermore, changing the sequence of the upper part of stem I while leaving the number of base pairs intact, or deleting the bulged-out A, did not reduce the yield of ternary complex (oligonucleotides U4-10 and U4-11, respectively, see "Experimental Procedures"; data not shown). We conclude that the length of stem I is critical for productive interaction of protein hPrp31 with the binary 15.5K-U4 snRNA complex.

Protein hPrp31 Does Not Require the Penta-loop of the 5′ Stem-loop of U4 snRNA for Binding—The hydroxyl radical footprint showed extensive protection of the entire penta-loop

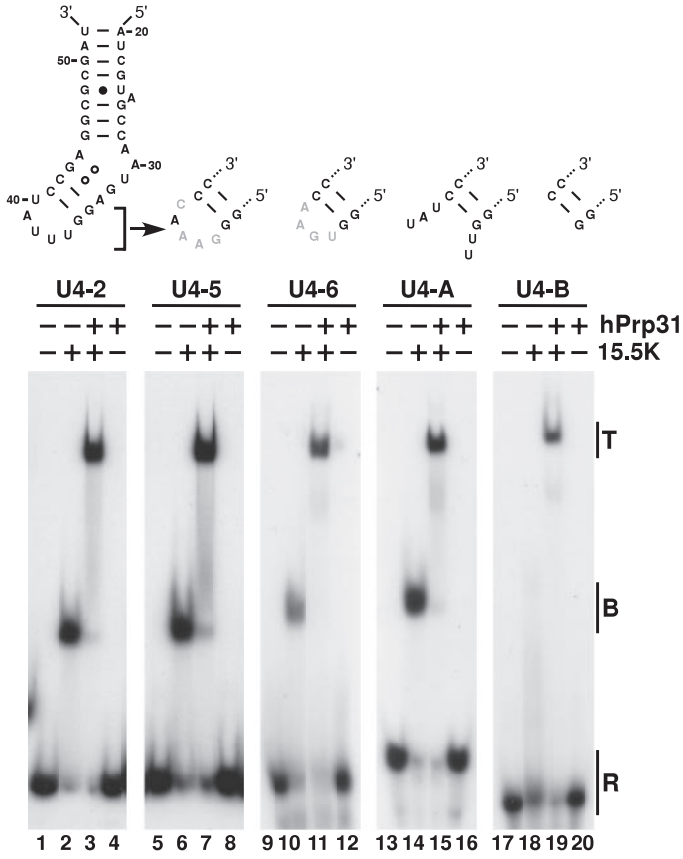

FIGURE 3. **Binding of protein hPrp31 to the U4 snRNP does not depend on sequence and structure of the terminal penta-loop.** Different U4 snRNA 5′ stem loop mutants were analyzed by EMSA for hPrp31 binding as in Fig. 2. U4-A and U4-B duplexes were generated by annealing equal amounts of ^{32}P-labeled 5′ and unlabeled 3′ halves of the 5′ stem-loop before the assay. Schematic representations of the sequences and the structures of each U4 snRNA 5′ stem loop mutation (derived from the U4-2 RNA) are indicated above the respective panel with nucleotide changes shown in *gray*. For other penta-loop variations see text for details; labeling is the same as in Fig. 2.

and protein hPrp31 in the ternary complex (Fig. 1). We therefore next tested a number of different penta-loop mutants and assayed their effect on stable ternary complex formation in the above experimental system. We used the U4-2 mutant (Fig. 2), which has wild-type properties while being closest in length to the other mutants, as a reference with which to compare the mutants. Replacement of the penta-loop by an unrelated sequence did not impair complex formation (U4-5, Fig. 3). Replacement with a tetranucleotide loop (UGAA in U4-6, Fig. 3; other loops tested were UUCG and GCAA) likewise had no effect (Fig. 3, *lanes 3* and *11*, and data not shown).

We therefore introduced a more dramatic change into the backbone of the penta-loop by opening it up. An RNA duplex with dangling ends was created by annealing the 5′ (20–37) and 3′ (38–52) halves of the U4 snRNA 5′ stem-loop (U4-A, Fig. 3). Unexpectedly, this duplex still allowed efficient ternary complex formation (Fig. 3, *lanes 3* and *15*). Finally, we simply deleted the penta-loop by annealing 5′ and 3′ halves lacking the penta-loop sequence entirely (U4-B, Fig. 3). The stability of this duplex was dramatically impaired (as assayed by EMSA; data not shown); it no longer supported the formation of significant amounts of stable binary 15.5K-U4 snRNA complex (Fig. 3, *lane 18*) compared with the reference

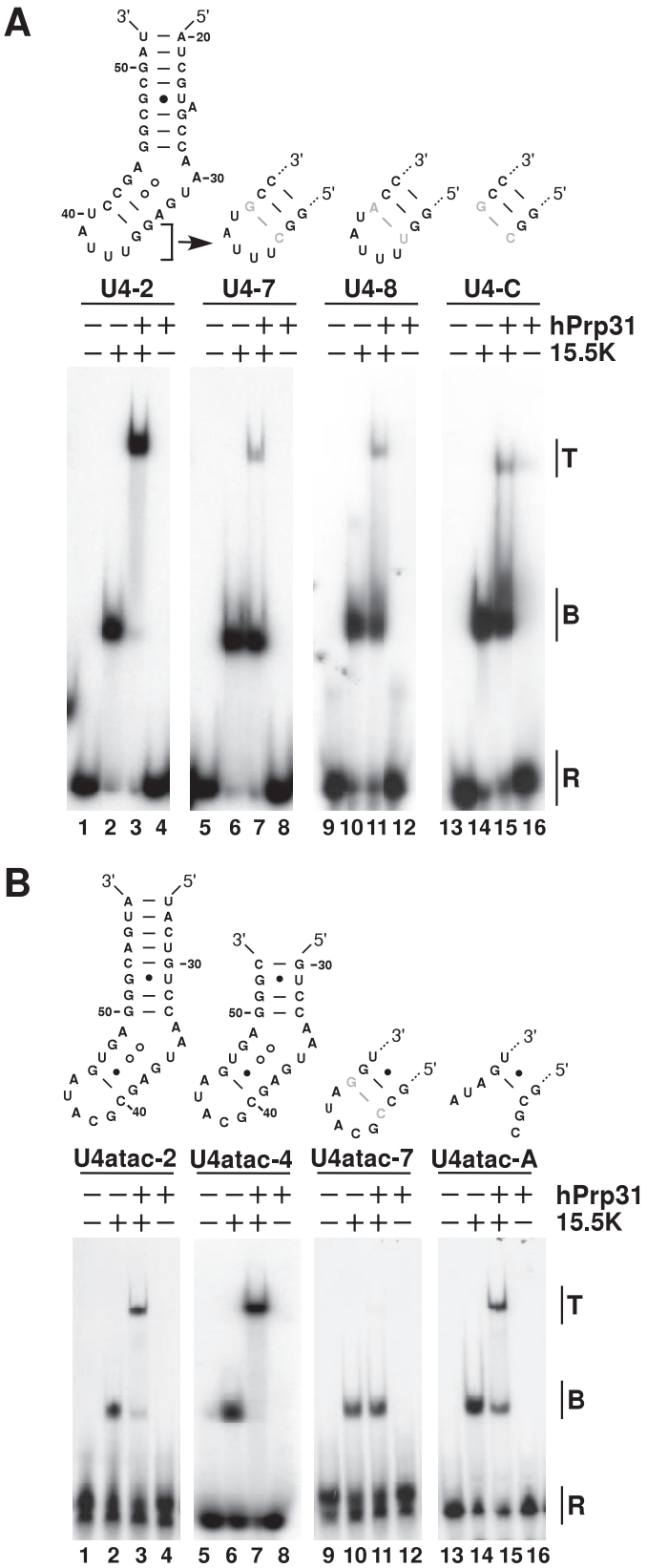

FIGURE 4. **Length of stem II is critical for hPrp31 binding to U4 and U4atac snRNP.** EMSA analysis of the interaction between protein hPrp31 and the 15.5K-U4 snRNA (*A*) or the 15.5K-U4atac snRNA (*B*) complex containing mutant RNAs was performed as described in Fig. 2. The U4-C and U4atac-A mutants were prepared as U4-A and U4-B in Fig. 3. Schematic representations of the sequences and the structures of each U4 or U4atac snRNA 5′ stem loop mutation (derived from the U4-2 or U4atac-2 RNAs, respectively) are indicated above the respective panel with base pair exchanges shown in gray; labeling is as in Fig. 2. In *B*, the experiment shown in *lanes 5* and *6* was performed separately from that in *lanes 1–4* and *7–16* but under identical conditions.

U4-2 (*lane 2*) and all other mutants tested (*lanes 6*, *10*, and *14*). However, this duplex did support the efficient formation of stable ternary complex upon addition of protein hPrp31 to the binary complex assembly mixture (*lane 19*). Protein hPrp31 therefore appears to stabilize transient interactions of protein 15.5K with U4 snRNA in the ternary complex. We conclude that the penta-loop *per se* is not essential for protein hPrp31 binding to the 15.5K-U4 snRNA binary complex.

Protein hPrp31 Requires a 2-Base Pair-long Stem II of the 5′ Stem-loop of U4 snRNA for Binding—In addition to the penta-loop, protein hPrp31 had also been found to be in contact with the adjacent short stem II (Fig. 1). Since the penta-loop was not required for stable ternary complex formation, we next investigated the function of stem II in ternary complex formation.

To this end, we first extended stem II by a single C-G base pair (U4-7, Fig. 4*A*). Strikingly, this mutant no longer supported ternary complex formation in our assay system (Fig. 4*A*, *lanes 3* and *7*; ~90% reduction of ternary complex formation). Extension by a U-A (U4-8, Fig. 4*A*) base pair had a similar effect (Fig. 4*A*, *lane 11*). Not surprisingly, identical results were obtained by increasing the length of the stem to 5 or 7 base pairs (oligonucleotides U4-12 and U4-9, respectively, see "Experimental Procedures"; data not shown). Furthermore, deletion of the penta-loop with retention of the three base pair long stem II (U4-C, Fig. 4 *A*) was similarly detrimental to ternary complex formation (Fig. 4*A*, *lane 15*). The inhibitory effect of this additional base pair was thus independent of the presence of the penta-loop. A shortening of stem II was not deemed feasible because the already weak binding of protein 15.5K to the penta-loop deletion (mutant U4-3) would be further compromised by interfering with the 15.5K binding site (22, 23). Taken together, these data show that the length of two base pairs of stem II is critically important for the association of protein hPrp31 with the 15.5K-U4 snRNA binary complex.

In summary, these results show that both stem I and stem II are required for stable binding of protein hPrp31 to the 15.5K-U4 snRNA binary complex, to form the ternary hPrp31-15.5K-U4 snRNA complex. While stem I could not be shortened without loss of binding activity, stem II, unexpectedly, could not be elongated, irrespective of the sequence of stem II. The penta-loop structure was, however, not required for stable ternary complex formation.

RNA Requirements for hPrp31 Interaction with the 15.5K-U4atac snRNA Binary Complex—Protein hPrp31 is known to interact also with the 15.5K-U4atac snRNA binary complex (14); not surprisingly, a U4atac-2 stem loop, analogous to U4-2 (Fig. 4, *A* and *B*), supported efficient ternary complex formation (Fig. 4*B*, *lane 3*). Inspection of stem I structures of U4atac snRNAs reveals that they differ markedly from the corresponding structures in U4 snRNA in that they do not have a bulge and at least one stem I (*Drosophila melanogaster*) is only 7 base pairs long (see supplemental Fig. S2; for the human case compare U4-2 with U4atac-2 in Fig. 4, *A* and *B*,

respectively). This suggests that stem I requirements may be less stringent in U4atac than in U4 snRNA. In fact, ternary complex formation was unaffected in a mutant that had the upper 4 base pairs of U4atac-2 deleted (U4atac-4, Fig. 4*B*, *lane 7*). In contrast, stem II of U4atac snRNA is of the same length but has different base pairs (compare U4-2 with U4atac-2 in Fig. 4, *A* and *B*, respectively). We therefore investigated whether the strict length restriction for stem II for ternary complex formation in U4 snRNA also applies in the different sequence context of U4atac snRNA. Strikingly, insertion of a C-G base pair in stem II (U4atac-7, Fig. 4*B*) completely abolished hPrp31 binding (Fig. 4*B*, *lane 7*). Furthermore, opening up the penta-loop (U4atac-A, Fig. 4*B*) did not impair hPrp31 binding to the U4atac 5′ stem loop (Fig. 4*B*, *lane 11*). Thus the determinants for the stem II length restriction are sequence-independent and are conserved between the divergent U4 and U4atac snRNAs.

DISCUSSION

In this work, we have investigated the binding of hPrp31 to the U4 snRNA 5′ stem-loop complexed with protein 15.5K and have obtained, using hydroxyl radical footprinting analysis, detailed information on the protection of the RNA by these proteins. RNA mutational analysis revealed essential features of the RNA requirements for the RNA-protein interaction in the U4 and U4atac snRNAs. Finally, our results contribute to an understanding of how hPrp31 binds specifically to U4 and U4atac snRNP and proteins NOP56 and NOP58 to the box C/D snoRNPs. These issues are discussed in turn below.

Global RNA Shielding by hPrp31 of the U4 snRNA in the Ternary hPrp31-15.5K-U4 snRNA Complex—In our quest to understand the interaction of hPrp31 with the U4 snRNA in the ternary complex we initially investigated direct protein-induced shielding of RNA by hydroxyl radical footprinting (Fig. 1). On both sides of the kink-turn, directly neighboring the 15.5K binding site (FP 1 and FP 2), discrete footprints were located. First of all, hPrp31 induces protection of the complete penta-loop structure and of all the stem II nucleotides that were not in contact with protein 15.5K (FP A). Second, on the other side of the kink-turn, only a few nucleotides of stem I were found protected by hPrp31 (FP B). An additional protection was found between hPrp31 and nucleotides G18 and C16 in the U4 snRNA context (FP C), where this region is part of a helix extending stem I (24). If we assume an ideal RNA helix, then the distance of 11 base pairs between FP B and FP C would imply that both footprints are located on the same face of the RNA helix, some 25 Å apart. In the context of the U4/U6 snRNA duplex, FP C is located at almost identical nucleotides, albeit somehow enlarged. To obtain a relative placing of FP B and FP C in the U4/U6 snRNA context that would be comparable with its placement in the U4 context, a continuity of stem I with helix II is required. This can only be achieved by a coaxial stacking of stem I onto the U4/U6 helix II. In fact, this has recently been proposed for the U4/U6 three-way junction comprising helix II, stem I, and helix I (25). Satisfyingly, the location of FP C is in good agreement with cross-links that were previously identified between hPrp31 and U4 snRNA nucleotides G18/U19 in the context of the U4/U6.U5 tri-snRNP (14). These two nucleotides are located directly next to the FP C.

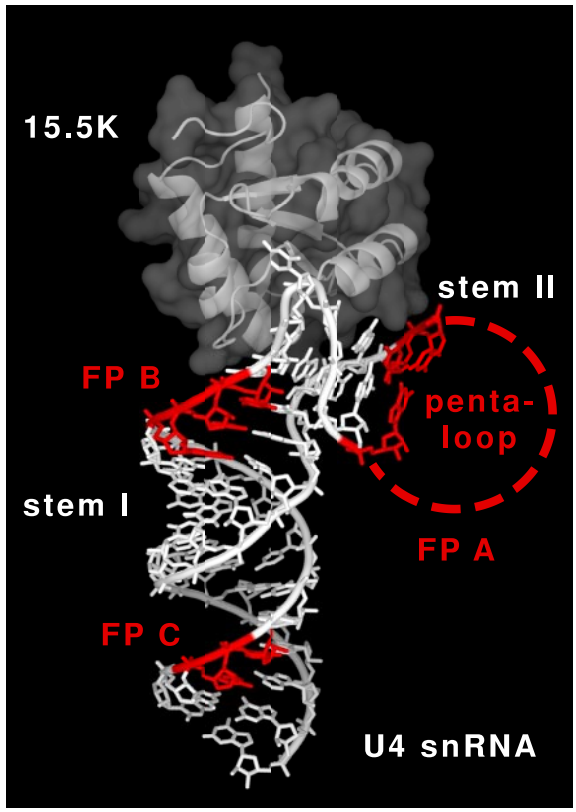

FIGURE 5. **Model demonstrating contact sites of hPrp31 in the U4 snRNA-15.5K binary complex.** The model is based on the structure of the 15.5K protein in complex with a U4 snRNA 5′ stem-loop fragment, of which only nucleotides G26–G35 and C41–C47 were resolved in the crystal (23). The stump of stem I was extended by an unrelated idealized 12-base pairs-long RNA A-helix (the 5′ and 3′ complementary strands were 5′-CCUUCCUUGGAU-3′ and 5′-AUCCAAGGAAGG-3′, respectively). Footprints of hPrp31 in the U4 snRNA are indicated in *green* and are labeled FP A–C as described under "Results" and in Fig. 1. The *dashed line* indicates the terminal penta-loop which was not resolved in the crystal structure (23). The figure was prepared with the PyMol program.

Interestingly, our data are fully compatible with extensive comparison by RNA structure probing of the U4/U6.U5 tri-snRNP (containing hPrp31) and 10 S U4/U6 di-snRNP (lacking it) (26). For example, sites of strong cleavage by the double-strand-specific nuclease V1 were observed on stem I of U4/U6 snRNAs, irrespective of the presence or absence of hPrp31 (U4-G48 and, to a lesser degree, U4-C49), indicating that hPrp31 makes contact with stem I from one side only. This is further corroborated by discrete V1 cleavages at the 5′ end of stem I (U4-U21 and U4-C22), which were only observed in the absence of hPrp31 (26). In sum, these observations suggest that what we observe in the U4/U6 complex is directly applicable to the situation in the tri-snRNP.

Having identified the hPrp31 sites of protection from hydroxyl radicals in the U4 snRNA, we next speculated how the experimental binding sites could be arranged in three-dimension. For this purpose we resorted to the crystal structure of the 15.5K-U4 snRNA kink-turn (23). Through the formation of the kink-turn, stem II would be forced into an angle of about 30° relative to the helix axis of stem I (23). Stem I of the U4 snRNA 5′ stem-loop was now simulated by adding an idealised RNA

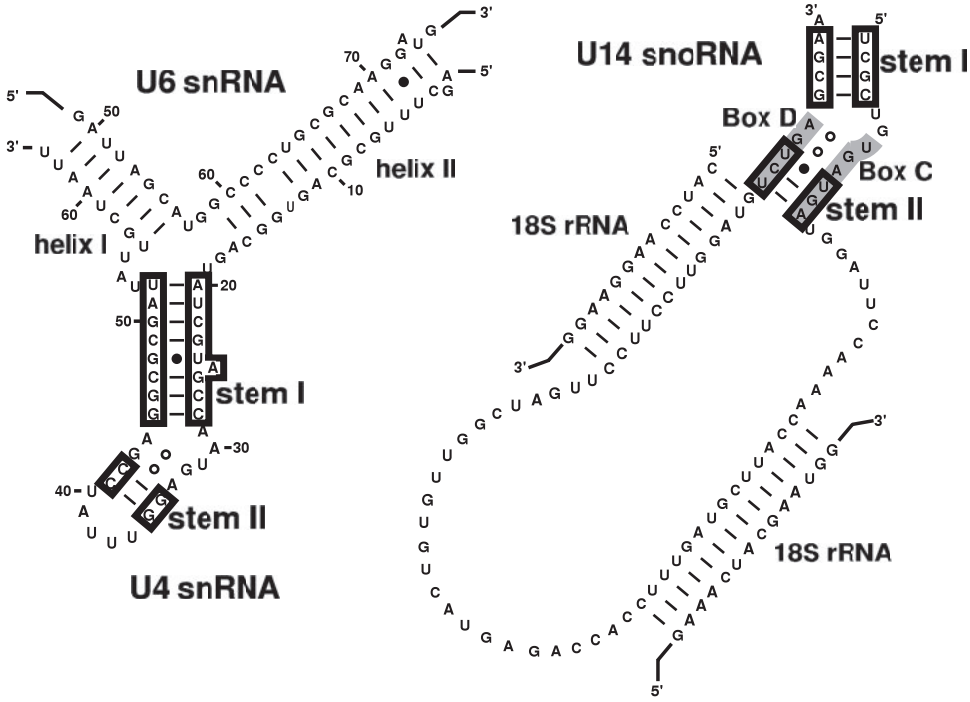

FIGURE 6. **Schematic representation of the secondary structures of the U4/U6 snRNA and the U14 box C/D snoRNA.** The secondary structures of the U4/U6 snRNA (7) and U14 box C/D snoRNA (21, 27) are shown on the *left* and *right*, respectively. The two flanking stems I and II that together with the internal loop form the kink-turn are *boxed* in *black*. The conserved boxes C and D of the U14 snoRNA are shaded in *gray*. Base pairing interactions are the same as described for Fig. 1. The two helical structures formed between U14 snoRNA and the 18 S rRNA, which are essential for modification and processing are indicated.

A-form helix on to the stump of stem I in the crystal structure (Fig. 5). The experimentally determined protection sites were then mapped onto this model. Clearly, hPrp31 would touch the RNA in two patches, namely FP C and FP B, and at the same time it would bury whatever is exposed of stem II and the penta-loop in FP A (Fig. 5). From this model it becomes clear that hPrp31 approaches the 15.5K-U4 snRNA complex from one side only.

From the footprint we concluded that the relative orientation of hPrp31 to the RNA elements should be topologically equivalent between U4 snRNA and the U4/U6 snRNA duplex. If we think of the entire helix II and the coaxially stacked stem I as a single idealized RNA helix, then hPrp31 would shield the upper half of this big helix in FP C and FP B. This asymmetric binding would be translated into an asymmetric architecture of the whole particle. Though highly simplified, this model, in conjunction with the three-way junction model (25), provides readily testable constraints for the relative positioning of helix I of the U4/U6 snRNA duplex. This is currently being done by a variety of different approaches.

RNA Structural Requirements for Binding of hPrp31 to the 15.5K-U4 snRNA Binary Complex—A systematic mutational analysis revealed a highly differentiated requirement for the various RNA structural elements. A battery of U4 snRNA 5′ stem-loop mutants was assayed by EMSA for overall kinetic stability in ternary complex formation.

The complete 5′-terminal nucleotides (A1-U19), including those nucleotides that were contacted in the ternary complex (FP C), were found to be inessential. Though we previously observed no binding of the U4-2 (14), we believe that the difference in the assay systems accounts for this discrepancy. The previous pulldown assay involved a wash step at 500 mM salt, whereas in the present study the ternary complex was never subjected to more than 150 mM salt. In addition, the EMSA may detect kinetically stable complexes, which cannot survive the continuous dilutions of a pulldown assay.

Although the region shielded by hPrp31 next to the kink-turn is only very short (FP B), the 8-bp stem I was required for efficient hPrp31 binding to the binary complex in U4 snRNA. Neither removal of the upper part (five base pairs and the bulge), nor a bottom part consisting of four base pairs only, were tolerated (Fig. 2). A comparison of the different U4 snRNA 5′ stem-loops (supplemental Fig. S2) revealed that stem I is invariably 8 base pairs long. In contrast, a U4atac snRNA mutant with a stem I of only 4 base pairs (U4atac-4, Fig. 4*B*) supported efficient ternary complex formation. This must reflect an intrinsic property of the U4atac stem II and/or the penta-loop, which presumably binds hPrp31 more strongly, thereby circumventing the stem I requirement. Consistently with this idea, the initial interaction of protein 15.5K with the U4atac stem-loop was found to be stronger than that with the U4 5′ stem-loop (22); in addition, hPrp31-RNA cross-link yields in the hPrp31-15.5K-U4atac snRNA complex were much higher than in the hPrp31-15.5K-U4 snRNA complex (14).

Paradoxically, the large penta-loop structure, in which every ribose is shielded upon hPrp31 binding, is itself not a prerequisite for binding. This is the more surprising because one of the two UV cross-linking sites of hPrp31 on U4 snRNA maps to this region (14). Furthermore, in the U4atac snRNA penta-loop, which has a different sequence, similar cross-links are observed to hPrp31 (18), indicating that various contacts can be detected by UV cross-linking, depending on which base is in contact with the protein surface. Since this RNA-protein interaction pattern is not required for primary recognition in the ternary complex, it is presumably important in one of the dynamic functional states of U4 snRNA during the splicing cycle.

Another interesting feature of the penta-loop deletion mutant (U4-B) is that the binding of protein 15.5K to the truncated U4 5′ stem-loop structure was hardly detectable by EMSA, even though ternary complex formation was unaffected (Fig. 3, *lanes 18* and *19*). Thus it appears that transient and/or weak interactions of this protein with the mutant RNA are dramatically stabilized by interaction with hPrp31. Since it is known that hPrp31 binds neither the U4 snRNA (14) nor the 15.5K protein (our pulldown data, not shown) on its own, it appears as if hPrp31 simultaneously recognizes

the protein and RNA components of the transient binary complex. It is conceivable that hPrp31 simultaneously recognizes a composite RNA/protein surface of the binary 15.5K-U4 snRNA complex.

The length of stem II turned out to be of crucial importance for hPrp31 binding to the 15.5K-U4 snRNA and 15.5K-U4atac snRNA binary complexes. Binding was completely abolished by extending the stem by a single Watson-Crick base pair, irrespective of the presence or absence of the penta-loop (Fig. 4). These findings imply that hPrp31 can accommodate at most two base pairs of stem II in a binding site; a third base pair interferes sterically with productive binding. This hypothetical binding site thus contributes significantly to the specificity of the RNA-protein recognition in both the U4 and U4atac snRNA contexts. Interestingly, a phylogenetic comparison (supplemental Fig. S2) of the penta-loops from human, rat, mouse, chicken, and fly revealed that in both U4 and U4atac snRNAs no Watson-Crick base pair can be formed at the bottom of the penta-loop. This restricts the length of stem II strictly to two base pairs, lending further support to the model for molecular recognition outlined above. Exceptions are found only in nematodes, which could form a C-G pair at the bottom of the penta-loop, and in yeasts, which have a six-membered ring. These latter may have special requirements, in that the six-membered ring closes with an A-U base pair (supplemental Fig. S2). Alternatively, a specialized loop structure may prohibit formation of the A-U base pair in yeast.

Different and Shared Binding Requirements between hPrp31 and the Box C/D snoRNP Proteins NOP56 and NOP58—The nucleolar box C/D snoRNPs and the spliceosomal U4 snRNP share a common core RNP structure, consisting of the 15.5K protein bound to an internal RNA loop. Protein hPrp31 selects exclusively U4 snRNA, whereas NOP56/NOP58 selects only the box C/D snoRNP structure, with the ultimate result that two particles with completely different biological properties and functions are produced. The RNA binding requirements of hPrp31 suggest how this discrimination may be achieved on a molecular level. RNA structures flanking the 15.5K-induced kink-turn are undoubtedly the important RNA structural determinants (Fig. 6) that are differentially recognized by the different Nop domains. As outlined above, the stem II binding site of hPrp31 can only accommodate a two base pair long stem, probably with further restrictions imposed by the closeness of the 15.5K protein. In contrast, in the box C/D snoRNA this stem is a conserved three base pair long stem with a closing A-U pair (27). In fact, the addition of a Watson-Crick base pair to the U4 or U4atac stem II (mutants U4-7 and U4-8, Fig. 4) resulted in the loss of hPrp31 binding. Therefore, stem II can be viewed as a major discrimination element. Moreover, box C/D snoRNAs have strict sequence requirements for stem II, where a central G-C base pair was found to be essential (27).

Acknowledgment—We thank Markus Wahl for critically reading the manuscript and for help in preparing Fig. 5.

REFERENCES

1. Burge, C. B., Tuschl, T., and Sharp, P. A. (1999) in *The RNA World* (Gesteland, R. F., Cech, T. R., and Atkins, J. F., eds) 2nd Ed., pp. 525–560, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
2. Will, C. L., and Lührmann, R. (2001) *Curr. Opin. Cell Biol.* **13,** 290–301
3. Jurica, M. S., and Moore, M. J. (2003) *Mol. Cell* **12,** 5–14
4. Bringmann, P., Appel, B., Rinke, J., Reuter, R., Theissen, H., and Lührmann, R. (1984) *EMBO J.* **3,** 1357–1363
5. Hashimoto, C., and Steitz, J. A. (1984) *Nucleic Acids Res.* **12,** 3283–3293
6. Rinke, J., Appel, B., Digweed, M., and Lührmann, R. (1985) *J. Mol. Biol.* **185,** 721–731
7. Brow, D. A., and Guthrie, C. (1988) *Nature* **334,** 213–218
8. Nilsen, T. W. (2003) *BioEssays* **25,** 1147–1149
9. Brow, D. A. (2002) *Annu. Rev. Genet.* **36,** 333–360
10. Staley, J. P., and Guthrie, C. (1998) *Cell* **92,** 315–326
11. Nilsen, T. W. (1998) in *RNA Structure and Function* (Simons, R. W., and Grunberg-Manago, M., eds) pp. 279–307, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
12. Makarova, O. V., Makarov, E. M., Liu, S., Vornlocher, H.-P., and Lührmann, R. (2002) *EMBO J.* **21,** 1148–1157
13. Schaffert, N., Hossbach, M., Heintzmann, R., Achsel, T., and Lührmann, R. (2004) *EMBO J.* **23,** 3000–3009
14. Nottrott, S., Urlaub, H., and Lührmann, R. (2002) *EMBO J.* **21,** 5527–5538
15. Makarov, E. M., Makarova, O. V., Urlaub, H., Gentzel, M., Will, C. L., Wilm, M., and Lührmann, R. (2002) *Science* **298,** 2205–2208
16. Klein, D. J., Schmeing, T. M., Moore, P. B., and Steitz, T. A. (2001) *EMBO J.* **20,** 4214–4221
17. Schneider, C., Will, C. L., Makarova, O. V., Makarov, E. M., and Lührmann, R. (2002) *Mol. Cell. Biol.* **22,** 3219–3229
18. Kühn-Hölsken, E., Lenz, C., Sander, B., Lührmann, R., and Urlaub, H. (2005) *RNA* (*N. Y.*) **11,** 1915–1930
19. Gautier, T., Berges, T., Tollervey, D., and Hurt, E. (1997) *Mol. Cell. Biol.* **17,** 7088–7098
20. Vithana, E. N., Abu-Safieh, L., Allen, M. J., Carey, A., Papaioannou, M., Chakarova, C., Al-Maghtheh, M., Ebenezer, N. D., Willis, C., Moore, A. T., Bird, A. C., Hunt, D. M., and Bhattacharya, S. S. (2001) *Mol. Cell* **8,** 375–381
21. Watkins, N. J., Ségault, V., Charpentier, B., Nottrott, S., Fabrizio, P., Bachi, A., Wilm, M., Rosbash, M., Branlant, C., and Lührmann, R. (2000) *Cell* **103,** 457–466
22. Nottrott, S., Hartmuth, K., Fabrizio, P., Urlaub, H., Vidovic, I., Ficner, R., and Lührmann, R. (1999) *EMBO J.* **18,** 6119–6133
23. Vidovic, I., Nottrott, S., Hartmuth, K., Lührmann, R., and Ficner, R. (2000) *Mol. Cell* **6,** 1331–1342
24. Myslinski, E., and Branlant, C. (1991) *Biochimie* (*Paris*) **73,** 17–28
25. Lescoute, A., and Westhof, E. (2006) *RNA* (*N. Y.*) **12,** 83–93
26. Mougin, A., Gottschalk, A., Fabrizio, P., Lührmann, R., and Branlant, C. (2002) *J. Mol. Biol.* **317,** 631–649
27. Watkins, N. J., Dickmanns, A., and Lührmann, R. (2002) *Mol. Cell. Biol.* **22,** 8342–8352